PHILIPS TECHNICAL REVIEW

N. V. PHILIPS' GLOEILAMPENFABRIEKEN

VOLUME 17

1955/1956

PHILIPS RESEARCH LABORATORIES – EINDHOVEN

CONTENTS

VOLUME 17, 1955/1956

LIGHT

TELECOMMUNICATIONS

X-RAYS AND NUCLEAR PHYSICS

SOUND

*) Articles marked with an asterisk * are brief reports.

MEASURING EQUIPMENT

MATERIALS AND THEIR RESEARCH

MISCELLANEOUS

*) Articles marked with an asterisk * are brief reports.

VOL. 17 No. 1, pp. 1-36 JULY 1955

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

MAINS-VOLTAGE STABILIZERS

by R. B. STEPHENS *).

621.316.722.1

Many types of electrical apparatus require a power supply with a very stable voltage. As the mains seldom satisfies this requirement, many types of auxiliary apparatus have been developed during the last two decades, to stabilize the mains voltage to a more or less constant value. This article deals with some of the existing types, and in particular with a new apparatus with some interesting improvements, which has been developed at the Philips X-ray Laboratory at Balham (London).

Introduction

It is well known that the voltage and frequency of the mains are subject to continuous fluctuations. In applications for which these fluctuations (and especially those of the voltage) are too large, matters can be improved by the use of a mains-voltage stabilizer. The requirements to be met by such apparatus may vary considerably according to the particular application. The principal requirements are discussed below.

a) The output voltage of the stabilizer should be as far as possible independent of the input voltage. The "factor of improvement", i.e. the ratio of the voltage fluctuations at the input to those at the output, should be very high for certain purposes, often as much as 100 (e.g. for the calibration of kWh-meters). In other cases a factor of 10 will be ample. An example of this is the stabilizing of the filament voltage of electron tubes, which is worth while for large installations as this results in a longer working life of the tubes.

b) The output voltage should not be dependent on the mains frequency. Although the frequency of most lighting mains averaged over a few hours is kept reasonably constant (in view of the use of synchronous clocks), short-term fluctuations of a few cycles per second may still occur.

c) Some applications require that the variations in the output voltage caused by variations in the load or load power factor, remain small. Where the load remains constant, this requirement is of course unnecessary.

d) Some mains-voltage stabilizers produce an output voltage which is by no means perfectly sinusoidal. For some purposes this cannot be permitted (e.g. in the calibration of A.C. instruments).

In this connection it may be noted that in most cases (e.g. for the heating of filaments) a constant *r.m.s.-value* of the voltage is required. Even if this last requirement is satisfied, a considerable distortion of the waveform of the output voltage (and also a fluctuation of its peak value) can still occur due to the presence of harmonics. If the load consists of a rectifier this is impermissible, since the rectified voltage is often a function of the peak value rather than of the r.m.s. value of the alternating voltage.

e) A further criterion in the choice of a stabilizer is the speed of regulation. If the mains voltage of the load fluctuates in sudden jumps, the output voltage will, as a rule, deviate for a short time from the desired value. The time necessary to reduce this deviation to 1/e of its maximum value, is called the recovery time. In some cases this recovery time is not critical, e.g. in the

*) Philips Balham Works, London.

stabilization of the filament voltage and high tension of an X-ray tube used for making long exposures. Here we are concerned only with the total X-ray dose (i.e. integrated over the time). If, however, short exposures are to be made, then the recovery time of the stabilizer has obviously to be short too.

The large number of different types of stabilizer now commercially available is due to the fact that satisfying all requirements at the same time is difficult and therefore often uneconomical; for this reason it is usual to design a stabilizer so that it meets only those requirements essential to a given application.

After a brief description of two types of stabilizer now in common use, a more thorough treatment will be given of another system, developed by the Philips X-ray Laboratory at Balham (London) which combines the advantages of both these types.

"Magnetic" stabilizer

For the "magnetic" stabilizer use is made of the non-linear relationship between magnetic induction (B) and field strength (H) in ferromagnetic materials. For a soft-iron core, for example, this results in a similar non-linear relationship between the applied alternating voltage (V) and the alternating current (I) passing through the choke, as indicated in *fig. 1*. The field strength is, in fact, proportional to the current, whereas the voltage is determined by dB/dt. It is obvious that, if the voltage is sinusoidal, the current will not be, and vice versa. As regards the r.m.s. value of voltage and current, however, the relationship of fig. 1 applies.

Roughly speaking, the choke can be said to behave as a self-inductance that becomes smaller as a higher voltage is applied to it. Such a choke shunted with a capacitor forms a resonant circuit, which, if fed with an alternating voltage of constant frequency, will resonate at a certain value of the applied voltage. This is illustrated by *fig. 2*. Here

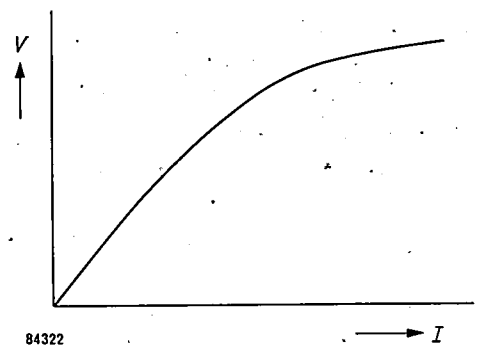


Fig. 1. Relationship between r.m.s. current and voltage for a choke with soft-iron core.

again the relationship between the current through the choke and the applied voltage is given by curve *1*. The straight line (*2*) shows the same relationship for the condenser current. The latter is in anti-phase

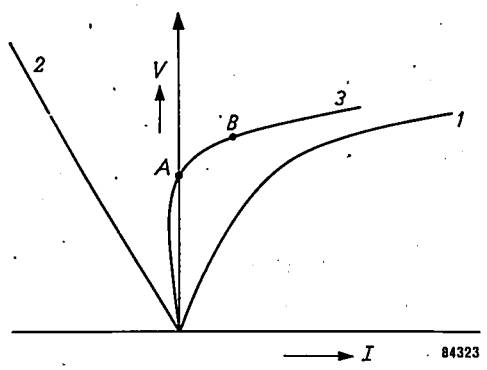


Fig. 2. Relationship between current and voltage for *1*) choke (as in fig. 1); *2*) condenser; *3*) choke and condenser in parallel.

with the choke current, so it has been drawn in the negative x-direction. The sum of both currents (curve *3*) passes through zero at a certain value of the applied voltage (point A). Here the circuit is in resonance. As the voltage becomes higher than the resonance value, the current consumed by the resonant circuit will rapidly increase. Conversely, we may express this by saying that in this range (about point B) the voltage across the resonant circuit is only slightly dependent on the current flowing through it. A resonant circuit of this kind, therefore, forms an excellent stabilizer when connected in series with an unsaturated and hence linear choke, which governs the order of magnitude of the current such that the resonant circuit adjusts itself to a voltage just above resonance (*fig. 3*,

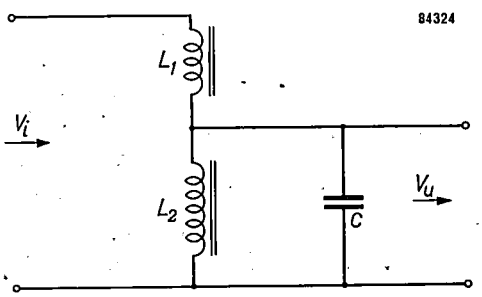


Fig. 3. Simple magnetic stabilizer. V_i = input voltage. V_u = output voltage, L_1 = non-saturated choke, L_2 = saturated choke. The resonant circuit L_2C is in resonance at one particular voltage.

operating point B of fig. 2). In this argument, hysteresis and other losses are neglected.

Without the condenser a stabilizing action would still remain. The difference between curves *1* and *3* of fig. 2, however, shows that the flat portion of the V-I relationship, necessary for good stabilizing, will occur at a smaller current value if a condenser is used.

Because of the fact that the voltage across the resonant circuit is so little dependent on the current passed through it, the output voltage is likewise little dependent on the current extracted from the output terminals.

Several improvements may be incorporated in the circuit of fig. 3. First of all an auto-transformer may be used for L_2, so that the output voltage V_u becomes as high as the nominal value of the input voltage V_i. L_1 may further be provided with an additional winding (*fig. 4*). The voltage V_1 produced by this winding is small compared with V_2 and of opposite polarity. The remaining small variations in V_2, caused by fluctuations in V_i, will then be compensated by the variations in V_1.

In the above description we have not considered the various complications that may occur. For a more thorough analysis reference can be made to the literature on this subject [1]).

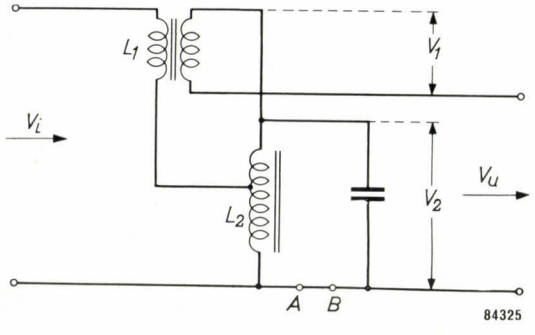


Fig. 4. Stabilizer similar to that of fig. 3, with compensation winding on L_1. Here L_2 is in the form of an autotransformer, so that V_u is equal to the nominal value of V_i.

Another version of this type of stabilizer employs one single iron core (*fig. 5*). The magnetic flux

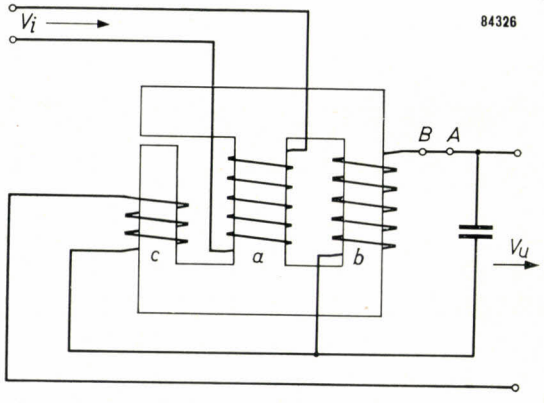


Fig. 5. Magnetic stabilizer with one core. The right limb becomes saturated earlier than the left limb with the air gap.

caused by the winding a, is distributed between the two outer limbs of the core. If the voltage V_i across a is increased, the greater part of the flux will initially pass through the right limb, because on the left the path is interrupted by an air gap. The right

[1]) W. Taeger, Spannungsgleichhaltungs-Schaltungen mit Eisendrosseln, Arch. Elektrotechn. **36**, 310-321, 1942. R. O. Lambert, Voltage regulating transformers, Electronic Engng. **15**, 384-387, 1943. H. A. W. Klinkhamer, Equivalent networks with highly-saturated iron cores with special reference to their use in the design of stabilizers, Philips tech. Rev. **2**, 276-281, 1937. and **6**, 39-45, 1941.

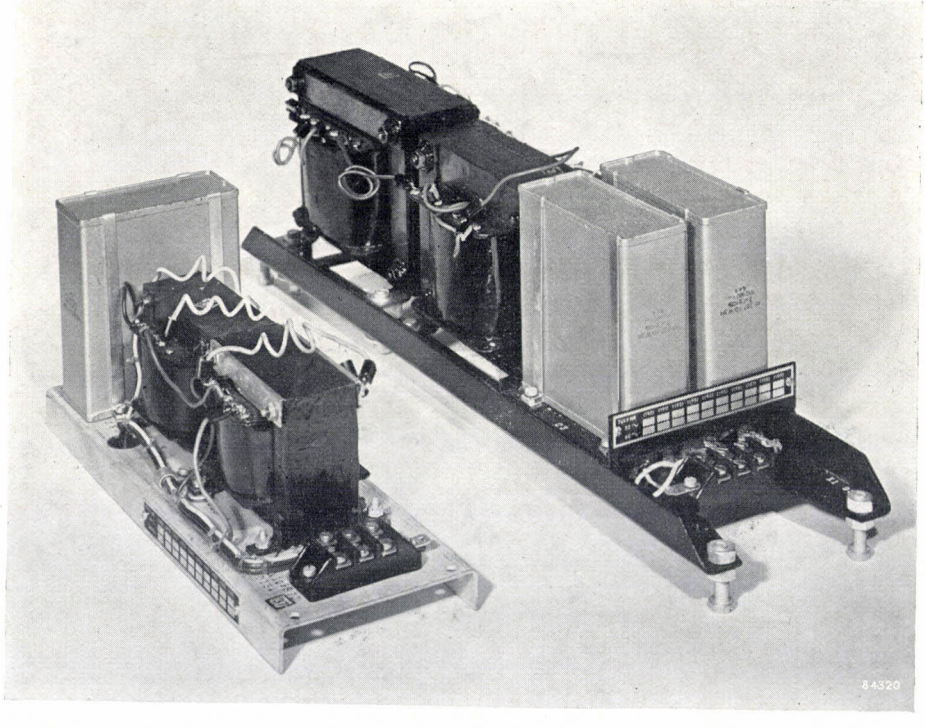


Fig. 6. Two magnetic stabilizers on the principle of fig. 4, one having an output of 80 VA and the other of 300 VA.

Table I

Stabilizer according to	Power output (VA)	Variations of output voltage for				Response time (seconds)
		10% variation of input voltage	5% frequency variation	Load variation between 10% and full load		
				$\cos\varphi = 1$	$\cos\varphi = 0.75$	
Fig. 6	80 300	< 0.2%	< 8%	< 1%	< 6%	0.02-0.04
Fig. 9	1000 2000 5000	—	—	< 0.2% (input voltage and frequency variations included)	< 0.5%	0.1-0.2
Fig. 16	150	< 0.1%	< 0.2%	< 0.2%	< 0.2%	0.02-0.04

limb, however, is thinner than the central one and thus becomes saturated at the higher values of V_i. Any further increase of V_i then cannot cause much change in the flux through the right limb. Any further increase of the flux through a will find its way through the left limb.

The effect of the capacitor in parallel to the b winding is analogous to that of fig. 3, although this is not at once apparent. Winding c again compensates for the remaining influence of V_i on V_u, and raises the factor of improvement. This winding may also be situated on the central leg.

The circuits of figs. 4 and 5 are both sensitive to frequency changes. In order to keep the oscillatory circuit in the region of resonance, it is necessary that at increasing frequency the self-inductance of the "saturated" coil becomes smaller. This means that the current through the coil and hence also the voltage across it, will increase. With these stabilizers a frequency change of 1% as a rule causes a variation of 1.5 % in the output voltage. Changes of load power factor also have a considerable effect on the output voltage. The regulating speed is high; a sudden voltage surge is compensated within one or two cycles. By choosing the right number of turns for the compensation winding it is possible to keep either the r.m.s. value, or the mean value, or the peak value of V_u constant. The shape of V_u is not sinusoidal, but depends on V_i and on the load current.

Fig. 6 shows two stabilizers, one for 80 VA and one for 300 VA (with circuits based on the principle of fig. 4). In *Table I* some of their properties are listed.

Stabilizers with feedback

The method to be discussed here (*fig. 7*) uses a detecting device B for determining the deviation of V_u from the nominal value. Between the input and output voltage of this detecting device there is a relationship as shown in *fig. 8*. V_1 is zero if V_u has the nominal value V_{u0}. This "error signal", after being amplified in A, is applied to a regulator C which adjusts V_u.

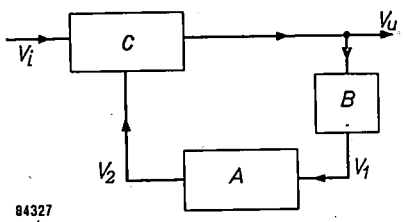


Fig. 7. Block diagram of a stabilizer with feedback. The deviation of V_u from its nominal value is converted into a voltage V_1 by the detecting device B. This error voltage is amplified by A and is used for regulating the output voltage by means of the controller C.

It will be clear that the constancy of V_u improves the more the error signal is amplified. In fact, it can be demonstrated than the factor of improvement (see *a*, p. 1) is equal to the overall amplification $+1$, the overall amplification being the total amplification in the closed circuit.

The non-linear element used to obtain a characteristic such as that shown in fig. 8, may take various forms. As a rule it has to react to the *r.m.s.-value* of V_u. In view of the definition of r.m.s.-value, this means that the output voltage of the controlling device should be a measure of V_u^2,

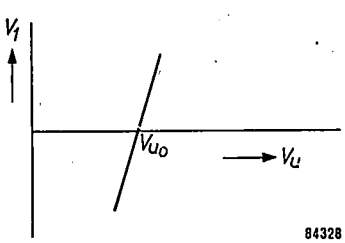


Fig. 8. Relationship between V_u and V_1 (fig. 7). V_1 is a measure of the deviation of V_u from its nominal value V_{u0}.

averaged over a certain time (at least one period). As a rule use is made of the quadratic relationship between the current through a resistor (proportional to V_u) and the power dissipated in it. The resistor temperature is a measure of the r.m.s. current; the averaging function is performed by its heat capacity. To ensure that the instantaneous current values over at least a whole cycle contribute about equally to the result of the averaging, the heat capacity should be so great, and/or the heat exchange with the surroundings so small, that the current (and therefore V_u^2) is in fact effectively averaged over a number of cycles. The delay produced in this way is characteristic of all mains-voltage stabilizers with feedback.

The stabilizers (5, 2 and 1 kVA) shown in *fig. 9* have a detecting device whose non-linear element consists of a diode operating in the saturation range. The filament of this diode (B_1, *fig. 10*) is connected, via transformer T_2, to V_u. The anode current is now highly dependent on the temperature of the cathode, and because the diode forms one of the branches of a D.C.-fed Wheatstone bridge, there exists a relationship similar to that of fig. 8 between

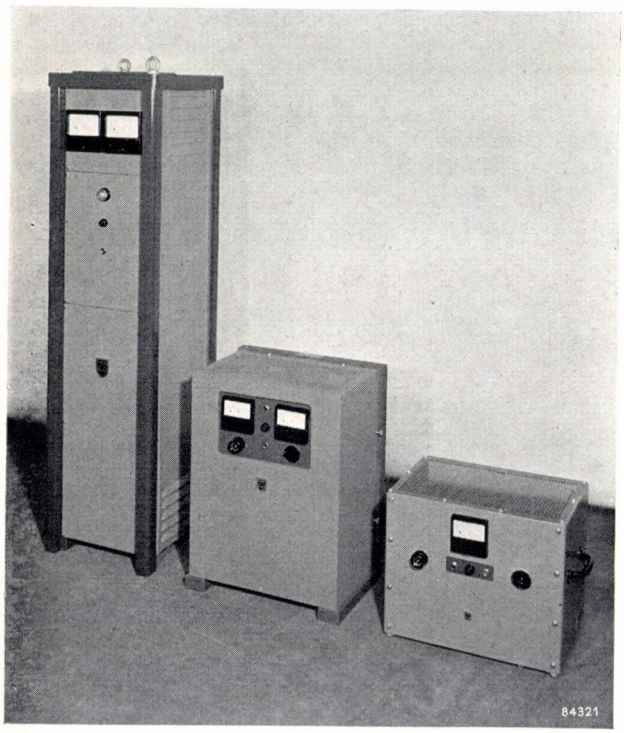


Fig. 9. 5, 2 and 1 kVA stabilizers on the principle of fig. 7.

V_u and the output voltage of the bridge. In this arrangement it is the filament that forms the squaring, averaging and delaying element.

The regulating part consists of the transductor *Td*, which is a choke whose self-inductance can be varied by passing a direct current through one of the windings. The step-up ratio of T_1 is in this arrangement a function of the self-inductance of *Td*. Here too, therefore, the saturation properties of soft iron are employed.

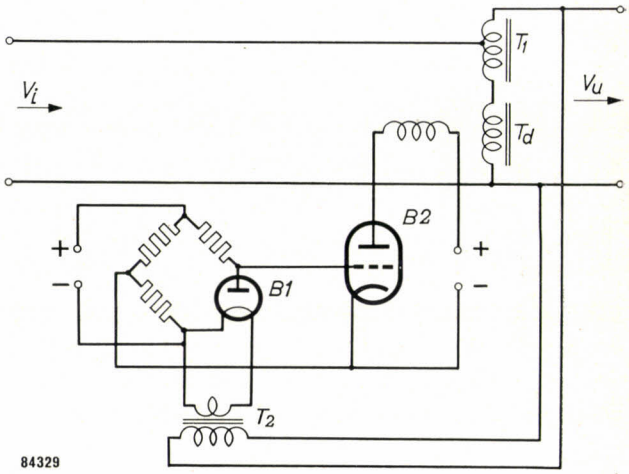


Fig. 10. Simplified circuit diagram of the equipments of fig. 9. The bridge incorporating the diode B_1 becomes unbalanced if the filament voltage of B_1 deviates from its nominal value; the bridge thus acts as detecting device.

Tube B_2 amplifies the error signal and controls the direct current through the transductor. An additional filter (not shown in fig. 10) suppresses the harmonics in the output voltage.

The stabilizers described here supply a more constant voltage than the "magnetic" stabilizers of fig. 6, as is evident from Table I. The speed of regulation, however, is somewhat lower. For many applications this is not important and this type of stabilizer is widely used.

A new fast-acting mains-voltage stabilizer

In the introduction, an application requiring a high speed of regulation was mentioned, viz. the supply of an X-ray tube used for making short-exposures, such as for diagnostic purposes. In this case, the high tension and the filament voltage of the X-ray tube must not deviate too much from their correct values: too high a voltage or too high an emission current would cause overloading and excessive blackening of the photographic material, while too low a voltage or current would cause insufficient blackening. The density of the photographic image is proportional to about the fifth power of the H.T. voltage, and to the eighth or tenth power of the filament voltage. The exposures are often very short (e.g. 0.02 sec.); hence a fast acting stabilizer is necessary.

The stabilizer to be described has been especially designed for those industrial and X-ray applications which require not only the regulation speed of the

magnetic stabilizers but also the frequency-independence and the accuracy of the feedback system. It is possible to combine these properties. To see how this may be done we need only consider what happens if a small, variable self-inductance is inserted between the points A, B in fig. 4 and 5. The introduction of this self-inductance will disturb the resonance. A system like this, however, as we have seen (fig. 2), will maintain a state that approximates closely to the state of resonance. The current through the saturated choke will rise, in order that the sum of the self-inductances in the resonant circuit remains substantially constant. Due to this, the voltage across the capacitor will also rise. The output voltage of the stabilizer can thus be controlled by means of an additional variable self-inductance. A certain output voltage will correspond to a given value of this self-inductance; this voltage is almost independent of the input voltage.

Here too, the variable self-inductance may be given the form of a transductor, energized by a feedback system similar to that shown in fig. 7. The result is that fluctuations of the output voltage, such as those caused by frequency fluctuations, are compensated by means of the transductor. It is true that this method of control is slow-acting, but the frequency variations concerned also occur slowly. The ordinary voltage fluctuations of the mains are rapidly compensated by the magnetic system.

This stabilizer also has a higher factor of improvement than that of the magnetic stabilizer alone. With regard to this better factor of improvement, the inertia of the feedback circuit does not constitute a serious drawback. This is due to the fact that the larger variations in the mains voltage (e.g. 10 V at a nominal mains-voltage of 220 V) do not as a rule occur suddenly, whilst the commoner smaller fluctuations are sufficiently stabilized by the magnetic stabilizer with its lower factor of improvement. Moreover, if the load or the load power factor is changed, the feedback will act so as to restore the output voltage to its correct value.

The detecting device

A convenient means of detecting small departures of an alternating voltage from its nominal r.m.s.-value is a thermistor bridge [2]), such as shown in *fig. 11*.

A typical voltage-current characteristic of a thermistor is shown in *fig. 12*, curve (*1*), whilst curve (*2*) shows the characteristic of a linear resistor.

[2]) Philips Tech. Rev. 9, 239-248, 1947.

If these two elements are connected in series, the resultant characteristic is of the form shown in curve (*3*), which by suitable choice of resistor has a flat portion of the curve AB wherein a very small change in applied voltage will produce a very large

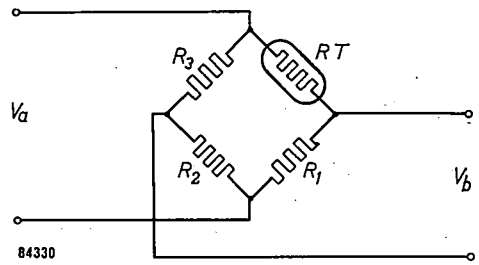


Fig. 11. Detecting device of the "Balham" stabilizer. V_a is proportional to the output voltage of the stabilizer. RT is a thermistor; its resistance is strongly dependent upon the temperature. At the normal value of V_a the bridge is just balanced. If V_a assumes a higher or lower value, the temperature of RT will change, so that the bridge becomes unbalanced. This causes a relationship similar to that of fig. 8 between V_a and V_b.

change in the circuit current, and consequently in the voltage developed across the linear resistor. This principle is employed in the bridge circuit of fig. 11; the resistors R_2 and R_3 are so chosen that the bridge is balanced if the system operates in the current range AB. Any small variation of the voltage supplied to the bridge (V_a) will then unbalance the bridge, and bring about a considerable variation of V_b. This applies not only to a direct voltage, but also to the r.m.s.-value of an alternating voltage, as the thermistor is incapable of following the rapid variations. The output voltage of the bridge will be either in phase or in antiphase with the supply voltage, according to whether the supply voltage is above or below the nominal value.

If a bridge of this type is used in a stabilizer, the peak shown in curve (*3*) of fig. 12 constitutes a difficulty, because of the fact that when the appara-

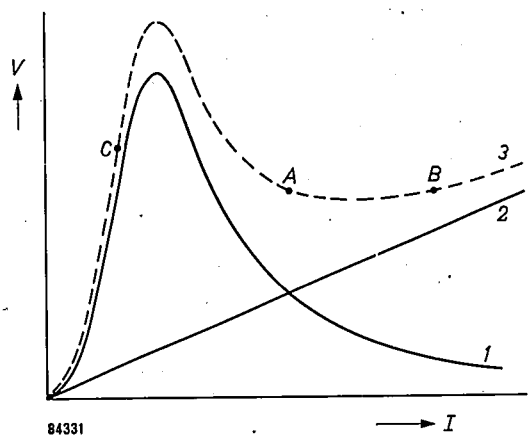


Fig. 12. Voltage-current characteristic of *1*) a thermistor, *2*) an ordinary resistor, *3*) a thermistor and resistor in series.

tus is switched on, the working point occurs somewhere in the vicinity of C. The bridge then produces a high output voltage which tends to increase the output voltage of the stabilizer, and thus to produce an upward shift of the working point. In this, however, it does not succeed, since the output voltage can only be varied by the transductor within certain limits. In order to raise the working point from C across the peak it would be necessary to raise the output voltage of the stabilizer (which feeds the bridge) for a short time above its nominal value, which, obviously, is not permissible.

In existing stabilizers with thermistor control special circuits with relays and electronic tubes are applied to overcome this obstacle [3]). In the apparatus under discussion the need for these special precautions has been obviated by the use of an *indirectly heated* thermistor, which is preheated by a spiral heating element surrounding it. The effect of this preheating is that the peak on the thermistor voltage-current characteristic is lowered considerably (curve *1*, *fig. 13*). If a resistor is

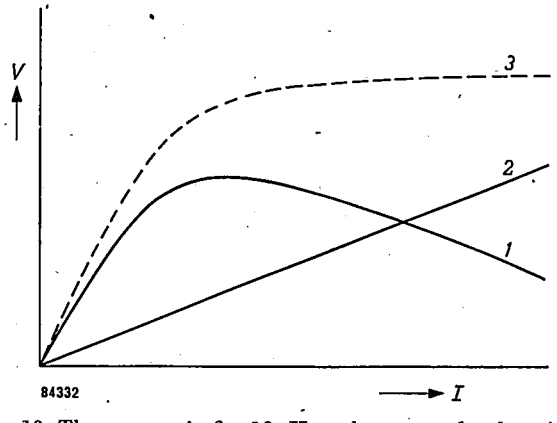


Fig. 13. The same as in fig. 12. Here, however, the thermistor is pre-heated by a separate heating element.

connected in series with such a thermistor the characteristic becomes as shown in curve (*3*). In this curve the long, flat portion necessary for high sensitivity has been retained, but as the maximum has disappeared, it is unnecessary to provide special circuits for moving the working point to a flat region.

The amplifier

If the output voltage of the stabilizer deviates from its nominal value, the thermistor bridge will produce a signal in phase or antiphase with the supply voltage. This signal must be used to alter the direct current through the transductor so as to restore the output voltage to its correct value. This is achieved by amplifying the bridge output voltage and applying it to the control grid of an output pentode (*fig. 14*) whose anode circuit incorporates the D.C. winding of the transductor *Td*. The screen

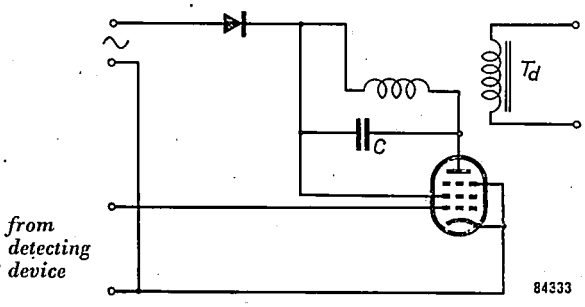


Fig. 14. Amplifier of the "Balham" stabilizer. A pulsating direct voltage is applied to anode and screen grid of the tube, so that this tube functions as a rectifier; the current through the D.C. winding of transductor *Td* in the anode circuit is dependent on the alternating voltage on the grid.

grid and anode of this tube are supplied with a rectified non-smoothed voltage in phase with the supply to the thermistor bridge. If the signal on the grid is in phase with the anode voltage, the mean anode current will increase by an amount dependent on the grid signal. Conversely, if the grid signal is in antiphase with the anode signal, the mean anode current will be reduced. The current flowing through the D.C. winding of the transductor is smoothed by the capacitor C.

This amplifier, which is at the same time a phase-sensitive detector, reacts to that component of the bridge output voltage that is in phase (or in antiphase) with the mains voltage. If the bridge is balanced for this component, there remains a low voltage with a phase difference of 90° with respect to the bridge voltage. This residual voltage is made up of components having the mains frequency and its third harmonic, and is produced due to the fact that the thermistor is not infinitely slow-acting, but changes its resistance perceptibly during one period of the mains frequency [4]). As the two components lead the mains voltage by 90°, they do not influence the current through the transductor. The filters used in previous thermistor circuits are, therefore, superfluous.

The complete circuit

The complete circuit of the detecting and regulating devices is given in *fig. 15*.

The transformer T_1, which is connected to the output voltage of the stabilizer, supplies the anode and filament voltage for the amplifying tube B_1, and in addition it also supplies the thermistor heater and the thermistor bridge. The latter consists of the potentiometer R_2, the resistor R_3 and the thermistor RT_2. The output voltage of the bridge is coupled to

[3]) G. N. Patchett, Precision A. C. voltage stabilizers, Part VI, Electronic Eng. 22, 499-503, 1950.

[4]) See the article referred to in [3]), Part I, pp. 374-377.

the control grid of an amplifying valve B_1. The anode supply of this valve is derived from the unsmoothed supply to the phase-sensitive detector via the smoothing network R_4-R_5-C_1-C_2.

Since the thermistor is a temperature-sensitive element, variations in ambient temperature would normally tend to cause drifts in the stabilizer output voltage. To compensate for such effects, the thermistor heater is shunted by another thermistor RT_1 of such a size that self-heating due to its own current flow is small. Increasing the ambient temperature produces a decrease in the resistance of RT_1 and a corresponding reduction in the heater current of RT_2. This compensation is sufficiently accurate in the temperature range of 10-50 °C.

In the event of failure of the anode current of B_1 for any reason, the self-inductance of the transductor and consequently the output voltage would rise to a high value. To prevent this, a relay RL_1 has been incorporated in the anode circuit of the tube. A contact RL_1c on the relay short-circuits the A.C.-winding of the transductor if the relay becomes de-energized. Since the short-circuiting of the transductor causes a low output voltage of the stabilizer, the load circuit is protected against excessive voltage in the event of tube failure.

The relay also facilitates the switching on of the circuit in the following way. When the stabilizer is connected to the mains, B_2 at first does not pass anode current due to its cold filament. Relay RL_1 is therefore not energized. Under these conditions thermistor RT_2 is provided with an excess heater voltage via winding S_3 and contact RL_1a, and is *rapidly* brought to its correct operating temperature. When the tube current reaches a value sufficient to close the relay, the normal heater voltage is applied to the RT_2 heater via contact RL_1b, and R_1 which is adjusted so that the current has the correct value.

The current at which the relay closes is adjusted by the parallel resistor R_8 to a value which does not produce a surge of stabilizer output voltage when contact RL_1c across the transductor is opened. Contact RL_1d disconnects R_8 when the relay is energized, enabling the latter to remain closed down to the lower safety limit of the tube current.

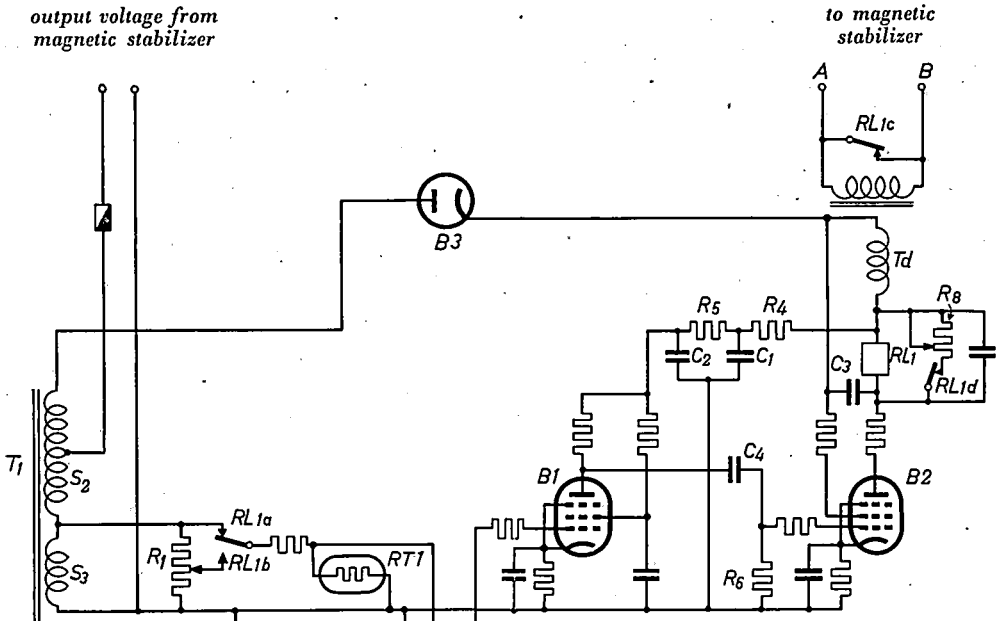
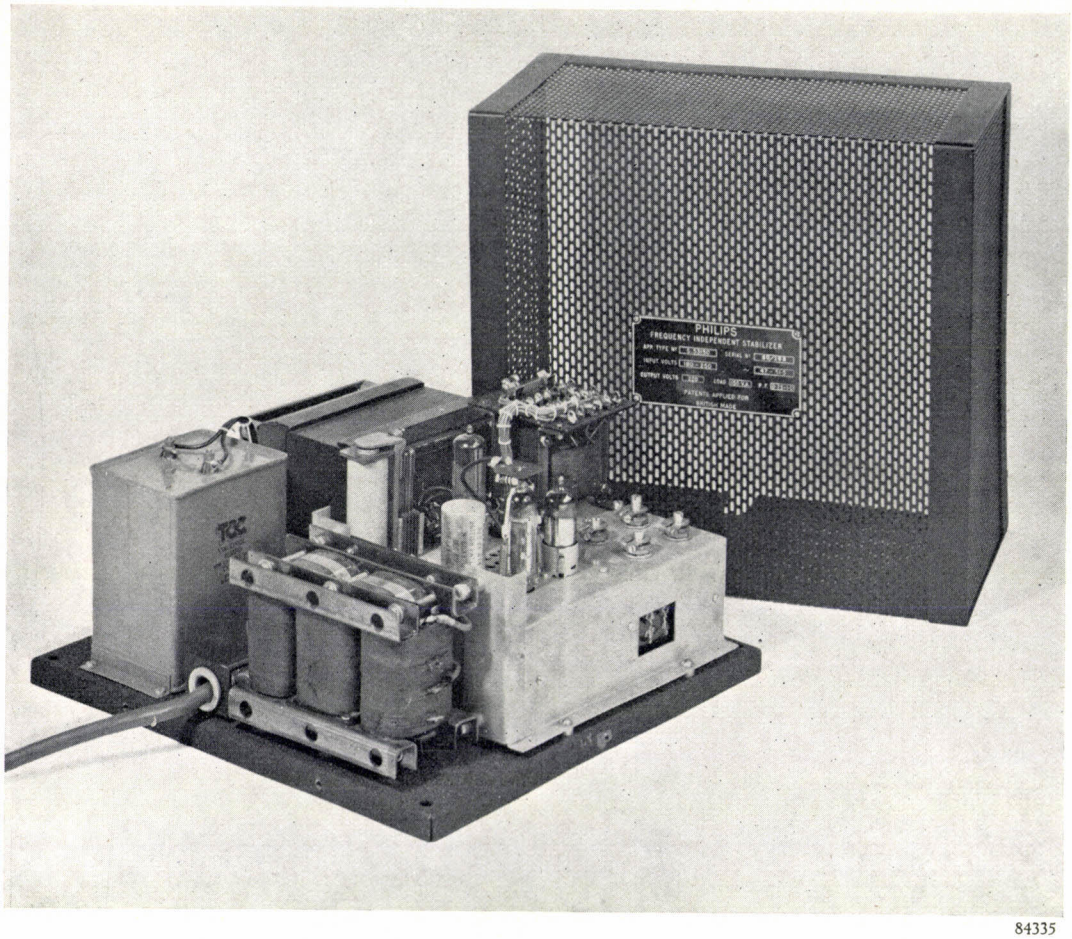


Fig. 15. Circuit diagram of the "Balham" stabilizer. The "magnetic" part is not shown. The A.C. coil of the transductor is assumed to be connected between the points A and B of fig. 4 or fig. 5.

The anode supply of B_1 is derived from the junction of RL_1 and the D.C. winding of the transductor. Hence a fraction of the anode voltage of B_2 is fed back to the grid via C_4, resulting in a negative feedback at frequencies within the passband of the network C_4R_6. This feedback, which is analogous to "velocity feedback" in a servo system, prevents low frequency oscillations in the loop of fig. 7. The presence of C_4 prevents D.C. feedback and thus the "improvement factor" is not reduced.

The inductance of the D.C. winding of the transductor, together with the parallel condenser C_3 and the internal resistance of B_2, causes a phase shift for sinusoidally varying signals which becomes larger at increasing frequency. The



84335

Fig. 16. Stabilizer for 150 VA output based on the circuit of fig. 15.

inertia of the thermistor produces a similar effect. For the frequency at which the total phase shift in the loop of fig. 7 is 180°, the total loop amplification should be smaller than unity to prevent oscillation. The above-mentioned frequency-dependent feedback via C_4 will counteract the phase shift due to C_3 and the transductor inductance, except at frequencies so high as to be attenuated sufficiently by the inertia of the thermistor.

The output voltage of the stabilizer can be adjusted within certain limits by means of R_7. The position of this potentiometer determines the value of the primary voltage of T_1 at which the bridge is balanced.

The complete voltage stabilizer is shown in *fig. 16*. The maximum load is 150 VA.

The stabilizer will compensate for any of the following changes:

Input voltage changes from 180 to 250 V.

Frequency changes in the range 47-51.5 cps.

Load changes in the range 0-150 VA.

Power Factor changes in the range 0.75-1.0.

Temperature changes from 10 °C to 50 °C.

For any combination of operating conditions within the above ranges, the output voltage will lie within $\pm$ 0.5 % of the nominal voltage.

For more limited changes the deviation will be less (see table I).

Summary. After surveying the requirements which mains voltage stabilizers should meet for various applications, this article gives a brief description of "magnetic" stabilizers and "feedback" stabilizers. The former operate on the basis of the non-linear relationship between magnetic field strength and induction in the soft-iron core of a choke. Although these stabilizers are very fast-acting, their applications are restricted by the fact that the output voltage is dependent on the mains frequency and the load power factor.

The feedback type of stabilizer, an example of which is also given, has the drawback in some applications of being rather more slowly acting, due to the fact that the voltage-sensitive element, incorporated in the feedback circuit is necessarily slow-acting. This type, however, achieves a high precision and frequency-independence. A new type of stabilizer is described, developed especially for an application requiring fast operation. It combines the favourable features of both above-mentioned types, because here the output voltage of an ordinary magnetic stabilizer is kept constant with the aid of a simple feedback system. The voltage-sensitive element is a bridge circuit incorporating a thermistor. Indirect heating of the latter prevents difficulties during warming up. The influence of the ambient temperature is compensated by a second thermistor. A stability to within $\pm$ 0.5 % over the temperature range 10-50 °C can be achieved for mains-voltage fluctuations between 180 V and 250 V, frequency variations between 47 c/s and 51.5 c/s, and load variations from no load to full load. The recovery time amounts to between one and two cycles of the mains frequency.

SUPPLY UNITS FOR AIRFIELD LIGHT BEACONING SYSTEMS

by Th. HEHENKAMP.

621.326.062:628.971.8:629.139.1

Resonance phenomena are essential features of high-frequency technique, but rarely find application in power engineering. Nevertheless, a resonant circuit occasionally offers the best solution to a particular power problem, for example that of supplying a constant current to lamps in series. The principle of resonant circuits — which are linked to such names as Boucherot and Steinmetz — is now more than 60 years old; it does not often happen in electrical engineering that a principle so long known and so much used in some fields has yet been neglected in others.

To enable aircraft to land safely, in darkness or in mist, it is necessary to equip the airfields with effective light beacons. An article describing in detail the system of beaconing employed on modern airfields appeared in a recent issue of this Review [1]). As explained there, the beacon lights are fitted with low-voltage incandescent lamps (6 to 30 V), since this permits the use of a compact filament, whereby a more concentrated light beam can be attained.

A direct supply to these lamps in parallel is impracticable, since it would necessitate unduly thick cables. On the other hand, supplying the lamp direct in series, although possible, has the disadvantage that breakdown of one lamp would interrupt the entire chain [2]). Accordingly, every lamp in the system, whether supplied in series or in parallel, is provided with a separate transformer which matches it to the appropriate mains voltage or current. Such lamp transformers are made for different powers and in various types, according to the particular method of installation (buried, incorporated in the light itself, etc.). The transformers are fed from a special supply unit equipped with magnetic switches, enabling the required runway lights and their luminous intensities to be selected. Such supply units are housed in transformer stations located as close to the runways as possible and often below ground level. They can be operated either in the station itself or from the control tower, which is then equipped with a desk incorporating the necessary controls and monitor lamps.

Before discussing the details of this equipment, let us consider whether it is better to connect the primary coils of the individual lamp transformers in parallel, or in series.

[1]) J. B. de Boer, Philips tech. Rev. **16**, 273-286, 1954/55 (No. 10).

[2]) This can be prevented by connecting a special cartridge fuse in parallel with each lamp, a method sometimes employed in street lighting systems (see Philips tech. Rev. **6**, 105-109, 1941); however, it is not considered reliable enough for airfield lighting.

Parallel or series supply?

Of all the various considerations upon which the choice of the supply system should be based, the most important are capital expenditure and reliability.

Capital expenditure

To minimize capital expenditure as far as possible, it is necessary to employ thin cables. Consequently, an appreciable voltage drop must be tolerated; in general, the associated loss of energy is not important, since the costs arising from it are small as compared with the cable depreciation (one runway 2 miles long may require cable costing £ 5000).

With a parallel supply system, the toleration of a substantial voltage drop gives rise to a noticeable difference in luminous intensity between the lamps at opposite ends of the cable. In a series system, on the other hand, such a drop produces no difference in luminous intensity, since the current is the same in all the lamps. Although this is a very important advantage of the series system, it is not in itself reason enough to conclude that such a system is best for the present purpose. Before reaching such a conclusion, we must compare the two systems more precisely by determining the relative voltage drop in them when the cable cost, or, to simplify matters, the volume of copper (by far the biggest contribution to the cost of the cable), is the same in both (for a given copper volume, the number of cores has little effect on the cost).

Let us consider the case of a runway (length l), flanked on both sides by n lights each of nominal power P arranged at regular intervals; in practice, n is usually between 50 and 100. With a parallel supply system (*fig. 1a*), the voltage across the first lamp is E_p (suffix $_p$ for parallel) and that across the last $E_p - \Delta E_p$. Owing to the positive temperature coefficient of resistance of the tungsten filaments, the difference in current between the individual lamps will be appreciably smaller than the difference

in voltage. To simplify the following calculation, let us ignore current differences, that is, assume that the current supplied to each lamp transformer is $I_p = P/E_p$.

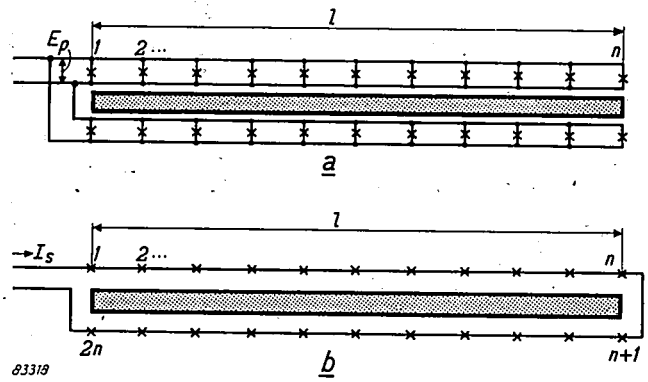


Fig. 1. Runway with n lights on each side. *a*) Parallel supply, *b*) series supply.

We then find that the potential difference ΔE_p is:

$$\Delta E_p = \frac{n\varrho l P}{A_p E_p},$$

where A_p is the cross-sectional area of one cable core, and ϱ the specific resistance of copper. Accordingly, the relative potential difference is:

$$e_p = \frac{\Delta E_p}{E_p} = \frac{n\varrho l P}{A_p E_p^2}. \quad \ldots\ldots \quad (1)$$

With a similar system of lights supplied in series (fig. 1*b*), with a current I_s, the voltage drop in the cable will be:

$$\Delta E_s = \frac{2\varrho l I_s}{A_s} \text{ (suffix s for series).}$$

The total voltage across the series system is then:

$$E_s = \frac{2nP}{I_s} + \Delta E_s.$$

Again introducing the relative voltage drop ($e_s = \Delta E_s/E_s$), we have:

$$e_s\,(1-e_s) = \frac{4n\varrho l P}{A_s E_s^2}. \quad \ldots\ldots \quad (2)$$

Since in a parallel supply system the runway is flanked on each side by two cores of cross-sectional area A_p, and in a series system by one core of area A_s on each side, the condition of equal copper volume is satisfied if $2A_p = A_s$.

From (1) and (2), we have, therefore:

$$e_p E_p^2 = \tfrac{1}{2} e_s\,(1-e_s) E_s^2.$$

Assuming that the supply voltages of the two systems are the same, the relationship between e_p and e_s becomes:

$$e_p = \tfrac{1}{2} e_s\,(1-e_s). \quad \ldots\ldots \quad (3)$$

This relationship is shown graphically in *fig. 2*.

Because in a parallel system the difference in voltage gives rise to a difference in luminous intensity between the lamps, e_p is limited, in the absence of special measures, to about 3% (the fact that the lamps themselves, the optical systems and their adjustment, exhibit a certain amount of spread is taken into account).

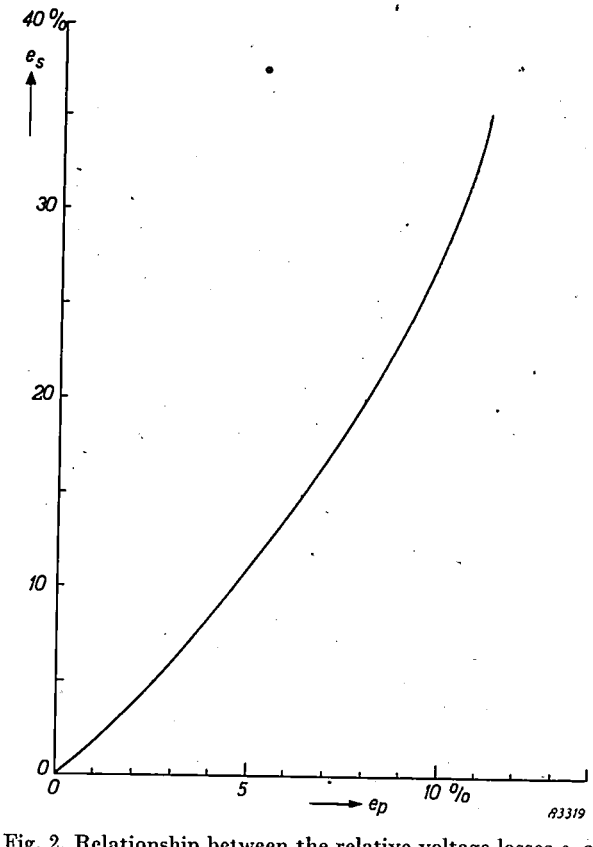


Fig. 2. Relationship between the relative voltage losses e_s and e_p, in a series and a parallel system respectively, the copper volume, mains voltage, and overall runway lighting power being the same in both cases.

According to fig. 2, $e_p = 3\%$ corresponds to $e_s = 6\%$ in a series system with the same copper volume and supply voltage. However, in a series system e_s may be taken very much higher (e.g. 25%), and a very considerable saving in cable costs thus procured without impairing the lighting, provided that the particular supply unit is designed for a correspondingly higher power level; e_s is then limited only by the last-mentioned condition, usually to about 25 or 30%.

To compensate for the fact that in a parallel system the voltage across the more remote lamps is lower than that across the nearer ones, the lamp transformers may be provided with suitable tappings. This enables the limit on e_p to be raised to

about 10%. (If it were any higher, the lamps would show too perceptible a difference in luminous intensity when operated at reduced power, owing to the fact that the resistance of the lamps is then so low as to increase the relative effect of the cable resistance considerably. The lamps are normally operated at reduced power to avoid glare, the full light yield being required only in foggy conditions by day). From fig. 2, $e_p = 10$ % corresponds to $e_s = 28$ % in a series system. If this is taken as the maximum limit, no further saving in cable costs will be possible; nevertheless, the series system is better in that it ensures a uniform light output, whereas in a parallel system with $e_p = 10$ % the irregularity of the lighting is only just within the limits of tolerance.

In many cases there are still stronger objections to the use of a parallel circuit. A high voltage supply is less acceptable in such a system than in a series circuit where there is never a high potential difference between adjacent points (except, of course, inside the transformer station). This makes problems of insulation and protection far easier to solve than in a parallel system.

Reliability

From the point of view of reliability, it is extremely important that the effects of minor local defects and breakdowns be reduced as far as possible. The transformers, and the actual lights, are often placed below, or only just above, ground level, so that various clips and pressure contacts (e.g. those of the lamp holders) are exposed to a corrosive atmosphere. Accordingly, the possibility of some contact resistance must always be taken into consideration. One of the advantages of series circuits whose current supply does not depend upon the load resistance is that contact resistances do not affect the current in the circuit, but merely cause a slight voltage drop across the bad contact. Moreover, a contact broken electrically but mechanically intact (e.g. that between the lamp cap and the lamp holder) may easily be restored by the high voltage applied to it when the installation is switched on. With a parallel supply, on the other hand, contact resistances present a far more difficult problem, since they may cause under-loading of the particular lamp or group of lamps, in some cases so much so as to extinguish it completely.

Interference of another kind is caused by short-circuits, as will now be explained. The lights may be damaged mechanically by aircraft or other traffic, or in the course of work carried out along the runway, in such a way as to cause a short-circuit. In a series system, the worst that can then happen is failure of the lights affected, the others continuing to operate normally. With parallel feeding, however, the possibility of shorting necessitates expensive safety measures; for example, each light must be provided with a fuse in series, and the positioning of such fuses usually presents an awkward problem. Moreover, it is difficult to ensure that they will function reliably, since, owing to the relatively high cable resistance tolerated for reasons of economy, the short-circuit current only slightly exceeds the normal operating current. Taking into consideration also the fact that it is simpler to standardise the supply units and transformers of a series system, and to compensate for changes taking place during or after the installation of the system, it will be evident that in general a series supply is best suited to the requirements of airfield beaconing systems.

However, approach lights may be an exception to this general rule, particularly high power lights. In approach lighting, a large number of lights, consuming a great deal of power, are concentrated in a limited zone some distance from the runway, so that the risk of damage to them or to the cables is relatively small. These lights and the associated transformers may therefore be raised quite a long way above ground level; hence the operating conditions of this part of the installation are more favourable than those of the actual runway lights. Accordingly, the lights are less likely to become defective and are easier to inspect and maintain. Moreover, the fitting of the fuses required in a parellel system is relatively simpler. Here, three-phase mains, with the lights distributed evenly between the three phases, is often the best solution.

Boucherot's constant-current bridge circuit

To employ a series system to full advantage, it is essential to make the supply current independent of the load. This can be accomplished by means of a transformer with a movable secondary coil, a method already used in street lighting for several decades. However, another method, involving no moving parts, has become more popular during the past 15 years. It is based on a bridge circuit originally designed by Boucherot (1890) and developed by Steinmetz [3]).

In the ideal case (exact resonance and no losses), this circuit exhibits two special characteristics:

1) the output current does not depend upon the size and nature of the load impedance, and
2) with resistive loading, the input current is in phase with the supply voltage from the mains.

We are most concerned with the first characteristic, since the second merely indicates that with resistive loading the power factor is unity.

Theory of the constant current bridge (neglecting losses)

To demonstrate the above characteristics, let us now consider a bridge circuit with impedance Z_1 in

[3]) See, for example, R. R. Miner, Resonant-type constant-current regulators, Trans. Amer. Inst. El. Engrs. 58, 822-829, 1939.

one pair of opposite arms, impedance Z_2 in the other pair of arms (*fig. 3a*), impedance Z in the diagonal link and a supply voltage E_0 across the other diagonal.

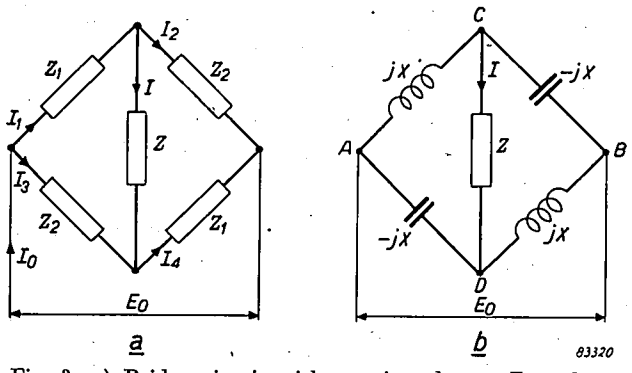


Fig. 3. *a*) Bridge circuit with two impedances Z_1 and two impedances Z_2 in the arms, and a load impedance Z in one of the diagonals.
b) Putting $Z_1 = +jX$ and $Z_2 = -jX$ in *a*), we have a Boucherot bridge, in which the current I in Z is independent of Z.

From the equations (for notation see fig. 3*a*)

$$I = I_1 - I_2$$

and

$$E_0 = I_1 Z_1 + I_2 Z_2,$$

after eliminating I_1, we have:

$$I = \frac{E_0 - I_2(Z_1 + Z_2)}{Z_1}. \quad \ldots\ldots \quad (4)$$

We now choose Z_1 and Z_2 such that

$$Z_1 = -Z_2 \quad \ldots\ldots \quad (5)$$

(implying that the two are in resonance at the mains frequency).

From (4), we then get:

$$I = \frac{E_0}{Z_1}, \quad \ldots\ldots \quad (6)$$

showing that the output current I is in fact independent of the load impedance Z.

We find that the current from the mains (I_0) is:

$$I_0 = I_1 + I_3,$$

and that:

$$I_1 Z_1 + IZ - I_3 Z_2 = 0.$$

Eliminating I_1 from these two equations, we have:

$$I_0 = \frac{I_3(Z_1 + Z_2) - IZ}{Z_1},$$

which becomes with the aid of formulae (5) and (6):

$$I_0 = -\frac{Z}{Z_1^2} E_0. \quad \ldots\ldots \quad (7)$$

To satisfy the resonance condition (5) fully, Z_1 and Z_2 must be entirely devoid of losses, that is, pure reactances: $Z_1 = jX$ and $Z_2 = -jX$ (fig. 3*b*). Accordingly, the denominator in formula (7), Z_1^2, is real; hence it is seen that the input current will be in phase with the supply voltage (E_0) if the load impedance (Z) is a pure resistance. This, then, demonstrates the second characteristic.

The circuit considered has yet another unusual feature (this time an undesirable one), namely that the currents in the four arms, I_1, I_2, I_3 and I_4, can not be individually determined; application of Kirchhoff's laws to the junctions and loops of the bridge produce a set of interdependent equations. However, the sums of the currents in opposite arms are determinate:

$$I_1 + I_4 = I\left(1 - \frac{Z}{Z_1}\right) \quad \ldots\ldots \quad (8)$$

and

$$I_2 + I_3 = -I\left(1 - \frac{Z}{Z_1}\right). \quad \ldots\ldots \quad (9)$$

Provided that formulae (8) and (9) hold good, the currents in the individual arms may assume any value. This will be evident from the fact that the impedance of circuit $ACBDA$ (fig. 3*b*) is zero, so that any current can flow in this circuit without affecting the voltages across the diagonals.

Another point to bear in mind in this connection is that the individual currents referred to are indeterminate only at resonance and with circuit elements devoid of losses, as will be seen from the following. In general:

$$I_1 + I_3 = I_2 + I_4 \quad \ldots\ldots \quad (10)$$

and

$$I_1 Z_1 + I_2 Z_2 = I_3 Z_2 + I_4 Z_1.$$

Hence we have:

$$(I_1 - I_4) Z_1 = -(I_1 - I_4) Z_2. \quad \ldots \quad (11)$$

Now, if $Z_1 \neq -Z_2$ (owing to a resistive term in Z_1 and Z_2, or a deviation from resonance, or a combination of the two), it follows from formulae (11) and (10) that:

$$I_1 = I_4 \quad \text{and} \quad I_2 = I_3 \quad \ldots\ldots \quad (12)$$

It will be seen, then, that the individual arm currents associated with any variation from formula (5) are determinate, these currents being indeterminate only when formula (5) holds; at the same time, this formula must hold good to give the circuit the required constant current characteristic, (equation 6). In practice, this contradictory state of affairs necessitates a compromise, as will now be explained.

Theory of the bridge taking losses into account

The practical bridge circuit differs from the purely theoretical one so far considered, in that the circuit elements are not entirely devoid of losses. The capacitor losses are usually insignificant, but for economic reasons the losses in the coils cannot be reduced till their effect is made negligible. Hence it is necessary to ascertain the precise nature of this effect.

With circuit elements not entirely devoid of losses, condition (5) cannot be fully satisfied, and, as we have already seen, failure to satisfy this condition would eliminate the indeterminacy of the currents. However, practical experience has shown that the currents may still be unduly dependent on circumstances: one capacitor voltage may be, say, 5 times the other. This is, of course, highly undesirable, since it would necessitate over-dimensioning the capacitors and coils to a very considerable extent.

This can be avoided in several ways. From formula (12), assuming that formula (5) does not hold good, the currents in the coils are theoretically equal and 180° out of phase, and it is only by chance that in reality these currents usually differ.

Identical currents can be procured by placing the two coils on a common iron core (*fig. 4a*). A variant of this method is to couple the coils by means of a 1 : 1 transformer, thus maintaining the capacitor voltages the same (fig. 4*b*).

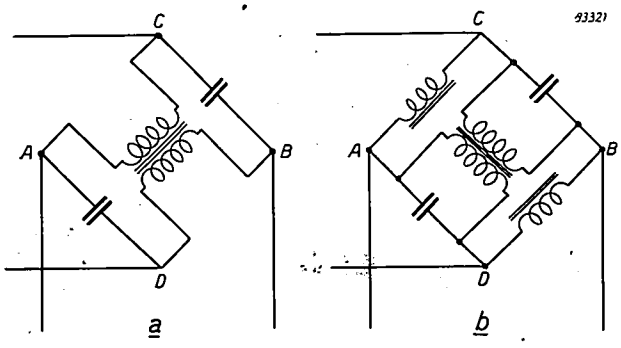


Fig. 4. The indeterminacy of the four arm currents in the Boucherot bridge can be eliminated by fitting the two coils on one core (*a*), or by coupling the capacitors across a 1 : 1 transformer (*b*).

Another method — employed in Philips equipment — involves a deliberate deviation from condition (5), by detuning slightly, apart from the inevitable deviation caused by the losses. Not only does this method enable the indeterminacy of the arm currents, which persists despite losses, to be avoided, but it also precludes all possibility that a simultaneous failure of several lamps will cause the remainder to be overloaded (a possibility to which we shall refer again in the section on lamp transformers). The permissible degree of detuning depends of course, on the extent to which it detracts from the constancy of the output current. In view of this, the calculation of the output current will now be repeated, on the assumption that neither aspect of condition (5) is fulfilled.

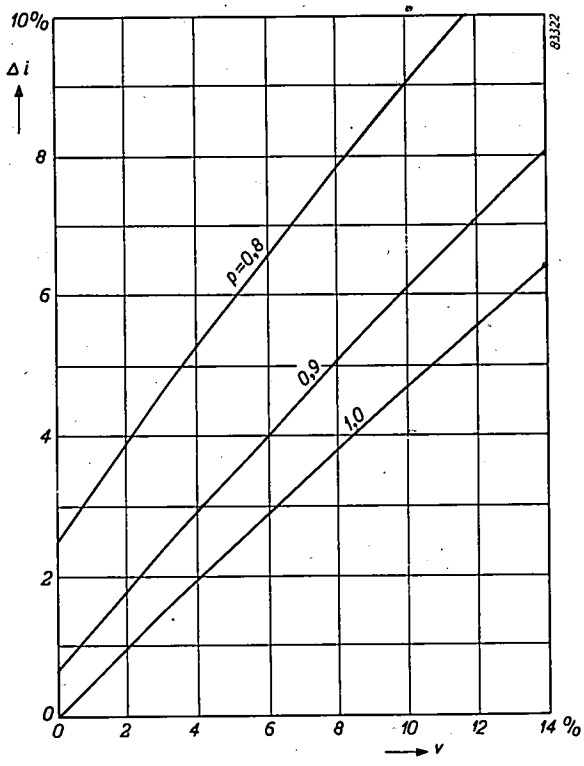


Fig. 5. The relative current difference Δi plotted against the relative loss v in the Boucherot bridge, for various values of the detuning p. Almost identical curves hold good for the reciprocal values of p.

The general equation defining the output I is:

$$I = \frac{Z_2 - Z_1}{ZZ_1 + ZZ_2 + 2Z_1Z_2} E_0 \quad . \quad . \quad . \quad (13)$$

The load impedance Z, formed by the series system of lamp transformers each loaded with an incandescent lamp [4], is for all practical purposes a pure resistance (R). Now, if Z_2 is made up of a coil with self-inductance L and a resistance r, and Z_1 of a capacitor C, assumed to be free of losses, and taking ω as the angular frequency of the mains, we have:

$$Z = R, \qquad Z_1 = \frac{1}{j\omega C}, \qquad Z_2 = r + j\omega L.$$

Next, we introduce two parameters, the one, $v = r/R$, referring to the losses, and the other, $p = \omega/\omega_0$ (where $\omega_0^2 = 1/LC$), to the detuning, so that (13)

[4]) The case of unloaded transformers (defective lamps) will be discussed later.

may be written:

$$I = \frac{E_0}{R} f(p,v).$$

The current reaches a maximum, I_{max}, in the event of a short-circuit ($R = 0$, $v = \infty$). The relative current difference, $\Delta i = (I_{max} - I)/I$, is here adopted as a measure of the deviation from constant current output.

Fig. 5 shows Δi plotted against v for various values of p. It is seen from this diagram that the increase in Δi produced by detuning from $p = 1$ (resonance) to $p = 0.9$ or 1.1 will be small, provided that v is not unduly large. For $v = 6\%$, for example, Δi will increase from 3% to 4%, but any further detuning will cause a sharp rise in Δi.

Now, detuning to $p = 0.9$ is enough to reduce the indeterminacy of the arm currents appreciably, particularly if combined with one of the methods illustrated in fig. 4.

Description of a practical supply unit

On the basis of the theoretical arguments so far considered, Philips have designed a series of supply units ranging from 1.5 to 25 kVA. *Fig. 6* is a simplified circuit diagram. Here, *ACBDA* is Boucherot's bridge; the input diagonal, *AB*, is connected to

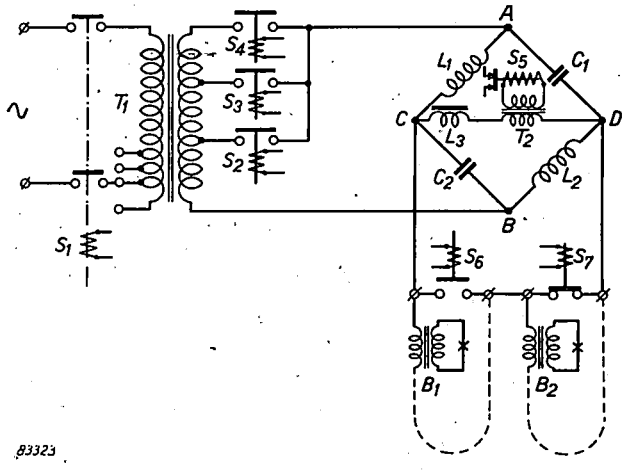


Fig. 6. Simplified circuit diagram of a low power supply unit. S_1 main switch; T_1 isolating transformer; S_2, S_3, S_4 luminous intensity selectors (varying in number from 0 to 5); *ABCD* Boucherot bridge; L_3 choke to limit the output voltage when the runway circuit is interrupted; T_2 current transformer; S_5 relay, which, when energised, opens mains switch S_1; B_1, B_2 runway light circuits; S_6, S_7 runway selectors.

the H.T. winding of a transformer, T_1, whose primary is supplied from the mains. The primary coil is provided with tappings, to enable adjustments to be made according to the average local mains voltage (it is seen from formula (6) that I varies in proportion to the supply voltage (E_0) of the bridge).

It is usual to make the luminous intensity of the installation variable in a number of steps; T_1 must therefore also be provided with tappings on the secondary. The bridge is connected at will to any of the secondary tappings by means of magnetic switches — so-called "luminous intensity selectors" — S_2, S_3 and S_4. It will be evident that only one selector may be energised at any given moment, since otherwise transformer T_1 would be short-circuited; hence the selectors are interlocked.

The output terminals of the bridge, *C* and *D*, may be connected to the circuit of runway lights either direct, that is, if the voltage across the series circuit is less than 1500 V, or via a transformer (not shown in the diagram), if a higher voltage is employed. It is often possible to operate two or more circuits alternately with one supply unit to light, say, the two directions of a single runway, or two alternative runways. These circuits are then connected in series and shunted separately by magnetic switches, known as runway selectors (fig. 6 shows two circuits and therefore includes two runway selectors, S_6 and S_7). On runways the circuits must operate one at a time; hence the runway selectors must also be interlocked. On taxiways, on the other hand, the lights must often be able to operate simultaneously in various combinations.

Interlock system

Various methods of interlocking magnetic switches are known; most of them involve auxiliary contacts designed to interrupt the connections to the switch energising coils. Where a large number of switches are to be interlocked, the number of auxiliary contacts required will be very large, and the risk of breakdowns therefore considerable.

In view of this risk, we have developed a system in which a high degree of reliability in service is ensured by minimizing the number of auxiliary contacts. It is based on the principle that the impedance of the energising coil of a magnetic switch is very much higher when the switch is closed than when it is open. It will be seen from the diagram (*fig. 7*) that one end of each energising coil is connected to a particular point *P*, which, in turn, is connected, via a capacitor (C_1) and a coil (L) in series, to one pole (*1*) of the mains supplying the control voltage. The other end of each coil can be connected to pole *2* of the mains via the push button (D_1, D_2, ...) appropriate to the particular switch. The ordinary change-over contacts, each in parallel with the associated push button, are indicated by *OC*, and the auxiliary contacts, to shunt capacitor C_1 across a low resistance R (almost a short-circuit), by *CC*. The series circuit C_1-L is resonant at the mains frequency; hence its impedance is low.

Now, let us assume that button D_1 is pressed. The coil of switch S_1 is then energised normally, so that it closes the main contact of S_1 (not shown in the diagram), and transfer contact *OC* and auxiliary contact *CC*. With the closing of *CC*, capacitor C_1 is "shorted" across resistance R (acting as a damping resistor to suppress the contact spark). For practical purposes, only coil L is then left between points *P* and *1*. (Owing to the

voltage drop in L, the voltage across the coil of S_1 would then be about 10% lower than the control voltage. In order to prevent S_1 from opening, even in the most unfavourable circumstances, a special capacitor C_2 is provided which, together with the coil of S_1, forms an impedance higher than that of the coil alone.)

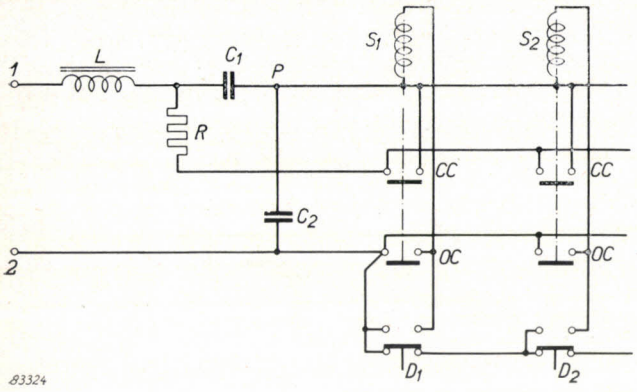


Fig. 7. Interlock mechanism to preclude the possibility that two or more switches (with energising coils S_1, S_2 .....) will be closed at the same time. *1*, *2* terminals of the mains supplying the control voltage; L-C_1 resonant circuit; R damping resistor; CC short-circuit contacts of C_1; OC transfer contacts C_2 auxiliary capacitor; D_1, D_2 ..... push buttons.

Next, button D_2 is pressed. This brings the coil of S_2 in parallel with that of S_1, and the two coils together in series with L. Since switch S_2 is still open, the impedance of its coil is low, and most of the control voltage is therefore dropped across L; hence the coil of S_2 is not sufficiently energised to close this switch, and the energising current of S_1 is so weakened that it opens. However, the shunt across C_1 is then interrupted; the impedance of circuit L-C_1 decreases and the energising current in S_2 becomes strong enough to close it.

Any number of switches may be employed, but it is impossible to energise more than one of them at a time. The pressing of any one of the buttons opens the switch already closed, and closes the one controlled by the particular button. As will be seen from the oscillogram shown in *fig. 8*, there is an interval of $3\frac{1}{2}$ cycles between the opening of the one switch and the closing of the other.

To preclude all possibility that two or more switches will be closed at the same time (owing to two or more buttons being pressed simultaneously), the push buttons are connected in series in the usual way.

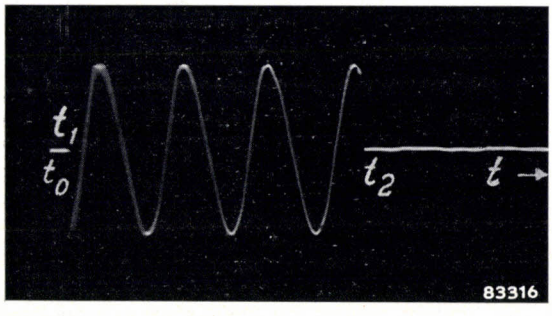


Fig. 8. Oscillogram referring to fig. 7. At $t = t_0$, the push button of S_2 is pressed, with S_1 already closed. After a certain delay, depending upon the design of S_1, the contacts of S_1 open, that is at $t = t_1$. A 50 c/s voltage then appears on the oscilloscope; this voltage is suppressed when S_2 closes (at $t = t_2$). The interval between t_1 and t_2 is $3\frac{1}{2}$ cycles.

A closer analysis shows that it is advisable to detune the L-C_1 circuit slightly with respect to the mains frequency, since this reduces the adverse effect of the control cable resistance considerably. In fact, it enables this effect to be made much smaller than in ordinary interlock circuits, so that it is possible to employ relatively longer and/or thinner cables for the control current.

Limiting the output voltage

A break in the output circuit, say, owing to the cable being damaged, would cause a considerable increase in voltage. To avoid this, an auxiliary circuit comprising a choke (L_3) and a current transformer (T_2) is connected between output terminals C and D (fig. 6). Coil L_3 is so constructed that the current produced in it by a normal voltage between C and is negligible. However, if the voltage increases, owing to a break in the external circuit, the core of L_3 becomes saturated, thus limiting the voltage to less than twice its nominal value (*fig. 9*), despite the fact that the auxiliary circuit then carries the full current.

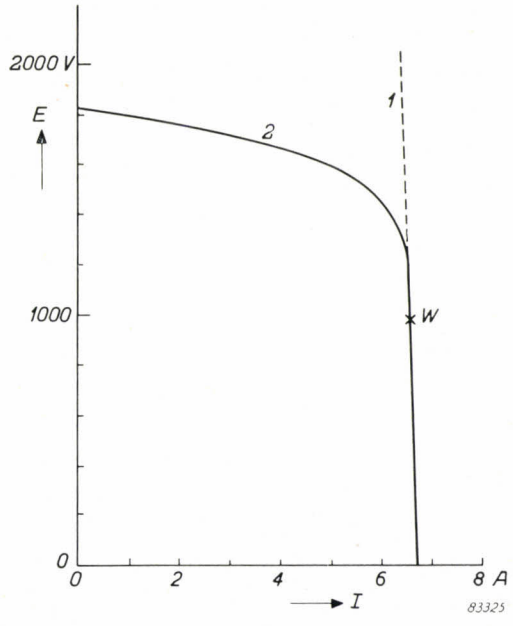


Fig. 9. The output voltage A plotted against the current in the external circuit I, (*1*) without limiter L_3-T_2 (fig. 6), (*2*) with the limiter in circuit. W full-load working point.

This current then energises a relay (S_5) via a current transformer. The break contact of this relay, in series with the energising coil of mains switch S_1 (fig. 6), therefore opens; thus, any break in the output circuit isolates the entire supply unit from the mains.

Mechanical design

All the components of the supply unit are accommodated in a single cabinet. At the front of this

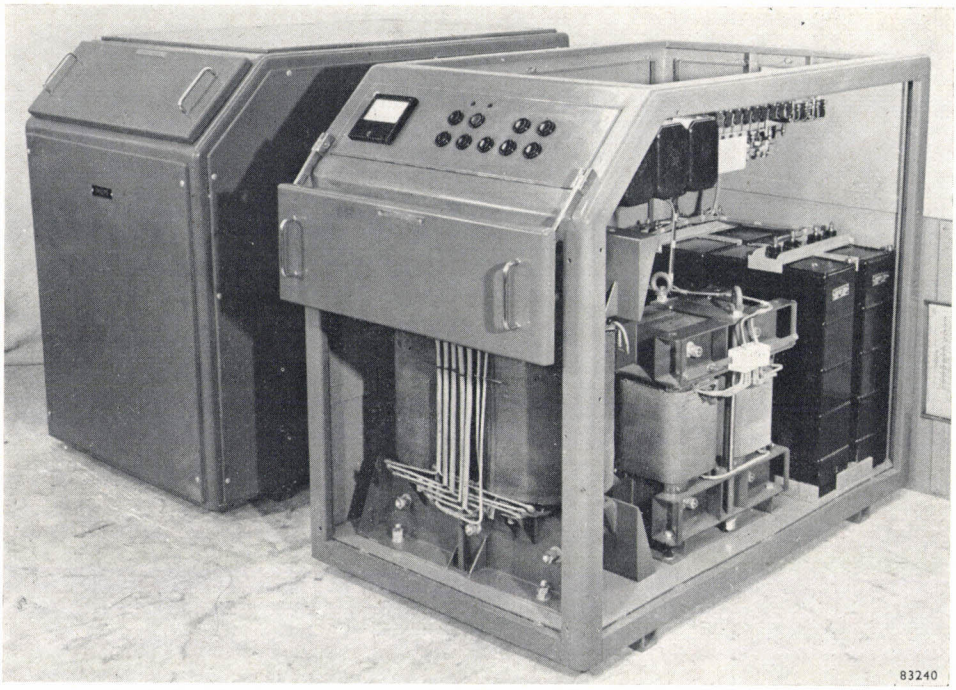


Fig. 10. Two 25 kVA, 20 A supply units, the one on the right with the cover open and the front and sides removed. Note the ammeter and the push buttons on the sloping panel. The isolating transformer, one of the bridge-arm coils and the capacitors can be seen at the bottom, and the magnetic switches at the top of the unit.

cabinet is a control panel (*fig. 10*) with an ammeter to indicate the output current, and push buttons to operate the mains switch, luminous intensity selectors and runway selectors.

The panel is provided with a cover fitted with a door contact to close the remote control circuit; accordingly, local control is possible only with the cover open, and remote control only with the cover closed.

Remote control of the different light systems on an airfield usually operates from the control tower which is then equipped with a panel or desk on

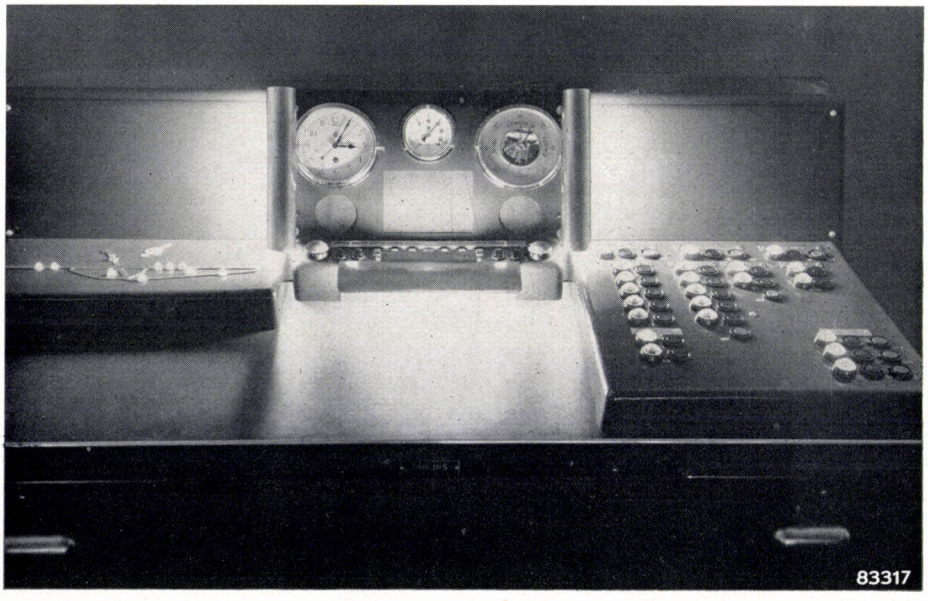


Fig. 11. Control desk. Left: plan of airfield with monitor lamps at points corresponding to the positions of the real runway lights. Right: push buttons and signal lamps of the main switches and the runway and luminous intensity selectors. Centre: panel with clock and meteorological instruments. The spaces on either side are for telecommunications equipment.

which all the light controls are conveniently grouped. Monitor lamps enable the operator to see at a glance which of the systems are in use, and with what luminous intensity, at any given moment. The efficiency of this arrangement can be enhanced by mounting the monitor lamps on a plan of the airfield at points corresponding to the positions of the real lights.

A control desk so arranged is shown in fig. *11*.

The lamp transformers

As stated in the introduction, each lamp in a series system is connected across a separate transformer. The electrical properties of these transformers are highly important if full advantage is to be taken of the special features of the series system, and the reliability of the installation as a whole depends to a very large extent upon their mechanical design.

Electrical characteristics

As regards the electrical properties, firstly, the current in the lamp must remain as far as possible constant, despite a certain amount of spread in the lamp voltage as between individual lamps of the same type and the possibility of contact resistances in series with the lamps. Accordingly, the transformers must possess the characteristic typical of current transformers.

Secondly, the voltage across the lampholder must not build up unduly when a lamp fails or is removed from its holder; otherwise, the replacing of a lamp would involve a certain amount of danger.

Both these requirements are fulfilled by a transformer having the characteristic shown in *fig. 12* (secondary voltage plotted against secondary current). The curve is so steep at the working point that a 10% additional resistance in the secondary circuit reduces the current by only about 0.5%; moreover, the no-load voltage is limited to 3 times the normal operating voltage, this being accomplished by ensuring that the iron transformer-core is completely saturated when there is no load.

However, this saturation is associated with distortion of the primary current; hence the output current of the supply unit is distorted whenever the load is withdrawn from one or more of the lamp transformers. Experience has shown that this increases the r.m.s. value of the current, which is, of course, undesirable, in that, if several lamps happen to fail in a relatively short time, the current in the circuit may increase considerably and so impose a heavy overload on the remaining lamps. Although at civil airports, where the lamps are inspected regularly and defective ones are quickly replaced, this effect is not very important, on military airfields the possibility that many lamps will fail simultaneously, say as a result of enemy action, must certainly be taken into account. Here, then, the authorities will probably require that the r.m.s. current in the series circuit be made as far as possible independent of the percentage of defective lamps. This can be accomplished by detuning the bridge circuit slightly (i.e. by allowing p to deviate slightly from 1), a method which, as we have already seen, is also effective in eliminating the indeterminacy of the arm currents. However, good matching of the supply units to the lamp transformers is then essential, since the proper detuning of the bridge depends upon the transformer characteristic.

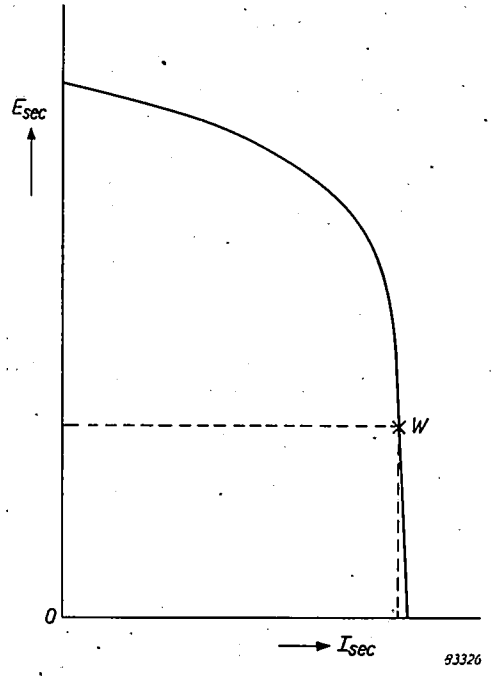


Fig. 12. The secondary voltage (E_{sec}) plotted against the secondary current (I_{sec}) of a lamp transformer loaded with a variable resistor. The primary winding of this transformer is included in a series circuit supplied by a Boucherot bridge. The curve should be steep at the working point W, and the voltage at $I_{sec} = 0$ should not be unduly high.

Mechanical construction

The lamp transformers are designed to operate with lamps ranging from 30 to 500 W, and vary in design according to the particular purpose.

Transformers of the simplest type may be employed for building into runway lights or mounting in a closed box above ground level. The transformer, completely enveloped in moisture-repelling compound, is then placed in a metal box provided with ceramic terminal insulators (*fig. 13*).

Fig. 13. Three transformers of the simplest type mounted in a flush runway light (as shown in fig. 21 of the article referred to in note [1]). The cover, only the bottom of which is seen in the photograph, contains one omni-directional top light and two beam lights (one for each landing direction).

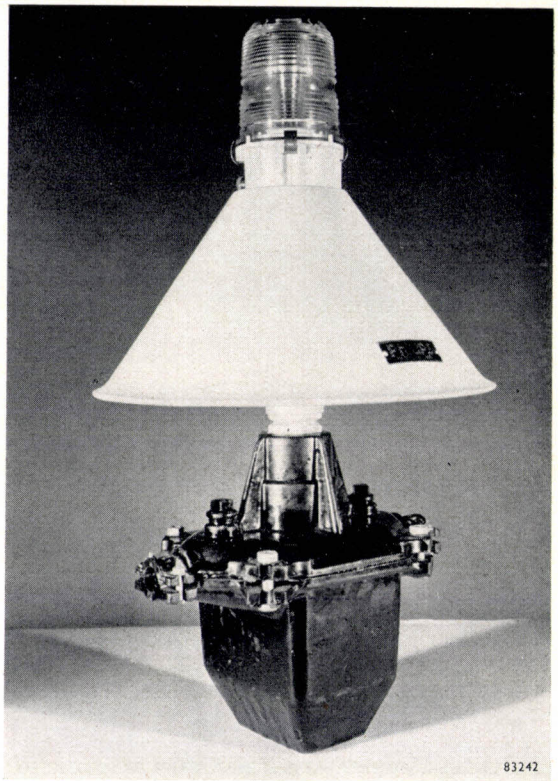


Fig. 14. Taxiway light, mounted on a transformer with cast iron case filled with compound. Note bushings for connection to armoured cables.

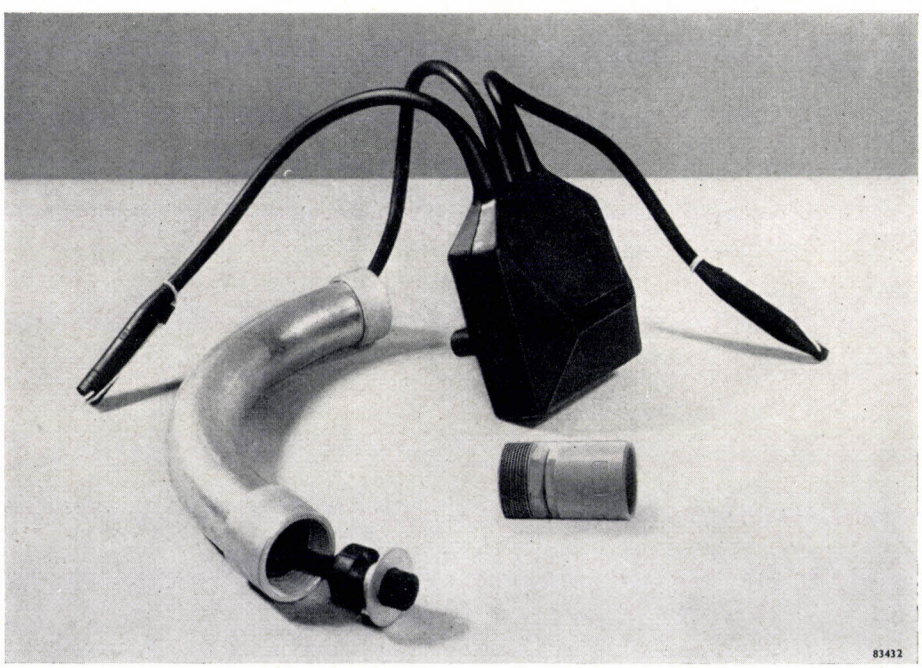


Fig. 15. 200 W, 6.6 A lamp transformer for circuits operating with 5000 V supply. It is completely enclosed in rubber, moulded in one piece with the insulation of the leads. The secondary lead (passing through the curved pipe) is fitted with a lamp socket. If the light is knocked over, the safety joint, that is, the small weakened pipe seen in front of the transformer, breaks (see page 282 of the article referred to in note [1]) and the plug is pulled out of the secondary socket. The leads on the extreme right and left are the primary leads. Note the end sleeves employed to protect them from corrosion in storage or in transit.

However, the requirements imposed on transformers to be mounted in pits close to the runway lights, or completely buried, are more stringent. The covering must then resist the action of corrosive elements in the ground; hence the transformer is placed in a heavy, cast-iron box with a removable cover, and the outside of this box is coated with a protective asphalt compound. The transformer itself is delivered from the factory completely enveloped in compound, that is, leaving only the porcelain terminal insulators exposed. When it is installed on the airfield, the cables are connected, the cover is secured, and the remaining space inside the box is filled with compound. When non-armoured cable insulated with a synthetic rubber such as "Neoprene" is employed, the transformer box is provided with coupling nut bushings; for armoured cables, bushings of the type usually employed with cable sleeves are fitted (*fig. 14*).

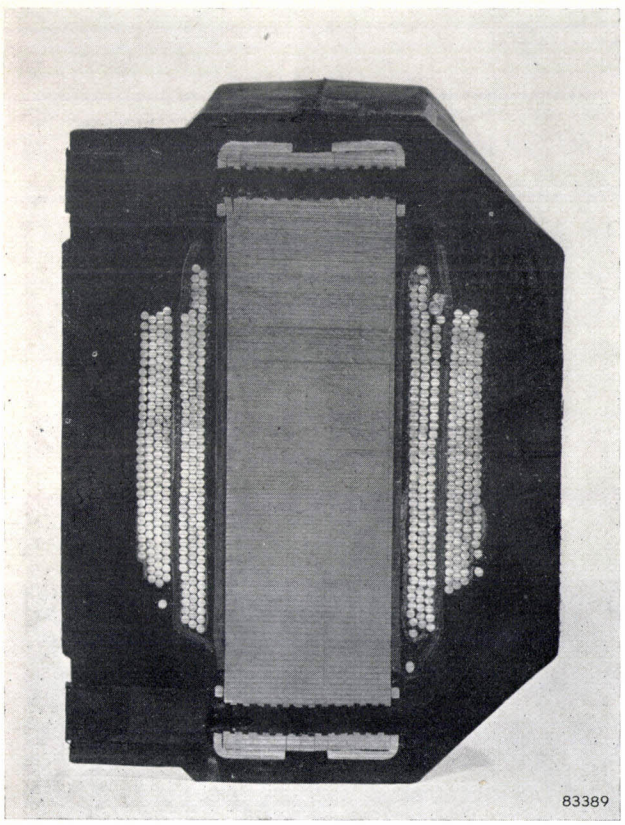


Fig. 16. Cross-section of the rubber-cased transformer shown in fig. 15. It will be seen that the turns are completely enclosed in rubber, which has penetrated into the spaces between them.

Another transformer for use underground, recently developed in America, is hermetically sealed in moulded rubber [5]). Transformers of a similar type are marketed by Philips (*fig. 15*). Here, the waterproof rubber cover is moulded in one piece with the rubber insulation of the leads, which are short lengths of cable with plug sockets vulcanised on to the ends. The transformer is thus protected against corrosion of any kind, and nevertheless convenient to handle. In fact, the mounting of such transformers is so quick and simple an operation that it can be entrusted to unskilled personnel. *Fig. 16* shows a cross-section of the transformer.

The quality of this protective covering is illustrated by the result of a special test to which it was subjected. This test consisted in raising the transformer daily to the working temperature and then cooling it abruptly by plunging it in water. Although this very stringent test was continued for several weeks, insulation resistances of the order of thousands of megohms were afterwards measured at a test voltage 3 times the normal operating voltage of the transformer.

[5]) L. C. Vipond, No drinks for transformers, Aviation Age **17**, 56, November 1952.

Summary. For airfield light beacons incandescent lamps of voltages ranging from 6 to 30 V, each connected to a separate transformer, are employed. In the present article it is shown that such transformers are better connected in series than in parallel, since in a series circuit the voltage drop in the supply cables causes no variation in luminous intensity between the individual lamps connected to the transformers; hence it is possible to employ very much thinner cables and so considerably cut down on cable costs, which constitute an important part of the total capital investment. Another advantage of the series system is that it is more reliable (effect of contact resistances less noticeable; separate lamp fuses may be dispensed with).

The Boucherot bridge circuit is a suitable current source for a series system. It has two convenient features; the output current does not depend upon the nature or size of the load impedance, and with resistive loading the power factor is unity.

However, the indeterminacy of the individual currents in the four bridge arms is a disadvantage. Measures to eliminate this include slight detuning of the bridge.

Philips employ the Boucherot bridge in a series of airfield light-beacon supply units ranging from 1.5 to 25 kVA. Certain features of these units, including the method of limiting the output voltage when the external circuit is interrupted, and a new interlock system, are described.

Finally, the lamp transformers are discussed. The characteristic of such a transformer must precisely match that of the associated supply unit. One type of transformer, insulated entirely with rubber, is specially good for its corrosion resistance.

ELECTRONIC CONTROL OF INDUSTRIAL PROCESSES

by H. J. ROOSDORP.

621-523.8:621.317.733.083.4.078

Referring to a previous article giving a general review of regulating systems for industrial processes, we shall deal here with the various forms of electronic controllers and their advantages with respect to non-electronic control systems.

Introduction

In many industrial processes the quality of the final products is dependent on various physical or chemical quantities, e.g. temperature, pressure, acidity, etc. In such processes it is necessary that these quantities are adjusted to the most favourable value and that this adjustment is maintained by checking and controlling the production process in a suitable manner. This can be done either manually by an operator or automatically, without human intervention. In the former case the quantity to be regulated has to be measured and read from some scale before it can be adjusted to its correct value. Strictly speaking such measurements — or at least such readings — are not necessary when the process is automatically controlled. If the installation is functioning properly, the value of the controlled quantity should remain practically equal to what is desired. In fact, there are a great number of such non-indicating controllers, which operate without showing the value of the controlled quantity. In most cases, however, indication of the controlled quantity is desirable, since it gives a continuous check on the progress of the operation. Indeed it is often convenient to have a record of the controlled quantity as a function of the time; recording controllers are therefore frequently used.

An electrical instrument (automatic potentiometer) of robust construction, for measurement and recording in industrial processes has been described earlier in this Review [1]). The present article deals with some of the ways in which this instrument can be used for automatic control. For a study of the general aspects of automatic control we may refer back to another article in this Review [2]). For the sake of simplicity, the present article is mainly concerned with one specific process, viz. maintaining the temperature of a gas-fired furnace. The discussion, however, is equally valid for entirely different processes.

In the automatic control of furnace temperature a "closed loop" is formed by the furnace, the temperature-measuring instrument, the controlling unit and the correcting element (in this case the valve regulating the gas supply). In most cases an actuating mechanism of some sort will be required between the automatic controller and the valve, e.g. an electric motor or an electro-pneumatic valve positioner. *Fig. 1* gives a schematic representation of the closed loop thus obtained. The controlling

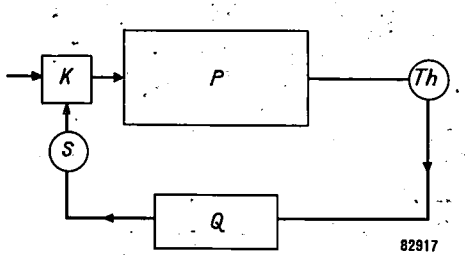


Fig. 1. Diagram of a closed loop automatic control system. The process to be controlled is represented by *P*. *Th* measuring element; *Q* controller; *S* motor element; *K* correcting element.

unit *Q* establishes the functional relationship between the temperature of the measuring element *Th* and the position of the gas valve *K*, i.e. the "control action" of the system. A controller of this kind may be mechanically, electrically, hydraulically or pneumatically operated. Each of these systems has its specific merits and drawbacks. Electric and more particularly, electronic systems are finding increasing application. This is essentially due to the following reasons. Electricity is almost universally available, so that no other sources of power have to be installed (e.g. compressed air supply). Furthermore, most measurements are in any case carried out electrically, so that it is often simple to extend the measuring circuit with a

[1]) H. J. Roosdorp, An automatic recording potentiometer for industrial use, Philips tech. Rev. **15**, 189-198, 1953/54.

[2]) H. J. Roosdorp, On the regulation of industrial process, Philips tech. Rev. **12**, 221-227, 1950/51.
See also, e.g., W. Oppelt, Kleines Handbuch technischer Regelvoränge, Verlag Chemie, Weinheim, 1954; and D.P. Eckman, Principles of Industrial Process control, John Wiley, New York 1945.

circuit for electrical control. Transmission of electric signals is virtually instantaneous, which may be important if large distances (e.g. in excess of 100 m) have to be bridged. The elements used in electric circuits (mainly resistors and capacitors) to obtain the required control action have substantially constant values, so that the functioning of an electric controlling circuit is restricted to linear operations involving simple mathematics. Moreover, these circuit elements are standard radio components, everywhere obtainable and easy to handle. The design of the amplifiers used in these circuits may draw on the experience gained in telecommunications. For these and other reasons, the control action with an electronic controller is achieved in a relatively simple manner. As a final argument it should be mentioned that the maintenance and testing of electric and electronic instruments and, if necessary, the replacement of parts, are usually simple routine work.

Sometimes it is desirable that the position of the valve depends not only on the deviation of the measured quantity from the desired value, but also on the time for which this deviation has existed, or on the rate at which the deviation increases or decreases (integral or derivative control respectively, cf. [2])). Futhermore it may be necessary that the desired value of the controlled condition changes as a function of time (programme control). All these requirements can be satisfied by electronic controllers in a simple way.

If no work of a mechanical nature is required of the correcting element, a wholly electric system can be used. An example of this is the regulation of an electric furnace with the aid of transductors or gas-filled tubes. If, however, as in the case considered here, the position of a mechanical valve has to be altered, a motor element (valve motor S in fig. 1) has to be used, which may operate electro-mechanically, pneumatically or hydraulically.

In the following we shall not be concerned with the nature of this mechanism, but only with the electrical part of the controlling unit.

Automatic control by means of the automatic potentiometer

The working principle of the automatic potentiometer/bridge is as follows: a contact fitted on a carriage is moved by a motor along a resistance slidewire incorporated in the measuring circuit, the motor adjusting this contact to a position corresponding to the value of the measured quantity. For automatic control, a voltage is required whose amplitude and polarity corresponds to the deviation of the measured quantity (in this case the furnace temperature) from the desired value. Such a voltage can be obtained by means of a second slidewire, equal in length to the measuring slidewire and with a movable contact fixed to the same carriage as that on the latter. On the controlling slidewire, a second contact is fitted which can be set by hand to a position corresponding to the desired value of the controlled condition. With a voltage E_1 applied across the controlling slidewire, there is a voltage E_x between its two contacts (*1* and *2* in *fig. 2*) that is proportional to the difference between the actual furnace temperature and the desired furnace temperature. This voltage can be used to determine, via an actuating mechanism, the position of the gas valve.

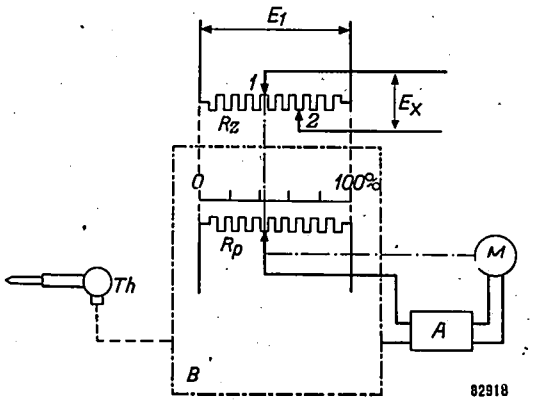


Fig. 2. Use of the automatic potentiometer for control of the temperature of a furnace. B measuring circuit; Th thermo-element; R_p measuring slidewire; A amplifier; M drive motor; R_z control slidewire. The contact *1* is mounted on the same carriage as the contact of the measuring slidewire. Contact *2* is set to a position corresponding to the desired temperature.

In the following we shall deal with some circuits for electronic controllers and what can be achieved with them in industrial processes.

Two-step control

One of the simplest systems of control is that in which the valve or correcting element has two positions only, viz. a position A, which, if maintained, would raise the furnace temperature T to above the desired value T_0, and position B which would cause the temperature to drop below T_0. The valve has to be in position B if T is higher than T_0 and in position A if T is lower than T_0. Each time the desired value is passed, the valve changes its position, so that the temperature continues to oscillate between two values, one higher and one lower than T_0.

This can be effected with the circuit shown in *fig. 3*. The position of the contact *2* of the controlling slidewire R_z corresponds to the desired value of the temperature; that of contact *1* to the actual

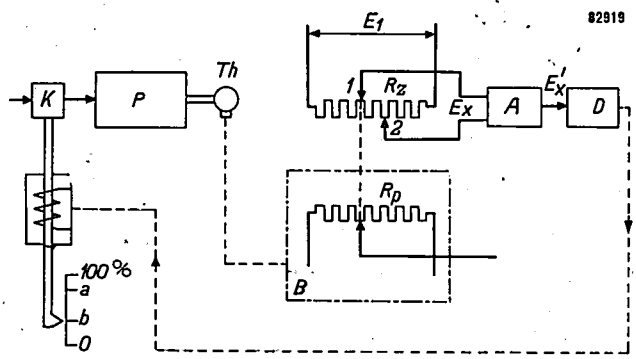


Fig. 3. Circuit for two-step control of the temperature of a furnace. *A* amplifier; *D* phase-sensitive detector. The other letters have the same significance as in figs. 1 and 2.

value of the temperature. An alternating voltage E_1 of mains frequency, is applied to the controlling slidewire. The voltage E_x between the contacts *1* and *2* (the error voltage), after being raised to a value E_x' by the amplifier *A*, is applied to a phase sensitive detector *D*. The circuit diagram of the latter is shown in *fig. 4*. The alternating voltage E_x' is added to an alternating voltage E_k and applied to the grid of a triode. An alternating voltage E_a is applied to the anode of this tube. E_a, E_k and E_x' have, of course, the same frequency. E_k and E_a are in anti-phase and have such values that, when E_x becomes zero, anode current ceases to flow through the triode. E_x' and E_a are either in phase or in anti-phase, according to whether the contact *1* is to the left or to the right of contact *2*. If both voltages are in phase, anode current flows through the tube during each half-period and the armature of relay *Re* is attracted.

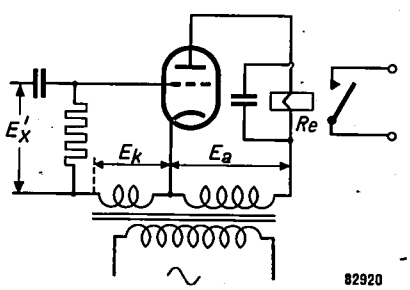


Fig. 4. Diagram of a phase sensitive detector. *Re* Relay.

If on the other hand E_x' is in anti-phase with E_a, the tube passes no anode current and the relay is not energised. Whether or not the relay is energised is thus an indication of the direction in which the temperature deviates from the desired value. This relay operates the correcting element, e.g. a gas valve, via a motor element and sets it to either position *A* or position *B*. The circuit is preferably so arranged that the releasing of the armature of *Re* corresponds to a safe position in the process to be regulated; in the present example, therefore, it corresponds to the closed position of the gas valve. Should the controller break down, so that E_x' remains zero, there is then no chance of the temperature rising to a dangerous value.

The positions *A* and *B* of the gas valve correspond to a % and b % respectively ($a > b$) of the maximum opening y_{max} of the valve. The inherent temperature oscillations are smaller according as the difference between a and b is less. In practice, the extent to which this difference can be reduced is restricted by the fact that at $y = y_{max}\, a/100$ the temperature must be able to rise above T_0 under all circumstances, that is to say, even under the most adverse conditions, e.g. the lowest ambient temperature and the lowest gas pressure. Conversely, at $y = y_{max}\, a/100$ the furnace temperature must be able to drop below T_0 even at the highest ambient temperature and the highest gas pressure. The neccessary difference between a and b is therefore, as a rule, dictated by local conditions. In a special case a can be 100 % and b 0 %. Such a system is termed on-off control. A system like this obviously involves fairly large oscillations in the temperature, but it has the advantage that in many cases the construction is relatively simple.

The magnitude of the oscillations may be limited by selecting the most suitable position of the measuring element with regard to the source and dissipation of the heat. If the measuring element reacts (via the furnace temperature) to the position of the gas valve with only a slight time lag, excessive fluctuations may be avoided.

Multi-step control

If for a given process the external conditions (ambient temperature, gas pressure etc.) vary to such an extent that a two-step control would necessitate an excessively large differential between a and b, the following system can be used. The differential between a and b is made just large enough to ensure a properly functioning control for small fluctuations of the external conditions. A third setting *C* of the valve is provided to take care of those situations in which the small differential between a and b cannot cope with the wide variation in external conditions. If, for example,

the ambient temperature, were so high that, even with the valve in the B position, the temperature continued to increase, the third position C of the valve would be set at a more closed position than B, i.e. $c<b$. Thus, if the temperature should exceed a certain value T_1 ($>T_0$) the controller causes the valve to move to the c position, where it stays until the temperature has dropped below T_1. The setting c is of course fixed at such a value that under all condition, the temperature T can drop below T_1 when $y = y_{max}\, a/100$. The net result is that at a high ambient temperature, control will take place about T_1, just as with two-step control. If the ambient temperature falls, control again takes place about T_0. This is known as three-step control. *Fig. 5* is a sketch showing how such a system can

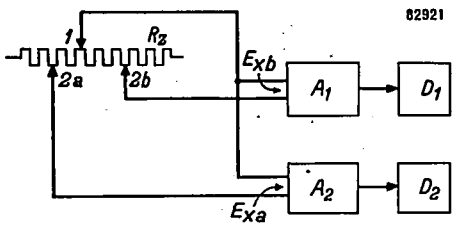


Fig. 5. Part of a circuit for three-step control. R_z controlling slidewire; A_1 and A_2 amplifiers; D_1 and D_2 phase sensitive detectors. Contact *1* moves with the contact of the measuring slidewire. Contacts *2a* and *2b* are adjusted by hand.

be realized. The controlling slidewire R_z is now provided with two manually adjustable contacts, *2a* and *2b*. The contact *1* is again moved together with the contact of the measuring slidewire. The two voltages E_{xa} and E_{xb} are applied, via amplifiers, to separate phase detectors D_1 and D_2. The position of contact *1* determines whether both the phase detector relays are open, both closed or one open and one closed. These three possibilities determine, via a motor element, the three positions of the gas valve. *Fig. 6* shows the effect of a three-step control on the furnace temperature T as a function of the time t when the ambient temperature is rising. Initially T oscillates regularly about the value T_0, but after some time it rises and oscillates about T_1. The three-step control discussed is the simplest case of a multi-step control. Clearly this system may be extended in principle to a control system with any number of steps. With multi-step controls, oscillations in the furnace temperature due to large changes in external conditions can be kept reasonably small.

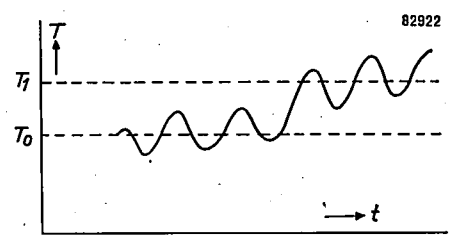


Fig. 6. The temperature T of a furnace as a function of the time t for a system of three-step control, the external conditions being such that the furnace temperature has a tendency to rise continuously.

Proportional control

Both two-step control and multi-step control are examples of discontinuous control systems. In these systems the gas valve (or any other correcting element) can be set only to a number of discrete positions. This limitation does not exist in continuous control systems. One such continuous control is proportional control. This method may be considered as an extreme case of a multi-step control system with an infinite number of steps at infinitely small intervals. The displacement of the correcting element is then proportional to the deviation of the furnace temperature from the desired value. (The time required for the movement of the correcting element is disregarded here.)

Proportional control could be realized electrically with a gas valve operated by a solenoid to which a voltage is applied proportional to the difference between the actual and the desired temperature. This construction, however, has the drawback that the relationship between the valve opening and the voltage applied to the solenoid is likely to be disturbed by forces acting upon the valve (e.g. frictional forces). A further drawback is that considerable power must be expended merely to maintain the position of the valve.

A system less subject to these drawbacks is shown schematically in *fig. 7*. Here again we have the controlling slidewire, R_z, supplied with a voltage E_1. The contact *1* of R_z is moved by the carriage of the measuring slidewire (not shown in the diagram). The voltage E_x tapped off on this slidewire is added to a second voltage E_y, which is obtained from the two parallel slidewires R_y and R_c, supplied with a voltage E_2. The contact *4* of R_y is mechanically linked to the correcting element (gas valve). Contact *3* of R_c can be set by hand. E_1 and E_2 are alternating voltages derived from the mains via a transformer.

The difference between the two voltages E_x and E_y is applied to the amplifier A_1. The output voltage of the latter controls the motor M_1 of the valve in such a way, that owing to the displacement of the

contact *4*, the voltage difference $E_x - E_y$ becomes zero. Since (see fig. 7)

$$E_x = (x - x_0)\, E_1$$

and

$$E_y = (y - y_c)\, E_2,$$

it follows that

$$y - y_c = \frac{E_1}{E_2}\,(x - x_0) \quad . \quad . \quad . \quad . \quad . \quad (1)$$

The displacement of the gas valve with respect to the position corresponding to $y = y_c$ is thus proportional to the distance between contact *1* and *2*. Since there is usually a linear relationship between the furnace temperature and the displacement of contact *1*, proportionality between the displacement of the valve and the temperature deviation of the furnace is thus achieved. The proportional control factor E_1/E_2 can be varied by the choice of E_1 and E_2.

In general, x can reach the value x_0 only when y_c is made equal to the valve position y. This condition, however, is possible only if the external conditions are such that for $y = y_c$ the furnace temperature does indeed become equal to the desired value. This can be realized by adjusting y_c by means of contact *3*. If, after this, any change occurs in the process to be controlled or in the external conditions, another position of contact *3* would be necessary in order to make $x = x_0$ when $y = y_c$. If, however, the contact is left at its original setting, then x will differ from x_0 by an amount which remains small so long as the proportional control factor has a large value.

In practice it is not possible to make the proportional control factor arbitrarily large. Because of the inevitable time lag that always exists between a displacement of the gas valve and its complete effect on the furnace temperature, too great a value of the proportional factor may create an unstable condition in which the furnace temperature is subject to large oscillations [3]). This restricts the choice of the proportional control factor. The large drift from the desired value which can occur if the factor is too small can, of course, be corrected by adjusting contact 3, but this may be most inconvenient if it has to be done too frequently. By adding to the proportional control a circuit for integral control, adjustment of contact *3* is no longer necessary, as will be demonstrated presently.

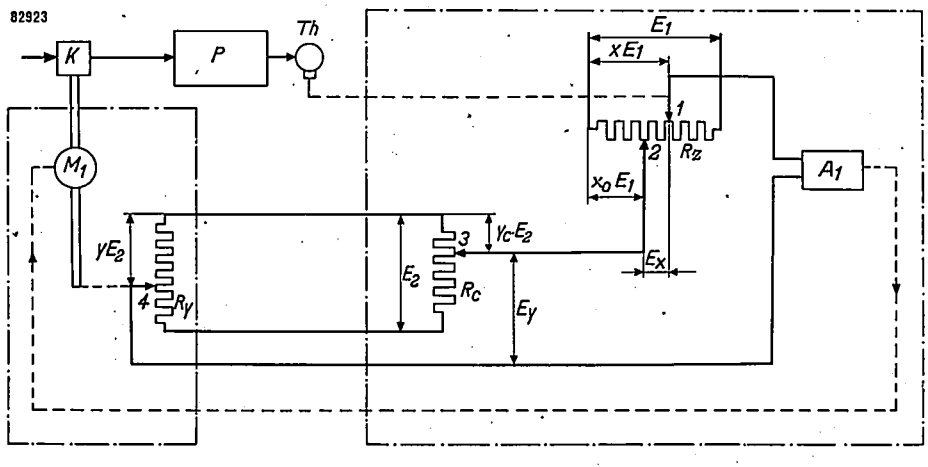


Fig. 7. Schematic lay-out of a proportional control system. The right-hand part (surrounded by chain line) is situated near the measuring element, the left-hand part near the correcting element (valve).

Floating control and integral control

With proportional control there is a fixed relationship between the deviation of the furnace temperature and the position of the correcting element. In the case of floating control, however, the final control element is continuously displaced until the furnace temperature no longer deviates from the desired value. The rate of displacement may be constant, or it may have two or more fixed values, or it may be proportional to the value of the deviation. In the latter case the total displacement of the correcting element is proportional to the time integral of the temperature deviation. This

[3]) This is completely analogous to the fact that an amplifier with negative feedback can go into oscillation if the feedback is large and if a large phase shift occurs at certain frequencies.

special case of floating control is called integral control.

An example of a floating control system is shown schematically in *fig. 8*. The error voltage E_x is

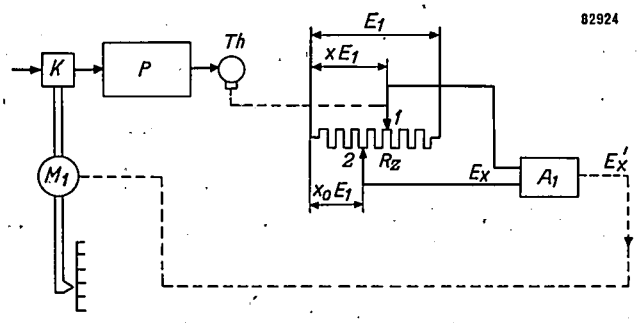


Fig. 8. Scheme for floating control. The motor M_1 actuated by the output voltage E_x' of the amplifier A_1, changes the position of the correcting element K.

applied to the amplifier A_1. The output voltage E_x' of this amplifier controls the motor M_1 by which the final control element is driven. If the speed of the motor is made proportional to the applied voltage, an integral control is obtained (also known as proportional speed floating control).

The circuit may also be so arranged that the speed of the motor remains constant (single-speed floating control). In this case only the direction of rotation of the motor, is governed by the direction of the furnace temperature (i.e. on the sign of of the deviation $x - x_0$). This can be effected by applying the output voltage of the amplifier A_1 to a pair of phase detectors, whose supply voltages are in anti-phase. (*fig. 9*). Whether relay Re_1 or relay Re_2 is energized depends on the sign of $x - x_0$, so that the motor is switched on with the desired direction of rotation. If $x - x_0 = 0$, which means that the controlled variable has the desired value, both relays are without current and the motor remains at rest.

With floating control the correcting element is at rest only when the deviation of the furnace

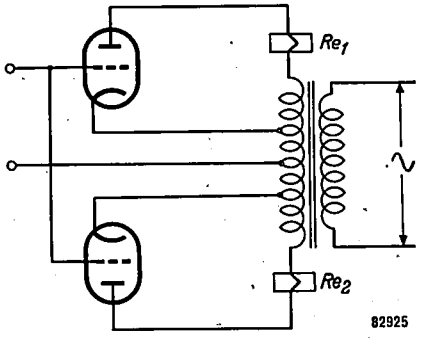


Fig. 9. Combination of two phase detectors, for use with a floating control with constant motor speed.

temperature is zero. It therefore avoids the disadvantage of proportional control, namely, that in spite of the control action, there remains a temperature deviation. An advantage of proportional control, however, is that once a change in the controlled quantity occurs, the valve or correcting element very quickly moves to the position corresponding to the new situation. (The time of adjustment is usually limited only by the speed of the valve itself.) In the case of floating control, the speed of the motor element is relatively slow. This means that if the changes in the controlled quantity occur rapidly, the final control element cannot quickly enough reach the position necessary to establish zero deviation of the controlled variable from the desired value. Thus also with floating control, a sudden change in the process results temporarily in excessive deviations of the controlled quantity. Raising the speed of the motor element may throw the system into oscillation. For this reason floating control is generally used in combination with proportional control. A floating control of the integral type is usually chosen in such cases.

Proportional and integral control

Fig. 10 shows a simplified diagram of a control system in which the displacement of the final control element (y) depends both on the deviation from the furnace temperature (therefore upon $x - x_0$) and on the time-integral of this deviation. The voltage E_y, which depends on the position of the correcting element, is now not connected directly in series with E_x, but via the capacitor-resistor coupling C-R. The voltage E_1 and E_2 are direct voltages in this circuit [4]). Again, the correcting element is adjusted in such a way that the input voltage to the amplifier becomes zero. From fig. 10 we have:

$$E_x = iR.$$

Furthermore

$$E_y = E_2(y - y_c) = \frac{1}{C}\int i\,dt + iR$$

and

$$E_x = (x - x_0)\,E_1.$$

From these equations it follows:

$$y - y_c = \frac{E_1}{E_2}\left\{(x - x_0) + \frac{1}{RC}\int_0^t (x - x_0)\,dt\right\}. \quad (2)$$

[4]) The direct voltage applied in this case to the amplifier A_1 will usually first be converted by means of a vibrator-converter into an alternating voltage, so that a normal A.C. amplifier can be used (Cf. e.g. Philips tech. Rev. **16**, 117-122, 1954/55, (No. 4).

In this way we see that the displacement of the correcting element can be separated into two terms, viz. a term $(x-x_0)E_1/E_2$, which is proportional to the deviation from the furnace temperature, and a term that is proportional to the time-integral of this deviation. The latter term will be larger the smaller the value of the product RC orarily in a fixed position). Adding the two terms we see that the initial change in $y-y_c$ due to the proportional control action is doubled by the integral term after a period of time given by

$$t = RC.$$

This is called the integral action time.

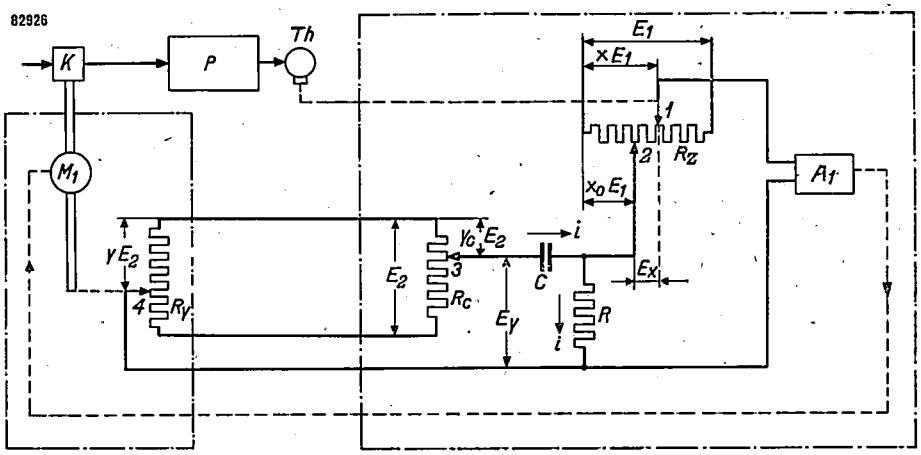


Fig. 10. Circuit for proportional + integral control. By interchanging the elements C and R, a circuit for proportional + derivative control is obtained.

The significance of this RC-product may be clarified by the following considerations (*fig. 11*).

If, starting from the condition where $x = x_0$, a sudden change in x occurs, this is accompanied by

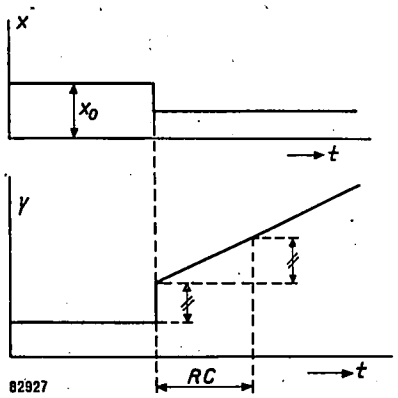


Fig. 11. Variation of x and y with time, due to a sudden change in the controlled variable x, for a system of combined proportional and integral control.

a sudden change in y to the effect that $y-y_c = (x-x_0)E_1/E_2$. Because of the integral term in (2) however, $y-y_c$ is further increased by a term

$$\frac{E_1}{E_2} \cdot \frac{x-x_0}{RC} t$$

(it is assumed that the furnace reacts so slowly that during the interval considered $x-x_0$ remains virtually constant or that contact 1 is kept temp-

From (2) it is self-evident that the correcting element is displaced as long as there is any deviation present between x and x_0. Hence, with the system considered here, the controlled variable does in fact attain the desired value.

Proportional and derivative control

In our discussion of proportional control we have already pointed out that the proportional control factor cannot be made indefinitely large because of the risk of instability. In many cases the stability may be improved by using a circuit in which the correcting element is given a displacement proportional to the time derivative of the deviation from the controlled condition in addition to the displacement due to proportional action. A circuit suitable for this can be derived from fig. 10 by interchanging the capacitor C and the resistor R. By a similar argument to that used for integral control it can be shown that the displacement of the final control element is given by the equation:

$$y-y_c = \frac{E_1}{E_2}\left\{(x-x_0) + RC\frac{\mathrm{d}}{\mathrm{d}t}(x-x_0)\right\}. \quad . \quad (3)$$

Here too, the latter term depends upon the RC-product. The significance of this product can be established as follows. Suppose that the system, initially in equilibrium ($x = x_0$), is subjected to a change such that x varies at a constant speed,

i.e. dx/dt is constant. Let the value of dx/dt be k (see *fig. 12*). The correcting element is then given a sudden displacement, due to the derivative action, of value $y-y_c = RCkE_1/E_2$. It also gets a displacement from the proportional action, and this increases with time since the variable itself $(x-x_0)$ has been assumed to be changing linearly with time; this displacement is given by $y-y_c = ktE_1/E_2$. These two displacements reach the same value after a time $t = RC$. This period is called the derivative action time.

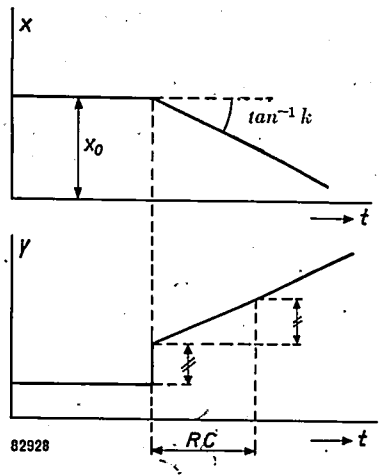


Fig. 12. Variation of x and y due to a change in x at constant rate for a system of combined proportional and derivative control.

Proportional, integral and derivative control

The properties of the two previous circuits can be combined as in the circuit of *fig. 13*. The two RC-networks, R_1C_1 and R_2C_2, are now both incorporated. The displacement of the correcting element is then given by the equation

$$y-y_c = \frac{E_1}{E_2}\left\{(x-x_0)\left(1+\frac{R_2}{R_1}+\frac{R_2C_2}{R_1C_1}\right)+\right.$$
$$\left.+\frac{1}{R_1C_1}\int_0^t (x-x_0)\,dt + R_2C_2\frac{d}{dt}(x-x_0)\right\}. \quad (4)$$

The integral and derivative actions are given by the same terms as in equations (2) and (3), but the proportional action is now given by the term

$$\frac{E_1}{E_2}(x-x_0)\left(1+\frac{R_2}{R_1}+\frac{R_2C_2}{R_1C_1}\right),$$

so that in this case, the proportional action also depends on the value of R_1, C_1, R_2 and C_2.

Time-proportional control

The continuous control of large electric furnaces requires large and elaborate control devices, such as thyratrons, variable transformers and transductors. For this reason a discontinuous control action is often preferred, in which the furnace is simply switched on and off at intervals. A far simpler apparatus (as a rule a magnetic switch) can than be used, but this has the disadvantage that quite considerable oscillation in the furnace temperature may occur. It is, however, also possible to obtain the effect of a continuous control, whilst maintaining the simplicity of a discontinuous control. This can be achieved by switching the applied power either on and off or between two specific values, at a constant frequency. The ratio between the periods during which the furnace is switched on and switched off is made continuously variable and dependent upon the difference between the actual and the desired furnace temperature. The relationship between these

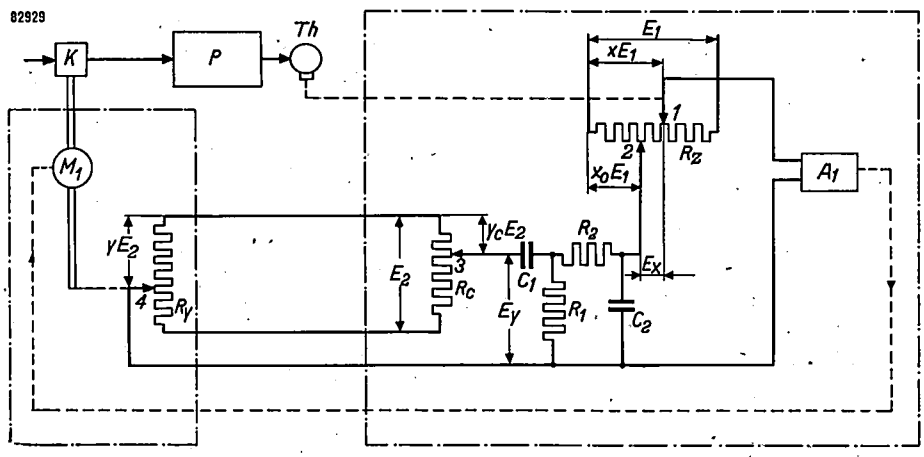


Fig. 13. Layout of circuit for proportional, integral and derivative control.

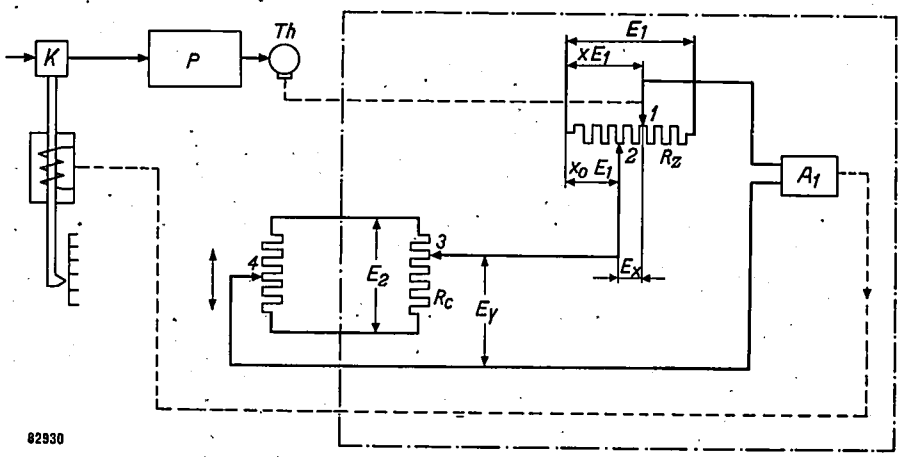


Fig. 14. Diagram of a circuit for control by time-proportional switching. The contact *4* is moved up and down periodically at a constant rate.

two values may be proportional, integral or derivatative, or it may be a combination of these.

Fig. 14 shows a simplified diagram of a circuit for obtaining a proportional control in this way. Here again the voltage E_y is added to the voltage E_x of the controlling slidewire. Contact *4* is now moved periodically up and down, so that the voltage E_y shows a similar periodic variation. In *fig. 15* the voltages E_x and E_y are plotted as functions of time. As regards the latter voltage, if contact *3* is placed in the middle, E_y will vary symmetrically, about the zero line. The instants when the input voltage of the amplifier passes through zero are given by the points of intersection of the E_x and E_y curves. At these instants the input voltage changes its phase by 180° which, by means of a phase detector and a magnetic switch, causes the furnace to be switched alternately on and off. As can be seen from fig. 15, the ratio of the time t_1 during which the furnace is switched on, to the time t_2 during which it is switched off, is dependent upon E_x. The mean value of the applied power is thus continuously dependent upon the deviation of the furnace temperature. The rate at which switching occurs (i.e. the period of the E_y oscillation) should be so selected that the fluctuations of the furnace temperature caused by it, are sufficiently small.

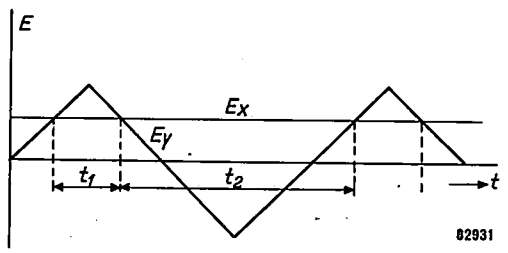


Fig. 15. Voltages E_x and E_y of fig. 14 as functions of the time. t_1 and t_2 represent the times during which the furnace is switched on and switched off, respectively.

As in the system of proportional control dealt with above, a certain deviation from the controlled condition after a change in the process will likewise remain with a circuit such as that shown in fig. 14. Here too, this can be corrected if necessary, by re-adjusting contact *3*.

Programme Control

In the foregoing the value of the controlled condition was assumed to be constant. In some cases, however, it is necessary to vary the desired value as a function of the time. One way in which this could be achieved is by incorporating in the control unit a mechanism by which contact *2* is displaced in accordance with a given time function, viz. the "programme" of the process. Generally, however, a method is preferred which does not require any special mechanism to be built into the controlling unit. An example of this is the circuit shown in *fig. 16*. A separate apparatus, the programme transmitter, contains a second slidewire R_p in parallel

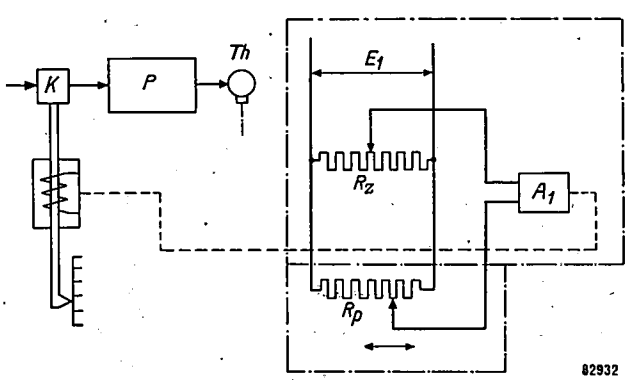


Fig. 16. Circuit for programme control. The slidewire R_p is incorporated in the programmer which is separate from the potentiometer.

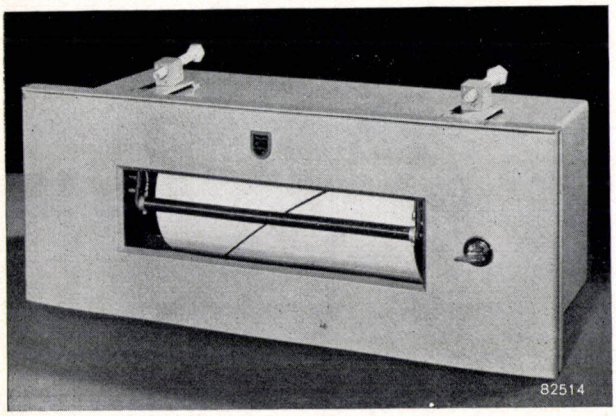


Fig. 17. The Philips programmer, type PR 7211.

with the controlling slidewire R_z, The contact *2*, which serves for setting the desired value is now mounted on this second slidewire R_p and is moved along it in accordance with the required time function. The Philips programme controller has a rotatable cylinder of insulating material mounted parallel to R_p, carrying a conducting wire on its surface running along the graph of the required time-function. At one spot this wire is in contact with the potentiometer R_p. The cylinder is rotated at uniform speed by an auxiliary motor, so that the point of contact moves according to the desired time function. The programmer is illustrated in *fig. 17*.

Alternative control systems

A number of the control systems described in the foregoing may also be realized in other ways, still using the automatic potentiometer for the measurement. An on-off control, for example, can be realized by using the carriage of the measuring slidewire to operate an on/off contact at a certain point. There is the drawback, however, that in this case a greater force is necessary for displacing the measuring carriage at the position corresponding to the value of the controlled condition. In this way the sensitivity of the measuring bridge is impaired. Furthermore, for making and breaking a contact, a certain displacement of the carriage is always required, so that there would be a permanent differential between the points of switching on and switching off. In the control system already described, the carriage has to overcome only relatively small frictional forces and the sensitivity of the measuring bridge is uniform. By using a high gain amplifier, the differential between switching on and switching off may be reduced virtually to zero.

Systems for continuous control may also be derived mechanically, pneumatically or hydraulically from the automatic potentiometer. An integrating and derivative system can be realized, for example, pneumatically or hydraulically by means of capillaries and volume elements. These, however, require more maintenance than the corresponding electrical elements (resistors and capacitors). In pneumatic systems, for example, it is necessary to use dry, oil-free and dust-free air to minimize rust and clogging or freezing in orifices and capilliaries.

The use of an electric control system does not exclude the use of hydraulic or pneumatic final control elements. An electrically controlled hydraulic or pneumatic valve positioner or similar device is then used.

Summary. A survey of the various types of automatic control actions and how these may be achieved electronically for the control of industrial processes. The various circuits described are suitable for use in conjunction with the automatic potentiometer/bridge described in a previous article. The various methods of control may be divided into systems for discontinuous and for continuous control. The discontinuous system most widely used is the two-step control. If the conditions under which the process takes place vary considerably, this method can be extended to a three-step or multi-step control. When the number of possible positions of the regulating unit is made very large, this system becomes a continuous control system, e.g. proportional control. One objection to the latter is that the controlled variable may not remain exactly at the desired value. This objection may be overcome by using a floating control system, of which integral control is a special form. The drawback inherent in integral control, viz. that the correcting element does not react instantaneously to a sudden disturbance in the process conditions, may partly be obviated by combining integral and proportional control. In order to improve the stability when the proportional control factor is large, derivative action may be added to proportional control. The quasi-continuous control of electric furnaces may be achieved with advantage by means of time-proportional switching; with simple circuitry the ratio of the on and off times of the furnace may, for example, be made proportional to the deviation of the furnace temperature from the desired value. Finally, a brief description is given of programme control and of alternative non-electronic control systems used in conjunction with an electronic automatic potentiometer.

LINEAR ELECTRON ACCELERATORS FOR DEEP X-RAY THERAPY

621.384.62:615.849

Along with the development of linear electron accelerators for nuclear physics research, similar machines have been designed and built for medical purposes. Linear accelerators are capable of producing an intense narrow beam of electrons having energies of several millions of electronvolts [1)][2)]. When impinging on a suitable target, the electrons generate X-rays of high penetrating power, so providing a very efficient source for deep therapy.

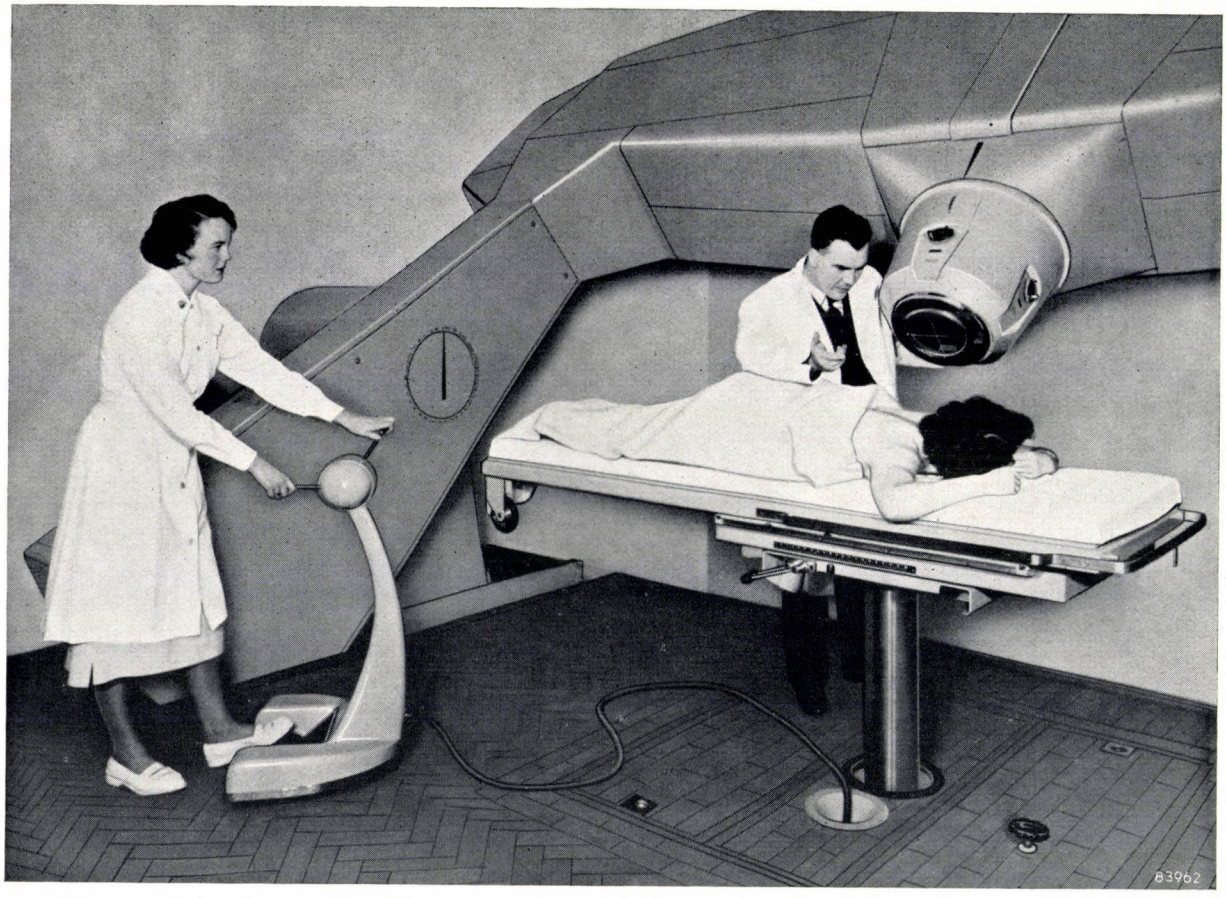


Fig. 1. Medical linear accelerator generating 4 MeV X-rays, installed at Newcastle Genera Hospital (England). This was the first of a series of 4 MeV-accelerators to be used for deep X-ray therapy.
The photograph shows a patient being set up for treatment. The double-ended gantry houses the accelerator with accessories and X-ray head. The assistant on the left adjusts the height of the couch and the angle of the gantry to provide the correct point and angle of incidence of the radiation.

The photograph *fig. 1* shows a 4 MeV medical linear accelerator designed and built at the Mullard Research Laboratories, Salfords (England), to machine by mounting the accelerator on a large double-ended gantry, the X-ray beam being directed towards the axis of rotation where the patient is

Ministry of Health specification. The machine was installed in Newcastle General Hospital during August 1953 and has been in use for the treatment of patients since December 1953. Since the X-rays generated by electrons in this energy range are emitted chiefly in a forward direction, the electron beam itself must be swung round in order to provide for different angles of irradiation of a patient. This facility has been provided with the Newcastle

1) D. W. Fry, The linear electron accelerator, Philips tech. Rev. **14**, 1-12, 1952/53.

2) C. F. Bareford and M. G. Kelliher, The 15 million electron-volt linear electron accelerator for Harwell, Philips tech. Rev. **15**, 1-26, 1953/54.

located. *Fig. 2* may serve to illustrate the general lay-out. A more detailed description of this equipment [3]) will appear in a future issue of this Review.

Fig. 3 gives an impression of an even larger medical linear accelerator producing 15 MeV X-rays. This machine has recently been installed in St. Bartholomew's Hospital Medical School, London, by a team of scientists of the Mullard Laboratories headed by T. R. Chippendale under the direction of P. E. Trier [4]). The accelerator is similar to that already described in this Review [2]). Swinging the whole accelerator round for changing the angle of irradiation is not a practical proposition for a machine of this size. In this case, therefore, the accelerating waveguide is mounted in a fixed horizontal position and the electron beam emerging from it is bent through 90° by a powerful electromagnet before hitting the target. The X-ray head housing the target and the electromagnet can be rotated about the horizontal axis of the machine to enable the angle of irradiation to be varied.

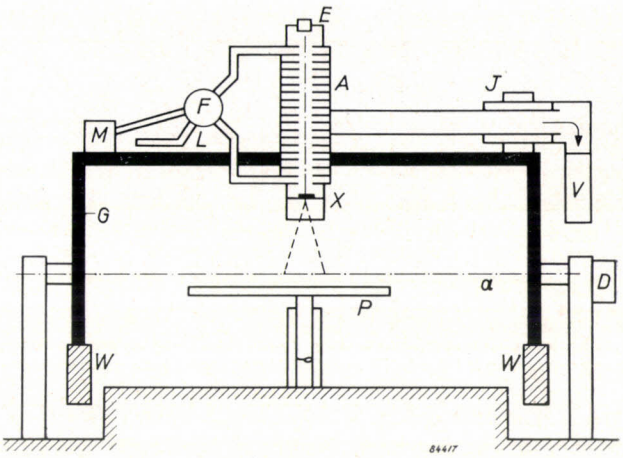


Fig. 2. Schematic representation of the machine shown in fig. 1. *A* is the accelerator proper (corrugated waveguide 1 metre long), mounted in gantry *G* rotatable about the axis *a* by means of an oil motor *D*. *E* electron gun, *X* X-ray head, *M* magnetron, *F* feedback bridge, *V* diffusion pump, *J* rotating vacuum joint, *W* counterweights, *P* patient's couch, *L* water load.

[3]) A preliminary account was given by T. R. Chippendale and M. G. Kelliher, A linear accelerator for X-ray therapy, Discovery **15,** 397-404, 1954 (No. 10).

[4]) A 15 MeV linear accelerator for medical use, Electronic Engineering **26,** 527-528, Dec. 1954 (No. 12).

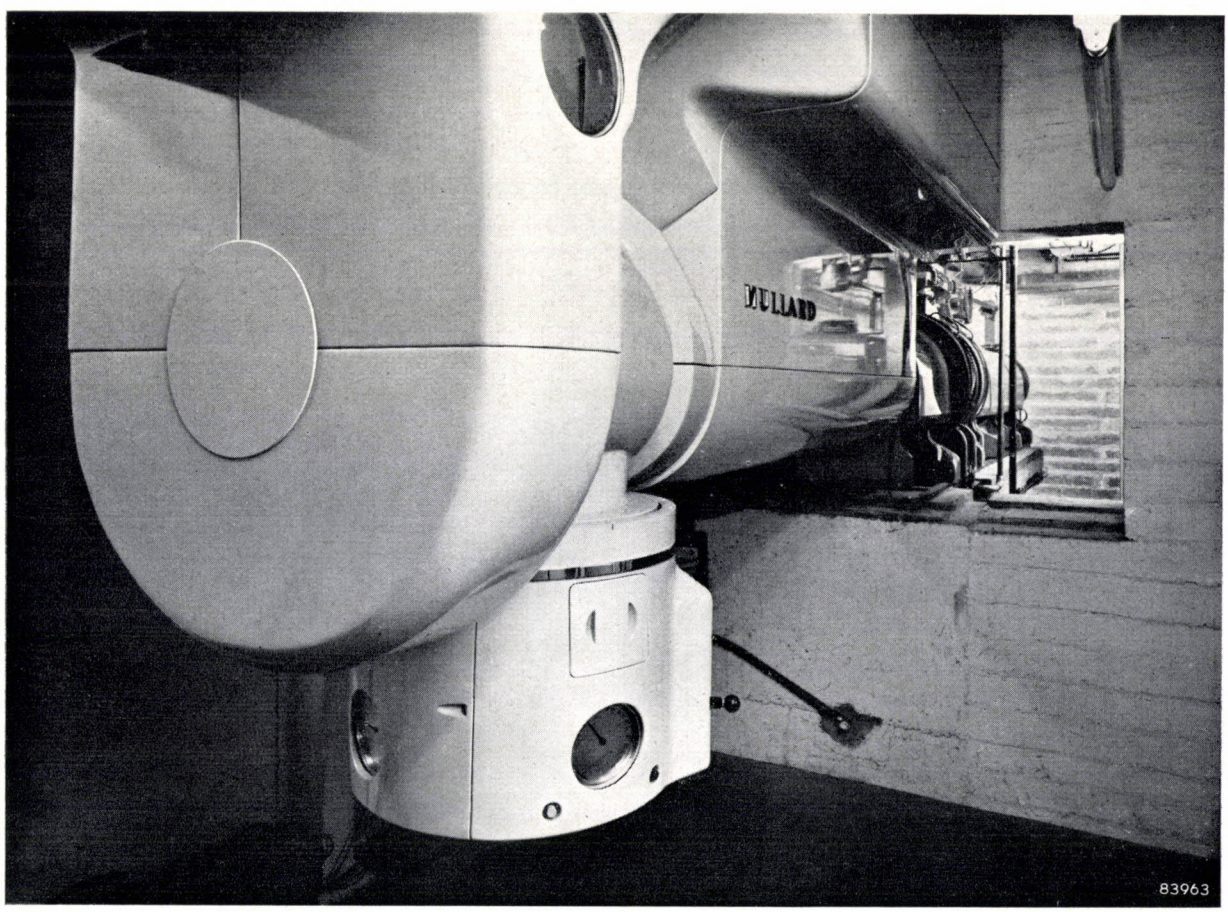


Fig. 3. 15 MeV linear accelerator installed in its temporary building at St. Bartholomew's Hospital Medical School, London.

In fig. 3 it is seen that the target end and X-ray head of the machine are supported from above to permit the positioning of patients underneath. The angular position of the X-ray head is indicated by the scale at top centre. The X-ray beam field size, which may be adjusted using the control buttons on the head, is indicated on the two lower scales.

Four of the six metres length of the accelerator are housed in a separate room which is divided from the treatment room by a concrete wall 1 metre thick. The gap in the wall visible in the phtograph and permitting a complete view of the accelerator is normally closed with concrete blocks.

T. R. CHIPPENDALE.

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

Reprints of these papers not marked with an asterisk * can be obtained free of charge upon application to Philips Electrical Ltd., Century House, Shaftesbury Avenue, London W.C. 2.

2161: R. Vermeulen: Sound recording. General review, presented at the first I.C.A. Congress on Electro-accoustics, Delft, 1953 (Acustica 4, 17-21, 1954.)

General considerations on sound recording. The following topics are dealt with; improvements in gramophone records, gramophone recording and reproduction, tape recorders; information theory in relation to reproduction of music, fidelity of the reproduction, stereophony, artificial reverberation.

2162: J. Rodrigues de Miranda: The radio set as an instrument for the reproduction of music (Acustica 4, 38-41, 1954).

In connection with the reproduction of music by radio the following desiderata are put forward: 1. Distortion should be decreased as the high frequency range increases. 2. Means are desirable for adjusting bass response, for cutting off and for gradual attenuation of treble. 3. The electro-acoustical engineer and the architect should cooperate closely. The cabinet must be rigid, the loudspeaker placed forward; its cloth chosen carefully. 4. The speaker should be chosen in accordance with the acoustical properties of the cabinet. The audio-frequency characteristic of the set should be carefully determined.

2163: W. K. Westmijze: Application of the reciprocity theorem to magnetic reproducing heads (Acustica 4, 50-52, 1954).

It is shown that the output of a magnetic reproducing head can be calculated if the field distribution in front of the head is known for the case when the latter is energized. Application of this method enables us to predict the response curve of a wide-gap head and to explain the existing difference between wide-gap and narrow-gap measurements.

2164: A. G. Th. Becking and A. Rademakers: Noise in condenser microphones (Acustica 4, 96-98, 1954).

It is pointed out that the mechanical resistance, used for damping a simple mechanical oscillator, acts as the source of a fluctuating force with spectral intensity 4RkT. In the case of a condenser microphone this results in a noise pressure on the diaphragm, which lies in the region of audible sound pressures. Experimental results, obtained with two microphones, agree well with the theory.

2165: A. G. Th. Becking: Methods of measurement for hearing aids (Acustica 4, 143-145, 1954).

The different ways of defining the input sound pressure on a hearing aid are discussed. In addition, a comparison is given of the specifications for acceptable hearing aids in some European countries.

2166: J. Volger, J. M. Stevels and C. van Amerongen: The dielectric relaxation of glass and the pseudo-capacity of metal-to-glass interfaces, measured at extremely low frequencies. (Philips Res. Rep. 8, 452-470, 1953, No. 6).

See **R 232.**

2167*: H. Bremmer: On a phase-contrast theory of electron-optical image formation. (Electron Physics, Circular 527, Nat. Bur. Stand., March 1954, 145-158).

The scattering of electrons by an object, as applied in electron microscopy, can be dealt with by means of an integral equation, which follows from the Schrödinger equation corresponding to the electrostatic potential field inside the object. This integral equation can be solved by a series, the consecutive

terms of which may be interpreted as the contribution due to repeated scatterings of the incident electron waves by the object. Summation of the main contribution (for very small electron wavelengths) of these terms leads to a wave function in the object plane (this is a plane immediately behind the object), which in every point depends only on the potential distribution along the straight line joining this point with the electron source. From the values of the wave function in the object plane it is possible, by well-known mathematical methods, to derive the behaviour of the wave function in the space behind that plane. In this way it is possible, in principle, to investigate also the image formation.

2168: J. Haantjes: Die Fernsehübertragung der englischen Kröningsfeier nach dem Festland (Fernmeldetechn. Z. **7**, 129-133, 1954, No. 3). (Television transmission of the English coronation festivities to the continent; in German.)

See Philips techn. Rev. **15**, 297-306, 1953.

2169: A. M. Kruithof and A. L. Zijlstra: Les propriétés élastiques d'un certain type de verre (Verres et Réfractaires **8**, 1-13, 1954, No. 1). (The elastic properties of a certain type of glass; in French.)

The disappearance of internal stresses in glass at constant temperature, turns out to be dependent on three factors: the viscosity, the instantaneous elasticity and the elastic after-effect. The first part of the article deals with these factors for a certain type of glass. Special attention is paid to the elastic after-effect. This can be described by a so-called delayed elongation which, after the sudden application of a constant stress, can be represented as a function of time by the sum of two exponential factors.

In the second part, a formula is derived for the disappearance of internal stresses in a rod of stabilized glass at constant temperature, which is suddenly stretched to a specified exent by the application of a constant tensile force. As a result of the mutual effects of viscosity, elasticity and elastic after-effect the resultant stresses disappear. Their disappearance can be described by the sum of three exponential functions. This formula is tested experimentally. The results agree very well with the theory.

2170: E. J. W. Verwey: Das Kräftespiel zwischen Teilchen in lyophoben Kolloidsystemen (Kolloid Z. **136**, 46-52, 1954). (The effect of forces between particles in lyophobic colloids; in German.)

It appears from the theory that the behavior and particularly the stability of lyophobic colloids is in accordance with the balance between the Van der Waals-London forces of attraction and the repulsion forces of the electrolytic double layer. The nature of these forces is examined more closely. The results of the theory point to an elegant explanation of the well known Schulze-Hardy rule, and thereby a further extension of our knowledge of van der Waals-London forces.

2171: H. J. G. Meyer: Theory of radiationless transitions of F centres (Physica **20**, 181-182, 1954).

Calculations are reported, based on the Huang-Rhys model, on the probality of radiationless transitions in F centres. By using a very accurate approximation formula for modified Bessel functions of large index and large argument, an expression is derived which allows a simple physical interpretation for the cases of high and low temperature and which is easily calculated numerically. It is shown that in principle with this model large probabilities for radiationless transitions can be obtained.

2172: J. Volger: Electrical properties of ceramics, Part I. Semiconductors (Research **7**, 196-203 1954).

A review is given of some ceramic products of value in the electrical industry, most of the materials considered being oxides of the transition metals. Some introductory remarks on their crystal structures and on the occurrence of certain substitutions and imperfections in the lattice are followed by a discussion on their electronic conductivity and in particular on the relation between the conduction mechanism and the occurrence of certain transition metal ions in different valency states. The influence of the polycrystalline nature of these materials on conduction and on a number of second-order conduction effects (e.g. Hall effect and frequency dependence of resistivity and magneto-resistance) is briefly discussed. Some applications of ceramic oxidic semiconductors are mentioned.

2173*: C. J. Bouwkamp: Diffraction theory (Rep. Progr. Phys. **17**, 35-100, 1954).

A critical review is presented of recent progress in classical diffraction theory. Both scalar and electromagnetic problems are discussed. The report may serve as an introduction to general diffraction

theory although the main emphasis is on diffraction by plane obstacles. Various modifications of the Kirchhoff and Kottler theories are presented. Diffraction by obstacles small compared with the wavelength is discussed in some detail. Other topics included are: variational formulation of diffraction problems, the Wiener-Hopf technique of solving integral equations of diffraction theory; the rigorous formulation of Babinet's principle, the nature of field singularities at sharp edges, the application of Mathieu functions and spheroidal wave functions to diffraction theory. Reference is made to more than 500 papers published since 1940.

2174: J. Volger: Electrical properties of ceramics, Part II. Dielectries and ferromagnetics (Research **7**, 230-235, 1954).

After some remarks on ceramics as insulating materials, the use of ceramic dielectrics for use in condensors is discussed. The factors favouring a high value of the dielectric constant are considered and in particular the occurrence of such high values with $BaTiO_3$. A number of interesting properties of soft ferromagnetic ceramics than are given, all of the ferrite or ferroxcube type. The saturation magnetization values are discussed in connection with the theory of *Néel*. Attention is given to the h.f. properties of ferroxcube, i.e. the various resonance and relaxation phenomena which occur. Finally, a hard ferromagnetic ceramic called magnadur (also known as ferroxdure) is dealt with. This is a hexagonal oxidic compound of iron and barium with an extremely large coercive force. It may be used for permanent magnets which must withstand large demagnetizing fields.

2175: R. Vermeulen: Stereophonic reproduction (Audio Engineering **38**, 21, 1954, No. 5).

Recapitulation of the work of K. de Boer on the mechanisms of binaural and stereophonic sound phenomena. The perception of direction is discussed as being due to differences both in the time of arrival of the sound and in the intensity. The differences between binaural and stereophonic listening are outlined; in the latter case the sound image is less well defined. It is pointed out that the aim is not to try to reconstruct the original sound field, but to deliver the correct sound to each ear to simulate a sound source from a certain direction. Some empirical rules for the placing of microphones and loudspeakers are given. Finally mention is made of the part played by head movements in the location of the sound image by the listener.

2176: J. L. H. Jonker: Ten-volt effect with oxide-coated cathode (Electronic Engineering **26**, 282, 1954).

According to a theory given earlier, the anomaly in the diode characteristic that occurs at an anode potential of 10 V should be attributed to the fact that the space-charge contribution of electrons reflected from the anode then shows a minimum (see Philips Res. Rep. **2**, 331-339, 1947). This theory is confirmed by measurements on the reflection and secondary emission of a surface that had been maintained for some time facing an emitting oxide cathode.

2177: B. Combée and A. Engström: A new device for micro-radiography and a simplified technique for the determination of the mass of cytological structures (Biochim. et Bioph. Acta **14**, 432-434, 1954).

A sealed-off X-ray tube with a very thin Be window is used as a source of soft X-rays for quantitative historadiology by means of contact micro-radiography. A procedure is described whereby the mass of the specimens may be determined without the necessity of reference exposures. The resolution of the image obtained with this tube is about 0.5 μ.

2178: H. de Lange: Relationship between critical flicker-frequency and a set of low-frequency characteristics of the eye (J. Opt. Soc. America **44**, 380-389, 1954)

Measurements of the critical flicker frequency of the eye as a function of both the average luminance of the test field and the time variation of this luminance were recorded by plotting the "ripple ratio" r against the critical frequency; r is defined as (amplitude of first Fourier component)/(average luminance) of the stimulus. It is shown that with constant average luminance, the points observed for various time functions fit into one smooth curve, which for low luminance is monotonic. At high luminances the curve shows a minimum in r at a critical frequency of about 9 c/s. This means that the eye has a maximum sensitivity to flicker at this frequency.

2179: L. A. Æ. Sluyterman and B. Labruyère: Side reactions in the polymerization of α-amino acid N-carbonic anhydrides. Titration data of polyglycine and polyalanine (Rec. trav. chim. Pays-Bas **73**, 347-354-1954).

Polyglycine and polyalanine obtained by polymerization of their N-carbonic anhydrides often

appeared to certain more acid groups than animo groups. This is ascribed, at least partly, to the formation of hydantoin groups. A small amount of 2.5-diketopiperazine was isolated from polyglycine. A discussion of the results and their bearing upon the calculation of molecular weights is given.

2180: F. A. Kröger, H. J. Vink and J. van den Boomgaard: Controlled conductivity in CdS single crystals (Z. phys. Chem. **203**, 1-72 1954, No. 1/2).

The electrical properties of single crystals of pure CdS or of CdS containing gallium, indium, antimony, chlorine or silver depend markedly on the atmosphere of preparation. Crystals subjected to an oxidizing atmosphere (i.e. sulphur vapour) are insulators or semi-conductors, showing photoconductivity; crystals subjected to a reducing atmosphere show quasi-metallic, electronic conductivity, For crystals doped with Ga or Cl, the number of carriers in the reduced crystals is constant over a wide range of temperatures and equal to the concentrations of foreign ions over a wide range of atmospheres (controlled valency). The optical properties (absorption, fluorescence) vary also with the atmosphere, absorption bands in the yellow part of the spectrum appearing in oxidized, but not in reduced crystals. A general theory is developed by extending Schottky and Wagners theory of lattice imperfections along the lines indicated by Schottky (1935), By means of this theory it is possible to calculate the dependence of the concentration of the various kinds of lattice imperfections, donors, traps, and acceptors, on the concentration and nature of the impurities and the reducing power of the atmosphere. Applying this theory to CdS, a satisfactory agreement with the experiments is obtained.

2181: J. te Winkel: Lijnversterkers voor draaggolfsystemen op coaxiale kabels (De Ingenieur **66**, E.61-65, 1954, No. 25). (Line amplifiers for carrier telephone systems on coaxial cable; in Dutch.)

A brief survey of the principles governing the design of these amplifiers.

2182: E. Havinga and J. P. L. Bots: Studies on Vitamin D, I. The synthesis of vitamin D_3 3 C^{14} (Rec. trav. chim. Pays-Bas **73**, 393-400, 1954).

In connection with the study of the metabolism of vitamin D, vitamin D_3 has been synthesised with a C^{14} atom in place of the third carbon atom. The preparation of this vitamin D_3 3 C^{14} is described. Special attention is paid to the technique of the photochemical conversion of 7-dehydrocholesterol 3 C^{14} (small quantities) and to the isolation of the labelled vitamin D_3 from the irradiated mixture. Some biological experiments are briefly described.

2183: G. W. Rathenau and G. Baas: Croissance des grains, observée par microscopie electronique à emission (Métaux, Corros. Ind. **29**, 139-150, 1954 (No. 344). (Grain growth observed by means of the emission electron microscope: in French.)

Changes in structure in metals and alloys at elevated temperatures can directly be observed applying emission electron microscopy. With the instrument used for these investigations a resolving power of about 1000 Å is obtained.

Rolled face-centred cubic NiFe alloys have been used to study grain growth in a texture. This investigation includes growth of one large grain of deviating orientation in a well-pronounced texture as well as grain growth in an imperfect texture. In both cases only high energy boundaries move at measurable speed. The results are discussed in terms of interfacial tension effects.

An example of grain growth hindered by inclusions is given for a Cr Ni steel specimen, containing inclusions of σ-phase.

Grain growth accompanying the transformation $\alpha \rightarrow \gamma$ in a SiFe alloy has been described. In this case the rate of diffusion determines the rate of growth of the γ grains.

The austenite-pearlite transformation in eutectoid carbon steel has been directly observed. The existence of an orientation relationship between austenite and pearlite is improbable as shown by the observations. Grain-boundary displacements in the austenite have been observed near a growing colony of pearlite.

VOL. 17 No. 2, pp. 37-68 AUGUST 1955

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

STUD WELDING WITH WELDING CARTRIDGES

by W. P. van den BLINK *), E. H. ETTEMA *) and P. C. van der WILLIGEN.

621.791.75

The number of studs welded every day is roughly estimated at over one million, and will certainly increase considerably during the next few years. The importance of an inexpensive and efficient stud welding process, as developed by the writers of the present article, will therefore be evident.

Introduction

Stud welding may be defined as the welding of a rod-shaped object (pin, stud, bolt) to, say, a plate, at right angles, without employing a separate welding electrode. It can be done by means of a resistance weld or by employing the stud itself as a welding electrode, that is, by drawing an arc between the stud and the work.

Steel studs joined to steel surfaces at right angles have a vast number of applications. Some examples are the securing of covers, the assembly of various parts of machines, flange connections, the fixing of facing material to a steel surface (say, wooden planking on a steel deck), or the anchoring of pipe brackets in ships and other steel constructions. The boiler-making industry offers another example, viz. the fixing of studs, in large numbers, to heat

*) Philips' Welding Electrode Factory, Utrecht (Holland).

exchanger tubes, to increase the overall area of heat transfer. The frontispiece shows studs being fixed inside a steel tank at the "N.V. Werkspoor", Amsterdam, by the recently developed process of stud welding described in the present article.

Stud welding is much quicker and cheaper than the conventional method of fixing studs, that is, screwing threaded studs into drilled and tapped holes. Moreover, such holes weaken the plates and may cause leaks, objections which apply equally to rivets. Hence it is not surprising that such methods are now being superseded more and more by various stud welding processes.

Resistance welding

In resistance welding, stud and workpiece are pressed together and a strong current is passed through them. The contact resistance between the two results in heat being generated at the point of contact, so that the material is softened. When the material is suitably plastic, the stud and the work are clamped together between the jaws of a press. How much of the cross-sectional area of the stud is actually welded depends upon the current and the pressure employed. The use of a press imposes certain limitations on the size and shape of the workpiece. On the other hand, this method of welding is better than others in that it does not form a collar of weld metal around the base of the stud (*fig. 1*); this is an important advantage for some purposes (e.g. joining flanges).

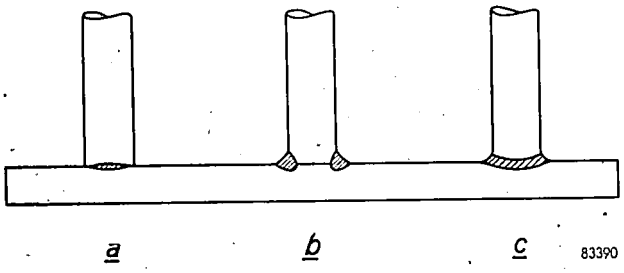


Fig. 1. Difference between the welds produced by the various methods of stud welding.
a) Resistance weld.
b) Arc weld produced with separate welding electrode.
c) Weld produced by drawing an arc direct between stud and plate.

The conventional method of stud welding

Studs can be welded, of course, with an ordinary welding electrode. Although this method generally produces a better joint, and is less limited in scope (e.g. as regards the size of the workpiece) than resistance welding, it requires much time and skill and is therefore rather expensive. In view of this fact, another method, that of drawing an arc direct between the stud and the work, was much used in the ship-building and boiler-making industries during the second world war. Usually, the stud is pressed against the work by a mechanism in the holder (or gun), the current is switched on, and a mechanism at once retracts the stud a fraction of an inch to draw the arc. After a pre-determined interval, in which a suitable amount of molten steel accumulates on the surface of the work and on the end of the stud, a separate unit connected to the gun switches the current off and releases a spring in the gun to press the stud into the molten steel.

For a $\frac{1}{2}$-inch stud, the welding time is of the order of $\frac{1}{2}$ second and the welding current 500-1000 A. Frequently, and invariably in the case of vertical and overhead welding, a porcelain ring ("ferrule") is fitted round the base of the stud to minimise loss or spattering of the molten metal. When the stud is welded, this ring is broken off.

Although quite practicable, this method of stud welding has certain serious disadvantages, largely owing to the fact that the stud, also employed as the welding electrode, is bare; D.C. has therefore to be used, and, since the currents are very heavy, an expensive rectifier is required.

Bare electrodes cannot be welded with alternating current of normal voltages (approx. 80 V), since the arc would then be unstable and liable to extinguish prematurely. Moreover, without special precautions, electrodes not enclosed in a slag-forming coating produce welds of inferior quality. The precautions referred to consist in employing studs made of a special type of steel, and, in some cases, applying deoxidizing agents to the ends of them. It follows that, at least for many purposes, it is necessary to employ studs specially made to suit the particular welding conditions. Also, the equipment required to move the stud in the above-mentioned manner and control the current is rather complex.

It will be evident, then, that the search for a simpler stud welding process is well worth while. The principles and application of a new method evolved as a result of this investigation will now be explained. Some measurements carried out to analyze the new process, and a number of tests on welds produced by it, will then be described. To conclude this article, one or two important practical details will be mentioned.

Principles of the new method

In 1945, the Philips Laboratories at Eindhoven developed what is known as the contact arc-welding

process [1]). It involves the use of welding electrodes whose coating is very thick, and is rendered semi-conductive by including in it an appreciable part of the weld metal as a powder. Such welding electrodes ignite and re-ignite very readily; moreover, they are suitable for touch-welding. The new method of stud welding is based on the same principles as contact welding [2]).

Let us now consider what happens when a partly used contact welding electrode, that is, one whose tip contains a deep "cup" formed during welding (*fig. 2*), is applied to a workpiece and the voltage is switched on. The current then flowing through the cup to the workpiece builds up very rapidly, owing to breakdown effects between the individual grains of iron powder in the material forming the

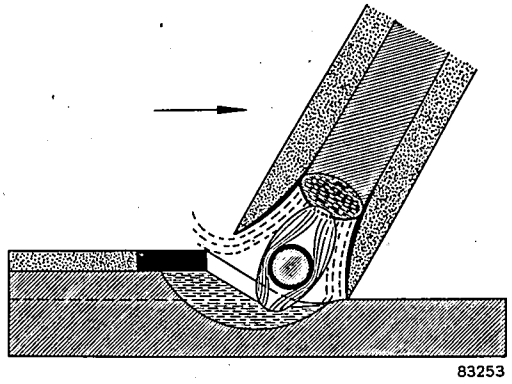


Fig. 2. A deep "cup", formed at the tip of a contact welding electrode during the deposition of metal, prevents the electrode from "freezing" to the work.

cup: the grains fuse together and so produce a rapid drop in resistance. In a very short time, then, the temperature at the point of contact rises enough to strike an arc. Once formed, the cup remains intact throughout the time the electrode is fused, keeping the core wire a certain distance from the work and thus preventing the two from "freezing" together.

Now, if the entire coating of a contact welding electrode, other than a small portion at the tip, be stripped off, and if this electrode is then pressed against the workpiece at right angles (*fig. 3*), the arc produced when the current is switched on will be quite normal at first; however when the remaining portion of the coating is almost consumed, the core wire suddenly breaks through the cup and freezes to the weld metal. This is precisely what is required in stud welding. The fusion time of the electrode is governed by the dimensions and composition of the coating, as well as by the welding current.

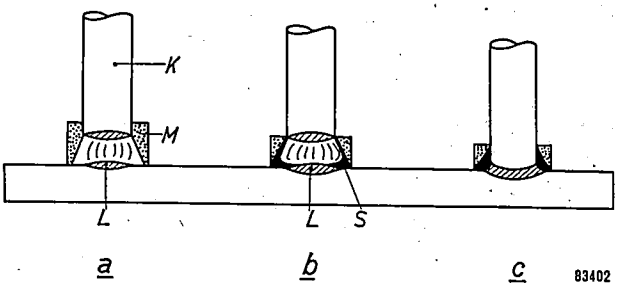


Fig. 3. The different stages in stud welding with a partially stripped contact welding electrode having a "cupped" tip. *M* portion of coating left on the rod, *K* core wire, *L* molten pool of metal, *S* slag.
a) Immediately after the striking of the arc.
b) Just before the electrode drops through the cup. The coating and the core wire have partly melted. A pool of molten metal and a certain amount of slag have been formed.
c) After the electrode has dropped through the cup. The weld metal is covered with slag and the remnant of the cup around the weld helps to shape it.

The basic idea of the new method of stud welding is to make the "coating" separately in short sections, cup-shaped and similar in composition to the coating of the contact electrodes. Fitted to the end of a stud, such a section ("welding cartridge") would enable it to be welded in the same way as the stripped contact electrode described above. These ideas, after many experiments, have led to the simple method now to be described.

The new method of stud welding

The shape of the welding cartridge

A diagram showing the welding cartridge, attached to a stud *A*, as placed on the work to which the stud is to be welded, is seen in *fig. 4*. The cartridge is in two parts, viz. an igniter element, *B*, made of

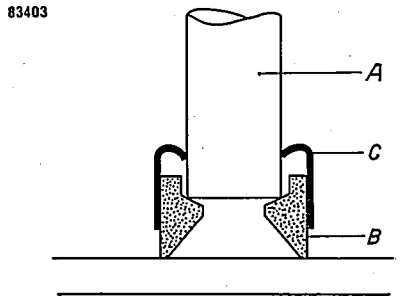


Fig. 4. The welding cartridge, comprising a semi-conductive igniter element *B* and a cardboard collar *C* to hold it on the stud to be welded *A*. During welding, the ridge of the igniter element supporting the stud melts, thus enabling the stud to be pressed into the pool of molten metal.

[1]) P. C. van der Willigen, Philips tech. Rev. **8**, 161-166, and 304-308, 1946; Sheet Metal Ind. **26**, 155, 1949; Schweissen und Schneiden **2**, 270, 1950.
[2]) The idea of basing a simple stud welding process on the principles of contact arc-welding came from Dr. H. Bienfait of the Philips Laboratories at Eindhoven.

semi-conductive material, and a cardboard collar *C*, fitted round the igniter. The igniter has a cup-shaped cavity similar to that acquired by a contact welding electrode, and an inside ridge to support the stud to be welded. One or two studs ($\frac{1}{2}$-inch), together with the associated welding cartridges, will be seen in *fig. 5*. The welding process is very

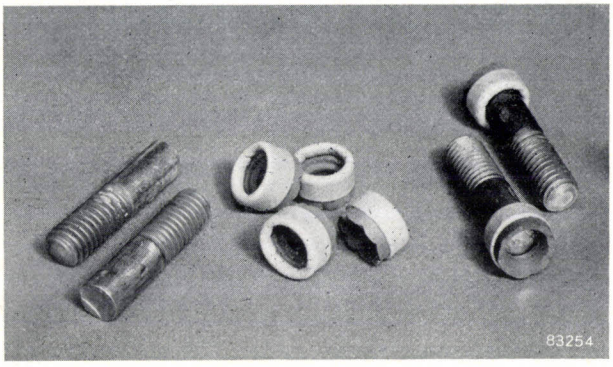


Fig. 5. Threaded studs ($^1/_2$-inch) and the corresponding welding cartridges. Two studs with cartridges fitted are shown on the right of the picture.

much the same as that of the stripped contact electrode already described. The heat generated by the arc melts the igniter from the inside; when the supporting ridge has melted away, the stud (which is under a certain pressure) shoots into the molten weld metal, the arc is shorted and the stud "freezes" to the work. Whilst the inside of the igniter is melting, its outside remains intact, thus keeping the arc length more or less constant until the stud breaks through the cup.

The cardboard collar serves two purposes. Firstly, it prevents the rather brittle semi-conductive material of the igniter from disintegrating too soon during the welding process and thus shortening the welding time unduly. Secondly, it holds the welding cartridge on the stud. To enable it to do so, the top edge is curled inwards to fit fairly tightly round the stud. The cardboard is stiff enough to prevent the cartridge from slipping off during handling of the stud.

The use of a cardboard collar is possible by virtue of the fact that the heat generated during the burning of the arc (about 0.5 sec.) does not reach the outside of the igniter until the stud is welded. The cardboard, having served its purpose, then catches fire.

The composition of the welding cartridge

The first experiments were carried out with welding cartridges similar in composition to the coating of a contact electrode. However, it was found that very much stricter requirements as regards freedom from gas and non-sensitivity to moisture must be imposed on the igniter than on the coating of welding electrodes, owing to the fact that no after-heating occurs as is necessarily the case in normal arc welding; the weld metal therefore sets very quickly. Consequently, the probability that any gas occluded in the metal will be able to escape is very much smaller than in ordinary arc welding [3]). The requirement that the igniters be free from gas necessitates heating them during manufacture to a temperature higher than normal for welding electrodes (baking). This must be done in a neutral or reducing atmosphere, to avoid oxidation of the metal powders in the material.

Experiments with different coating-constituents have shown that a low melting point should be avoided. This is not surprising, in view of the fact that the stud is released as soon as the supporting ridge in the igniter melts. If the igniter has a very much lower melting point than the stud, the latter will be released before the end of it is fully molten, and therefore pressed into the welding-pool prematurely.

However, it was found that an igniter composed partly of titanium dioxide (TiO_2) and partly of binding agents, deoxidising metals, and so on, meets all the requirements imposed by this method of welding. The desired conductivity is not obtained by including a certain amount of iron powder, as in contact electrodes. Instead, since the igniter must be baked in a neutral or reducing atmosphere, we take advantage of a well-known characteristic of TiO_2, that is, that in a reducing atmosphere it loses part of its oxygen content and in the process becomes somewhat conductive (semi-conductive). Since the conductivity thus imparted to TiO_2 depends upon the degree of reduction attained under the particular baking conditions, it can be regulated accurately within certain limits. The actual conductivity of the igniter is very critical; if it is too low, the arc will not strike, and if it is too high, too much of the welding current passes through the igniter, causing it to melt prematurely.

The welding gun

In principle, welding with the cartridges described here can be carried out by clamping the stud in a holder (very similar to the type employed in ordinary welding) and pressing it against the work, the welding current then being supplied through the holder. In practice, however, this method has serious objections:

[3]) See J. D. Fast, Causes of porosity in welds, Philips tech. Rev. **11**, 101-110, 1949/50.

1) It is impossible to place the studs really squarely on the work by hand.
2) When a number of studs of equal length are to be welded, it is very difficult to get them all at exactly the same height, since some of the studs are pressed deeper into the molten pool than others.
3) In the welding of studs the current cannot be switched on and off simply by moving the holder onto or off the work as in normal hand welding.

To avoid these difficulties, it was necessary to design a special holder, or welding gun. The principle of this device is shown in *fig. 6* and its operation explained in the caption.

The stud welding procedure is as follows. The studs to be welded are inserted in the welding gun, which is then adjusted to suit the particular stud length. When a number of studs of the same length (tolerance, say, ± 0.2 mm) are to be welded, only this first adjustment is, of course, necessary. Next, a welding cartridge is fitted on the stud, the gun is applied to the appropriate spot on the workpiece and pressed firmly against it. The current is then switched on by means of a switch built into the pistol grip. When the arc (operating period roughly 0.5 sec.) is extinguished and the weld completed, the switch is released, cutting off the short-circuit current. The welding gun is then retracted from the work and the fragments of the welding cartridge are removed from the stud. The interval between switching on and switching off is roughly 2 sec. *Fig. 7* shows the welding gun,

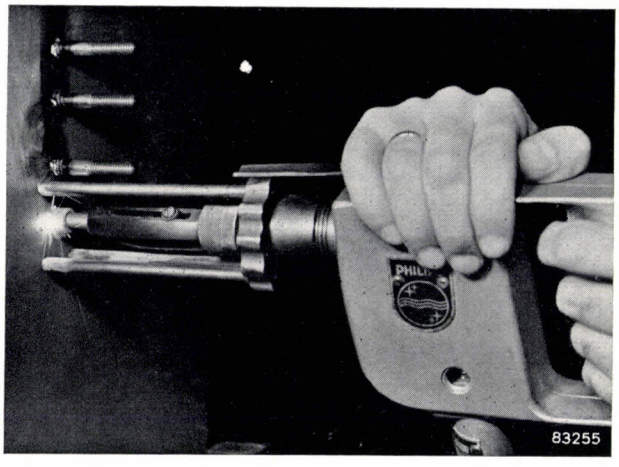


Fig. 7. Welding a $^3/_8$-inch stud to a vertical wall by the new method with a welding cartridge.

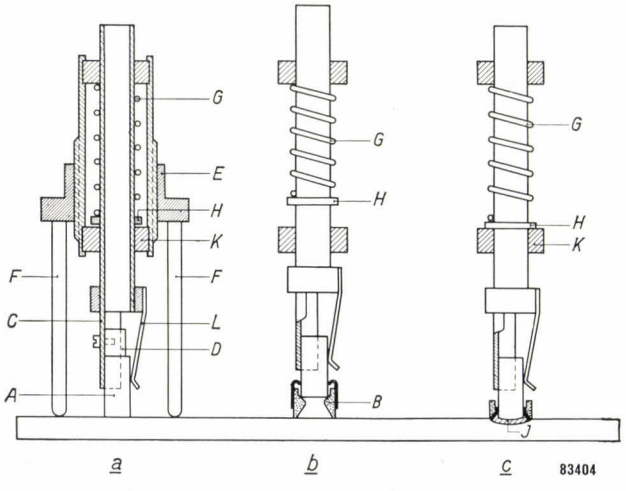


Fig. 6. Diagram to demonstrate the operation of the welding gun.
a) The stud *A* is clamped in a stud holder *C* by a flat spring *L* and rests against a stop *D*, adjustable to different stud-lengths (coarse adjustment). A threaded ring *E* (fine adjustment) enables the bottom face of the stud to be adjusted to a level about 0.5 mm beyond the plane passing through the ends of the supports *F*. This can be done with the aid of a simple jig, e.g. a flat plate about 0.5 mm thick with a hole, somewhat larger in diameter than the stud, at the centre; this plate is so positioned against the supports that the end of the stud just enters the hole. A spring *G* presses the stud on to the workpiece.
b) A welding cartridge *B* is fitted on the stud. As the gun is pressed against the workpiece, collar *H* is pushed upwards, thus compressing spring *G*. The current is then switched on and an arc is struck. When the part of the igniter supporting the stud melts, the stud is pressed on to the workpiece by spring *G*.
c) When once the stud is welded, collar *H* rests on stop *K*; this ensures that each stud is pressed equally far into the polten pool of metal. The amount of "excess" weld metal at *J* depends upon the distance to which the stud originally protudes (0.5 mm).

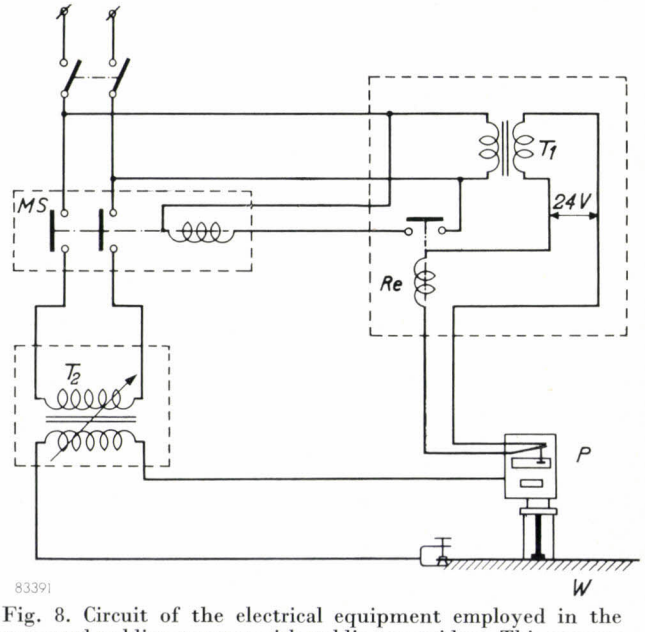


Fig. 8. Circuit of the electrical equipment employed in the new stud welding process with welding cartridges. This process enables alternating current to be used, which means that ordinary welding transformers (open-circuit voltage roughly 75 V) can be employed. A micro-switch built into the grip of the welding gun *P* operates the coil of a relay *Re*. Since the energising voltage of this relay is only 24 V (transformer T_1), the micro-switch can be touched without risk, even when defective. The relay operates a magnetic switch *MS* controlling the welding transformer T_2. *W* is the workpiece to which the stud is to be welded.

with stud and cartridge fitted, being used to weld a stud to a vertical plate. *Fig. 8* is the circuit diagram.

Comparison of the new technique with the conventional method

Comparing the new method of stud welding with the one usually employed hitherto, we see that the disadvantages associated with the latter are now eliminated. Firstly, the new method incorporates all the advantages of coated, as compared with bare welding electrodes. Amongst other things, the coating serves a metallurgical purpose (that is, not only in producing slag to react with the molten metal and the impurities in it, but also in introducing alloying constituents into the weld), thus enabling studs of ordinary commercial steel to be employed; hence it is no longer necessary to employ studs made of special steels. Another important task of the coating is to reduce the amount of nitrogen absorbed from the atmosphere. Filing tests on welded joints have shown that hardness owing to the formation of nitrides in the weld is very much less noticeable in the new, than in the old process. Moreover, the coating also contains substances which emit electrons readily when heated, thus stabilizing the arc. This arc stabilization enables low-frequency (say, mains-frequency) alternating current, and a relatively low voltage, to be employed. The advantages of A.C. welding (inexpensive welding equipment, less arc "blow") need not be pursued further here.

Secondly the welding cartridge itself strikes the arc and keeps the length of it constant throughout the welding period. Also it automatically leaves a clear path between stud and work at the end of the prescribed arc period. Accordingly, no special means have to be built into the welding gun and the associated equipment to achieve this behavior. The equipment is therefore relatively simple and inexpensive. It will be seen from fig. 8 that apart from a simple mechanism to grip the stud and press it home, the welding gun itself contains only one switch. The associated equipment, apart from the welding transformer T_2, includes only a relay *Re* with a transformer T_1 (both housed in the same box) and a magnetic switch *MS*.

Measurements on the process

To study the welding process stage by stage, the voltage, the current and the distance between stud and work were plotted as functions of time with the aid of a recording instrument. *Fig. 9* shows a diagram recorded by this instrument during the welding of a $\frac{1}{2}$-inch stud to a plate. Note the three distinct stages, which are revealed most clearly by the voltage diagram.

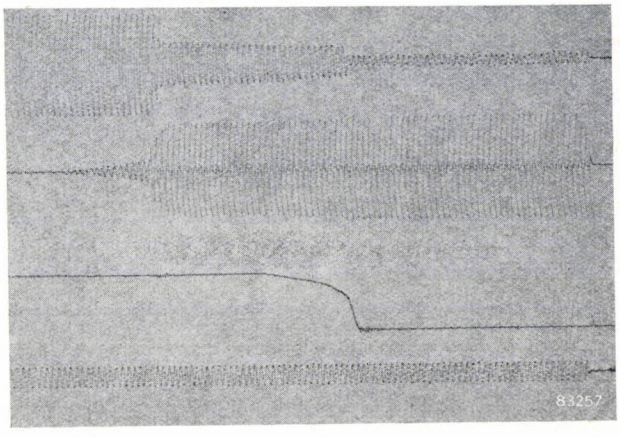


Fig. 9. Diagram recorded during the welding of a $\frac{1}{2}$-inch stud. Starting at the top we have, plotted against time: the voltage, the current, the displacement of the stud, and a constant 50 c/s A.C. voltage which serves as a time scale.

During stage one, the voltage is almost equal to the open-circuit voltage of the welding transformer (about 80 V). The current (ignition current in the conductive igniter) builds up gradually. Stage two begins with the striking of the arc and continues while the arc burns. The voltage drops to the arc voltage (about 40 V) and the current increases to roughly 700 A. The inside of the igniter body begins to soften about 0.4 sec. (20 cycles) after the striking of the arc; the softening is manifested by a gradual sinking of the stud (see third curve in diagram). When the whole of the stud-supporting ridge inside the igniter has melted, the stud is suddenly thrust into the molten pool of metal; this is indicated by the steep slope of the displacement curve in the diagram.

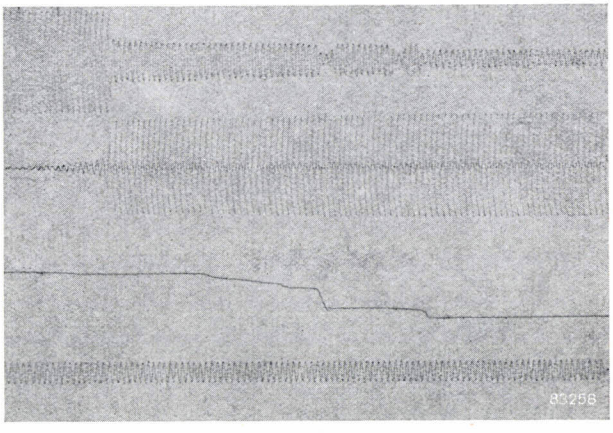


Fig. 10. As fig. 9, but recorded during the welding of a $\frac{1}{2}$-inch stud with a coarse thread on the end to be welded. The stud then drops intermittently, thus producing an inferior weld. Premature melting of the screw thread is a possible cause.

The stud and the plate are thus joined, and the current then increases slightly to the short-circuit value. It can then be switched off, and the gun can be withdrawn. However, it is advisable to wait some tenths of a second after the arc is extinguished before withdrawing the gun, so that the metal has time to solidify.

Diagrams such as that in fig. 9 are very useful for studying the welding process, in that they depict the effects of various variables. For example, *fig. 10* shows the trace obtained during the welding of a ½-inch stud, provided with a rather coarse thread at the end to be welded, under conditions otherwise similar to those of fig. 9. The most interesting feature of fig. 10 is the displacement curve, showing that here the stud drops intermittently instead of smoothly. It is possible that the thread resting on the supporting ridge inside the igniter begins to melt before the ridge itself has melted. Consequently the stud is enabled to drop slightly, causing a short-circuit as indicated by the voltage diagram. After about 0.02 sec. the arc is re-established, only to be extinguished again after a further interval of roughly 0.1 sec; it is again restored, and the stud then drops still lower, producing another short-circuit, and remains where it then is. The total displacement is less than in the preceding diagram, and investigation has shown that, in consequence, the stud is not properly welded to the plate. In practice, the operator can hear from the irregular sound of the arc that the welding is not proceeding satisfactorily.

The requirements imposed on the finish of the stud to ensure a good weld are not stringent. The end face should be round, smooth and free from burr and should be at right angles to the stud. *Fig. 11* shows the equipment employed to record the above diagrams.

Properties of the weld

Since the greater part of the weld metal comes from the stud and the plate, the properties and composition of the weld are governed partly by the composition of these components. However, deoxidizing agents incorporated in the igniter ensure that the weld metal solidifies as "killed" steel [4]) (*fig. 12*).

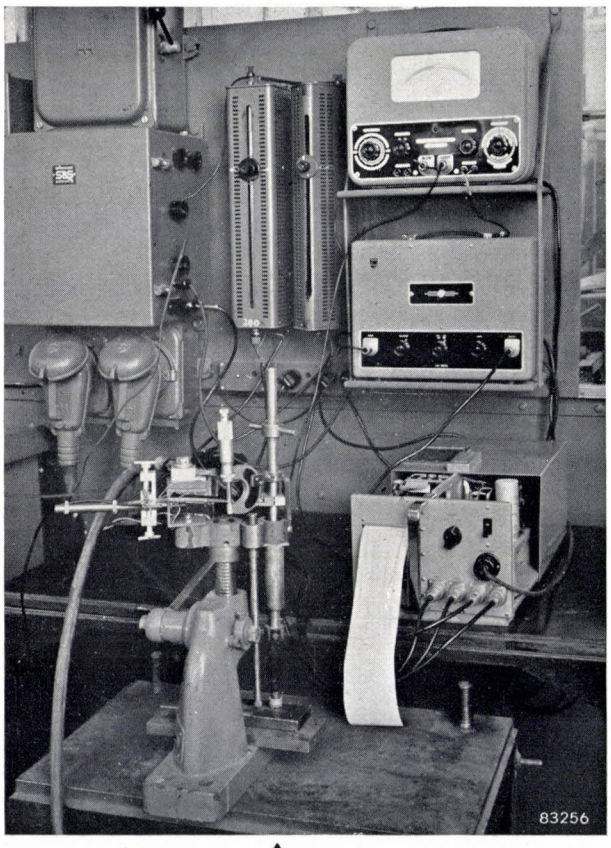


Fig. 11. The equipment employed to record variations in the voltage and the current, and in the distance between the stud and the work, as a function of time during the welding process. The ordinary welding gun was not employed during these measurements. The arrows indicate the position of the welding cartridge.

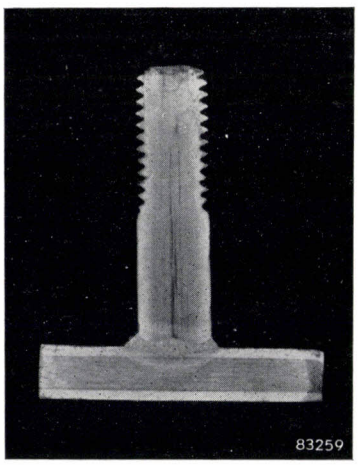


Fig. 12. Cross-section of a ½-inch stud welded to a plate by the cartridge method of stud welding. The slag enclosures revealed by etching show that the stud and the plate are both made of rimmed steel [4]).

The process described here is intended primarily for the welding of studs to unalloyed steel (C-content roughly 0.2%). Depending on the particular requirements to be met by the weld, it is advisable with hardening types of steel to pre-heat the parts to be welded, in order to reduce the cooling rate, and hence limit the hardening.

[4]) "Killed" steel is steel which generates very little gas when it sets after casting. The gas generated during the setting of rimmed steel causes segregations and porosity.

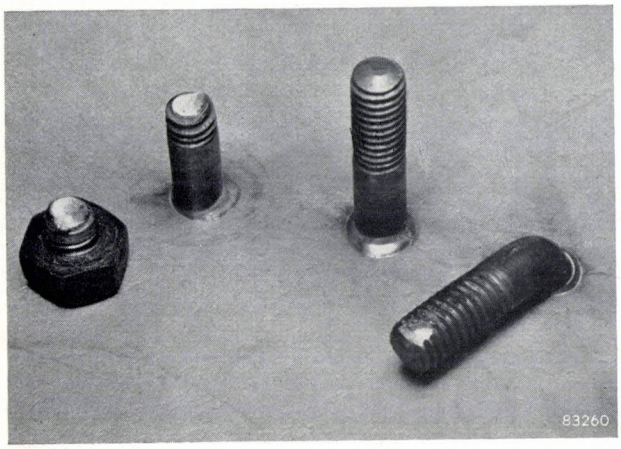


Fig. 13. Simple tests to determine the mechanical strength of the joint between stud and work. Right: a stud bent through 90° by hammering. Left: a stud sheared off by turning a nut against a metal sleeve (not shown in the photograph) fitted round the stud. The part of the stud broken off, with the nut still on it, is seen on the extreme left of the picture.

The simplest method of testing the strength of the weld is to bend the stud, say, by hitting it with a hammer. Experience has shown that a stud of mild steel can be bent in this way to an angle of 90° without cracking (*fig. 13*). With studs or plates of more brittle materials the plastic deformation possible is less.

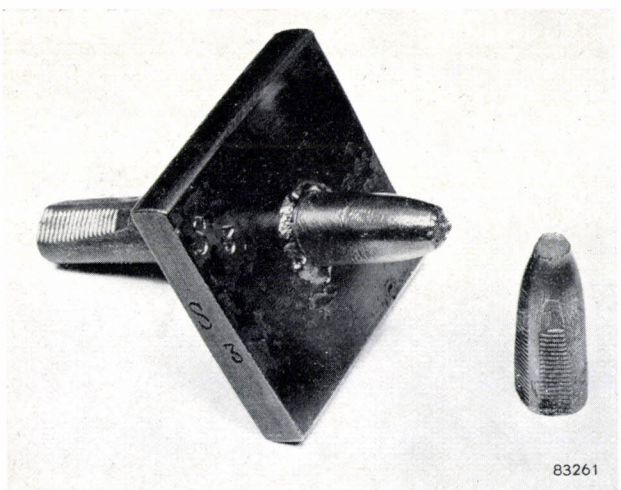
and a plate of mild steel, the weld itself will invariably remain intact and the stud will break where it is threaded, that is, unless the thread itself is sheared off (fig. 13).

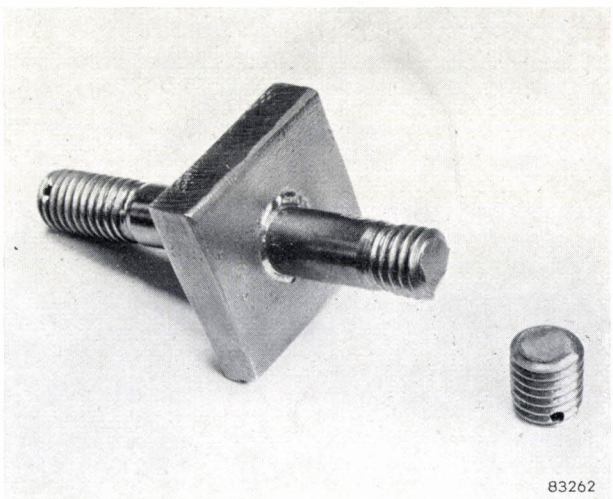
Studs without thread can also be subjected to tensile tests by welding two of them end to end to opposite sides of a plate and testing this assembly on a tensile machine. As in the other test, the weld itself normally remains intact and fracture occurs in another part of the assembly (*fig. 14a*).

In a similar way, fatigue tests can be carried out. A series of test pieces, pre-stressed under a tension of 0.25 kg/mm² (0.16 tons/in²), were subjected to alternating stresses of amplitude varying, as between the different tests, from 10 to 20 kg/mm² $6\frac{1}{2}$-13 tons/in²); for practical reasons, the studs used were threaded. These test pieces also fractured in the thread, that is, clear of the weld (fig. 14*b*). The fatigue limit of mild steel studs was found to lie in the region of 15 kg/mm² (10 tons/in²).

Some practical details

Welding cartridges are now made to fit four stud diameters viz. $^1/_2''$, $^3/_8''$, $^5/_{16}''$ and $^1/_4''$. The tolerance on the $^1/_2''$ studs is ± 0.3 mm and that on the others ± 0.2 mm; if the stud is too thin it will drop through





a *b*

Fig. 14. *a*) Result of a tensile test on two studs welded end to end to oposite sides of a plate.
b) Result of a fatigue test.

Another simple test is to shear off a partly threaded stud by fitting a short metal sleeve round the welded stud, screwing a nut on the stud and tightening it against the sleeve. If the nut is turned far enough, the stud ultimately breaks. With a stud

the cartridge too soon, thus shortening the welding time unduly, and if it is too thick the welding time will be too long.

With an ordinary welding transformer, the current settings required for the four stud-diameters

specified above are 750, 500, 450 and 325 A, respectively. These values apply to welding on relatively thick plates. For thin plates (thinner than about half the stud diameter), the above values should be reduced.

The open-circuit voltage of the welding transformer should be at least 70 V. As we have already remarked in connection with hardening effects, the requirements imposed on the steel to be welded and the material of the stud are not very stringent. Excessive slag enclosures should be avoided, since they may make the weld unduly porous. However, the sulphur content, for example, need not be particularly small, and, provided that the weld is not required to be above average toughness, sulphur alloyed free-cutting steel can also be welded by this technique.

Summary. A new technique for welding studs perpendicularly to a workpiece is described. It involves fitting a welding cartridge on the end of the stud to be welded. This cartridge is made of semi-conductive material; its special functions are: to strike the arc, fix the distance between stud and workpiece and regulate the welding time. The cartridge also fulfils the metallurgical and other functions of the coating of an ordinary welding electrode. By virtue of the metallurgical effect of the cartridge on the weld, it is possible to employ studs of ordinary commercial grade steel. The stabilizing effect of the cartridge on the arc enables alternating current to be employed. The welding gun to position the studs correctly on the work incorporates only simple devices to hold the stud and press it against the work, and a switch to turn the current on and off. A relay and transformer, together with a magnetic switch to control the welding transformer, are all the auxiliary equipment required.

AN X-RAY TUBE FOR MICRORADIOGRAPHY

621.386.1:778.33:539.23

For some decades various investigators have been working on the problem of making X-ray photographs of very thin objects, such as insects, sections of animal and plant tissue, metal foils, paper, etc. In view of the small X-ray absorption in such thin objects, very soft X-rays must be used in order to get sufficient contrast; the X-rays must therefore be generated by a rather low voltage X-ray tube (3—10 kV). However, the softness of the X-rays means that very thin tube windows are then necessary e.g. thin aluminium foil. Owing to the difficulty of making such windows vacuum-tight, it has hitherto been necessary to work with a continuously-evacuated tube [1][2]).

Following up earlier work [3][4]), it has proved possible to make a vacuum-tight beryllium window only 50 μ thick. This is sufficiently transparent to X-rays in the range between 1.5—5 kV (2—15 Å). This window has been used in a small 5 kV sealed-off X-ray tube, a cross-section of which is shown in *fig. 1* and a photograph in *fig. 2*. The tube is incorporated in a small portable apparatus, the CMR 5, which includes the H.T. generator, controls and the camera.

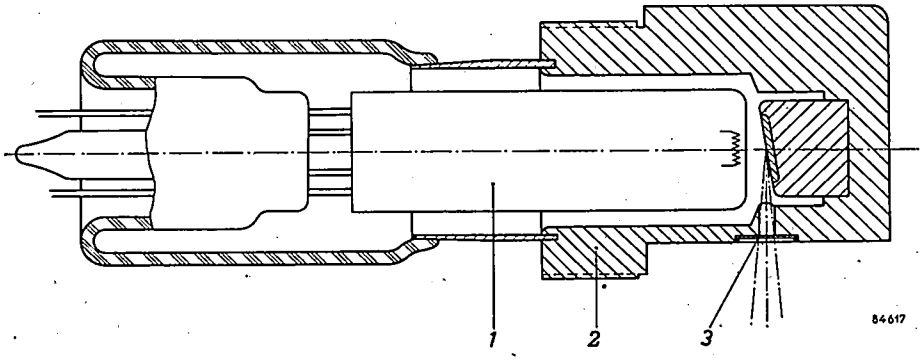


Fig. 1. Cross-section of the 5 kV X-ray tube used in the CMR 5 apparatus. *1* cathode with tungsten filament; *2* anode with tungsten target; *3* beryllium window, 50 μ thick. Length of tube approximately 8 cm.

The sample (mounted on a formvar film across a ring) is placed in contact with the photographic emulsion in the camera: whence the term *Contact Microradiography* (CMR), which has been coined

[1]) P. Lamargue, L'historadiographie, Presse Méd. 478 (1936)
[2]) A. Engström, Quantitative micro- and histo-chemical elementary analyses by Roentgen absorption spectrography. Acta Radiol. Suppl. 63 (1946).
[3]) B. Combée en P. J. M. Botden, Special X-ray tubes, Philips tech. Rev. **13**, 71-80, 1951/52.
[4]) P. J. M. Botden, B. Combée en J. Houtman, An experimental X-ray apparatus with a midget X-ray tube, Philips tech. Rev. **14**, 165-174, 1952/53.

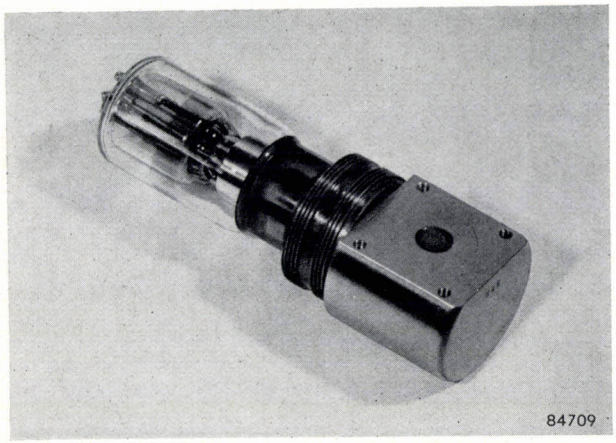


Fig. 2. Photograph of the 5 kV X-ray tube.

for this technique. The full-size radiograph thus obtained is afterwards enlarged optically some hundred times in a light microscope.

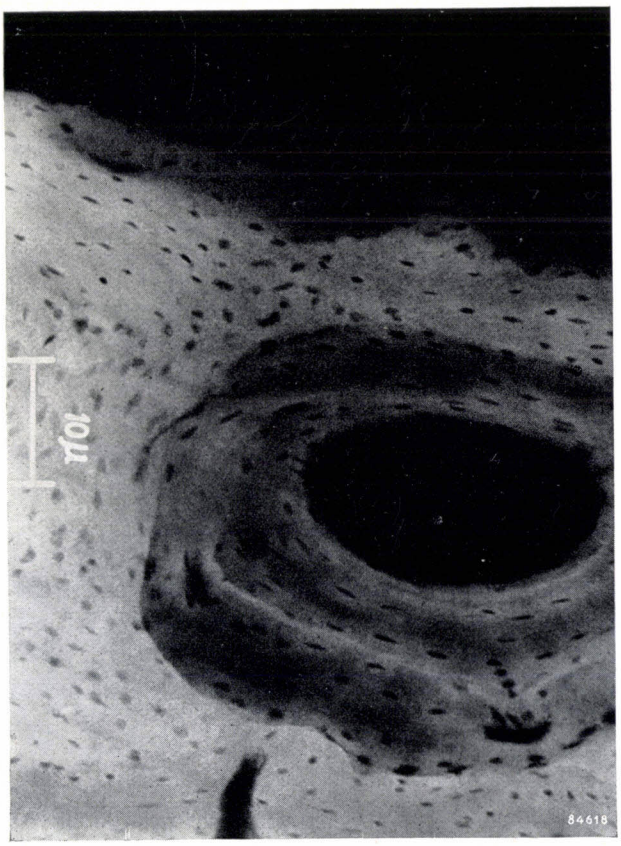


Fig. 3. Bone tissue. To make such a photograph it is unnecessary to de-calcify the bone as in light microscopy; hence it is possible to obtain information on the distribution of mineral salts in the bone (5 kV, 3 mA. Exposure 25 min.)

The resolution achieved with this apparatus is between 0.5 and 1 μ. It is true that this is not quite so good as that obtained with optical microscopes, but because of the much better transparency of materials to X-rays, it is possible to examine many objects which are outside the scope of optical methods.

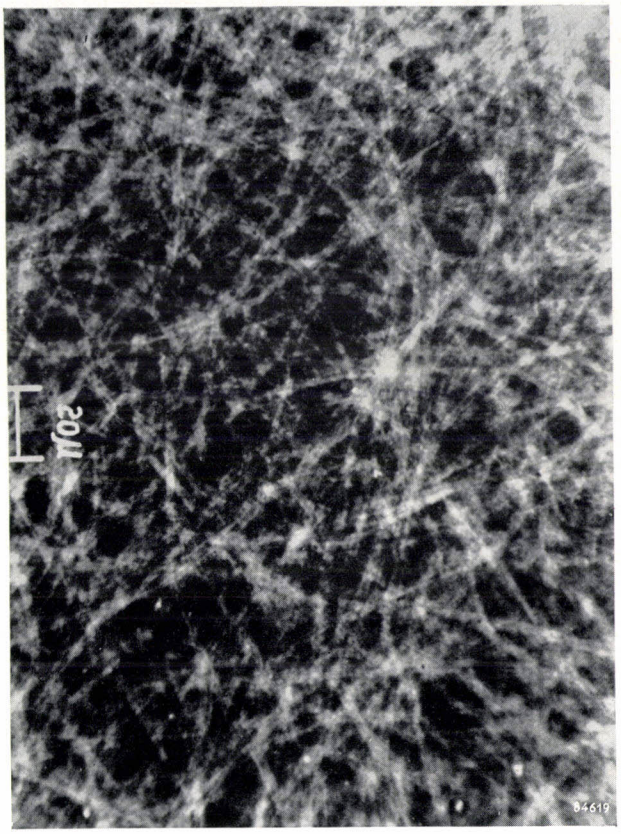


Fig. 4. Paper. The internal texture of the paper is rendered visible by the penetrating power of the X-rays (5 kV, 3 mA. Exposure 10 min.)

The present apparatus, whose price is comparable to that of a good optical microscope, thus brings microradiography within the reach of many workers for whom the construction of an elaborate continuously-evacuated tube was out of the question. X-ray microradiography undoubtedly has a very useful role to play in the region between optical and electron microscopy.

In a later article in this Review these questions will be discussed in some detail.

Meanwhile, the micrographs shown here (figs 3 and 4) illustrate a few of the possible applications of the technique.

B. COMBÉE

THE EM 75 kV, AN ELECTRON MICROSCOPE OF SIMPLIFIED CONSTRUCTION

by A. C. van DORSTEN and J. B. Le POOLE *).

621.385.833

The electron microscope in its present form is still a complicated and costly instrument. This restricts the number of potential users. Attempts have therefore been made by some designers to simplify the instrument, while retaining as far as possible its valuable features. In this connection, the aims of simpler construction and simplified operation run parallel to the aim of reduced costs.

The magnetic electron microscope described here is basically simpler than most instruments of this type, particularly with regard to the electrical equipment. It does not contain a single amplifying tube. It has, however, apart from other valuable properties, a high resolving power (at least 100 Å).

Introduction

In addition to the EM 100 kV, a magnetic electron microscope with a very high resolving power, Philips have for some time been making a smaller and simpler electron microscope, the EM 75 kV. In developing this instrument it was postulated that a slightly lower resolving power would be acceptable, viz. about 100 A. It was found possible to make the instrument so much cheaper and so much simpler to use that electron-microscopic research may now be expected to be brought within the reach of a much wider circle of scientific research workers than hitherto.

The basic principle of the new apparatus is the use of an objective lens with an extremely short focal length. In this article it will be shown that in this way an electron microscope can be built which, for a given resolving power, sets relatively low requirements as to the maintenance of constant current and voltage, in other words, the aberrations of the image which result from electrical fluctuations are relatively small. By analogy with the optical microscope we may call such aberrations "chromatic" (they arise because the adjustment of the microscope is not correct with respect to the velocity of the electrons, i.e. the wavelength of the electron beam).

It is perhaps of interest to mention in this connection that small chromatic aberration resulting from a short focal length is not a property of electron lenses alone but is equally applicable to the case of optical magnification by means of a glass lens. It is particularly true when one is considering a single uncorrected lens. This was the secret of the microscopes made by Antoni van Leeuwenhoek in the 17th Century: he reduced the faults of the single glass lens to a minimum by making lenses of minute dimensions, with focal lengths as short as possible.

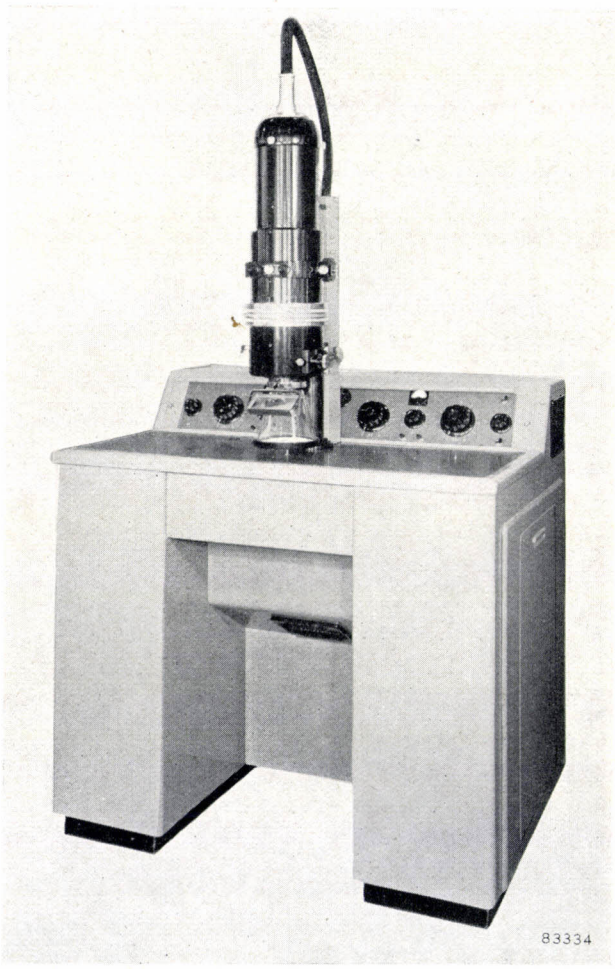


Fig. 1. The small electron microscope EM 75 kV (for 75 kilovolts, type 11981). The electron beam traverses the vertically mounted microscope tube from top to bottom. The desk houses the pump installation and the entire electrical equipment.

*) Applied Physics Service, Netherlands Organization for Industrial Research, and Technische Hogeschool, Delft.

The principles and several practical forms of the magnetic electron microscope have previously been described in this Review[1][2][3]. In this article we shall mainly focus attention on those points in which the new instrument differs from others of this sort. The instrument is shown in *fig. 1*. It is a two-stage microscope, containing three lenses: condenser, objective and projector. We shall now consider in more detail the objective lens, which contains the clue to the new instrument, and the projector.

Focal length and chromatic aberration

A magnetic electron lens consists of a magnetic field, rotationally symmetrical about the axis of the incident electron beam (z-direction, *fig. 2*) and active over a section, generally short, of the

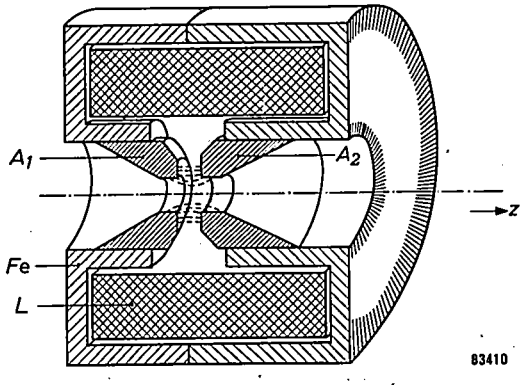


Fig. 2. Magnetic electron lens. L coil, *Fe* iron jacket, A_1, A_2 pole pieces. The electrons move close to the z-axis, which is the axis of rotational symmetry of the magnetic field of the lens.

electron path. The field is generated by a current-carrying coil of so many ampere-turns (in some types the field is obtained by means of a permanent magnet which is equivalent to a number of ampere-turns). The coil is surrounded by a housing of ferro-magnetic material and in the centre are pole pieces of similar material. In this way the field is concentrated in the small gap between the pole pieces, the electron beam passes through a small circular hole drilled through the centre.

The focal length f of such a lens for electrons which have been subjected to an accelerating voltage V, is given approximately by the proportion relationship (see e.g. the article quoted in [1]):

$$\frac{1}{f} \propto \frac{1}{V} \int_{-\infty}^{+\infty} B^2 \, dz. \quad \ldots\ldots (1)$$

B is the induction on the z-axis, along which the integration is carried out. The formula for f actually consists of a series of terms of alternating sign and increasing powers of B^2: in the case of a "thin" lens of moderate power the first term is sufficient.

Small fluctuations ΔV of the voltage V result in changes Δf in the focal length. From (1) it follows that:

$$\Delta f = f \frac{\Delta V}{V}. \quad \ldots\ldots (2)$$

A similar equation applies to the fluctuations ΔI of the energising current I, which cause fluctuations in the magnetic induction B at all points of the field. If we now consider in particular the objective lens, we can say that the object plane, i.e. the plane which is sharply focused on the fixed fluorescent screen, is not fixed in space but moves to and fro along the axis (while the principal plane of the lens remains more or less in the same place). The object placed in the microscope (specimen) will thus in general be situated at a distance Δv in front of or behind the object plane(*fig. 3*). In this way the "chromatic" blur appears either in visual observation or on the photographic record.

If the magnification is sufficiently large, the object plane nearly coincides with the focal plane; thus

$$\Delta v \approx \Delta f. \quad \ldots\ldots (3)$$

What was stated above is now qualitatively confirmed. In any case, the chromatic lack of definition will become greater according as the object plane is displaced. Now, from equations (3) and (2), Δv is proportional to f. Therefore, by giving the objective a very short focal length, one can permit relatively large fluctuations of voltage and current for a given chromatic error.

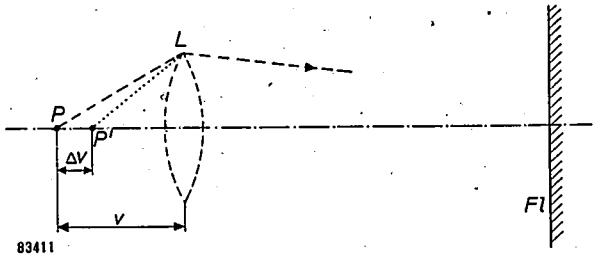


Fig. 3. If the acceleration voltage of the electrons or the excitation current of the objective L vary, the image on the fluorescent screen Fl is not that of the object plane P, but is the image of a plane P' displaced by Δv from the object (v is the distance between object and objective lens).

[1]) J. B. Le Poole, A new electron microscope with continuously variable magnification, Philips tech. Rev. **9**, 33-45, 1947/48.

[2]) A. C. van Dorsten, W. J. Oosterkamp and J. B. Le Poole, An experimental electron microscope for 400 kilovolts, Philips tech. Rev. **9**, 193-201, 1947/48.

[3]) A. C. van Dorsten, H. Nieuwdorp and A. Verhoeff, The Philips 100 kV electron microscope, Philips tech. Rev. **12**, 33-51, 1950/51.

In order to get a more accurate picture of this relationship, the chromatic blurring can be analysed into two components: a geometrical error, δ_g, and a diffraction error, δ_b. It can be shown that the former is given by:

$$\delta_g = 2a_0 \Delta v, \quad \ldots\ldots \quad (4)$$

and the latter by:

$$\delta_b = 3\sqrt{\lambda \Delta v}. \quad \ldots\ldots \quad (5)$$

where λ is the wavelength of the electron beam and $2a_0$ the angular aperature of the electron pencil on the object side. This aperture has a different value for each point of the object, depending on the local thickness and density of the specimen. In fact, this is just what makes the image in the electron microscope visible, i.e. gives contrast to the image: the scattered pencils of electrons originating from the various parts of the object have various angular apertures, so that the objective aperture intercepts a different fraction of each, resulting in differences in intensity when they are united in the image plane. For composite (not uniformly thin) objects, such as are usually studied with the electron microscope, "illuminated" by a practically parallel beam of electrons, we can assume a maximum value of $a_0 = 3 \times 10^{-3}$ radians.

The origin of geometrical blur is illustrated in *fig. 4*. The image of the object, due to the change of focal length, does not lie in the plane of the screen, but at a distance Δb from it (b is the image distance, v the object distance). A point in the object is therefore represented on the screen by a small circle of diameter $2r_i \approx 2a_i \Delta b$, $2a_i$ is the (very small) angular aperture of the electron pencil on the image side. If the magnification be M ($= b/v$), a detail in the object of dimensions $2r_i/M$ cannot be distinguished in the image from the blur circle. This is the blur δ_g, related to the object ("resolving power"). By differentiating the geometrical optical lens formula

$$\frac{1}{v} + \frac{1}{b} = \frac{1}{f}$$

it follows that the relationship between Δb and Δv is:

$$\Delta b = M^2 \Delta v.$$

Hence, using the law of Helmholtz-Lagrange $a_i = a_0/M$, one obtains equation 4:

$$\delta_g = \frac{1}{M} 2r_i = \frac{1}{M} 2a_i \Delta b = 2a_0 \Delta v.$$

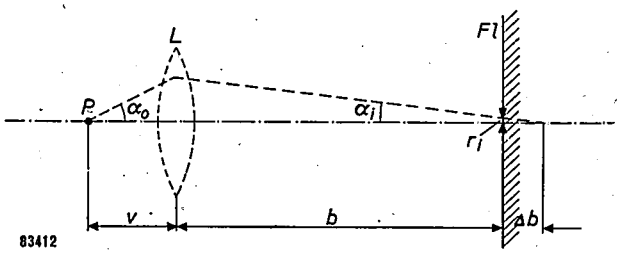


Fig. 4. Illustrating the geometrical blur, δ_g. P specimen, L electron lens, Fl fluorescent screen. (The projector lens can here be neglected.)

The origin of the chromatic diffraction blur can most simply be demonstrated for a "non-transparent" object detail with a sharp edge (*fig. 5*). Behind this detail the scattering is so great that practically no electrons pass through the aperture; beside the object detail the nearly parallel incident rays pass

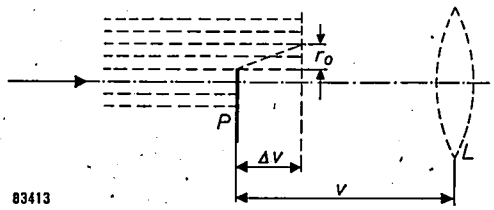


Fig. 5. Illustration of the diffraction blur, δ_b. A parallel electron beam falls on an opaque object P from the left. The "shadow" which the object throws on the real object plane at a distance Δv, is sharply reproduced on the fluorescent screen by the lens L. The shadow itself is not sharp as a result of diffraction of the electron rays at the edge of the object.

uninterrupted. In the object plane (P' in fig. 3) we have a shadow of the object, and this shadow (which is reproduced truly in the image plane) would be exactly similar to the object at a very small distance Δv, were it not that diffraction effects occur at the edge of the object due to the wave nature of the electron beam. The first order diffraction occurs at a distance r_0 from the edge. We can determine r_0 by applying Huygens' principle which states, for this case: if an object stands in a plane wave causing parts of the wavefront to be absent in some plane beyond the object, then a maximum disturbance will occur at those parts of that plane where rays passing through a point on the edge of the object have a path difference of the order of λ or greater with respect to the undisturbed wave. Thus from fig. 5:

$$\sqrt{r_0^2 + (\Delta v)^2} - \Delta v = \lambda,$$

i.e.

$$r_0^2 = 2\lambda \Delta v + \lambda^2.$$

Since $\lambda \ll \Delta v$ (otherwise the displacement Δv itself could be neglected), we obtain $r_0 = \sqrt{2\lambda \Delta v}$. The value $2r_0$ gives approximately the width of the Fresnel diffraction fringe associated with the image of a point object, i.e.:

$$\delta_b = 2r_0 \approx 3\sqrt{\lambda \Delta v}.$$

Two particles can be resolved separately although their respective diffraction rings partially overlap. The distance, centre to centre between the smallest resolvable particles is, according to Rayleigh's criterion, about $\frac{1}{2}\delta_b$.

From (4) the geometrical blur δ_g is proportional to Δv, and therefore also proportional to the focal length f (see equations 2 and 3); from (5) the diffraction blur δ_b is proportional to $\sqrt{\Delta v}$, i.e. proportional to the root of the focal length. Considered quantitatively with the value of the angular aperture (a_0) already given, and for a wavelength λ corresponding to an acceleration voltage $V = 75$ kV and for an objective of focal length $f = 0.8$ mm (as achieved in the EM 75kV), we find:

a) that the geometrical blur δ_g which must be taken into account in practice, is larger than the diffraction disc δ_b;

b) that for a geometrical blur of 70 Å, the relative variation of the voltage, $\Delta V/V$ may not exceed 1.5×10^{-3}. Of course the voltage is not required to remain permanently constant to this degree, but only long enough after focusing to expose the film. Taking into account other causes of lack of sharpness which are not dependent on voltage fluctuations (spherical aberration, astigmatism of the lens, etc.), the above value of δ_g permits a resolving power better than 100 Å.

Since the focal length of a magnetic lens is dependent on the quantity I^2/V (see equation 1), the requirement of a constant value of the lens current I is twice as strict as for V; in the case under consideration, therefore, the maximum permissible value [4]) of $\Delta I/I$ is 0.75×10^{-3}.

This can be quite easily attained with a simple magnetic stabilizer, consisting of a transformer with saturated core. Electronic stabilization is not necessary for this purpose.

In following this explanation the reader may very well have wondered why we do not use permanent-magnetic lenses to avoid fluctuations. The construction of such lenses with modern permanent magnetic materials is quite feasible [5]), and the problem of keeping the magnetic induction in the lenses constant is then solved. The problem of keeping constant the acceleration voltage, which is technically more difficult and more expensive to solve, remains, however, and is even made more difficult, since it is not possible to make objective lenses with very short focal lengths when using permanent magnets, and the influence of chromatic aberration is therefore relatively large.

Design of an objective of very short focal length

It has been stated above that the objective of the EM 75 kV has a focal length of 0.8 mm. Is it not possible to make this still shorter in order to make the sensitivity to fluctuations still smaller? We shall show that the value given is just about the limit of what can be achieved without undue expense — and that, after all, is the whole point of the thing.

[4]) Fluctuations of I as well as of V are caused by variations of the mains voltage. Since an increase in I has the opposite effect on the focal length ot that of an increase in V, one might suppose that fluctuations of I and V would compensate each other partically to such an extent that it would be sufficient to limit relative fluctuations of the mains voltage (to which I and V are proportional) to 1.5×10^{-3}. This reasoning, however is, incorrect since the power supplies for the lens currents and the acceleration voltage involve different time lags, so that correlated fluctuations ΔI and ΔV do not occur simultaneously.

[5]) B. von Borries, Kolloid Z. **114**, 164-167. 1949. J. H. Reisner and E. G. Dornfeld, J. appl. Phys. **21**, 1131-1139, 1950.

The parameters which are important in this connection are: the choice of the steel for the pole pieces, the length S of the gap between the pole pieces (*fig. 6*), the diameter D of the cylindrical hole in the pole pieces and the excitation of the coil.

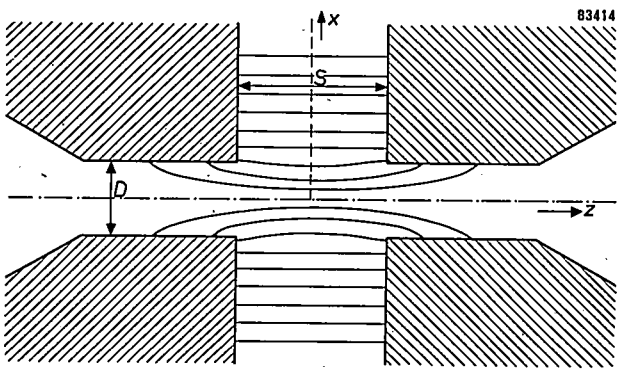


Fig. 6. The configuration of a magnetic lens is characterised by the diameter D of the bore through which the electrons pass (z-direction), and the width S of the gap between the pole pieces.

From equation (1) it may be deduced that a short focal length will be obtained if the induction B is made as great as possible over as long a section as possible of the electron path. For a given lens, B is made greater at all points if the excitation is made stronger. The limit to this is set by the magnetic saturation of the steel. The induction in the pole pieces must remain so far below saturation that the permeability μ is still large, e.g. 1000. At greater induction, for example, so close to saturation that the permeability falls below 50, the types of steel suitable for this use show local magnetic inhomogeneities. Since the induction immediately outside the iron is $1/\mu$ times that in the iron, the field acting on the electrons will also be inhomogeneous and lack the necessary rotational symmetry. Such defects of the field can not only make the image astigmatic but can also result in small deviations of the electron pencil. This latter effect means that the axis of the lens is not straight and that therefore no adequate centring independent of the excitation can be obtained. The consequences of inadequate centring will be discussed later.

The cobalt-iron alloy which we have used has a very high saturation, so that we have been able to raise the excitation to give an induction in the gap between the pole pieces of about 1.8 Wb/m² (18000 gauss). At the critical points in the magnetic circuit μ is then still greater than 1000.

The problem is thus reduced to the core: the choice of the width of the gap S and the bore D.

Consider now the variation of the magnetic

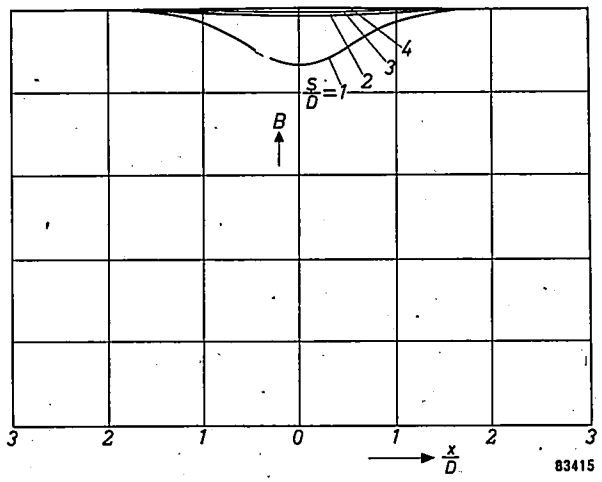


Fig. 7. Variation along the x-axis, see fig. 6, of the magnetic induction B, for lenses with differing values of S/D (the change of field with respect to x/D depends only on this ratio). It has been assumed that each lens is excited so that the maximum induction occurring in the gap reaches a given value (at which no troublesome saturation of the steel of the pole pieces occurs).

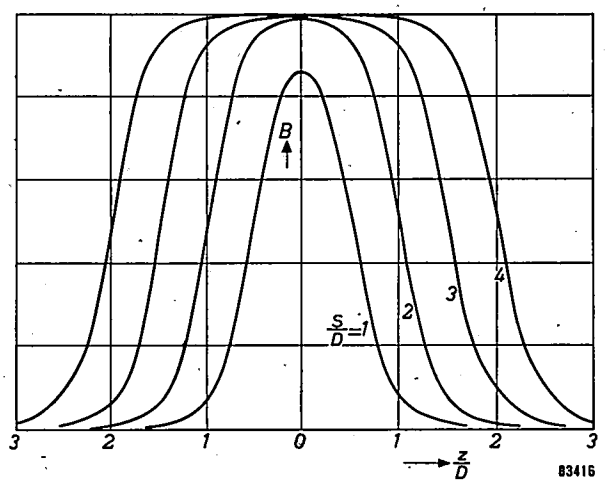


Fig. 8. Variation of the magnetic induction B along the z-axis, for various values of S/D. The induction at $z = 0$ is equal to that at $x = 0$ in fig. 7; the field here is at a saddle point.

field in two mutually perpendicular directions, viz. along the axis of the lens (the z-axis) and perpendicular to it along an arbitrary radial axis in the plane of symmetry of the lens (x-axis, see fig. 6). The variation of the field is dependent only on the proportion S/D and can be found either by calculation or by measurement [6]). In *fig. 7* the variation along the x-axis has been drawn and in *fig. 8* that along the z-axis, both for a series of values of S/D.

It can be seen that the field in the centre of the lens exhibits a saddle point: for the x-direction, the induction at this point is a minimum, for the z-direction, maximum. Consider that the excitation is raised to the level which gives the maximum induction in the gap, a magnitude of 1.8 Wb/m^2 (fig. 7). From fig. 8 it appears that for a given width of gap S the bore D must be made small, or for a given D the width S must be large in order to achieve a strong magnetic field along as great a section of the electron path (z-axis) as possible. A more complete picture is given by *fig. 9*. Here, for each combination of values of S and D, the focal length corresponding to the maximum permissible excitation is plotted (see above). The figure makes it clear that for very different values of D, the same small focall ength can be realised if S is only made large enough. Now the number of ampere turns needed to produce the required induction of 1.8 Wb/m^2 is approximately proportional to the gap width S (the gap presents by far the greatest resistance in the magnetic circuit). To achieve an economical construction, one is therefore forced to make the bore as small as possible.

A limit is set to this by the optical curvature of the field. The electron beam must have a cross-section large enough to "illuminate" a section of the object which is sufficient to fill the fluorescent screen at the smallest magnification with which the micro scope is to be used. To avoid an excessive curvature of the field, the cylindrical hole in the

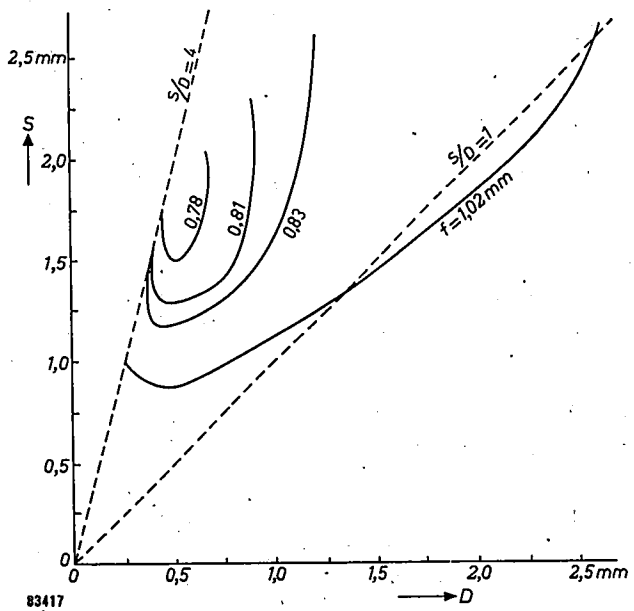


Fig. 9. Focal length f, calculated for lenses of various bore D and gap S when the lens is excited so as to give a maximum induction in the gap of 1·8 Wb/m^2. The acceleration voltage of the electrons is fixed at 75 kV. Points with equal f have been joined in the diagram. (The curves turn upwards slightly at the left-hand side and appear to bend backwards a little at the right; it is not clear whether these effects are real or arise from the approximations in the calculation, but they are of little importance in the S-D range under consideration.

[6]) For a method of measurement, see e.g. A. C. van Dorsten and A. J. J. Franken, Philips tech. Rev. 15, 52-55, 1953/54.

pole pieces must be made wide with respect to the necessary cross-section of the beam. On the basis of these considerations [7]), the objective of the EM 75 kV was given a bore of $D = 0.5$ mm.

The gap width S in this lens is 2 mm. These values give the focal length of 0.8 mm already mentioned.

There is not much point in making S greater than 2 mm; this may be seen from fig. 9; the focal length decreases to a limiting value (and even appears to increase again). At $S = 1.3$ mm the minimum value has already been nearly reached. The further ampere-turns needed to make $S = 2$ mm are thus, in fact, not justified merely to obtain the small additional decrease in f, but rather to allow room between the pole pieces to insert the specimen and objective aperture (see below figs 12 and 15).

Let us consider for a moment the decrease in f to a limiting value mentioned above. Fig. 8 does not give any indication of such behaviour: for a fixed D and increasing S/D, high inductions approaching 1.8 Wb/m^2 are obtained over increasingly long sections of the axis, which according to equation (1), should lead to still smaller values of f. But for such "thick" (or rather, long) lenses as we are now considering, we can no longer apply equation (1) which had been simplified to a single term. To understand the behaviour of f for increasingly long lenses, we can best follow the path of an electron which, for example, enters the magnetic field from the left and parallel to the axis. In *fig. 10* we see the paths for three lenses. The electron crosses the axis of the lens at some point. In the case of an *objective* lens one should imagine the object placed at (or, rather, very close to) the point of intersection mentioned above and a scattered electron leaving the object and travelling to the left along the path drawn. The effective focal length is found by the dotted construction. It is seen in the figure that for strong lenses theobject has to be placed some distance within the lens ("immersion lens"), and it can be furtherseen that an extension of the lens has thus very little effect on the focal length of the objective [8]).

In using such thick lenses we must also consider whether our basic assumption is still valid: viz. that the chromatic aberration of the objective is smaller according as the focal length is shorter. Variation of the excitation current I or the acceleration voltage V, which can be combined into an "excitation factor"

$$k = \frac{(\text{Amp. turns})^2}{V},$$

have, for a given lens, an analogous effect to the

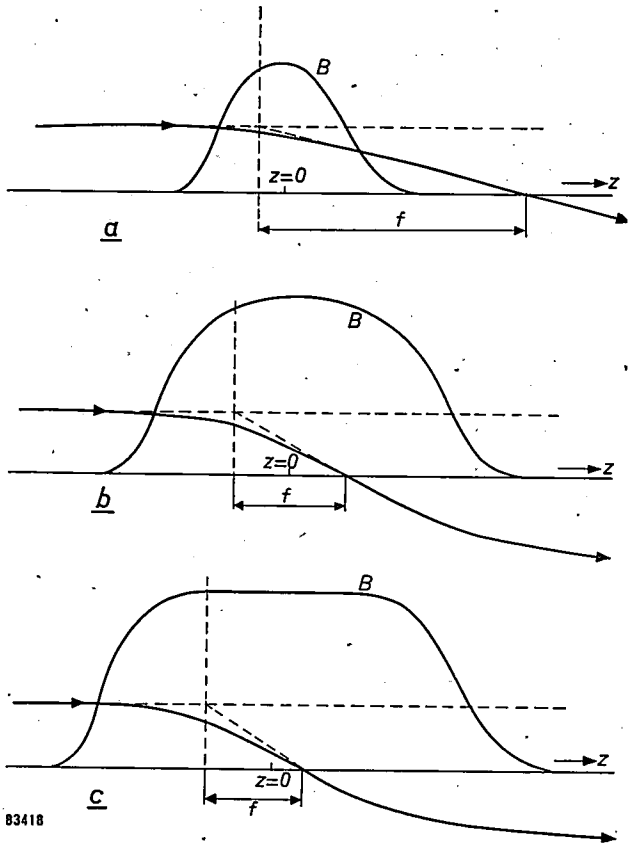


Fig. 10. Path of an electron coming from the left and entering the magnetic field of an electron lens parallel to the axis. The change of induction B is also shown. *a*) Relatively "thin" lens (S/D small). *b*) Longer lens. *c*) Very long lens.
The focal length, i.e. the distance between the focal point and the principal plane is found, for the case of an objective, by drawing the tangent to the electron path at the point where it intersects the axis and finding its intersection with the produced incident path of the electron; this latter intersection defines the position of the principal plane (drawn with a broken line). (For use as a projector, on the other hand, the focal length is found by substituting the linear path of the electron outside the field for the tangent. By producing this linear path backwards one can find the focal point and the principal plane respectively from the intersections with the axis and with the produced incident path.)
The component of the path perpendicular to the plane of the drawing (rotation of the image) has been neglected it is not of importance for the consideration of the focal point.

[7]) For a fluorescent screen 90×90 mm and the minimum magnification of 1200×, the largest object diameter is a good 100 μ, i.e. about 20% of the bore. In our case, the corners and centre of the object do not lie more than 1.5μ outside the curved object surface. This is about the same Δv as calculated from equation (4) for $\delta_g = 70$Å and $\alpha_0 = 3 \times 10^{-3}$ radians.

[8]) If the object is situated very deep within the objective, the electron rays, before falling on the object (coming from the right in fig. 10), must pass through a large section of the field of the objective, which then behaves (undesirably) as a condenser. This is a further reason for not making S too large. In our case the "condenser action" is not disturbing: the extremely small angle of the (nearly parallel) illuminating electron beam is increased only about 15% and so remains small with respect to α_0 (see equation 4).

variation of the length of the lens which has been discussed above. For a particular lens with various excitation values k, figures are obtained analogous to fig. 10*a* and *b* (except that the *shape* of the field pattern does not change). From this it is plausible that at large values of k, changes in this quantity have relatively less effect on f. This does not mean that chromatic aberrations, which result from fluctuations in k, are correspondingly smaller. In fact, the variation of k now results also in a displacement of the principal plane of the lens: though f may alter less, an extra lack of sharpness occurs because the lens, as it were, is displaced with respect to the fixed object [9]). The resulting chromatic blur (only δ_g, which is the more important) is plotted in *fig. 11* as a function of k for a number of objectives with differing values of S/D.

From fig. 11 it can be deduced that the conclusions drawn from fig. 9 do not have to be revised: for the chosen bore D, a wider gap S (provided the same induction is obtained which means choosing k roughly proportional to S^2) results in smaller aberration, and there is little purpose to be served by making $S>2$ mm. However, it is of great importance, as shown in fig. 11, that the lens, once constructed with $S/D = 4$, be always excited to the permissible limit, i.e. where $k \approx 170$.

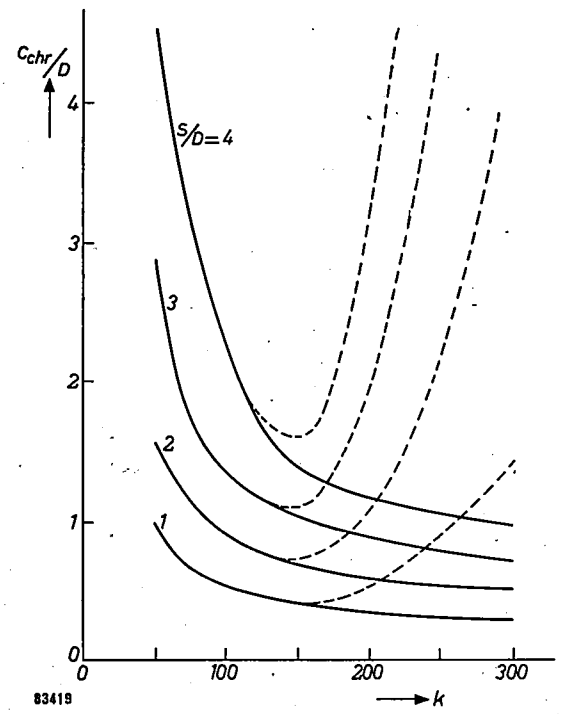


Fig. 11. Chromatic blur for a number of lenses with different S/D, as a function of the excitation factor k. C_{chr} is the proportionality factor between δ_g and $2\alpha_0$, which must be substituted in equation (4) for Δv in more accurate calculations. C_{chr}/D has been plotted to give generally applicable curves. The acceleration voltage is here fixed at 80 kV. These curves may only be used up to that value of k which gives the maximum permissible induction in the gap between the pole pieces.
The full lines apply to the lens when used as an objective ("immersion lens"), the broken lines to use as a projector.

[9]) Furthermore, the accompanying change of magnification also plays a part. For this see: J. B. Le Poole, Some designs in electron and ion optics, Diss. Delft 1954, p. 47.

Design of the projector lens

Returning for a moment to fig. 10: in a projector lens, the electrons traverse the whole of the magnetic field. If we follow the path, in fig. 10*b* or c, of an electron coming from the left, we see that the path, after crossing the axis of the lens, is bent back towards the axis again. The electron will thus leave the lens making a smaller angle with the axis than at the point of intersection. This means that the focal length of the lens as projector is longer than as objective. This effect becomes more marked as the lens is made longer or, for a given lens, as the excitation k is increased. Both measures, although they have little further effect on an objective, beyond a certain point will result, for a projector, in an increase in the focal length [10]).

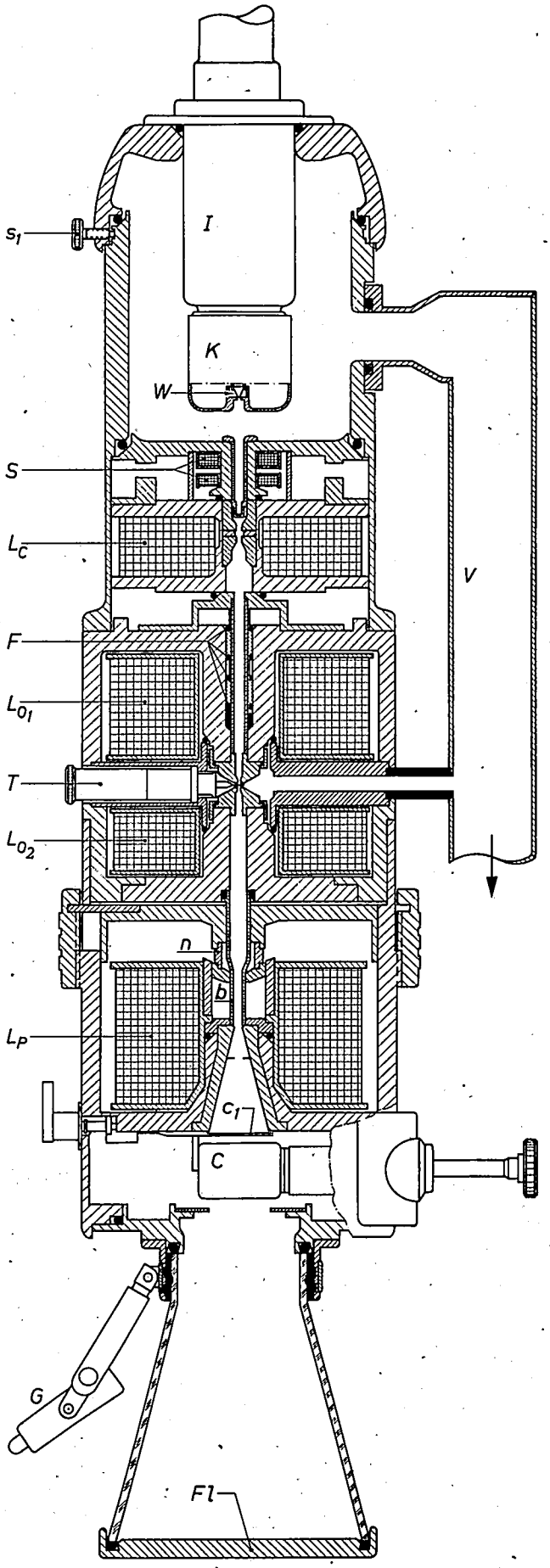
This is accompanied by a larger chromatic aberration; see the broken lines in fig. 11. It appears from this diagram that the abberration is a minimum for a particular excitation factor k_0, which, surprisingly, has about the same magnitude, $k_0 \approx 160$, for projector lenses of all possible forms (various S/D). Moreover, in passing through this "achromatic point" of the curve, the "cushion" distortion which exists with weak excitation of the projector lens is transformed to "barrel" distortion. Thus, at the achromatic point, not only is the influence of fluctuations of the lens current and the acceleration voltage smallest, but the (isotropic) distortion is zero. The projector of the EM 75kV has been designed for use near this point.

Construction of the electron lenses

In *fig. 12* a simplified cross-section of the microscope tube is shown. Of the component parts, which are specified in the caption, we shall first discuss the electron lenses.

The condenser lens is a normal magnetic lens with inserted pole pieces. A diaphragm 0.2 mm in diameter fixes the highest attainable intensity of the electron beam. Variation of the excitation

[10]) If the field is sufficiently long, (or the excitation strong enough) the electron may cross the axis again, and even a third time, etc.; between these cases the focal length becomes infinite.



controls the focal length, and with it the angular aperture, of the beam falling on the specimen so as to give the image on the fluorescent screen the required degree of brightness. The lens, together with the electron gun, is mounted in a tube which can be moved bodily in a transverse direction (the adjusting screws for this purpose have not been drawn). In this way the electron gun and condensor lens combination can be accurately centred with the objective lens.

The strong excitation of the objective brings the object plane — as in the EM 100 kV [3]) — to a point between the pole pieces. It was thus possible to make the objective in two sections and to place the specimen holder (which also includes the objective aperture) between them. The two parts of the lens are screwed together. The sections of the core carry the pole pieces which face each other and are made with great precision from the cobalt-iron alloy already mentioned. The core in the upper half of the lens, contains two auxiliary coils which can be energised with alternating current so as to give an oscillating angle of incidence to the electron beam, facilitating the focusing of the image on the screen; this "wobble focussing device" was explained in the description of the EM 100 kV [3]). The lower half of the lens is noteworthy for its core which, with the attached pole piece, can be displaced sideways by screws (not shown in fig. 12). This enables the pole pieces to be centred extremely accurately with respect to one another.

The projector lens is characterised by the separation of the pole pieces which is variable over a wide range. The upper pole piece is mounted outside the vacuum and can be moved about 15 mm in a vertical direction by turning a milled ring which encircles the microscope tube (*fig. 13*). In this way, for constant excitation, the magnification of the lens can be continuously varied from about 10 × to about 90×. The magnification of the microscope ranges from about 11 000× down to 1200×, so linking up with magnification obtainable with optical microscopes.

The image remains sharply focused when the magnification is changed, and it does not rotate.

Fig. 12. Cross-section (simplified) of the microscope tube of the EM 75 kV. K cathode with filament *W*; *S* deflection coils for centring the electron beam L_C condenser lens; L_{O_1}, L_{O_2}; objective lens composed of two parts with specimen stage *T*; *F* auxiliary coils to assist in the focusing; L_P projector lens, *n* non-magnetic spacer, *b* brass tube carrying pole-pieces; *C* camera with operating rod (turned through 90° for reasons of clarity), c_1 shutter; *Fl* fluorescent screen; *G* observing lens; *V* vacuum connection. The cap of the tube with the bushing insulator *I*, which varries the cathode, can be moved laterally by means of three screws (s_1 etc.) for centring of a changed filament.

It is not practicable to vary the magnification of an electron microscope by regulating the excitation of the projector or the objective: this has been pointed out on an earlier occasion in this Review (see [1]), p. 40). We may recall here merely that this is a consequence of the desire for an image free of distor-

Fig. 13. The microscope tube of the EM 75 kV. By turning the milled ring which encloses the tube, the upper pole piece of the projector lens is displaced. The ring can make $^7/_8$ of a complete turn; this causes the pole piece to move 15 mm in a vertical direction, which alters the magnification by a factor 9. The wall of the vacuum tube is here formed by a thin brass tube (*b* in fig. 12). which makes a sliding fit in theb ore of the pole pieces, which are thus outside the vacuum.

tion and completely filling the fluorescent screen, even at the smallest magnification. In the case of the EM 75 kV, the excitation of the projector must be kept fixed in order to preserve the minimum sensitivity to fluctuations. The excitation of the objective is rigourously fixed because of the fixed position of the specimen (and thus the object plane). In larger microscopes, such as the EM 100 kV, regulation of the magnification is achieved by the use of intermediate lenses, the excitation of which can be varied over a wide range. For our simple microscope this method was not suitable, since the chromatic blurring caused by fluctuations of k which has been adequately reduced by the design of the objective and projector lenses, would be reintroduced in full via the intermediate lens (at least, for photography). The solution described with the adjustable pole pieces was therefore preferred; it is furthermore, less costly and serves its purpose admirably, due partly to the device described below.

The sliding pole piece is made with an extra magnetic gap (the sleeve *n* of non-magnetic material, fig. 12). This forms a sort of additional lens, which, in the highest position of the pole piece, when the strength of the projector and the magnification are at a minimum, reduces the strength of the whole by a further factor of nearly 2. When the pole piece is moved downwards and the lens is thus made stronger, the gap is gradually magnetically short-circuited by the soft iron guide in which the pole piece moves. The weakening additional lens is thus gradually cut out. This device makes it possible to increase the ratio between maximum and minimum projector magnification from about 1 : 5 for the normal construction with one gap, to about 1 : 9.

The variation of the gap (alteration of S at constant D) does not increase the aberrations appreciably. Thus we reap the benefit of choosing the achromatic point for the excitation of the projector, see fig. 11.

Special attention must be paid in this microscope to the centring of the electron lenses. Faulty centring can mean that small fluctuations of the lens currents or of the acceleration voltage cause the image to be displaced laterally (without loss of definition however). For direct observation this is of little importance, but for photography it is fatal. Especially because there is no electronic stabilisation of current and voltage, the instrument makes higher demands on the centring than many more complicated instruments.

The methods of centring — lateral displacement of lenses or pole pieces — have already been mentioned. It may be wondered whether the adjustment is not a difficult and time-consuming business. In fact, a simple and effective routine method has been worked out. The detailed instructions of the method are not within the scope of this article but it may be said that the adjustment is made according to various criteria which can be directly deduced from the electron-optical characteristics of the system. The relative positions of the optical axes of the two lenses is checked by means of the rotation of the image which occurs with changes of the excitation in magnetic lenses. The adjustment of the pole pieces in the objective can be checked further, by reversing the direction of the current in this lens (objective and projector are not connected in series in this microscope): when perfectly centred, the point about which the image rotates with changing excitation is the same for both current directions. The normal direction of the current in the objective and projector lenses is

chosen so that the rotation of the image is in opposite directions in the two lenses. Small fluctuations of the current and voltage then result in no appreciable rotation of the image, which could impair the photographic record.

Fig. 14 reproduces a micrograph made with the EM 75 kV, to give an impression of the image quality normally obtainable.

Fig. 14. Animal cellulose (Tunicine). Magnification 20 000 ×. Micrograph made with the EM 75 kV. (L. H. Bretschneider, Utrecht.)

Other components of the microscope body

Only a summary description of the other component parts of the microscope body will be given.

The electron gun does not differ in principle from that described for the EM 100 kV [3]). The filament is a V-shaped tungsten wire which is raised to a high potential (maximum 75 kV) with respect to the metal wall of the microscope body. The filament is enclosed by a cathode cap (fig. 12) which is held at a small negative potential with respect to the filament according to the principle of automatic grid bias. This ensures that there is no appreciable variation in the emission current and, at the same time, bunches the emitted electrons. The cathode is mounted as a separate unit in the cap of the tube in such a way that the filament can be easily renewed. The cap, mounted on the microscope body with a rubber vacuum seal, can be displaced laterally a few millimeters with respect to the column by means of three radially placed screws. Thus it is possible, after changing the filament, to centre the electron gun accurately with respect to the condenser. Centring is simplified by four deflection coils, situated between the electron gun and the condenser (fig. 12), which impart small angular deflections to the electron beam.

The specimen stage (*fig. 15*) consists of a metal block which can be moved in two directions perpendicular to the electron beam by means of two adjusting screws, to make it possible to scan the specimen. The specimen, as is customary, is placed on a mesh grid or plate mounted on a rod (as in the EM 100 kV). The specimen stage includes a vacuum seal through which the rod is inserted. A hinged movement of the stage displaces the specimen slightly in an axial direction in the microscope tube. This is necessary in order to keep the specimen in its place under all conditions: the short focal length of the objective reduces the tolerance for the axial position of the object to a few hundredths of a millimeter.

The objective diaphragm has been fitted into a cut-out portion of the specimen stage and, like the specimen, can be displaced in three directions.

The image on the fluorescent screen, of dimensions 90 × 90 mm, is observed in this microscope through the side of the tube (in contrast to the EM 100 kV where the screen is observed through the end of the tube as in a television set); see fig. 13. The conical end of the microscope tube is made of glass for this purpose; lead glass has been chosen, which completely protects the observers (the image can be observed by several persons at the same time) from the X-rays given off by the screen. For focusing, use can be made of a reading lens of magnification about $2^1/_2$ ×.

The image can be recorded on standard 35 mm

roll film (as in the EM 100 kV). The internal camera takes a strip of film sufficient for 40 exposures each 28 × 28 mm. The film is brought into the electron beam by pulling out the camera rod with the right hand. At the same time a window in the camera is opened and a shutter, fixed in the microscope tube, is closed. The film is then exposed by operating the shutter with the left hand. With this construction, the need for a shutter mechanism in the sliding camera has been obviated. The shutter is of a type used in some cameras and permits fairly accurately timed short exposures of the order of 1 second.

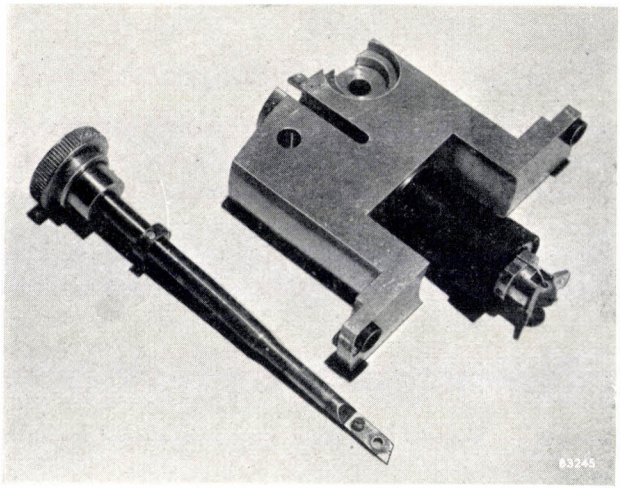


Fig. 15. The specimen stage and specimen holder (left). The cylindrical central section of the specimen stage contains a vacuum lock through which the specimen holder is inserted into the microscope tube.

After exposure, the rod is pushed in again, which automatically operates the transport of the film and the counting device. When the film strip is fully exposed, the camera is removed from the microscope and replaced by another loaded camera. This method makes it possible, by placing the loaded camera in a desiccater, to remove so much water vapour before use that evacuation of the microscope tube after changing cameras takes only about 20 minutes.

The vacuum system

With our simplified electron microscope, the need for a vacuum system will probably be felt to be the most important complication over the light microscope. Therefore, the greatest possible simplicity has been aimed at in this equipment, too. The first stage of the vacuum is obtained with a rotary oil pump, the high vacuum by means of an oil diffusion pump with glass wall. The element of the latter is a thick spiral of chrome-nickel wire, which is immersed in the oil to be heated. This method of heating is so efficient that a power of only about 60 watts is required. Since the walls are scarcely heated, no energy is lost and the important advantage is achieved that no water-cooling is required: the natural air-cooling keeps the walls cold enough for the oil vapour to condense. Another great advantage of the small power and accompanying small heat capacity of the heating element is that when evacuating the microscope tube (e.g. after changing film or filament) the vacuum valve may be opened at once: the tube is then pumped to a low vacuum through the oil diffusion pump — a most unusual procedure, which in our case can be applied without any objections since the sudden flow of air cools down the small quantity of oil in a few moments to a temperature where oxidation of the oil need no longer be feared.

The rotary pump has a vibration-free mounting, so that it may be allowed to run while working with the microscope without any deterioration of the sharpness of the micrographs. The high vacuum pump can also be allowed to continue working after the vacuum valve has been closed for the admission of air to the microscope. The operation of the vacuum system is therefore extremely simple.

The safety devices which normally play a large part in vacuum systems, have been reduced here to one automatic valve. This is situated between the rotary backing pump and the oil diffusion pump. The electrically operated valve remains open so long as the mains voltage remains across the motor driving the backing pump. If the motor is stopped or the voltage fails, the valve closes automatically as a result of the atmospheric pressure and, at the same time, the vacuum side of the backing pump is connected with the outside air. Oil from the rotary pump can therefore never be sucked into the vacuum system if the motor stops — an old trouble with many vacuum systems.

The instrument includes the usual precautions against arcing inside the microscope: a relay permits the application of the high voltage to the cathode only when a sufficient vacuum has been reached.

The vacuum is measured by a Philips manometer (Penning gauge).

The mechanical construction of the vacuum system is also worthy of mention. The oil diffusion pump has been suspended inside the desk of the microscope on a specially strengthened section which extends upwards as a vertical vacuum tube with internal diameter of 50 mm. This wide tube ensures a reasonable pumping speed and, thanks

to its very firm connection with the desk, performs an important function as stand for the microscope column: at its upper end, the tube carries a connecting block which is fixed with screws to the lower end of the microscope tube. The microscope is rigidly fixed only at this point, so that if vibrations occur, the whole instrument behaves as a rigid body. Such vibrations have therefore no influence on the photographs made and the instrument need not be placed in a vibration-free room or on a special foundation.

In the connecting block is the vacuum valve already mentioned and the tube is connected here by a rubber seal (thus not rigidly connected, to maintain the principle of suspension from one point) to an extension which brings the cathode space and the object space in direct connection with the high vacuum pump (fig. 12). This measure reduces the pumping time from the start to about 15 minutes; when changing the specimen, the vacuum is restored after 20 seconds pumping.

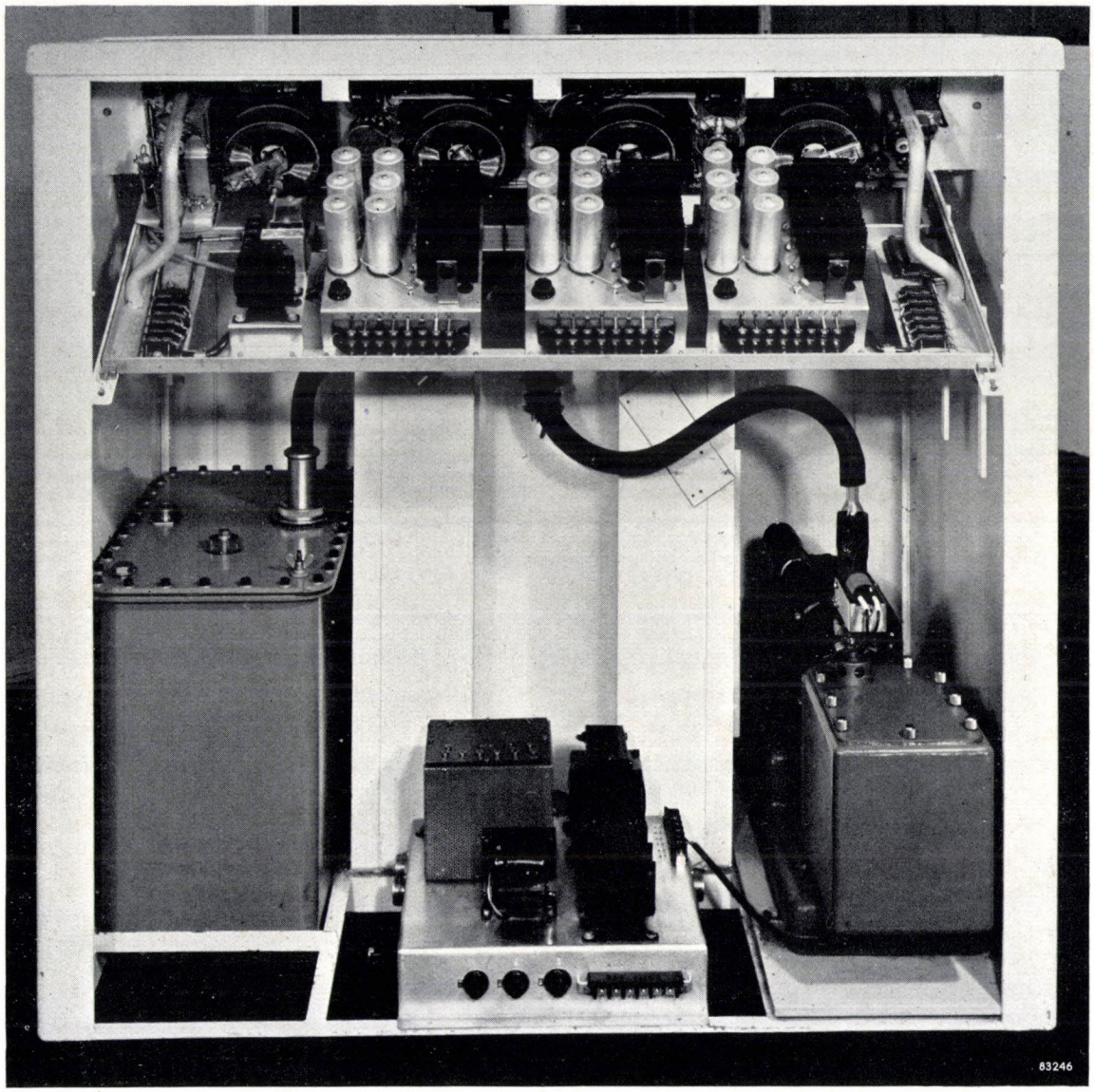


Fig. 16. The desk of the EM 75 kV seen from the back (back panel removed). Bottom left, tank containing the oil-immersed high tension generator, rectifier valves and filament current transformer. Bottom right, backing vacuum pump. Centre magnetic mains voltage stabilizer. Above, rectifiers and smoothing condensers, which supply the energizing current for the three electron lesnses. Behind them is the back of the operating panel with knobs and switches.

The electrical equipment

The acceleration voltage for the electrons is supplied by a small cascade generator with three rectifier valves. A variable transformer makes it possible to vary the high tension from 0 to 75 kV. The generator, together with the rectifiers, is fitted in an oil bath which can be hermetically sealed. The bath stands on the base of the desk where, in addition to the pumps, the magnetic stabilizer for the mains voltage and the power pack for the lenses, the deflecting coils and the vacuum gauge are housed (*fig. 16*).

The three lenses, condenser, objective and projector, are supplied independently by three power packs, each with their own variable transformer, selenium rectifiers and smoothing condensers. The currents for all the lenses can be continuously varied from zero to the maximum value — we have already seen that one of the uses for this is in the adjustment of the optical axis. The adjusting knobs are all mounted on the front panel of the desk (fig. 1). Here, too, is the switch for reversing the current in the objective lens.

It has been seen that in this electron microscope, no use has been made of electronic stabilization. The instrument does not contain a single amplifier tube, and apart from the rectifiers the microscope tube itself is the only "electron tube" in the apparatus. For an electron microscope with magnetic lenses this may be considered unique.

Summary. The small electron microscope EM 75kV has been constructed to give a resolving power of at least 100 Å. The instrument contains three magnetic lenses: condenser, objective and projector. The objective has a focal length of only 0.8 mm. This makes the "chromatic" aberrations relatively small. The same is true for the projector lens, which is excited in the "achromatic point" of its characteristic. The total result is that for the desired resolving power, fairly large fluctuations of the excitation currents I of the lenses and the acceleration voltage V of the electrons can be permitted, so large, in fact, that electronic stabilization is unnecessary; a simple mains voltage stabilizer of the magnetic type is sufficient, and the microscope contains not a single amplifier tube. Besides this great simplification of the electrical equipment, the vacuum system and mechanical construction have also been simplified without prejudicing the demands of reliable operation, short pumping time and insensitivity to mechanical vibration.

ELECTRONIC EQUIPMENT FOR THE CONTINUOUS MONITORING OF TURBINES

by C. van BASEL, H. J. LINDENHOVIUS and G. W. van SANTEN

621.317.39:621-135

Electronic measuring systems have several features which render their application very attractive outside the field of electrical engineering, for example in mechanical, physical and chemical applications. This article describes one of the latest applications, viz. electronic devices for the continuous control of the behaviour and running conditions of turbines.

Quantities requiring monitoring in turbines

Among the machines that operate under extremely severe physical conditions one should certainly include turbines; they operate at steam pressures of 80 atm or more and at temperatures of 500 °C, which, in combination with very high speeds of rotating parts, set up enormous stresses in the material. At the same time, to ensure high efficiency performance, only very slight tolerances are permissible. Since, moreover, these machines are very expensive, it is obvious that the running conditions of a turbine should be subject to continuous careful attention. Added to this is the fact that even seemingly insignificant defects in a turbine, or slight negligence in its control, may cause serious damage, not only to the machine but also to other things in the vicinity, quite apart from the indirect consequences of the breakdown of the energy supply — whether used to generate electricity or to propel a ship.

In order to illustrate where monitoring is necessary, consider *fig. 1*, which shows a simplified cross-section of the turbine of a turbo-generator. The rotor shaft, rotating at a speed of 3000 r.p.m., is fitted with a number of rows of moving blades. In between the moving blades are rows of guide blades fixed to the casing, which have the task of re-directing the steam on to further rows of moving blades. The rotor shaft is located radially by slide bearings (*1*) and in the axial direction by a thrust

bearing (Michell bearing, *2*). The places where the shaft passes through the casing are sealed steam-tight by labyrinth glands (*3*). The two parts of the shaft are connected by a flange coupling (*4*). The turbine rests on a foundation (*6*) to which it is rigidly fixed at a point (*5*) perpendicularly under the thrust bearing. Only at this point (because of the considerable expansion of the turbine) are casing, shaft and foundation mutually fixed.

Since the rotor possesses a smaller heat capacity and a greater coefficient of thermal conduction than the stationary, thick-walled casing, the former suffers faster changes in temperature and therefore expands more quickly. The small space between moving blades and fixed blades becomes even smaller (especially between the rows farthest removed from the common point of fixture *5*) and in the course of a rapid temperature change there is a risk

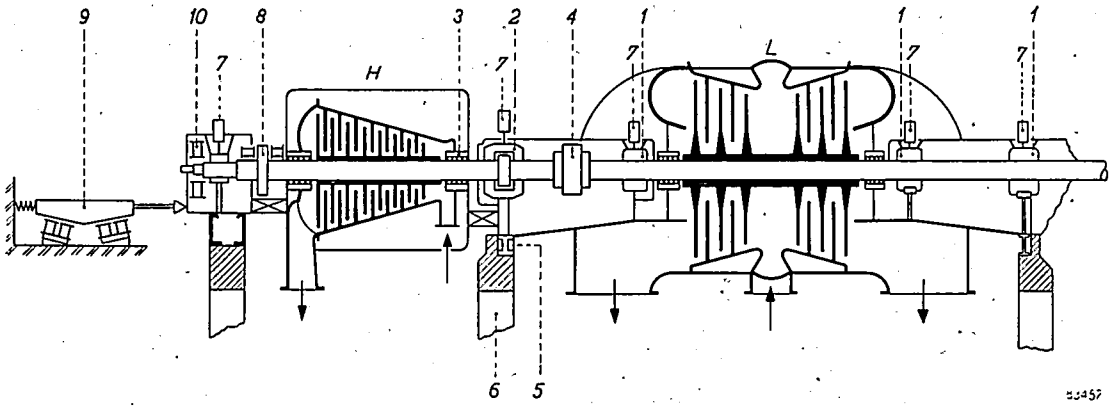


Fig. 1. Cross-section of a steam turbine. *H* high-pressure stage, *L* low-pressure stage (the arrows indicate the direction of the steam). *1* slide bearings. *2* thrust bearing. *3* labyrinth gland. *4* flange coupling. *5* point at which the casing is rigidly connected to the foundation *6*. *7* vibration pick-ups for checking the vibrations of the shaft bearings *1* and *2*. *8* disc for measuring relative changes in length of the shaft. *9* pick-up for measuring the absolute change in length of the casing. *10* disc for measuring the eccentricity.

The first quantity to be considered for measurement is the vibration amplitude in the bearings *1* and *2*. This measurement provides information regarding any unbalance of the rotor system, any play in the bearings or "fouling". A timely indication of any excessive vibration amplitude may prevent enormous damage. To give some impression of the magnitude of the amplitudes concerned it may be mentioned that for a given case in the literature [1]) an amplitude of 10 μ is considered satisfactory, whereas at 45 μ corrective action is urgently called for and at 80 μ it becomes absolutely imperative. (These numbers are applicable to a rotor speed of 3000 r.p.m.).

A second quantity of importance is the change in length of the shaft relative to the casing. For every variation in the steam pressure (which may be caused by a load variation, but is particularly evident during running up and slowing down of the machine) the quantity of heat supplied per second is changing and together with it the temperature of rotor and casing.

[1]) Measurements by T. C. Rathbone, quoted by I. E. Church, Power 103, July 1950, and K. Federn, Arch. tech. Messen 149, 1954 (No. 222).

of fouling, with all its serious consequences (such as damage to the labyrinth gland). Continuous monitoring of the relative change in length of the shaft with respect to the casing is therefore most desirable. The maximum permissible deviation is 1-2 mm.

Apart from this relative change in length it is often desired to measure the absolute change in length of the casing. The temperature of the casing in the high-pressure stage is 300-400 °C and the length 2-3 m, so that in heating up from cold the expansion may be 10-20 mm.

The fourth quantity to be measured is the eccentricity of the shaft. Particularly during running up and slowing down of the turbine it is possible that the heating or the cooling along the circumference of the rotor does not take place quite uniformly. The accompanying irregularity of expansion or contraction is manifested in a bending of the shaft. If this effect becomes very pronounced, there is again the risk of serious damage, caused by fouling between moving and stationary blades. For this reason the eccentricity should also be checked, particularly at low rotor speeds.

Among the methods for measuring the above-

mentioned quantities, electronic methods are particularly attractive for the following reasons:

1. By means of electronic tubes it is possible to obtain power amplification and consequently a high degree of sensitivity.
2. The virtually inertialess action of electronic amplifiers makes it possible to measure rapidly varying quantities as well.

Added to these are some of the general advantages of electronic measuring methods, viz. the fact that reading or recording may be done at a considerable distance from the object to be measured, and the fact that the properties of electrical circuits are easily adaptable to the circumstances obtaining in the turbine. In the case in question, moreover, it is important that the relative displacement and the eccentricity of the rotating objects can be measured without any mechanical contact with the object and hence without any wear on the detecting element and without the risk that in case of breakdown of the latter, parts of it could interfere with the turbine. We shall now deal more closely with an electronic set-up for the monitoring of turbines.

Monitoring the shaft-bearing vibrations

The amplitude of bearing vibrations is measured with the aid of an electro-dynamic vibration pick-up type PR 9260. As it is a considerable time ago that this instrument was discussed in this Review [2]), we shall once more give a brief description of it.

Fig. 2 shows a longitudinal section of the vibration pick-up. Two parts can be distinguished:

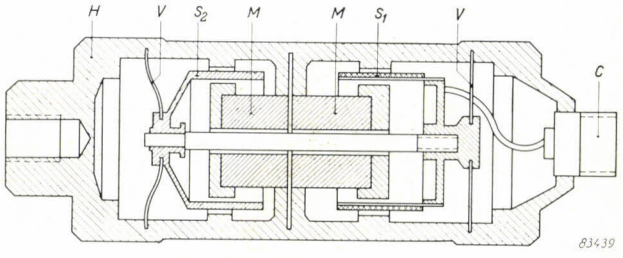


Fig. 2. Cross section of the electro-dynamic vibration pick-up PR 9260. *M* permanent magnet, mounted in the housing *H*. S_1 output coil (with connection contact *C*). S_2 damping coil (copper sleeve acting as a short-circuited winding). The coils are mechanically connected and suspended between the spring membranes *V*.

1. the housing *H* surrounding the rigidly fixed permanent magnet *M*, and
2. the spring membranes *V* connected by a central pin, and carrying the coils S_1 and S_2.

The housing is fixed on the vibrating object — in this case the cap of the bearing in question (*fig. 3*) — and in this way follows every motion of the latter. The spring-mounted pair of coils on the other hand, follows these motions to a far less degree, and hardly at all if the frequency is sufficiently high, in which case it may be considered as stationary which is a characteristic of a pick-up

[2]) J. Severs, Philips tech. Rev. **5**, 230-237, 1940; at the time the type number was GM 5520.

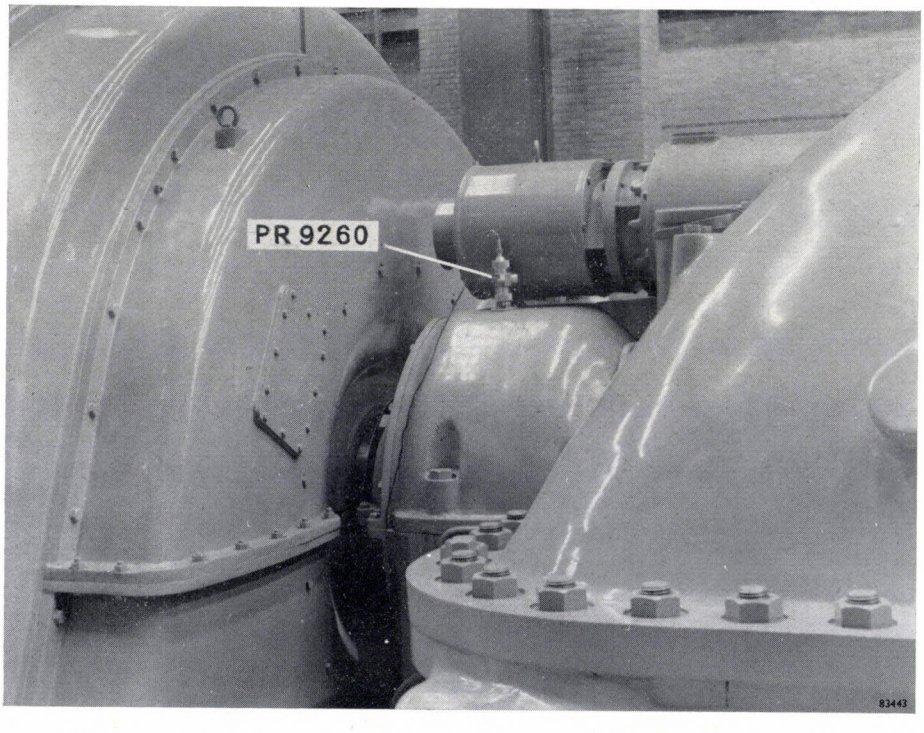


Fig. 3. The vibration pick-up PR 9260 fixed to the bearing cap.

of the seismic type. The magnet connected to the housing moves relative to the more or less stationary coils and induces in the latter a voltage proportional to the velocity of the displacement. Coil S_2 is short-circuited: this provides the "half-critical" damping (i.e. a damping half as great as the critical damping) required for a favourable frequency-response curve. Coil S_1 is connected to an output cable. The ratio of the voltage in this coil to the velocity, is plotted as a function of the frequency in *fig. 4*.

In order to obtain a voltage which is not proportional to the velocity of the displacement, but to the displacement itself, the output voltage of the pick-up has to be integrated over the time. This is easily effected by means of a single *RC*-network.

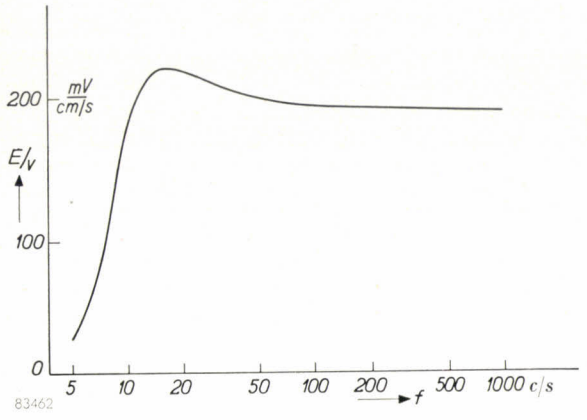


Fig. 4. Response curve of the vibration pick-up PR 9260; the ratio of the output voltage E to the velocity V is plotted as a function of the frequency f.

Fig. 5 shows a block-diagram of the electrical circuit. By means of a rotary switch a maximum of 12 pick-ups can be successively connected to the integrating *RC*-network. This network is followed by an amplifier, a rectifier and a smoothing filter. The direct voltage thus obtained is a measure of the amplitude of the vibrations and may be recorded by means of a recording millivoltmeter (*fig. 6*). In practice it has been found most suitable to select such a sensitivity that a vibration amplitude [3]) of 50 μ corresponds to full deflection of the pointer. The recorder is provided with an adjustable contact which operates an alarm installation as soon as one of the vibration amplitudes exceeds the pre-set value.

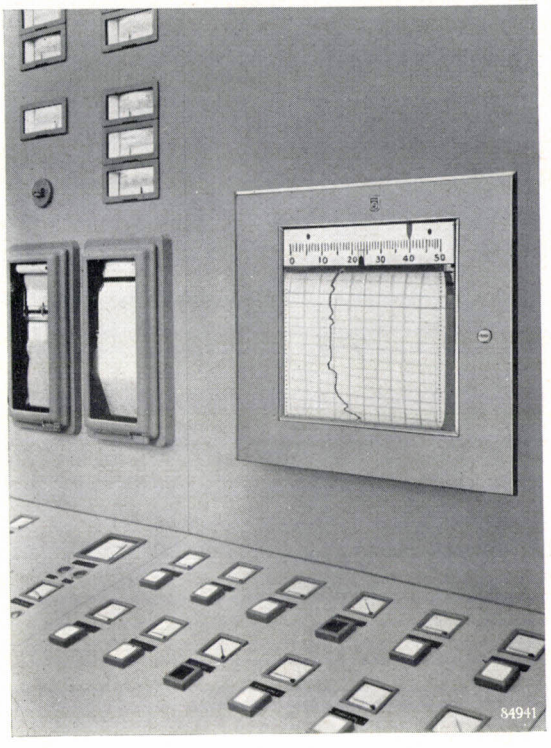


Fig. 6. The recording millivoltmeter type PR 2200 A/21.

[3]) That is, a peak-to-peak value of 100 μ.

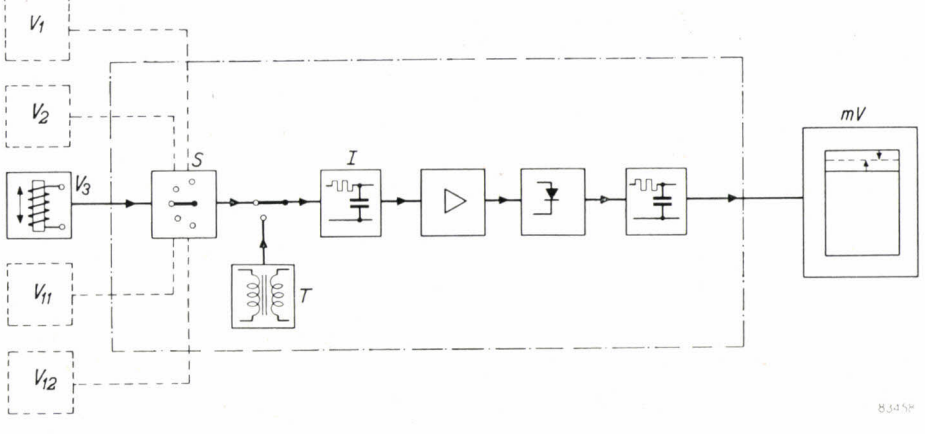


Fig. 5. Block diagram showing the connection of the vibration pick-ups $V_1 \ldots V_{12}$ (7 in fig. 1) alternately connected via a rotary switch S to the integrating network I. The output voltage of I is applied via an amplifier, a rectifier and a smoothing filter to a recording millivoltmeter mV (see fig. 6). T calibration transformer.

The proper functioning of the equipment may easily be checked at any desired moment by means of an accurately adjustable alternating voltage derived from a transformer (T, fig. 5) connected to the mains.

Monitoring changes in length and eccentricity

Relative change of length of shaft with respect to casing

The relative change in length of the shaft is measured with an inductive displacement pick-up. This instrument (*fig. 7*) consists of an armature (A) of a magnetic material situated between two U-shaped yokes of laminated nicalloy. Each of the yokes is provided with a coil. The two coils, of an equal number of turns, together with two equal resistances form a bridge circuit. One diagonal is fed, via a transformer (T_1), from a valve oscillator (frequency 500 c/s). The primary coil of an output transformer (T_2) is connected to the other diagonal.

When the armature is situated exactly mid-way between the two yokes, the self-inductances of the two coils are equally great and no voltage appears across the primary of transformer T_2. When the yoke is displaced to the right a distance x, the self-inductance of the right-hand coil increases,

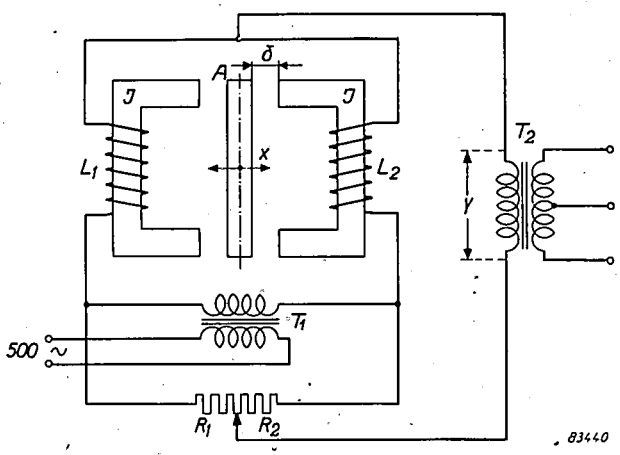


Fig. 7. Schematic diagram of inductive displacement pick-up. The coils L_1 and L_2, together with the resistors R_1 and R_2, form a bridge circuit fed by an valve oscillator (500 c/s) via transformer T_1. The bridge is balanced if armature A is located mid-way between the yokes J. If the armature is located to the left or to the right of the centre, a voltage y is produced across the primary of the output transformer T_2.

since here the air gap has become smaller, whereas the self-inductance of the left-hand coil decreases. The bridge circuit is now no longer balanced, and a voltage v is produced across T_2 which is nearly proportional to the displacement x over a wide range (*fig. 8*). If the armature were moved the same distance x to the left, then an equally large voltage would be produced across T_2, this time, however, in opposite phase. In order to distinguish between movement to the right and movement to the left, it is necessary to identify the phase, for which purpose a phase-sensitive detecting circuit is used.

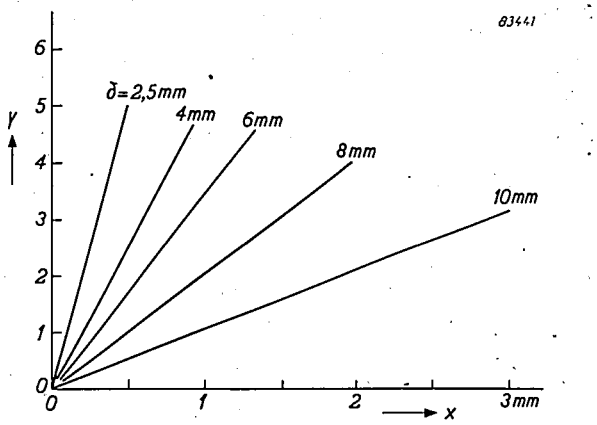


Fig. 8. The voltage y (fig. 7) plotted on an arbitrary scale as a function of the deflection x of the armature from the central position for various values of the air gap δ.

The manner in which a pick-up of this type is used for measuring the change of shaft length of a turbine is shown in *fig. 9*. A disc R (*8* in fig. 1) is fitted on the shaft, which has the same function as the armature A in fig. 7. The two yokes are rigidly connected to the casing.

The working principle needs no further elaboration, but two minor complications may be mentioned here. One is due to the fact that the turbo-gen-

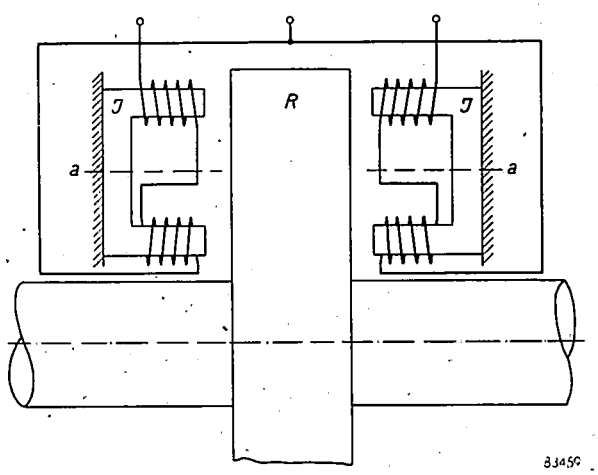


Fig. 9. The displacement pick-up of fig. 7 in a construction suitable for measuring the change in length of the shaft relative to the casing. The yokes J are rigidly connected to the casing. A disc R (of "Permenorm"), fitted on the shaft, functions as armature. (For the sake of clarity the yokes are shown rotated through 90° about line a-a.)

erator is surrounded by a magnetic and an electric field, both of which may impair measurements. The pick-up, therefore, should be carefully screened, magnetically as well as electrically. For this purpose it is housed inside a thick-walled steel box lined with mu-metal. Moreover, the coils are wound in such directions that the influence of an external magnetic field is reduced to a minimum.

The second complication arises from the fact that disc R rotates within the magnetic field of the coils, so that eddy-currents are generated in the disc, giving rise to amplitude and phase errors. To keep these errors as small as possible, the disc is made of a nickel-iron alloy (one of the many types of "Permenorm") with a high resistivity. Use of this material has the added advantage that it is easily machinable.

The air gap on either side of the disc is 10 mm. The maximum displacement that can be measured is 5 mm to either side.

Absolute change in length of the casing

The absolute change in length of the casing is also measured by means of a displacement pick-up (*9* in fig. 1). This is effected by a pick-up of which the armature is connected to the casing and the yokes to the foundation. Since here the displacements to be measured are fairly large (up to 20 mm), very wide air gaps would be required to obtain a linear scale, which would necessitate yokes of very great dimensions. This has been avoided by giving the armature a wedge-like shape (*fig. 10*), such that an axial displacement is converted into a 50 times smaller lateral displacement, the latter being measured in the manner described above. With this (somewhat unusual) reduction of the sensitivity, the air gaps can be confined to 5 mm, so that yokes of normal dimensions can be used. This construction (*fig. 11*) makes it possible to measure displacements up to 50 mm to either side.

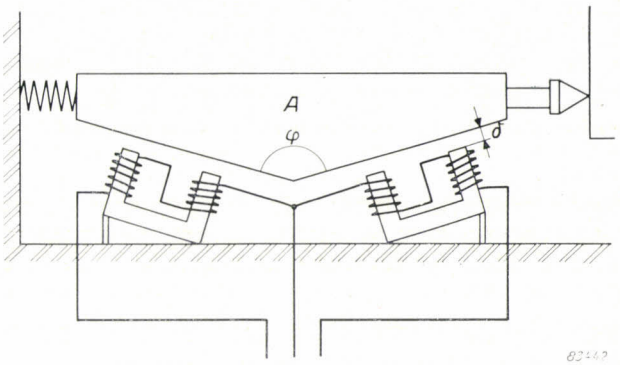


Fig. 10. Displacement pick-up for measuring the absolute change in length of the casing. The principle is the same as that of the pick-up of fig. 7, but owing to the tapered shape of armature A, a change in length of the shaft now corresponds to a 50 times smaller change of the air gap δ. The angle φ, actually about 178°, is shown smaller for the sake of clarity.

Eccentricity

Measuring the eccentricity is again effected by means of an inductive displacement pick-up. Its construction as shown schematically in *fig. 12*. The disc E, shown in this diagram, is located in position *10* in fig. 1 and is made of the same nickel iron alloy as disc R in fig. 9.

Any eccentricity of the shaft is manifested by the fact that the centre of disc E does not remain stationary, but describes a circle of radius e, say. Hence the distance from the circumference of the disc to the yokes changes periodically with an amplitude e.

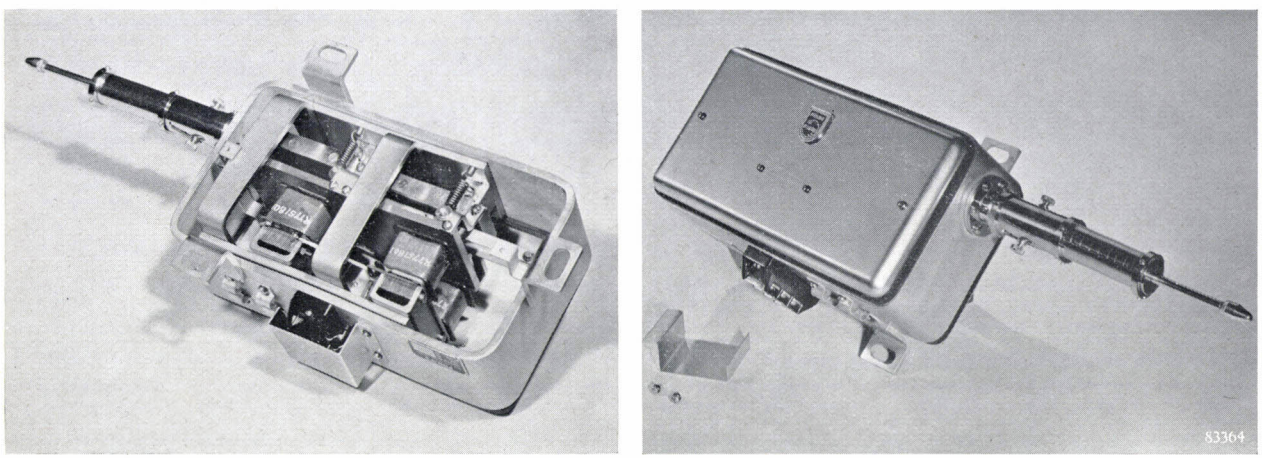


Fig. 11. The displacement pick-up of fig. 10 with the cover removed and closed ready for mounting. For calibration, the amount of the displacement can be read direct from a scale; this reading should correspond to the deflection of the recording meter.

In the two previous cases gradually changing quantities were involved; here, however, we have to deal with a dynamic phenomenon for the distance to be measured changes periodically with a frequency equal to the rotation speed of the shaft. The output voltage of the bridge, therefore,

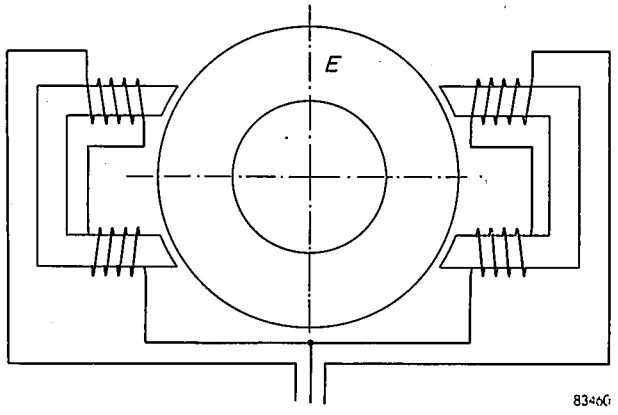


Fig. 12. Inductive displacement pick-up (cf. fig. 7) in a form suitable for measuring the eccentricity of the shaft. Here the armature consists of a disc E (of "Permenorm"), fitted on the shaft.

consists of a signal of 500 c/s, modulated with the (far lower) frequency of the shaft. This output voltage is applied via a demodulator to a recording measuring instrument. As long as the rotation speed is very small (below about 1 r.p.s.) — and checking the eccentricity is particularly important at low speeds of rotation, because of the uneven temperature distribution during running up and slowing down — the pointer of the recording meter closely follows the distance variations. As the rotor speed increases, the recorded amplitude decreases, until finally the pointer does not fluctuate at all, but only indicates the average position of the centre of the disc. This information, too, is valuable, since it is an indication regarding the thickness of the oil-film in the bearings.

If it is desired, however, to measure the eccentricity also at high speeds, then, just as when measuring the vibration amplitude, the demodulated output voltage has to be rectified and smoothed, after which the direct voltage thus obtained can be recorded.

The air gap between disc and yokes is adjusted to 2.5 mm; an eccentricity of 0.5 mm produces full scale deflection.

The electrical circuit

Fig. 13 shows a block-diagram of the circuit used for length and eccentricity measurements. A rotary switch driven by a small electric motor makes it possible to connect alternately six inductive pick-ups. The sensitivity and the zero point of each of these pick-ups is adjustable by means of its two potentiometers incorporated in the apparatus.

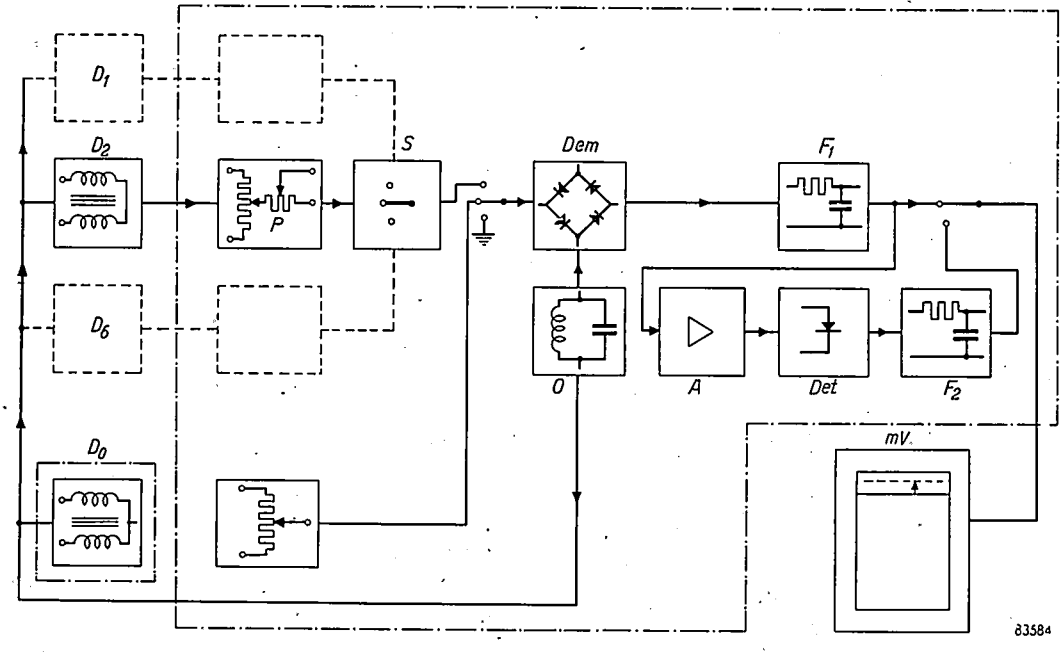


Fig. 13. Simplified block diagram of the circuit for monitoring changes in length and eccentricity. $D_1 \ldots D_6$ six displacement pick-ups. D_0 calibration pick-up. P potentiometers for adjusting the zero point and the sensitivity. S rotary switch. O oscillator 500 c/s. *Dem* demodulator. F_1 smoothing filter 500 c/s. A amplifier. *Det* detector. F_2 smoothing filter shaft frequency. (A, *Det* and F_2 only function for measuring the eccentricity at rotor speeds above 48 r.p.m.) mV recording millivoltmeter.

The demodulation of the output voltage of the bridge takes place in the phase-sensitive detector (*Dem*). The voltage thus obtained is smoothed of the 500 c/s component by a filter (F_1) and the remaining low frequency voltage is applied to a recording instrument. Only for measuring the eccentricity at high rotation speeds is the signal subjected to a subsequent rectification and smoothing as mentioned above, but provided with a micrometer screw, by means of which the armature can be displaced over accurately adjustable and known distances.

Those parts of the equipment which are sensitive to fluctuations of the supply voltage are fed from a stabilized voltage supply so that mains-voltage fluctuations cannot impair the results.

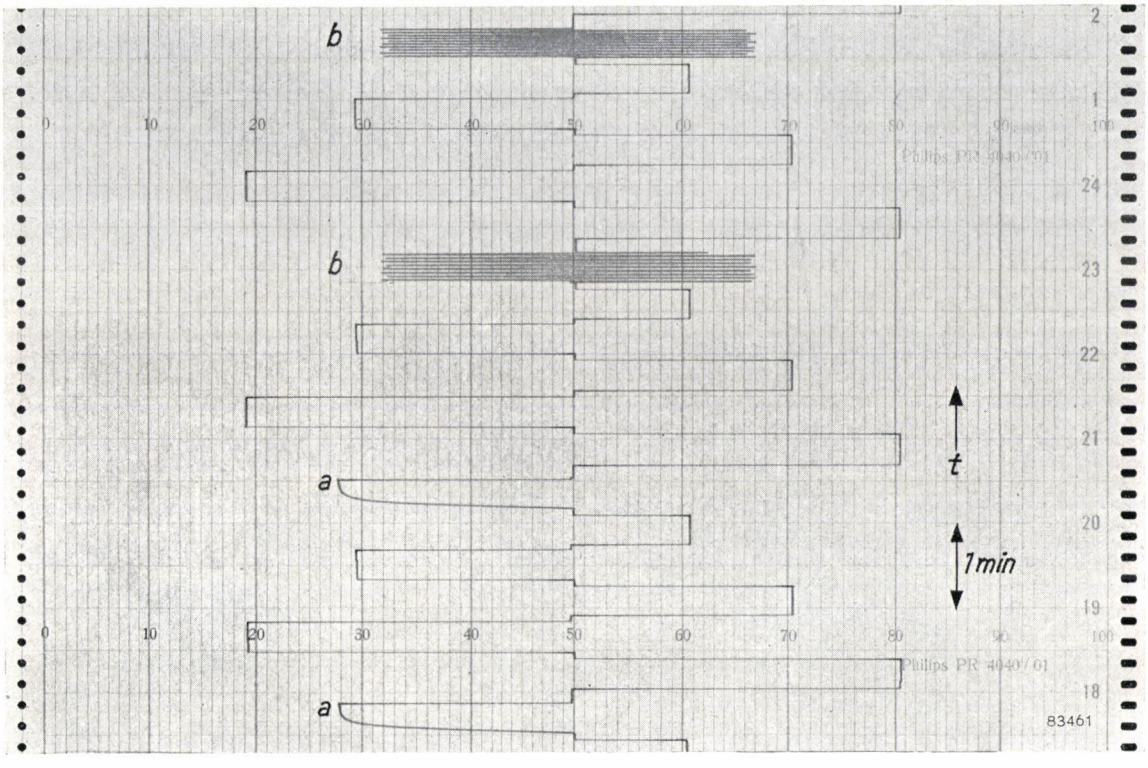


Fig. 14. Section of strip chart, showing the conditions during slowing down of a turbine. At *a* the eccentricity was recorded with the detector *Det* (fig. 13) functioning (rotor speed still above 48 r.p.m.); at *b* without this detector (rotor speed below 48 r.p.m.): the indicator now follows the demodulated alternating voltage. The intermediate rectangular deflections relate to changes in length.

thing as mentioned above. This rectifier (*Det*) may be switched on or off as required; sometimes the switch is operated by a relay, which is automatically energized as soon as the rotor speed exceeds a certain value (e.g. 48 r.p.m., corresponding to 0.8 c/s). The strip of recording chart, shown in *fig. 14*, represents the condition of a turbine passing the critical speed of 48 r.p.m. while slowing down.

The functioning of the equipment can at all times be checked by means of one or more standard pick-ups, of identical construction as the pick-ups

Summary. Turbines are very costly machines, working under severe conditions; moreover, only extremely small tolerances between the rotating and the stationary parts are permissible. The possibility of serious damage can be avoided by keeping a continuous check on certain quantities by means of an installation which gives an alarm if any of them exceeds its critical value.

The quantities to be monitored are the vibration amplitudes of various shaft bearings, changes in length of shaft and casing, and the eccentricity of the shaft. Electronic methods for measuring these quantities have the advantage of high sensitivity, the absence of inertia, and the ability to indicate and record at a distance.

The article describes a seismic electro-dynamic vibration pick-up for checking bearing vibrations, and various types of an inductive displacement pick-up for checking the change in length of the shaft relative to the casing, the absolute change in length of the casing and the eccentricity of the shaft.

ABSTRACTS OF RECENT SCIENTIFIC PUBLICATIONS OF 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 Philips Research Laboratory, Eindhoven, Netherlands.

R 234: I. Pelchowitch: A high resolution mass spectrometer with variable-bandwidth measuring circuits (Philips Res. Rep. **9**, 1-41, 1954, No. 1).

A high-resolution spectrometer of the magnetic-sector type is described. Following a general description of the instrument, a detailed analysis and circuitry of the electronic part are given. This is planned so as to include the possibility of varying the time of observation whilst retaining a high sensitivity in the measuring circuits. Accordingly the components units are designed for static and for dynamic operation. The ion-accelerating voltage supply can give a voltage sweep variable in starting-voltage value, amplitude and time dependence. It is also constructed to have a low effective impedance to minimize the disturbing effects of induction and leakage currents. A special DC-coupled wide-band electrometer amplifier is described covering the frequency range 0-1000 c/s. The useful bandwidth can then be selected by the succeeding measuring units in order to have a high signal-to-noise ratio. The circuits of an oscillograph display of the ion currents are given. An added feature of the instrument is an improved space-charge emission regulator used in connection with a modified Nier-type ion source.

R 235: I. Pelchowitch: A study of the evaporation products of alkaline-earth oxides (Philips Res. Rep. **9**, 42-79, 1954, No. 1).

The mass spectrometer described in *R 234* is used to study the evaporation phenomena in systems where the alkaline-earth oxides are coated on electrically heated metal ribbons. BaO evaporates mainly in the form of the oxide when it is coated on Pt and Ni or coated in admixture with SrO and CaO on Pt. When the BaO/Pt system is brought to high temperatures, an ion current of Ba_2O_2 can be measured. In the SrO/Pt and CaO/Pt systems the main evaporation product is the free element accompanied by the oxide and the singly ionized metal ions. The main evaporation product in the BaO/Ta system is free Ba accompanied by the expected amount of the oxide and by singly ionized Ba ions which can be measured at higher temperatures. When the experimental results of ion-current dependence on temperature are fitted to an equation of the form $\log I = -A/T + B$, I being the ion current at a certain M/e value measured at the sample temperature T °K, unexpected results are found. The evaporation-rate curve drawn using $\log_{10} I$ and $10^4/T$ as coordinates is not in all cases a single straight line. The evaporation curve of BaO from all BaO systems coated on Pt can be best described in the measured temperature range as two lines differing in slope joining at the temperature point $T = 1250$ °K. A critical behaviour in the evaporation-rate curve of free Ba evaporating from the system BaO/Ta seems to occur when this temperature point is reached. A different behaviour of the evaporation-rate curve of BaO is observed in the BaO/Ni system, the transition point at 1250 °K disappearing. We are then able to prove, however, that the Ni base metal evaporates through the BaO porous layer and that the evaporation-rate curve of Ni behaves critically in the neighbourhood of $T = 1250$ °K. A transition point in the evaporation-rate curve of free Sr is found in the system SrO/Pt. It is reversible with temperature and the temperature value is $T = 1600$ °K. In view of the unexpected results found in the evaporation-rate curves, resistance measurements of the systems are made using oxide layers pressed between metal electrodes. All the above-mentioned transition points appear clearly in the resistance behaviour with temperature. The difference between Pt and Ni as base metals is also indicated. Finally the experimental results are compared critically with data from the literature.

R 236: P. Zalm, G. Diemer and H. A. Klasens: Electroluminescent Zns phosphors (Philips Res. Rep. **9**, 81-108, 1954, No. 2).

In Part I, an account is given of the preparation of electroluminescent zinc-sulphide powders. The preparation is based on the assumption that a copper-rich layer on the surface of the grains is neccessary for electroluminescence.

In Part II, electrical and optical measurements on electroluminescent ZnS powders are described, including various oscillograms of suspended and binderless settled layers, and stroboscopic observation of the individual crystals by means of a microscope. The maximum efficiency is 1.5 ($\pm$ 0.2)

lm/W; the brightness at 600 V, 3000 c/s is about 1000 lm/m^2. The hypothesis of a barrier action at the surface of the grains is confirmed.

In Part III, the observations of Part I and II lead to the conclusion that electroluminescence is a kind of cathodoluminescence with "slow" electrons. The smallness of the efficiency of electroluminescent phosphors is compared with that of normal phosphors under cathode-ray excitation at low voltage and high current density. A discussion is given of an equivalent circuit for a suspended electroluminescent layer.

R 237: G. Diemer: Electric breakdown and light emission in Cds single crystals (Philips Res. Rep. **9**, 109-121, 1954, No. 2).

With measurements on D.C. breakdown in photoconducting Cds single crystals, activated with Cl, *I-V* characteristics are obtained very similar to those of arc discharges in gases. There is a Townsend region, giving rise to anode light, a region of negative slope where light phenomena occur pointing to a positive streamer discharge, and an arc region of thermal character.

R 238: W. F. Niklas: Die negative Ionenkomponente des Elektronenstrahles in Kathodenstrahlröhren, insbesondere Fernsehbildröhren (Philips Res. Rep. **9**, 122-130, 1954, No. 2).

The electron beam of a cathode-ray tube contains, apart from the electrons, also negative ions which inactivate the fluorescent screen. When in a cathode-ray tube with an electrostatic lens system a transverse magnetic field is superposed on the electrostatic field, it is possible to focus the ions on the screen in sharp points which are arranged according to the mass of the ions. A sharp focus is only obtained if the ions are generated at the same spot as the electrons because the lens system is arranged for sharp focusing of these electrons. An unsharp focus of the ions therefore indicates that the ions must have been generated elsewhere than the electrons. The mass of the ions is obtained by suitable calibration. By means of the degree of sharpness of the focus two ion components can be distinguished: ions generated in the remaining gases contained in the cathode-ray tube (gas ions), and ions from the cathode. By means of a shadow method (a fine gauze placed at different points in the cathode-ray tube) it is shown that the gas ions are generated between the grid and the first anode. By their mass and nature the cathode ions are identified as Cl^{37}, Cl^{35} and CN or C_2H_2, and the gas ions as O, CH, C. Reduction in the residual pressure results in the elimination of the gas ions; the intensity of CN (or C_2H_2) cathode ions reduces with increasing life of the tube.

R 239: H. G. Beljers: Faraday effect in magnetic materials with travelling and standing waves (Philips Res. Rep. **9**, 131-139, 1954, No. 2).

The ferromagnetic Faraday effect in a circular wave guide is described and the approximate theory for the angle of rotation of the plane of polarization is recapitulated. A few experiments are described for Ferroxcube and Ferroxdure samples and the experimental results are compared with the theoretical ones. Finally an experiment with a ferrite specimen in a resonance cavity is reported, demonstrating the positively and negatively rotating components of the mode of oscillation.

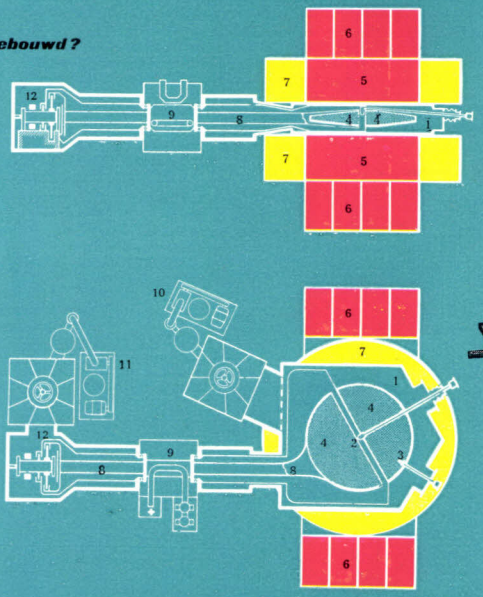
R 240: J. S. C. Wessels: Investigations on photosynthesis; the Hill reaction, Part I (Philips Res. Rep. **9**, 140-159, 1954, No. 2).

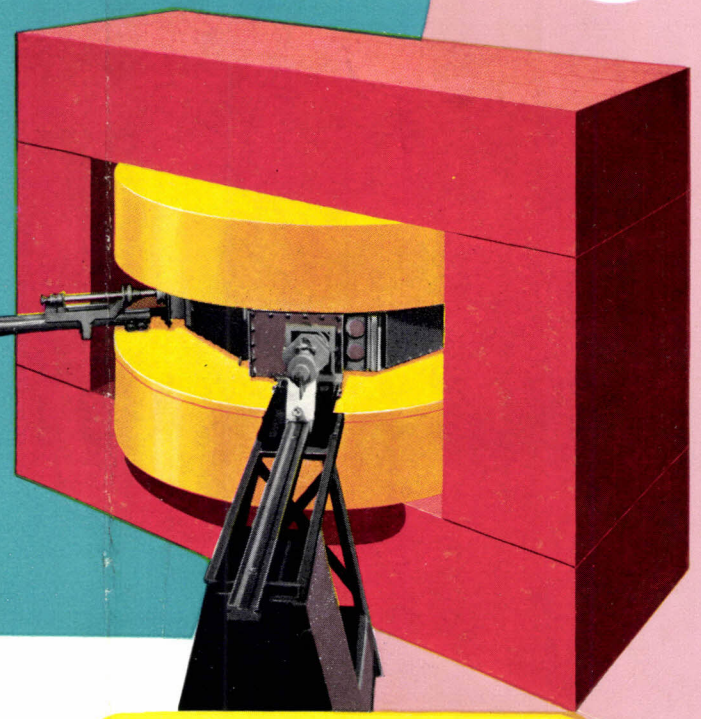
First part of a treatise on the photochemical reduction of various quinones and dyes by isolated chloroplasts, studied by measuring the redox potential during illumination. Apart from a general introduction, this part contains details of the experimental set-up.

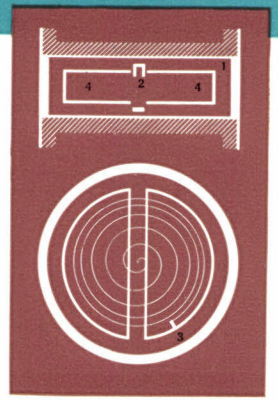
PHILIPS *Synchro Cyclotron*

Hoe is een cyclotron opgebouwd?

In principe bestaat een cyclotron uit een luchtledige versnellingskamer (1) in het middelpunt waarvan de ionenbron (2) is aangebracht. Aan de omtrek bevindt zich de trefplaat (3) waarop het te bombarderen preparaat wordt bevestigd. In de versnellingskamer bevinden zich ook twee D-vormige electroden (4) waartussen een hoogfrequente spanning heerst. Het geheel bevindt zich tussen twee magneetpolen (5) in een juk (6). Een constant magneetveld wordt door twee met gelijkstroom bekrachtigde spoelen (7) in stand gehouden. Een concentrische hoogfrequent leiding (8) verbindt een van de beide D-vormige electroden met de H.F. oscillator (9). De pompen (10 en 11) onderhouden een vacuum in de versnellingskamer en in het modulatorgedeelte.







Hoe werkt een cyclotron?

De ionenbron (2) schiet atoomkernen in de luchtledige versnellingskamer (1) met een relatief geringe beginsnelheid. Ten gevolge van de aanwezigheid van het magneetveld gaan deze deeltjes in cirkelvormige banen lopen en passeren daarbij telkens de spleet tussen de beide D-vormige electroden (4). Door een juiste keuze van de frequentie van het hoogfrequente wisselveld tussen deze electroden kan worden bereikt, dat een deeltje juist op het moment dat het de spleet oversteekt door de tegenoverliggende doos wordt aangetrokken. Zodoende wordt het deeltje bij elke rondgang 2 maal versneld. De steeds toenemende snelheid doet echter ook de centrifugaal-kracht toenemen. Hierdoor gaat het deeltje zich meer en meer naar buiten bewegen en hoewel hierdoor de baan steeds langer wordt, zal het deeltje toch steeds op het juiste moment de spleet blijven oversteken. De snelheid neemt nl. evenredig met de baanlengte toe en dus blijft de tijd van een omloop constant. De baan die het deeltje beschrijft wordt dus spiraalvormig. Op de buitenste baan van deze spiraal hebben de deeltjes de grootste snelheid en daar is juist de trefplaat (3) met het te bombarderen preparaat aangebracht. Het voordeel van deze werkwijze is, dat de deeltjes met behulp van een relatief lage spanning, een snelheid kunnen bereiken, die overeenkomt met een veelvoud daarvan.

Waartoe dient een cyclotron?

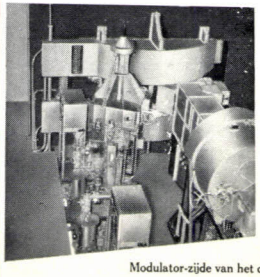
Een cyclotron is een apparaat dat gebruikt wordt om kleine geladen deeltjes (atoomkernen van waterstof, deuterium of helium) te versnellen. Met deze snelle, en dus energierijke deeltjes worden bepaalde stoffen gebombardeerd, met het doel daarin kernreacties te veroorzaken. Met het cyclotron kunnen dus onderzoekingen worden gedaan naar de opbouw van de atoomkernen. Tevens kan het apparaat worden gebruikt voor het produceren van radio-actieve isotopen, die bij het wetenschappelijk onderzoek, in de moderne industrie en in de geneeskunde steeds veelvuldiger worden toegepast.

Het synchro-cyclotron

Het linksonder beschreven principe van de werking van het cyclotron gaat bij versnelling tot zeer hoge energie niet geheel op. De massa van de deeltjes neemt dan nl. iets toe (relativiteits-theorie). Bovendien zal het magneetveld aan de buitenzijde van de polen iets zwakker zijn dan in het midden. De deeltjes raken hierdoor steeds meer achter op het hoogfrequente wisselveld tussen de electroden, naar mate de snelheid groter wordt. Het aantal gangen in de spiraalvormige baan en dus ook de eindsnelheid van het deeltje wordt hierdoor beperkt. Dit bezwaar kan worden ondervangen, door de frequentie van het hoogfrequente wisselveld te laten afnemen, naar mate de snelheid van de deeltjes toeneemt. Op deze wijze raken zij niet achter, doch zij blijven synchroon met het wisselveld de spleet oversteken.
Bij het synchro-cyclotron laat men de frequentie van het hoogfrequente wisselveld periodiek veranderen (frequentie modulatie). De deeltjes, die door de ionenbron worden afgeschoten op het moment dat de frequentie maximaal is, worden telkens gebundeld tot groepen die synchroon met het veld lopen. De rest van de deeltjes raakt uit de pas en gaat verloren. De trefplaat wordt dus telkens getroffen door een groep van met het veld synchroon lopende deeltjes, zodat daar een pulserende stroom heerst.

De ionenbron wordt gereed gemaakt

Het inzetten van een nieuwe trefplaat



Modulator-zijde van het cyclotron

PHILIPS SYNCHROCYCLOTRON

Purpose of a cyclotron

A cyclotron is an apparatus used for the acceleration of small charged particles (protons, deuterons or α-particles). These rapid particles, possessing a high energy are used to bombard certain substances with the object of causing nuclear transmutations in the latter. The cyclotron can, therefore, be used for research into the structure of atomic nuclei. The apparatus is also employed for producing radio-active isotopes, which are finding increasing application in scientific research, in modern industry and in medicine.

General construction

In its basic form a cyclotron consists of an evacuated chamber (1) in the centre of which the ion source (2) is placed. At the outside of the acceleration chamber is the target (3) provided with the preparation to be transmuted. Inside the acceleration chamber are also two D-shaped electrodes ("Dees") (4), in between which a H.F.-voltage prevails. The whole of this is located between two magnet poles (5) in a yoke (6). A constant magnetic field is maintained by two D.C.-energized coils (7). A coaxial H.F.-transmission line (8) forms the communication between one of the Dees with the H.F.-oscillator (9). The vacuum pumps (10 and 11) maintain a vacuum in the acceleration chamber in the modulator part.

Working principle

The ion source (2) emits atomic nuclei at a relatively low initial speed. Under the influence of the magnetic field the particles are forced to describe a circular path in the course of which they repeatedly pass the gap between the Dees (4). By a correct choice of the frequency of the H.F. alternating field between these electrodes, it is achieved that a particle, just at the moment when crossing the gap, is attracted by the opposite electrode. In this way a particle is twice accelerated during each rotation. The increasing orbital speed, however, also increases the centrifugal force, so that the radius of the orbit increases, so that the particles continue to cross the gap at exactly the right moment. This is due to the fact that the speed increases proportional to the length of the orbit so that the time of revolution remains constant. The particle, therefore, describes a spiral orbit. On the outer trajectory of this spiral the particles have their greatest speed and it is there that they hit the target (3) with the preparation to be transmuted. This procedure has the advantage that by means of a relatively low tension the particles can be accelerated to a speed, corresponding to a multitude of this tension.

The synchrocyclotron

The basic working principle of the cyclotron as given in the foregoing is not completely applicable when the particles are accelerated to a very great final energy, because of the relativitic increase of their mass. Moreover, as the particles get closer to the edge of the pole pieces, the density of the magnetic flux must necessarily be somewhat smaller than in the middle. Owing to this, the particles are lagging more and more behind the H.F. alternating field between the electrodes according as their velocity increases. All this restricts the number of revolutions in the spiral orbit and correspondingly limits the ultimate energy of the particle. This objection can be overcome by gradually reducing the frequency of the H.F. alternating field, according as the velocity of the particle increases. In this way lagging is avoided and the particles continue to cross the gaps synchronously with the alternating field. With the synchrocyclotron the frequency of the H.F. alternating field is periodically changed (frequency modulation). The particles emitted by the ion source at the moment when the frequency is maximum are each time bundeled into a stream running synchronously with the field. The remaining particles get out of step and are lost in the process. The target is, therefore, each time hit by a group of particles running synchronously with the field, which intermittant beam causes a pulsating current at this spot.

Captions

Preparing the ion source for operation.
Inserting a new target.
Modulator side of the cyclotron.

Attention

The other side of this sheet is a reproduction on a reduced scale of our 7-colour wall-chart of the cyclotron. Dimensions 80 × 120 cm (32" × 48"). This wall-chart is available in Dutch, French, German, English and Spanish. Price Hfl. 0.50 (1/— $ 0.15). For ordering, write to: General advertising Division, N.V. Philips' Gloeilampenfabrieken, Eindhoven, Holland.

79/001/A - 6:55

VOL. 17 No. 3, pp. 69-100 SEPTEMBER 1955

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 APPLICATION OF THE X-RAY IMAGE INTENSIFIER

For many years electronics has been engaged in the problems of amplifying very weak currents and voltages. In recent years, partly owing to the stimulus of television, a new branch of electronics, that is, the intensification of light, has emerged. The original object of this development was the conversion of long-wave, into short-wave light ("wavelength-transformation") which was investigated in the Philips laboratories at Eindhoven as long ago as 1934. Now, however, the emphasis is on luminance intensification with special reference to the intensification of weak fluorescent images.

The first article on the X-ray image intensifier to appear in this Review was published in 1952; since then, laboratory experiments and practical tests have produced much new and interesting information concerning this intensifier, and it is now considered worth while to publish one or two articles describing these developments. The main points considered here are: the minimum size of detail perceptible with the image intensifier; the optical problems involved; photography and cinematography with the image intensifier.

The present articles refer mainly to the existing type of image intensifier. Little is said as to the probable future trend of development in image intensification; this does not imply, however, that the present results are considered the last word in this field. On the contrary, we should emphasise that the subject is still developing and we hope to report further progress in due course.

I. GENERAL SURVEY

by M. C. TEVES. 621.386.8:616-073.75:621.383.8

By way of introduction, the purpose, principle and design of the X-ray image intensifier at present in regular production [1]) will be briefly re-stated.

The purpose of the image intensifier is to enable as much information as possible to be extracted from the fluorescent image of the particular object, for a given X-ray dose. The theoretical and practical factors governing the amount of information obtainable from the image are discussed fully in the second article of this series. It can be shown from these considerations that with ordinary fluoroscopy, the information in the screen depends on the X-radiation absorbed by the screen. The amount of radiation absorbed is closely related the dose to which the patient is exposed. However, the observer cannot extract all this information, owing to the weakness of the optical link between the fluorescent screen and the human detecting organ (that is, the retina of the observer's eye). The same applies to fluorography (miniature radiography), which likewise involves

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

an appreciable loss of light in an optical link, viz. that between the fluorescent screen and the film; the loss is so great, that only about 1% of the information latent in the screen is transmitted to the film.

Full-size radiography is very much better in this respect. The direct optical contact between fluorescent screen and film here prevents any loss of light. With regard to the first two methods referred to, an image intensifier tube considerably increases the amount of information obtainable with a given dose. Compared with full-size radiography, however, it offers only secondary advantages, namely that it enables the dose to be reduced. It has the further advantage that it enables cinematography to be employed.

Hence the main purpose of the image intensifier is to make good the light loss in the optical link between the fluorescent screen and the light detector (retina of the eye, photographic film or plate, or, possibly, the photo-cathode of a television camera tube).

Description of the image intensifier

The image intensifier is an evacuated glass tube containing a fluorescent screen on a thin aluminium base (*fig. 1*); in contact with the screen is a photo-cathode. X-radiation striking the screen makes it

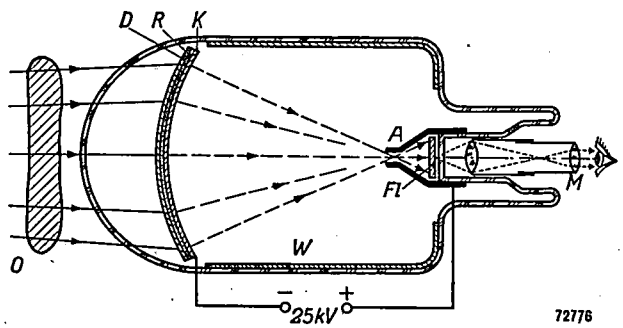


Fig. 1. Schematic cross-section of the image intensifier tube. *R* fluorescent screen receiving the X-radiation after it passes through the object *O* and the glass wall of the tube; *D* support carrying the fluorescent screen and the photo-cathode *K*. The fluorescence generated in *R* releases electrons from the photo-cathode. The "electron image" is reproduced, reduced in size, on the viewing screen *Fl* by the electric field between *K* and the hollow anode *A*. It is then observed through a simple microscope *M*. *W* is a conductive coating on the inside of the tube.

fluoresce and the light then releases electrons from the photo-cathode. The number of electrons so released from each point on the cathode is proportional to the luminous intensity of the fluorescent screen at that point. By means of an electric field, the electron image thus formed is reproduced, reduced 9 times in size, on another fluorescent screen, the viewing screen. Part of the energy of the electrons striking this screen is re-converted into fluorescent light to form a 9 times smaller facsimile of the image on the first fluorescent screen. This facsimile is then viewed through a simple eyepiece of roughly 9 × magnification, so that the image is seen in its original size, that is, roughly 13 cm in diameter, and upright, but about 1000 times brighter than before.

The luminance intensification arises from two factors (which, however, are not independent of each other). Firstly, an increase in the overall luminous flux (or "lumen intensification") due to the fact that the electrons from the photo-cathode are accelerated by the electric field: there is an accelerating voltage of roughly 25 kilovolts between the photo-cathode and the viewing screen. The higher the energies of the electrons striking the viewing screen, the more intense the fluorescence produced. Although only about 1 in every 10 light quanta falling on the first fluorescent screen releases an electron, and only about one tenth of the electron energy is converted into light on the viewing screen, the energy imparted to the electrons nevertheless results in the latter screen producing between 10 and 15 times as much luminous flux as an ordinary fluorescent screen viewing the same subject.

The second factor is the electron-optical reduction of the image size; it enables all the photo-electrons to contribute to the formation of the image, so that the amount of light generated does not depend upon the area over which these electrons are distributed. By employing a reduction of 9 times, we reduce the area within which the electron energy is concentrated by a factor of 9^2; hence the total luminous flux is emitted from an area about 80 times smaller than it would be with reproduction on a scale of 1 : 1. This, by definition, means an increase in luminance by a factor of 80. The total luminance intensification is the product of the lumen intensification and the gain from the reduction of the image size; with the tube under consideration, it is between 10 and 15 times 9^2, or from 800 to 1200. Thus the luminance is so increased as to make good all the light loss involved in the forming of the image.

In the conversion of a low-luminance image into a high-luminance one, special precautions are necessary to avoid loss of contrast or definition in the image owing to imperfections in the apparatus.

With the present image intensifier, sharpness is limited mainly by the thickness of the first fluorescent screen. However, it is also affected to some extent by the viewing screen. Blurring in the electron-

optical image forming system is almost negligible.

In principle, subtle contrasts in the low-luminance initial fluorescent image are not affected by the light-transformation (the γ of the image intensifier is unity). In practice, however, there is a slight, but unavoidable loss of contrast owing to "fogging", that is, a luminance contribution distributed more or less uniformly over the whole image. On the other hand, the increase in contrast sensitivity of the eye with increasing luminance far outweighs this slight loss. In photography, it can be made good by employing a film with a higher γ.

Although it is essential that the properties and possibilities of the image intensifier be fully investigated in the laboratory by means of "phantom tests" (see article II), its merits from the medical point of view can be determined only in actual medical practice. Such practical tests are being carried out in a number of places, e.g. in the Philips Health Centre at Eindhoven under Professor Burger and Dr. Feddema [2]), and in Maastricht by Dr. van der Plaats [3]).

Without particularizing unduly, we may quote the following examples of the usefulness of the image intensifier from these investigations (see article V).

In chest fluoroscopy, the intensifier gives good results with only one tenth of the normal X-ray dose. Apart from the fact that the resolution is at least as good under these conditions as in the direct image, the image intensifier enables the subject to be examined in a moderately lit room and without any preliminary adaptation of the eyes.

The intensifier is eminently suited for locating foreign bodies (e.g. metal particles), and for the routine examination of the setting of bone fractures.

For the examination of an oesophagus, stomach or colon into which contrast medium has been introduced, the investigation is considerably facilitated by the image intensifier.

An important use of the image intensifier is as an aid to visual positioning before the taking of an ordinary, full-size radiograph (spot film technique).

Medical experience has shown that in fluoroscopy, better results are obtained with, than without the image intensifier, especially in circumstances where the relatively small size of the field, i.e. 13 cm in diameter, is not a handicap.

As applied to fluorography, that is, photography of the viewing screen of the intensifier on film with a camera, the medical uses of the intensifier may be divided into two categories:

1) The taking of still photographs, singly or in series.
2) X-ray cinematography.

A great deal of information concerning both these uses has already been collected. It is found that the quality of a photograph taken with the image intensifier on fine-grain 35 mm film is very much the same as that of an ordinary full-size radiograph, although, with a suitable optical system, the X-ray dose required per photograph is a factor of 2-3 smaller than in full-size radiography. With regard to X-ray cinematography, it is enough to say that even with quite a long film, the X-ray dose is not heavy enough to endanger the patient; hence photographic X-ray examination can now be employed in physiological, as well as anatomical studies.

Briefly, then, we may safely say that the present image intensifier has demonstrated its value in many medical applications. However, it is still far from perfect. One of the practical improvements still required from the medical point of view is a larger image field. Moreover, the methods of presenting the image to the observer also require attention; this is important not only in fluoroscopy, but also from the point of view of cinematography. Another problem is how best to convey all the information in the film strip to the observer.

Finally, it should be pointed out that the image intensifier is also useful in industrial radiography. The relatively brighter image produced permits the visual examination of much thicker objects than has been possible hitherto. One or two examples are given in article VI.

[2]) J. Feddema, Image intensification. Some possible diagnostic applications in cineradiography, Brit. J. Radiology **28**, 217-220, 1955.

[3]) G. J. van der Plaats, De röntgendoorlichting in verband met nieuwe ervaringen met de beeldversterker, Ned. T. Geneesk. **97**, 1056-1063, 1953.

II. THE PERCEPTION OF SMALL OBJECT-DETAIL

by T. TOL and W. J. OOSTERKAMP. 621.386.8:616-073.75:621.383.8

In X-ray diagnosis, the principal aim is to obtain the desired information concerning the organ examined, with the smallest practicable X-ray dose to the patient. In general, direct visual examination of the fluorescent image involves a very much larger dose to the patient than the taking

of a single radiograph: in fluoroscopy the radiologist requires a certain amount of time to examine all the details of the image properly, whereas in radiography the time during which the patient is irradiated, is relatively short, and the radiologist has ample opportunity to study the film when once it is developed. On the other hand, a radiograph provides the radiologist with only one instantaneous picture and therefore tells him nothing about the condition of the particular organ at different moments; such information is important when moving organs or an unrepeated effect, say, the injection of contrast medium, are to be observed. Accordingly the relatively larger X-ray doses associated with fluoroscopy give, in general, more information. Both methods, the visual and the photographic, are still in use in X-ray diagnosis; they are complementary. We shall now consider in how far the two can be improved by employing an image intensifier; this involves investigating both theoretically and practically the minimum limits of contrast and detail that can be observed for a given dose in fluoroscopy and radiography, with and without the image intensifier.

Theoretical limit of detail perception

As stated in an earlier article[1]), the perception of small object-detail is limited, according to Rose [2]) and Sturm and Morgan [3]), by the fluctuations or noise in the number of quanta involved in the observation of the particular object-detail. In fluoroscopy, for example, such detail cannot be resolved unless the difference in luminance between it and the surroundings exceeds the natural luminance fluctuation. With a radiograph, a similar argument holds good for the local fluctuations in photographic density owing to the finite size and varying concentration of the silver grains in the picture.

The magnitude of the resultant fluctuations is governed mainly by the particular stage of the image transmission at which the average number of quanta or particles, $\overline{N}$, is smallest ($\overline{N}_{\min}$). The standard deviation of the actual values of $N_{\min}$ is $\sqrt{\overline{N}_{\min}}$. The contrast between two zones of different luminance, I_1 and I_2, may be defined as:

$$C = \frac{I_1 - I_2}{I_1}.$$

The fluctuations in luminance produce contrasts determined by $I_1 \propto \overline{N}_{\min}$ and $I_2 \propto \overline{N}_{\min} - \sqrt{\overline{N}_{\min}}$; hence $C_{\text{fluct}} = 1/\sqrt{\overline{N}_{\min}}$. The minimum contrast clearly perceptible despite these fluctuations is therefore $C_{\min} = k/\sqrt{\overline{N}_{\min}}$, where k is greater than unity; it will be shown from the experimental results that the actual value of k is roughly 3. Given $\overline{N}_{\min}$, then, it is possible to calculate this minimum contrast, and also, since $\overline{N}$ is proportional to the area of the detail observed, the minimum perceptible contrast as a function of the detail diameter (d). Here we have, then, a theoretical, quantitative limit on the detail perception. If $C_{\min}$ is plotted against d on logarithmic co-ordinates the resulting curves are straight lines.

Numbers of quanta involved in the different stages of image transmission

The numbers of quanta involved in the different image-stages in fluoroscopy with, and without the image intensifier are shown diagrammatically in *fig. 1*. The integration time of the eye, that is, the time during which the eye is able to co-ordinate a certain number of light quanta into a single light-impression, is assumed to be 0.2 sec. [2])

In fluoroscopy without the image intensifier, only a very small fraction (roughly 0.02%) of the light from the screen enters the pupil of the eye. In the complete image transmission chain, then, it is at the retina that the number of quanta per image-element is smallest; the screen must absorb 100 X-ray quanta to produce one effective light quantum on the retina. Hence the perception of detail is fundamentally limited by the relative fluctuations in the number of quanta effectively absorbed by the retina.

In fluoroscopy with the image intensifier, however, the number of light quanta is so increased by the 1000 times luminance intensification, as to exceed the number of X-ray quanta absorbed; here, then, perception of detail is limited by the number of absorbed X-ray quanta.

The smallest numbers of quanta are then a factor of 40 larger, and the relative fluctuations a factor of $\sqrt{40} \approx 6\times$ smaller than in fluoroscopy without the image intensifier; the theoretical minimum perceptible contrast (for a given detail size) is therefore likewise smaller by a factor of 6.

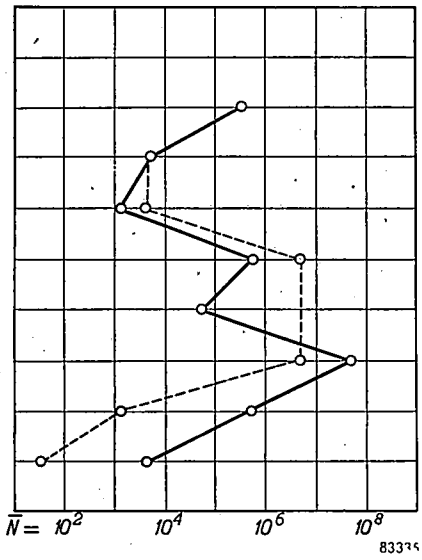
If the fluctuations in the number of X-ray quanta could be neglected, the factor by which the minimum perceptible contrast is reduced by the 1000 × luminance intensification would be very much larger [4]). In fact, such an improvement would be

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

[2]) A. Rose, The sensitivity performance of the human eye on an absolute scale, J. Opt. Soc. Amer. **38**, 196/208, 1948.

[3]) R. E. Sturm and R. H. Morgan, Screen intensification systems and their limitations, Amer. J. Röntg. Rad. Ther. **62**, 617-634, 1949.

[4]) H. R. Blackwell, J. Opt. Soc. Amer. **36**, 624, 1946.



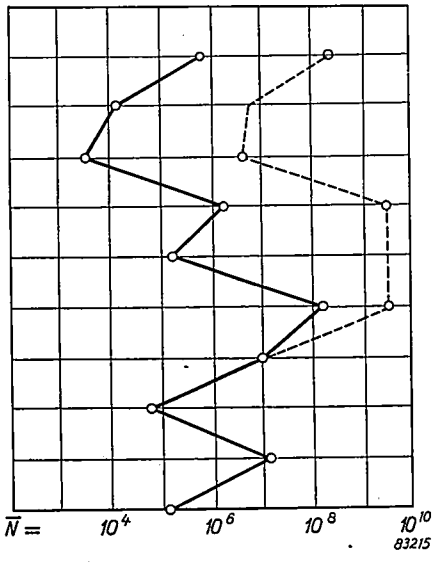
Stage of image transmission	Particles	$\overline{N}$ without I.I.	$\overline{N}$ with I.I.
On object	X-ray quanta	2.5×10^5	2.5×10^5
Transmitted by object	„	5×10^3	5×10^3
Absorbed by fluorescent screen .	„	4×10^3	1.2×10^3
Emitted by fluorescent screen .	Light quanta	4.5×10^6	6×10^5
Emitted by photo-cathode . .	Electrons		6×10^4
Emitted by viewing screen . .	Light quanta		6×10^7
In eye pupil	„	1.2×10^3	4×10^5
Absorbed by retina	„	3×10^1	4×10^3

Fig. 1. Average number of quanta or particles ($\overline{N}$), for a round detail 2 mm in diameter, effective for 0.2 sec, in different image-stages in fluoroscopy without the image intensifier (dotted line) and with it (full line).
Object: 8 cm "Philite" + stationary scatter grid. Distance to focus 90 cm. 40 kV, 1 mA.

obtained if the light from the screen were generated, not by X-rays, but by a very much larger number of relatively low-energy quanta.

The number of quanta for miniature radiography with and without the image intensifier are shown diagrammatically in *fig. 2*. Here, the time factor is not determined by the fixed period already referred to (0.2 sec), but depends upon the exposure time, since the photographic emulsion stores all the radiation imparted to it. However, the integration time of the eye is involved in the actual examination of the radiograph (last two points in fig. 2).

The radiograph (without image intensifier) refer-



Stage of image transmission	Particles	$\overline{N}$ without I.I.	$\overline{N}$ with I.I.
On object	X-ray quanta	2.2×10^8	7.5×10^5
Transmitted by object	„	4.5×10^6	1.5×10^4
Absorbed by fluorescent screen	„	3.5×10^6	3.5×10^3
Emitted by fluorescent screen	Light quanta	4.0×10^9	1.8×10^6
Emitted by photo-cathode . .	Electrons		1.8×10^5
Emitted by viewing screen. . .	Light quanta		1.8×10^8
On photographic emulsion . . .	„	1.6×10^7	10^7
In photographic emulsion . . .	Blackened grains	5×10^4	6×10^4
In eye pupil . . . } per 0.2 sec	Light quanta	1.5×10^7	1.5×10^7
Absorbed by retina } per 0.2 sec	„	2.5×10^5	2.5×10^5

Fig. 2. Average numbers of quanta or particles ($\overline{N}$), per detail 2 mm in diameter, effective in miniature radiography without the image intensifier (dotted line) and with it (full line).
Without the image intensifier: Fluorescent screen photographed with a mirror camera, $f/d = 1$, on 70 mm film ("Scopix Ortho"). Total reduction 7×. Exposure 40 kV, 180 mA sec.
With the image intensifier: Viewing screen of the image intensifier photographed through an optical system (two lenses: $f/d = 1.5$, $f = 55$ mm, and $f/d = 2.0$, $f = 75$ mm) on 35-mm film (Agfa "Fluorapid"). Total reduction 6.5 × (that is, reduction in image intensifier 9 ×, optical magnification 1.4 ×). Exposure 40 kV, 0.6 mA sec.
Object: 8 cm "Philite" + stationary scatter grid. Distance to focus 90 cm.

red to in fig. 2 was taken with the aid of a fast mirror camera. Owing to the low aperture ratio at the object side of this system, only a small fraction (roughly 0.4%) of the light from the fluorescent screen is photographically effective. Roughly 70 X-ray quanta must be absorbed by the fluorescent screen to produce enough light to cause the subsequent development of one silver grain in the film emulsion; hence the number of silver grains determines the theoretical limit of contrast. The precise relationship between the density fluctuations and the relative standard fluctuation in the number of grains per object-detail is not known. It is probably associated with the contrast (γ) of the emulsion. However, it is reasonable to suppose that unless the density is very large, its fluctuations will decrease as the number of grains per individual detail increases (that is, as the size of the grains decreases).

In radiography with the image-intensifier, as referred to in fig. 2, a roughly 1.4 times enlarged photograph of the viewing screen is taken on the film with the aid of an optical system. As in fluoroscopy, the increase in luminance compensates for the loss of light in the optical system. Hence each X-ray quantum gives rise to several silver grains (roughly 20) in the developed photographic emulsion. As in fluoroscopy with the image intensifier, then, the number of X-ray quanta absorbed by the initial fluorescent screen of the intensifier determines the theoretical limit of contrast.

With full size radiography, almost all the light from the two screens (so-called "intensifying screens") is effective, because the film is in actual contact with the screens. Measurements of the numbers of light quanta involved have shown that the smallest number (that is, the fluctuations) is governed in this case by the number of X-ray quanta absorbed by the screens.

Given the value of k, then, $C_{min} = k/\sqrt{\bar{N}_{min}}$ can be calculated from the measured numbers of quanta, for object detail of any size.

Experimental results

The next step is to determine by experiment to what extent the theoretically established threshold of perception is approached in fact.

Such measurements can be carried out with the aid of an X-ray phantom, as described in an earlier issue of this Review [5]).

The present measurements were carried out with a modified phantom provided by Prof. G. C. E. Burger. In principle, it comprises a number of plates of "Philite" (a phenol resin) one of which, known as the "phantom plate", contains several cylindrical holes of different diameters and depths. The differences in depth of the holes correspond to differences in absorption, and therefore to differences in intensity of the X-radiation passing through the

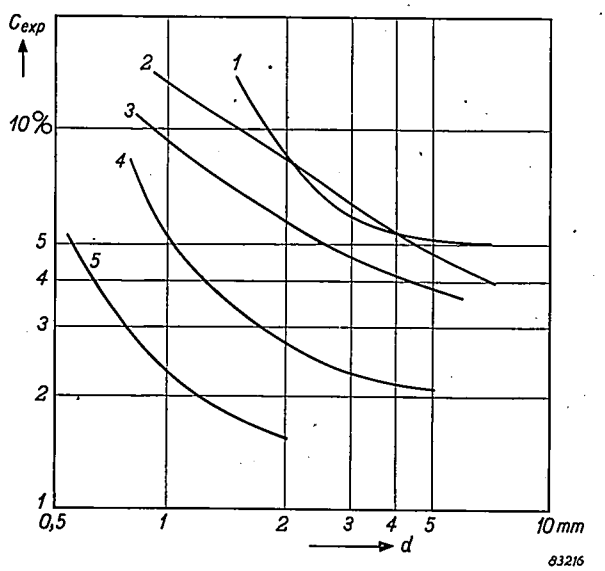


Fig. 3. Example of contrast-detail diagrams obtained by experiment.
Object: 8 cm "Philite" + scatter grid.
Distance to focus: 90 cm.
1: Fluoroscopy without image intensifier: 75 kV; 4 mA
2: Fluoroscopy with image intensifier: 67 kV; 0.1 mA
3: Fluoroscopy with image intensifier: 67 kV; 0.2 mA
4: Fluoroscopy with image intensifier: 67 kV; 3 mA
5: Full-size radiograph on "Curix": 67 kV; 3 mA sec

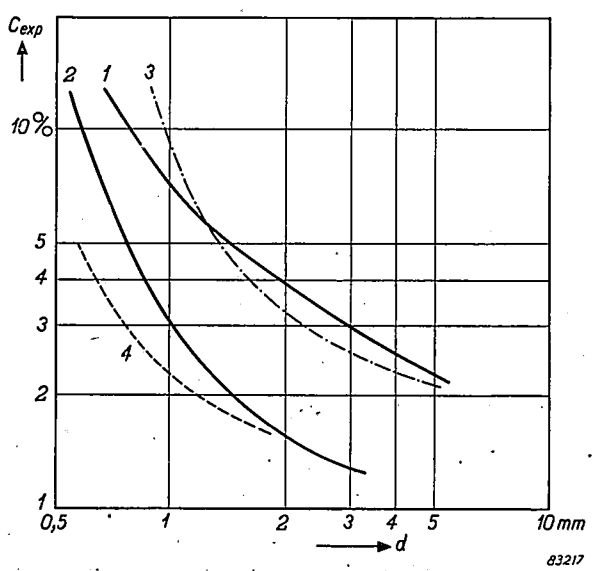


Fig. 4. Example of contrast-detail diagrams obtained by experiment, analogous to fig. 3.
Object: 8 cm "Philite" + scatter grid.
Distance to focus: 90 cm
1: Radiograph with I.I., "Fluorapid"; 67 kV; 0.2 mA sec
2: Radiograph with I.I., "Micro-File"; 70 kV; 5 mA sec
3: Miniature radiograph (mirror camera without I.I.), "Scopix"; 67 kV; 14 mA sec
4: Full-size radiograph on "Curix"; 67 kV; 3 mA sec

[5]) G. C. E. Burger, Phantom tests with X-rays, Philips tech. Rev. **11**, 291-298, 1949/50.

plate (frequently but misleadingly called "X-ray contrast"). A number of observers then indicate, either during fluoroscopy or on a radiograph of the phantom, which of the holes they are just able to see. Curves plotted from the depth and diameter of these just perceptible holes, as shown in *figures 3* and *4*, thus represent the boundary between perceptible and imperceptible contrast (C_{exp}) plotted against the diameter of the object detail.

Comparison of experimental and theoretical results

It is evident from figures 3 and 4 that in fact the measured curves are not straight, as predicted from the theory. However, the fluctuation theory holds good only for an ideal apparatus; it takes no account of imperfections, such as the invariable blurring of the image on fluorescent screens, the loss of contrast in the image intensifier, and the limited contrast sensitivity of the eye.

In most fluorescent screens, for example, the blurring exceeds 0.4 mm, and therefore precludes the resolution of detail below a certain size, depending upon the contrast.

What is the situation with regard to detail which is not unduly small, that is, say, 2 mm in diameter? Measurements to determine, for various fluoroscopic and radiographic methods, by what factor the experimental lower limit of perceptible contrast (C_{exp}) exceeds the fluctuation contrast (C_{fluct}) show that this factor is not constant (see last column of *Table I*). The extreme values occur in image intensifier fluoroscopy with X-rays of low intensity (2.2) and image intensifier radiography on film of low sensitivity (11.5); an intermediate value (4.5) is found for image intensifier radiography on sensitive film. The smaller the number of X-ray quanta employed (say, in image intensifier radiography on sensitive film), the higher C_{fluct} and the smaller the ratio C_{exp}/C_{fluct}. A ratio $C_{exp}/C_{fluct} = 4.5$ must be very close to the optimum. The lowest ratio, viz. 2.2, occurs in image intensifier fluoroscopy with X-rays of low intensity. However, it does not necessarily follow that $k = 2.2$ or, more precisely, that an X-ray contrast equal to 2.2 times the average relative standard fluctuation is invariably perceptible. The probability of a moment with a low fluctuation level increases with the period of observation, so that relatively longer examination may enable the observer to perceive a contrast lower than that corresponding to $C_{min} = k/\sqrt{\bar{N}_{min}}$ [6]).

In view of these arguments, we consider that $k = 3$ is a good approximation to the optimum value.

[6]) Incidental, favourable fluctuations are of no real value in radiography, because they never cover the entire field.

Table I. Minimum contrast observed (C_{exp}), minimum number of quanta ($\bar{N}_{min}$), the fluctuation contrast computed from it (C_{fluct}), and the ratio C_{exp}/C_{fluct}, for the observation of a detail 2 mm in diameter in a "Philite" phantom 8 cm thick, by different methods.

Method	kV	mA or mA sec	C_{exp} %	$\bar{N}_{min}$	C_{fluct} %	$\frac{C_{exp}}{C_{fluct}}$
Fluoroscopy without Image Intensifier	75	4	8	1.2×10^3	2.9	2.8
Fluoroscopy with Image Intensifier	67	0.2	4.4	2.5×10^3	2.0	2.2
Radiography with Image Intensifier on "Fluorapid" 35-mm film. Optical system $d/f = 1:4.5$	67	0.2	4	12.5×10^3	0.89	4.5
Miniature radiography without Image Intensifier; mirror camera	67	14	3.25	30×10^3	0.58	5.6
Fluoroscopy with Image Intensifier	67	3	2.8	37.5×10^3	0.52	5.4
Radiography with Image Intensifier on "Micro-File" 35-mm film. Optical system $d/f = 1:2$	67	1.5	1.55	100×10^3	0.31	5.0
Radiography with Image Intensifier on Pan F 35-mm film. Optical system $d/f = 1:4.5$	67	2	2.6	125×10^3	0.28	9.3
Full-size radiography	67	3	1.5	450×10^3	0.15	10.0
Radiography with Image Intensifier on "Micro-File" 35-mm film. Optical system $d/f = 1:4.5$	67	8	1.6	540×10^3	0.14	11.5

For the full-size radiograph and for the image intensifier radiograph on fine-grain film, the experimental values of C_{exp}/C_{fluct} are relatively high (>10). Here, then, the perception of a round detail 2 mm in diameter is virtually independent of the quanta fluctuations, but is governed almost exclusively by imperfections in the apparatus. Although difficult to eliminate such imperfections are not of a fundamental nature.

Direct perception of the quanta fluctuations

Confirmation of the above arguments is obtained by the fact that the X-ray quanta fluctuations can be observed direct in the X-ray image. In image intensifier fluoroscopy with X-rays of not unduly high intensity, the image exhibits a certain

amount of "noise". This noise is closely connected with the relatively small number of X-ray quanta involved in the forming of the luminous image. The higher the intensification factor of the image intensifier, the more readily are the fluctuations perceived by the observer, especially when the viewing screen is observed through an optical system of high magnification.

As already explained, in photographs taken with an image intensifier we generally find an appreciable difference between the experimental threshold of perception and the theoretical threshold determined by the quanta fluctuations (high ratio C_{exp}/C_{fluct}). Hence it is unlikely that the fluctuations will be perceptible in such radiographs. However, they are seen distinctly in image intensifier radiographs taken with a relatively very much smaller amount of X-radiation but a faster optical system. *Fig. 5* gives

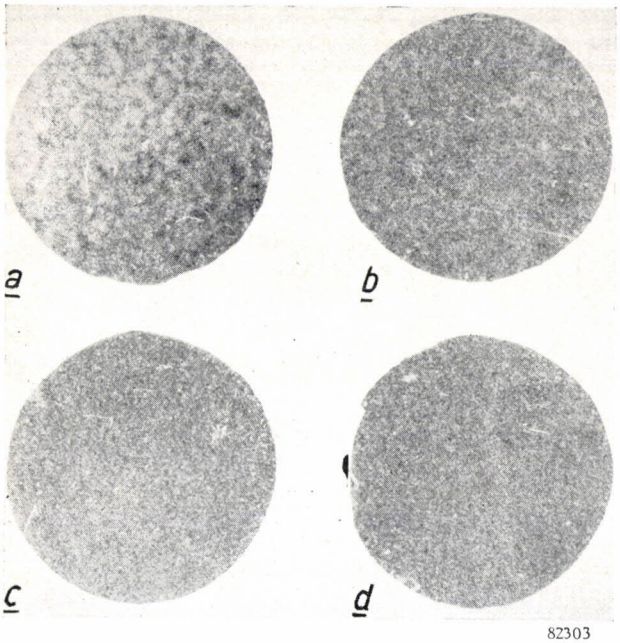


Fig. 5. Perceptible X-ray fluctuations in radiographs taken with the aid of the image intensifier. The photographs were taken with different stops, becoming smaller from *a* to *d*.

an example. Here we have four reproductions of parts of image intensifier photographs. They differ only in respect of the amount of X-radiation per photograph, which is in the ratio $a : b : c : d = 1 : 5 : 32 : 160$; this variation was made possible by varying the stop of the lens system.

It can be seen that the photographs differ quite appreciably. Photograph *a*, for example, taken with the maximum lens aperture of the system, shows a coarser structure than the others. The high intensification produced by the intensifier, combined with the use of the maximum stop, enabled this photograph to be taken with a relatively small number of X-ray quanta; the very noticeable local fluctuations in density arise from the fact that each X-ray quantum blackens an individual cluster of grains in the emulsion. Photograph *d*, taken with a very much smaller aperture, necessitated a very much larger number of X-ray quanta to expose the grains. Because the number of X-ray quanta and hence that of grain clusters in *a* is very much smaller, the relative statistical fluctuation in the number of individual aggregates is more noticeable. Hence the grain distribution of photograph *a* is appreciably less uniform than that of photographs *c* and *d*. A fine rose on a watering can sprinkles more evenly than a bucket.

Fundamental and practical advantages of the image intensifier

The advantages of the image intensifier as applied to fluoroscopy and radiography will be evident from fig. 1 and fig. 2. Without the intensifier, only a small proportion of the X-ray quanta are absorbed by the screen, on an average, in fact, only 1 in every 100, is effective; hence the relative value of the fluctuations is roughly 10 times the fluctuation inherent in the quanta absorbed. Accordingly, the detail perception is then fundamentally inferior to that obtained with the image intensifier, which makes all the X-ray quanta absorbed effective. Here, every quantum of radiation absorbed by the screen gives rise to a threefold stimulus in the retina, or exposes several silver grains in the film (e.g. 20—30), the precise number depending upon the type of film and the aperture ratio of the optical system.

Certain conclusions regarding the observation of not unduly small detail (one or two mm) with low contrast may be drawn from the values of C_{exp}/C_{fluct} established.

In fluoroscopy with the image intensifier, a very close approximation to the theoretical threshold of perception is reached experimentally; hence we cannot expect to gain very much from technical improvements when employing the conventional examination procedure.

Again, in image intensifier radiographs taken on 35-mm film with the aid of a fast optical system, the threshold of perceptibility is governed, in practice, by the fluctuations of the X-ray quanta. Consequently all that can be done to obtain more information in such a case is to increase the X-ray dose: a film of finer-grain may then be used. Ultimately, then, a limit is imposed by the imperfections of the apparatus.

Finally, let us consider one or two values taken

from figures 3 and 4 to demonstrate the advantages of the image intensifier technique. It is seen that the X-ray dose to observe a given threshold-contrast with the image intensifier is only 1/40 of that required in conventional fluoroscopy without the intensifier. The dose to take an image intensifier photograph on 35-mm film is a factor of 70 smaller than that required for a 70-mm miniature radiograph taken without the image intensifier, the film-sensitivity being the same in both cases and the X-ray images being reproduced roughly 7 times smaller on the film. There is very little difference in detail perception as between the two photographs. This is very important from the point of view of the filming of moving organs with the aid of the image intensifier.

With a very fine-grain film, the image intensifier enables us to take 35-mm photographs of a quality comparable with that of a full-size radiograph; moreover, if a fast optical system is employed, the dose required is several times smaller than in conventional radiography.

III. OPTICAL AIDS FOR THE IMAGE INTENSIFIER

by P. M. van ALPHEN. 621.386.8:616-073.75:621.383.8:535.82

It is neither necessary nor customary to employ optical aids in the examination of an X-ray image on a conventional fluorescent screen. However, to resolve the detail of the image, the eye must be properly adapted to the low luminance level of the screen.

With the image intensifier, however, precisely the reverse holds good. Although bright enough to be examined under ordinary room lighting, the visible image formed in the intensifier is so reduced by the electron optical system that optical instruments are required to enable the details to be perceived.

The image intensifier therefore necessitates a different approach, not only to fluoroscopy, but also to photographic work. The mirror camera, although eminently suitable for photographing the conventional low-luminance X-ray image, by virtue of its high aperture ratio, cannot be used without modification to photograph the relatively small image in the intensifier.

The problem of magnifying small objects for observation or photography is of course not new and many ways of doing it are known. The magnifying glass, microscope, telescope, camera, etc. are well-known examples. It is however, desirable to review the general requirements to be imposed on an optical system for use with the image intensifier. In so doing, we shall also discuss various designs examined during the development of the Philips image intensifier.

Quantities and concepts used in photometry

As an introduction to the above-mentioned discussion, let us consider the quantities employed as measures of light, quantities which are frequently misinterpreted.

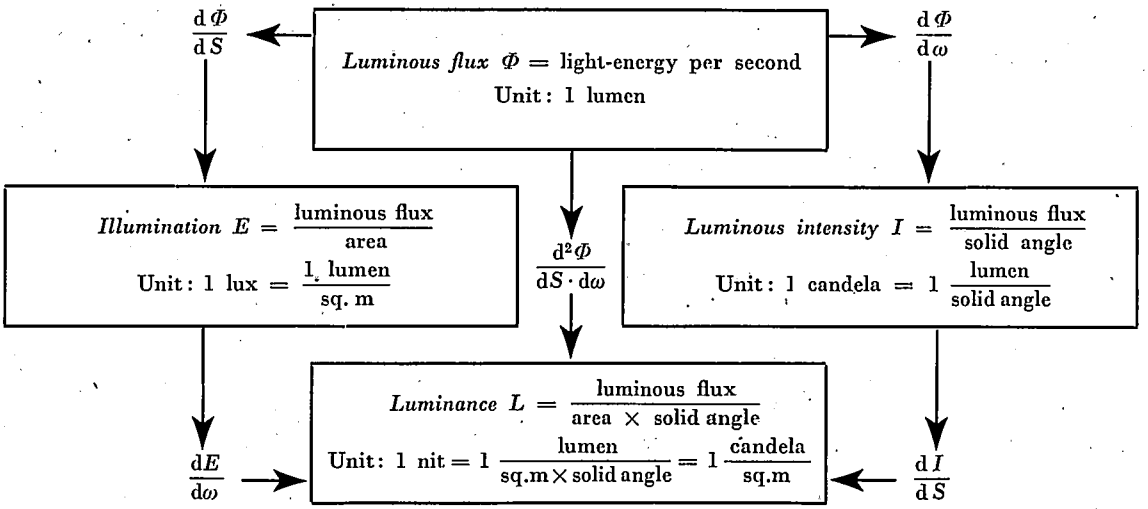
A light source radiates energy; that part of this radiant energy whose wavelength is within the visible spectrum forms the luminous flux (unit: the lumen).

Luminous flux, then, is the light proceeding from the entire surface of the light source in all directions. For certain problems it is necessary to determine the distribution of this flux in space; in other cases, we are more concerned with the flux-distribution as between different points on the surface of a light source or an illuminated surface.

In many cases we require the quantity luminous flux (throughout the solid angle) per unit area; in other cases we wish to know the luminous flux (from the whole of a surface) in a given direction that is, per unit solid angle. We may also go a step further and consider the luminous flux emitted by a given surface in a certain direction. Thus the overall luminous flux output may be divided according to area or direction, or area *and* direction. The quantities thus defined have acquired individual names, i.e. illumination level (or briefly illumination), luminous intensity and luminance (or brightness), respectively. They are shown diagrammatically in *fig. 1* and *Table I*.

How, then, are these quantities affected by such optical elements as lenses and mirrors? Since optical systems do not generate light, they do not increase the overall luminous flux, which therefore remains constant, apart from some absorbtion in the system. However, the illumination and luminous intensity do not necessarily remain constant. The luminous flux can be concentrated, with the aid of a lens or mirror, upon certain points (or areas), or in certain directions (solid angles). This concentration may produce substantial changes in both luminous

Table I. Diagram of the relationship between light quantities.



intensity and illumination[1]). The luminance however, is not affected; since the luminous flux per unit area and per unit of solid angular measure is associated with the flux density of the radiation, or radiance, it is a property of the light source. Hence it cannot be increased by means of an optical system.

If the luminance could be so increased, we would observe the very strange phenomenon of energy (in this case light) being transformed from a state of low concentration to a state of higher concentration without the performance of work. It is evident, then, from the reversibility of light radiation that luminance is affected only by absorption.

Let us now consider the relationship between luminous flux (Φ) and luminance (L) (*fig. 2*). It is seen that, at a given luminance level, the luminous flux increases with the solid angle and with the area of surface S. This holds good for a luminous, as well as for an illuminated surface. However, it should be borne in mind that L may also depend upon direction. In the simplest case of a radiator, whose luminance is constant in all directions, that is, a radiator obeying Lambert's law, the luminous flux is:

$$\Phi = \pi LS \sin^2 \tfrac{1}{2}\varphi \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where φ is the apex angle of the cone of flux.

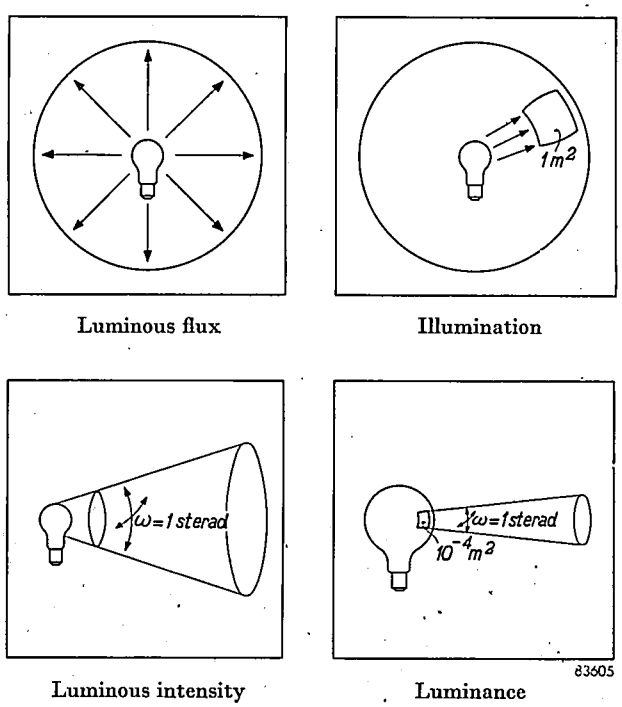


Fig. 1. The four photometric light quantities. Luminous flux is the overall light-output of the lamp. Illumination is the luminous flux incident upon 1 sq. m. Luminous intensity is the luminous flux radiated by the lamp per unit solid angle. If the light is simply freely radiating, light initially within a certain solid angle remains within it; hence luminous intensity is independent of the distance to the light source. Luminance (brightness) is the luminous intensity per unit surface area (sq. m or sq. cm) of the light source; hence it is also the luminous flux per unit area and per unit solid angle.

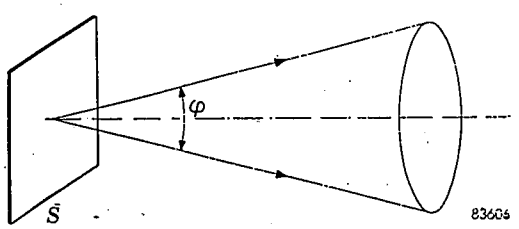


Fig. 2. The luminous flux Φ radiated within an angle φ by a surface with area S and luminance L in accordance with Lambert's law, is $\pi LS \sin^2 \frac{1}{2}\varphi$.

[1]) The simplest example of this is a candle with a plane mirror behind it. The mirror increases the the luminous intensity to two candles, but reduces the solid angle containing this luminous intensity to 2π, that is, the half of a sphere.

Application to optical systems

We shall now consider the above formula in four different cases:

1) the human eye;
2) a camera;
3) an arbitrary optical system;
4) the image intensifier.

The sensation produced in the human eye is governed by the strength of the stimulus imparted to the optic nerve in the retina. This stimulus depends upon the amount of luminous flux Φ entering the eye, for which formula (1) holds good. For the purpose of our argument, the area S of a retinal element may be considered constant (fig. 2). Let Φ be the vertical angle of a cone whose apex is on the retina and whose base is the pupil of the eye. For a high enough luminance level, this angle is constant. The light sensation, or subjective brightness is therefore governed entirely by the luminance L. However, it is here assumed by implication that the cone φ is completely filled by light rays, or, in other words, that the eye pupil is fully illuminated by the optical system through which it is looking. To satisfy this condition, the smallest aperture through which the light cone emerges from the optical system, that is, the so-called exit aperture must be larger than the pupil of the eye. Since the eye pupil distends according as the luminance decreases (*fig. 3*) it is necessary, therefore, in designing an optical system, to pay due regard to the luminance range for which it is to be employed (day or night glasses).

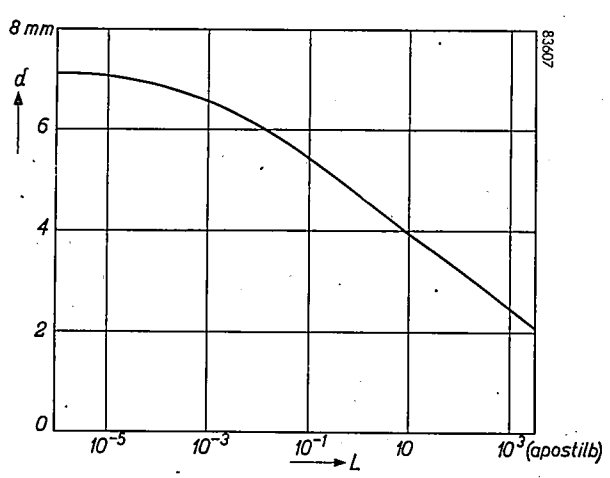


Fig. 3. Diameter (d) of the pupil of the human eye, plotted against the luminance L.

The luminous flux is also important in cameras; however, this statement requires elucidation. In referring to the luminous flux received by the optical system of the camera from the object, and projected onto the photographic film, we must state the area upon which the flux impinges, since the required exposure is shorter, the higher the illumination level $\Phi/S = \pi L \sin^2 \frac{1}{2}\varphi$. Given L, that is, the luminance of the object, we are still free to choose $\sin \frac{1}{2}\varphi$ as large as we please. When this has been done, we also establish Φ/S, and therefore the exposure; the latter does not vary with the focal length, but the size of the image is, of course, proportional to the focal length. On the other hand, if the luminous flux is fixed (e.g. because it is supplied by another optical system) minimum exposure may be obtained by reducing the size of the image on the film as far as possible. This is done by using a lens of suitable diameter and minimum focal length.

In practice, the problem is usually more complex than the above argument suggests, since not only the exposure, but also the sharpness of the picture, the grain of the emulsion, the price of the optical system and the variation of $\sin \frac{1}{2}\varphi$ with the magnification, are involved.

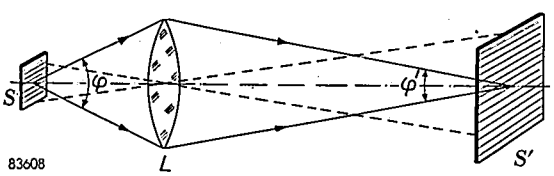


Fig. 4. Derivation of the sine condition. Lens L transforms area S into S' and angle φ into φ'.

By applying equation (1) to a lens or optical system, we obtain a simple derivation of a formula referred to variously as the Helmholtz, Huygens, or Lagrange formula, but more often called Abbe's sine law. Leaving absorption and reflection losses out of account, the luminous flux Φ' inherent in the image is equal to the effective flux intercepted by the lens (Φ; see *fig. 4*). Hence we have:

$$\Phi = \Phi', \quad \text{or} \quad \pi LS \sin^2 \tfrac{1}{2}\varphi = \pi L'S' \sin^2 \tfrac{1}{2}\varphi'.$$

We have already seen that the luminance remains constant; hence $L' = L$, and

$$\frac{S'}{S} = \frac{y'^2}{y^2} = \frac{\sin^2 \frac{1}{2}\varphi}{\sin^2 \frac{1}{2}\varphi'},$$

or

$$N = \frac{y'}{y} = \frac{\sin \frac{1}{2}\varphi}{\sin \frac{1}{2}\varphi'} \quad \ldots\ldots \quad (2)$$

where N is the magnification, and y' and y are the diameters of the image and the object, respectively. From the point of view of illumination, then, an optical lens is a transformer of area and angular aperture, the transformation invariably taking place in such a way that luminous flux and luminance remain unchanged, whereas luminous intensity and illumination may vary.

To come back to the image intensifier, it produces two different intensifications. Firstly, the high voltage on the tube imparts extra energy to the photo-electrons, thus making the fluorescence of the viewing screen brighter than that originally produced on the photo-cathode; this is known as the lumen intensification of the tube. Secondly, by virtue of the N-times electron-optical reduction, the light thus intensified proceeds from an area smaller than that of the photo-cathode, so that here (as opposed to purely optical reduction) an extra luminance-gain of N^2 is obtained. This factor contributes the most to the overall luminance intensification. The optical system through which the tube is then viewed must clearly be so designed as to enable the N-times smaller image to be seen without any loss of luminance.

Formula (2) enables us to determine the maximum magnification obtainable, without loss of luminance, by means of an optical system. As we have already seen, the whole area of the eye pupil (say, 8 mm in diameter) must be filled with light. The apex of the light cone is on the image observed, i.e. at the closest distance of distinct vision, say, 25 cm. Hence $\sin \frac{1}{2}\varphi = 4/250 = 0.016$. It will be evident that the maximum vertical angle of the cone intercepted by the optical system is 180°, which is consistent with $\sin \frac{1}{2}\varphi = 1$; hence the maximum magnification enabling the light to fill the whole eye pupil is $1 : 0.016 = 60 \times$. Any further magnification reduces the luminance of the observed image. In reality, however, the vertical angle of the intercepted light cone is invariably very much smaller than 180°; the maximum permissible magnification is therefore also smaller.

We shall now describe a number of different optical systems, outlining the conditions governing their use. It is necessary to point out, however, that one very important factor, that is, the convenience of the observer, is virtually ignored in these discussions. In practice, it is usually preferable to accept a slightly less perfect but easily adjustable system to one which is perfect only in theory.

Optical aids to fluoroscopy

The magnifying glass

The simplest means of observing the small image on the viewing screen of the image intensifier tube is an ordinary magnifying glass. This has not been used so far, however, for the reasons given below. To enable the image to be seen with both eyes, the magnifying glass must be employed as a reading glass. The magnification is then relatively weak; hence it would impose a limit on the permissible electron-optical reduction in the image intensifier. We have already seen that the luminance intensification, which is most important in fluorsocopic examinations, increases as the square of this reduction. A loss of a factor of 4 in electron-optical reduction is equivalent to a loss of a factor of 16 in luminance.

Within the required field of view (14 mm), a very much stronger magnification can be obtained by placing the glass to only one eye. However, both the eye and the glass must then be brought close to the viewing screen, a condition which cannot always be satisfied.

Moreover, the principal objection to the use of a magnifying glass is that the electron-optical system of the image intensifier produces an inverted image on the viewing screen; when the latter is observed through a magnifying glass, then, the image of the object is seen upside down. This is very distracting to a doctor unaccustomed to such image inversion in ordinary fluoroscopy. It can, of course, be corrected by means of image-erecting mirrors or a system of prisms, but they detract from the essential simplicity of the equipment; moreover, they preclude the use of a strong magnifying glass, since there is then no room to place them in the path of the rays.

Microscopes

These considerations have led to the idea of employing one lens to erect the image and another, a magnifying glass, for viewing it. This is essentially the same arrangement as in a microscope, where the objective functions as an image-erecting lens and the eyepiece is, in effect, a magnifying glass. Now, to restore the image to its original size we require a magnification of roughly 10 ×, which could be obtained with an eyepiece of, say, 5 ×, and an objective of 2 × magnification. The diameter of the exit pupil depends upon the aperture of the microscopic objective, which in a relatively weak objective, as here considered, is not very large. We employed a microscope having a numerical aperture of 0.07 and an exit aperture between 4 and 5 mm in diameter. It is smaller than the pupil of a dark-adapted eye, and would therefore be a possible cause of loss of luminance. However, practical experience has shown that with the 1000 × intensification, dark-adaptation of the eye virtually never occurs.

Binocular microscope

Viewing with both eyes is usually less fatiguing than with one eye. However, we wish to avoid the

binocular arrangement employed for ordinary microscopes of this type, that is, a semi-transparent mirror or prism distributing the available light evenly between both eyes, because it reduces the luminance at least by half. We are therefore compelled to double the light-gathering power of the system and employ an objective with a very large exit pupil, or, as in Greenough's microscope, two objectives, and two eyepieces, side by side.

an angled arrangement, a full description of which will be given in the following article (IV).

Another possibility will now be described. In it, the monocular microscope already described is retained for viewing, but two mirrors, parallel to each other and at roughly 45° to the path of rays, are introduced into their path. They produce a lateral displacement of the image, but do not rotate it (see IV). By making the viewing system rotate

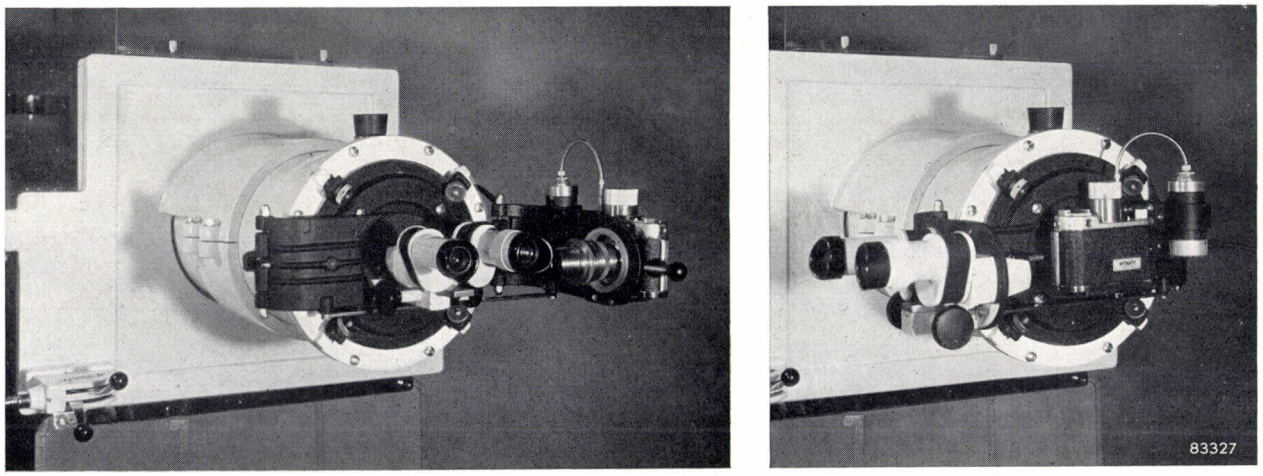


Fig. 5. Binocular microscope and camera, mounted so that they can be swung alternately into position in front of the image intensifier. The left-hand picture shows the microscope in use; in the right-hand picture the camera is in use.

Most of the microscopes of this pattern now on the market contain Porro image-erecting prisms; hence they are not very suitable for our purpose. Accordingly, we began our experiments with two monocular microscopes side by side at an angle of 18°. With this arrangement, however, it is rather difficult to adapt the system to individual eye-spacing (varying from 55 to 75 mm), without affecting the focus. Prisms are therefore used (despite some light-loss) to deflect the path of the rays slightly, thus enabling the distance between the eyepieces to be varied by rotating them about the optical axis, and differences in focussing as between individual observers to be compensated by adjusting the two eyepieces separately. The binocular microscope thus obtained, combined with a camera, is shown in *fig. 5*.

Movable mounting

The above mentioned arrangement is very useful for an image intensifier operated in a fixed position. With an image intensifier so mounted that it can be turned in all directions, however, we must pay more attention to the problem of bringing the eyepiece within easy reach of the eye. From this point of view, a monocular system, rotated by means of movable mirrors or prisms, is best. It is, in effect, about the axis of the image intensifier tube, the eyepiece can be adjusted to the eye-level of the particular observer (*fig. 6*). Moreover, by allowing the second mirror to tilt slightly about an axis perpendicular to the plane of the drawing, the

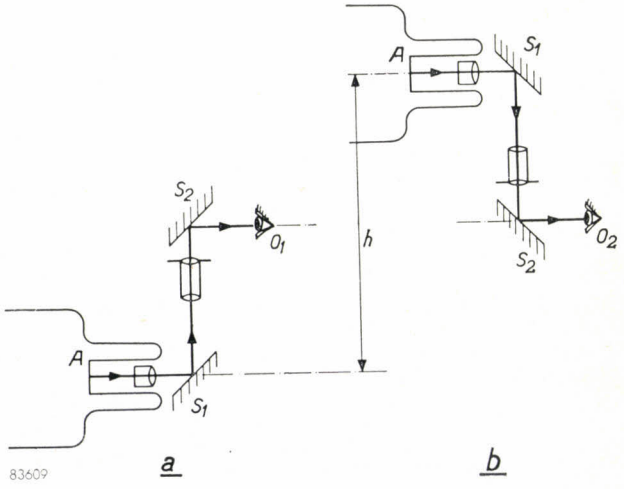


Fig. 6. Periscope optical system. (*a*) Two mirrors parallel to each other (S_1 and S_2) bend the path of rays, but do not change the position of the observed image. (*b*) By rotating the whole system about axis AS_1, the image intensifier can be raised or lowered a distance h without varying the level of the eye (O_1 or O_2). The direction of observation can be varied by tilting mirror S_2 about an axis perpendicular to the plane of the drawing.

viewing direction can be adjusted to a comfortable head position. With some movable systems, the thought of peering into the eyepiece tends to produce a slight feeling of apprehension: the observer is aware that it may hit him in the eye when moved. However, this can be avoided, either by providing a suitable head support (see IV), or by so designing the optical system that the point where the eye must be to observe the entire image, that is, the exit pupil of the instrument, is some distance away from the metal parts of the instrument. Moreover, if the exit pupil be as large as possible, the eye will not be restricted quite so much to one particular viewpoint.

With such an arrangement, it is not easy to find the proper position for the eye in relation to the system, if it happens that the eye is not in the path of the rays, the observer cannot see the image and does not know which way to move the head in order to see it. This problem has now been solved in the following manner (*fig.* 7). A piece of frosted glass is introduced into the path of rays at the point (B) where an image is formed by the objective (L_1). The glass is so frosted as to diffuse only slightly in lateral directions, and transmits mostly in the direction perpendicular to its surface, that is, along the axis of the system. This frosted glass screen is viewed through the eyepiece (L_2), then employed, in effect, as a magnifying glass, Looking obliquely through L_2 towards the frosted screen, the observer sees an image, but a faint one owing to the very limited diffusion of the glass in the non-axial direction towards his eye. However, he has only to move his head to perceive in which direction he should move it to increase the apparent luminous intensity of the image; in this way, the position of maximum image intensity is soon located. It lies precisely on the optical axis of the instrument, that is, on the same line as the above-mentioned exit pupil. When the frosted glass is withdrawn, then, the eye is exactly in the path of rays, and it requires no more than a slight forward or backward movement of the head to observe the entire image plane and its maximum luminance through L_2, which then functions as an eyepiece. If the correct eye-position is lost during the moving of the image intensifier, it is quickly found again by re-inserting the frosted glass in the path of rays. In some cases, where the X-ray dose makes it desirable, the frosted glass may be left in the system; this greatly facilitates the observation, but slightly reduces the luminance and sharpness of the image.

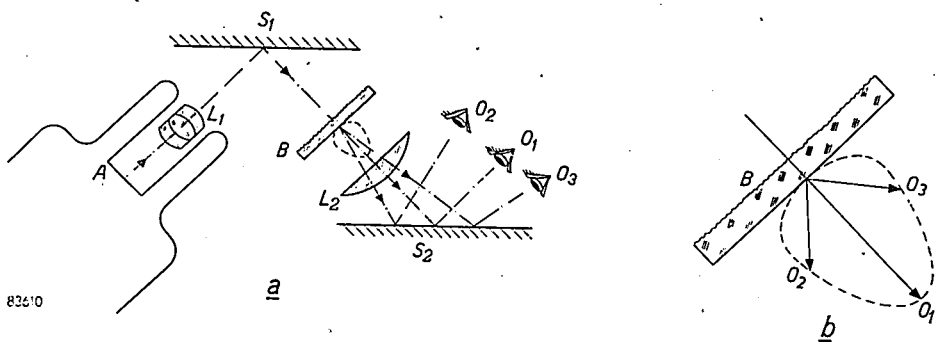


Fig. 7. Principle of the frosted glass view-finder (a). The fast lens L_1 projects an image of A upon mirror S_1, which reflects it into the only slightly frosted glass screen (B). The image on this screen is viewed through a system comprising a magnifying glass (L_2) and a mirror (S_2). The eye is attracted to position O_1 because from here the image is brightest. When the eye is kept in this position and the frosted screen is withdrawn from the path of rays, the image is seen in full brightness. L_1 and L_2 then operate as the objective and eyepiece of a microscope. The diagram (b) shows the diffusing pattern of the frosted glass.

Photographic optical system

There is a wide choice of cameras for recording the image on the viewing screen of the image intensifier. One or two general remarks on the subject may therefore be useful. We have first to consider whether we require single photographs taken at intervals, a series of photographs taken in rapid succession, or cine films to enable us to study movements or sudden effects. When this question is settled, we must decide upon a suitable picture size and choose the lenses with which to obtain it. In this connection, we must also decide on the type of film, and even consider the most suitable design for the image intensifier itself.

Let us now consider the question of film size. Given an image intensifier tube with a viewing screen of a certain luminance, we require an optical system with the highest possible light-gathering power to photograph it. Using the best optical system available, it is possible to intercept only a certain proportion of the luminous flux emitted by the viewing screen; hence it is not advisable to distribute this intercepted flux over a large area of film. Accordingly, we take the film as small as possible, that is, the smallest size consistent with the detail required and the resolving power of the particular photographic material. Another reason for choosing a small film size is that most of the fast optical systems now on the market are designed for miniature film.

Next, we must deal with the problem of forming

an image of an object (the viewing screen) on film of roughly the same size. However, photographic optical systems, and notably the ones considered for our purpose, are designed to form reduced images of large objects, that is, to produce images at the focus. They are less suitable for the formation of full-size images, because in so doing they fail to correct the aberrations in the image [2]). Moreover, with images formed on the scale of 1 : 1, the object and image distances are both equal to twice the focal length of the system; hence the angular aperture ($\frac{1}{2}\varphi$) is only half as wide then, as in the formation of an image at the focus.

At first glance it may seem that we can go a long way towards solving this problem by increasing the size of the viewing screen in the image intensifier tube, that is, by designing the intensifier so as to give a less drastic electron-optical reduction of the image. However, it is readily shown that in fact this would be a mistake, since it would deprive us of part of the luminance-gain from electron-optical reduction, which is one of the principal advantages of our system; although the particular lens would then operate under conditions more favourable as regards angle $\frac{1}{2}\varphi$ (the optimum choice

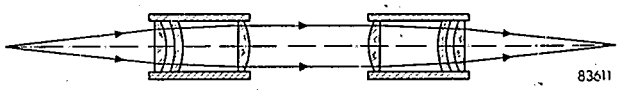


Fig. 8. Two lens systems in tandem: to form an image of roughly full size, two optical systems may be employed, the one with the object, and the other with the image in focus. The rays associated with any given image point are parallel between the two systems.

of optical system and film size is not affected by the supposed smaller electron-optical reduction in the image intensifier), the luminance (L) would decrease as the square of the diameter of the viewing screen. The final result, then, is that the film would receive less light (see equation 1).

However, the problem can be solved in another way, i.e. by the simple expedient of using, two ordinary photographic lens systems in tandem, so that the image produced at infinity by the one is focussed by the other (see *fig. 8*). Thus each lens is utilised in the manner best suited to it, the net angular aperture being twice that of a single optical system forming an image on the scale of 1 : 1.

[2]) This is illustrated most simply by a mirror system: with an infinitely long object distance the parabolic mirror is free from spherical aberration; with a finite object distance, however, it is necessary to employ the focal points of an elliptical mirror.

Although the path of rays between the two optical systems is parallel, they should be placed as close together as practicable; otherwise, rays oblique to the axis may be cut off (vignetting; see *fig. 9*). A certain amount of magnification can be procured

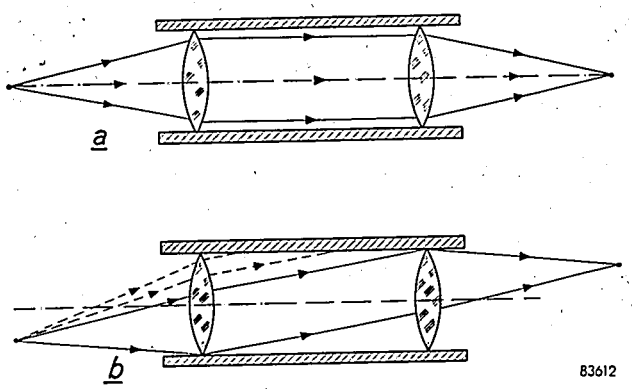


Fig. 9. Vignetting effect. The whole of the beam on the axis, that is, rays corresponding to the centre of the image, is transmitted (a), whereas oblique rays constituting the edges of the image are cut off sharply at one side (b).

with such a system by employing two lenses with different focal lengths, say, $f_1 = 50$ mm and $f_2 = 60$ mm. The magnification is equal to the ratio of the focal lengths: $N = f_1/f_2$.

Using a reflex camera, it is possible to keep the image to be photographed in view until immediately before the exposure. When filming the image, however, we require a view-finder to view the image continuously; also there is not very much space for a viewing system in front of or behind the lens. By placing a small mirror or prism between the two objectives, we gather a certain amount of light from

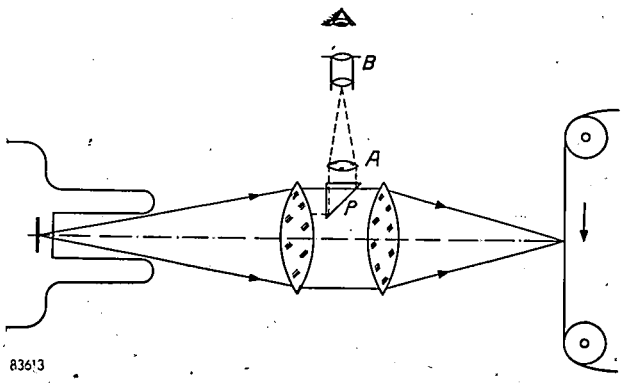


Fig. 10. View-finder for the tandem combination of lenses. Because prism P is placed in the parallel rays, it does not intercept much light. Looking through telescope AB at prism P, the observer sees the object apparently at infinity.

the image at infinity (*fig. 10*), enabling it to be viewed either with the naked eye or through a small telescope during the filming of the image. An alternative method is to provide the shutter of the film camera with a reflecting surface. Such measures have little effect on the photographic optical system.

IV. EQUIPMENT FOR SPOT FILM RADIOGRAPHY INCORPORATING AN IMAGE INTENSIFIER FITTED WITH A PERISCOPE OPTICAL SYSTEM

by H. VERSE *) and H. JENSEN **).

621.386.8:616-073.75:
621.383.8:535.82

The radiologist, long accustomed to working with radiographs and a fluorescent screen, is rather at a loss when first confronted with the picture provided by an image intensifier. The picture is brighter, but its location is unconventional. The relatively higher luminance of the image enables him to see that it is green; moreover, his eyes are not dark-adapted and the subjective contrasts appear quite different from those to which he is accustomed. Also, the relatively small field of view (13.5 cm) makes it more difficult for the radiologist to form an overall impression of the particular part of the body examined. Hence he may find that he requires the larger but fainter field of an ordinary fluorescent screen, as well as the image intensifier. A diagnostic apparatus specially designed to meet this need has been built in the X-ray factory of C.H.F. Müller at Hamburg [1]). It is intended to provide the radiologist with facilities for working either with an ordinary fluorescent screen, or with the image intensifier, and facilities for taking either direct full-size radiographs, or miniature fluorographs from the screen of the intensifier. Also it enables him to place the subject upright, prone, or in any intermediate position. The design includes several special features and the image intensifier itself is equipped with a novel viewing optical system to make viewing during the tilting of the subject as convenient as possible.

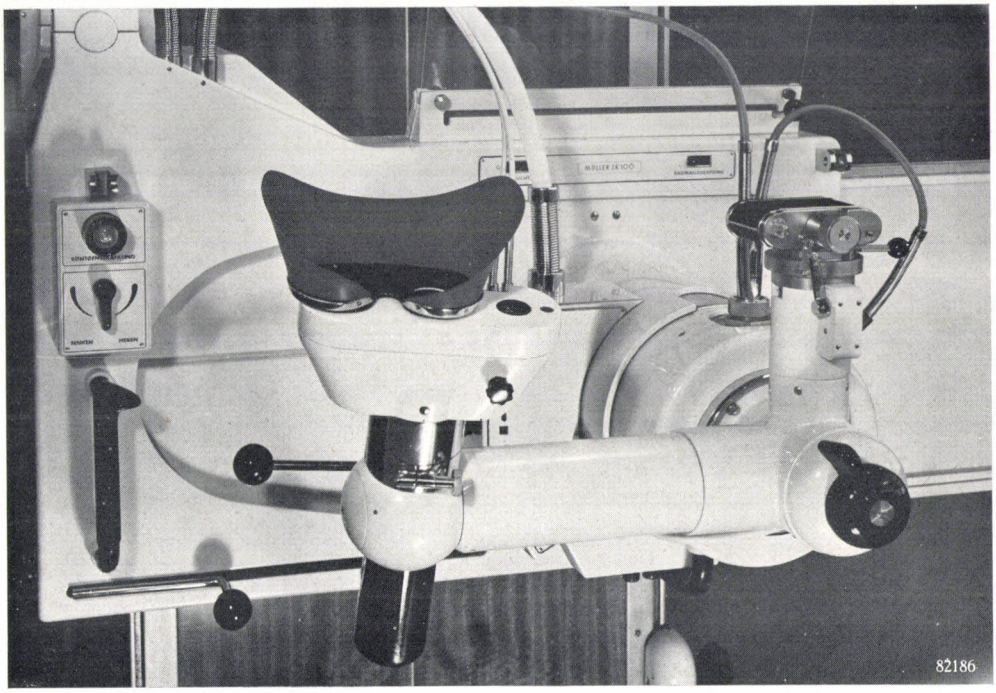


Fig. 1. Periscope viewing system of the "Müller" diagnostic apparatus ZK 100. The cover of the image intensifier tube, to which the system is attached, is seen on the right, and the rubber forehead-support with the eyepiece aperture on the left. The black knob in the bottom right-hand corner is used to deflect the light-rays when required into the camera immediately above it.

Principle of the observation

The viewing optical system (*fig. 1*) is so designed as to enable the observer to stand aside from the axis of the beam, that is, from the patient. The system can be rotated to vary its slant, and thus

*) C. H. F. Müller Aktiengesellschaft, Hamburg.
**) Allgemeine Deutsche Philips Industrie G.m.b.H., Hamburg.

[1]) H. Verse and H. Jensen, Ein Untersuchungsgerät mit Röntgenbildverstärker, Fortschr. Röntgenstrahlen **79**, 115-118, 1953.

enables the direction of view and the height of the eyepiece to be adjusted to the eye-level of the doctor carrying out the examination. From the optical point of view, the system is an angled, rotatable microscope with an optical magnification of 10 times, through which the image of the object is seen restored to its full size. Light rays from the viewing screen of the image intensifier tube

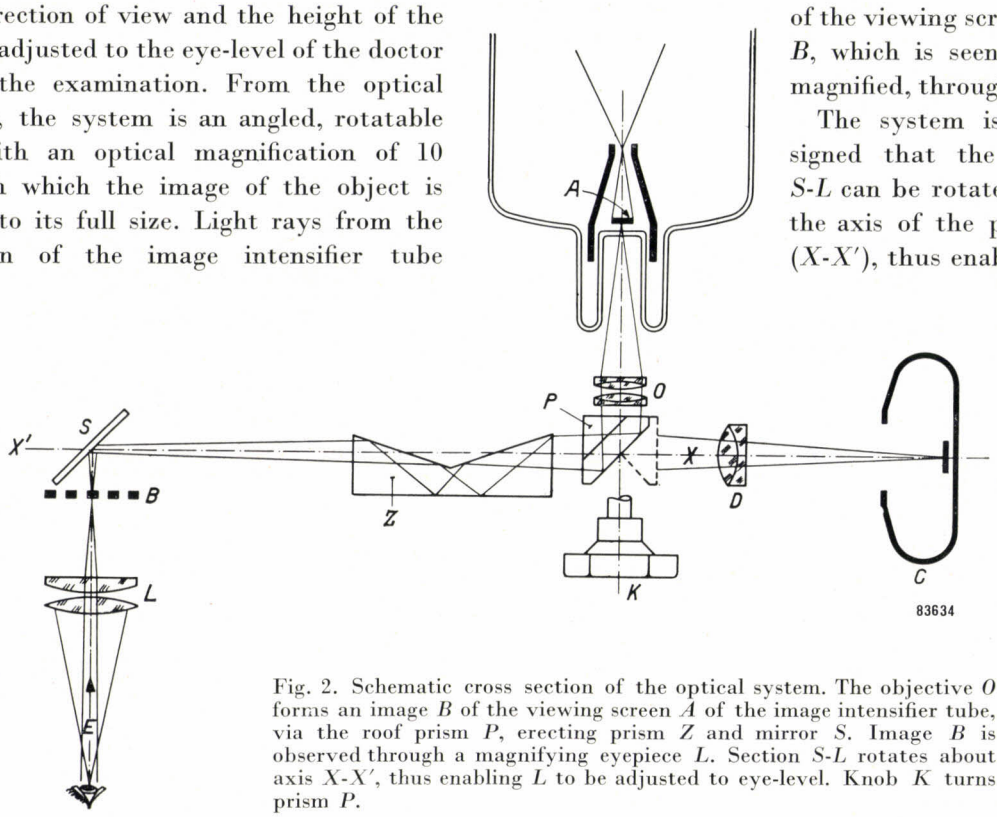


Fig. 2. Schematic cross section of the optical system. The objective *O* forms an image *B* of the viewing screen *A* of the image intensifier tube, via the roof prism *P*, erecting prism *Z* and mirror *S*. Image *B* is observed through a magnifying eyepiece *L*. Section *S-L* rotates about axis *X-X'*, thus enabling *L* to be adjusted to eye-level. Knob *K* turns prism *P*.

(*A*; *fig. 2*) fall upon an objective *O*, and are then deflected by a so-called roof prism *P* along the lateral axis of the periscope (*X-X'*). They pass through an erecting prism *Z* (see below) and are then reflected by mirror *S* towards the eye of the observer *E*. The objective *O* forms an enlarged image of the viewing screen *A* at *B*, which is seen further magnified, through lens *L*.

The system is so designed that the section *S-L* can be rotated about the axis of the periscope (*X-X'*), thus enabling the radiologist to examine the patient in different positions, using only one hand to vary the direction of view and the level of the eyepiece, the other hand remaining free for other adjustments associated with the examination (*fig. 3*). The rotation of section *S-L* about axis *X-X'* makes

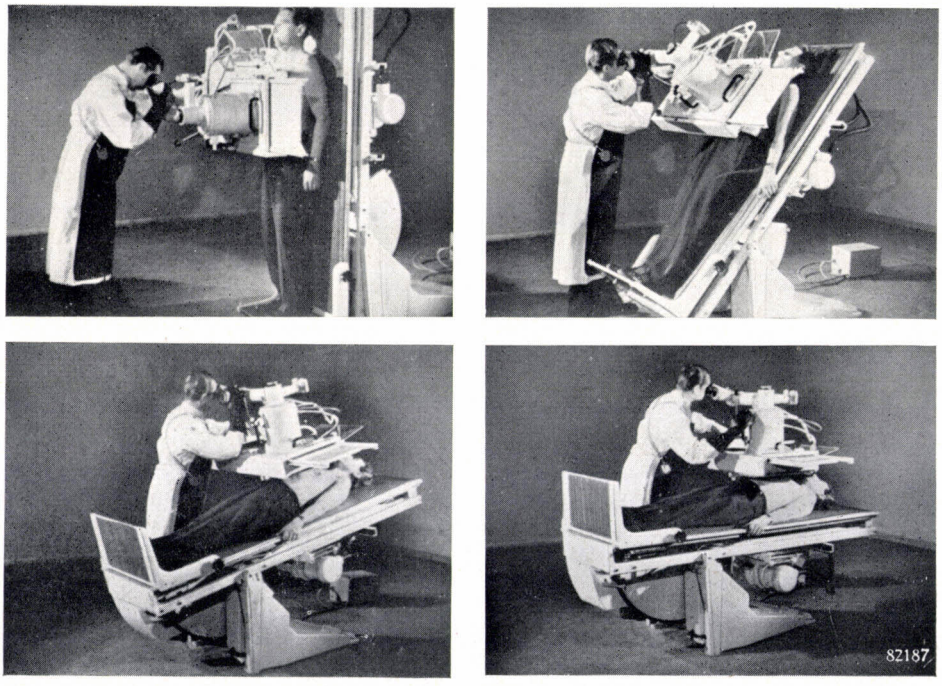


Fig. 3. Using the angled viewing system with the patient in the different positions employed in certain radiological procedures.

the erecting prism Z necessary since without this prism the image would also rotate in the field of view, and would be seen obliquely, or even upside down. Hence it would be difficult to identify the observed image at a glance.

The working of the above-mentioned erecting prism will now be described.

It is well known that the image obtained in a plane mirror is true to nature in all respects apart from the lateral inversion (left-hand and right-hand are reversed). Writing in such a reflected image is from right to left. However, the top and bottom do not change places: we do not see our own reflection upside down in a mirror.

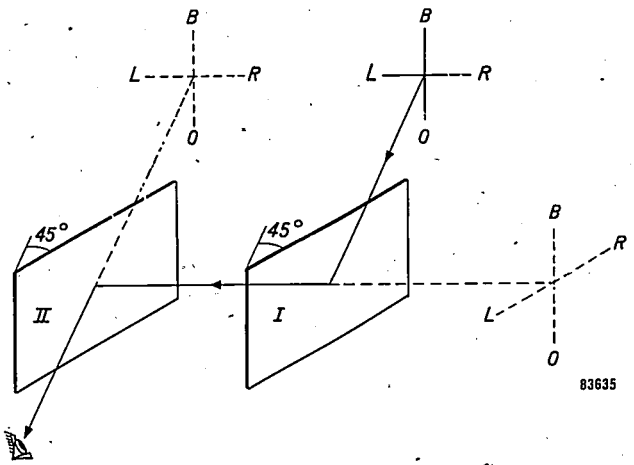


Fig. 4. The reflection at mirror I produces a virtual image in the form of "mirror writing"; it is re-converted to legible writing by reflection at mirror II.

With two parallel plane mirrors, the reversed image in the one is again reversed by reflection in the other, so that any writing in the final image is legible again. The second mirror also leaves the top and bottom of the image unchanged (*fig. 4*).

However, rotating the second mirror about axis X-X', as in the periscope already referred to, produces certain undesirable effects. Suppose that the second mirror is rotated through 90° (*fig. 5*). It then fails to restore the image in the first mirror where L and R are exchanged, to its orginal state; instead, it

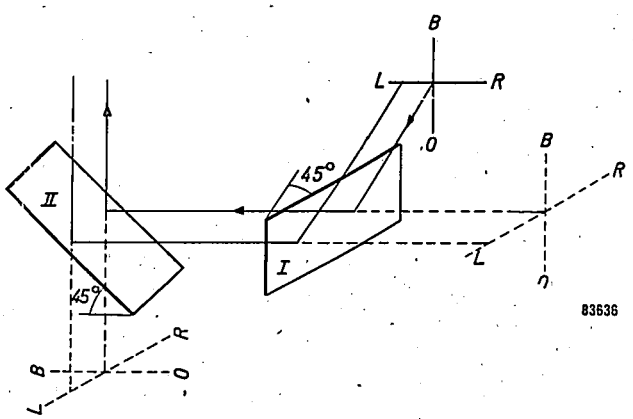


Fig. 5. To an observer standing with his back to us and looking down at mirror II, object $LBRO$ is apparently rotated anticlockwise through 90° in the field of view.

affects only the vertical direction (B-O) of the first reflection and so produces an image in which top and bottom, as well as left and right, are exchanged. The net result is that the image is once more legible, but it is rotated in the field of view through 90° about the direction of view, assuming, that is, that the observer keeps the image in sight during the

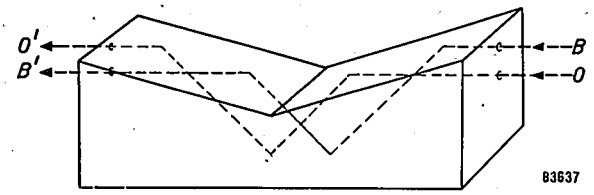


Fig. 6. Prism as described by J. Taylor (Brit. J. Phot. **14**, 348, 1867); it produces three reflections, thus exchanging the top and bottom of the image. If the prism be rotated with an angular velocity ω about direction OB', the image in the field of view will rotate with double this angular velocity, i.e. 2ω.

rotation of the second mirror merely by bending his head forward. (The direction of view as shown in fig. 5 has therefore turned through 90°.)

To compensate for this rotation of the image in the field of view, a prism system Z with three reflecting surfaces is employed (*fig. 6*) [2]). The number of reflections being odd, then, this system (like a single mirror) produces a reversed image. Now, if the prism is rotated through 90° about the axis of the light rays, the whole of the image in the field of view rotates through 180°. Hence the angular velocity of rotation of the observed image is twice that of the prism [3]). To compensate for the above-mentioned rotation of the image in the periscope viewing system, the prism must rotate at half the angular velocity of rotation of the line of sight (E-B).

There is one other point to mention. The image on the viewing screen of the image intensifier tube is inverted. In fig. 2, however, B is erect, because the objective (O) also produces image inversion; two parallel mirrors (S and P) produce no change, and the image rotation is compensated by

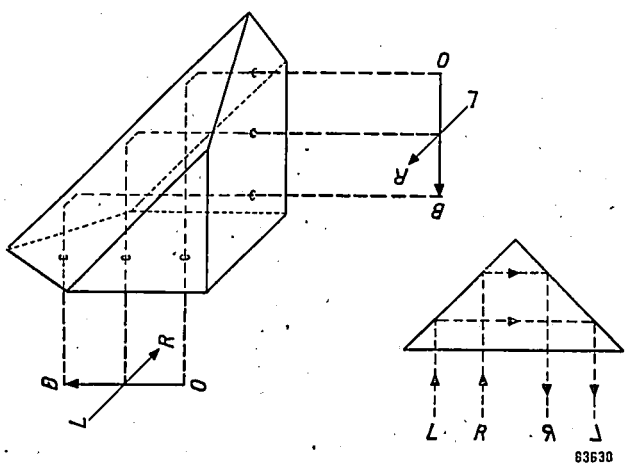


Fig. 7. Roof prism. The 45° prism inverts the image, that is, changes direction B-O; it is seen from the bottom right-hand diagram that a roof prism (two reflecting planes making an angle of 90°) also reverses the image laterally so that L and R are exchanged.

---

[2]) See Czapski-Eppenstein, Grundzüge der Lehre der Theorie der optischen Instrumente, Barth, Leipzig 1924, 3rd impression, pages 593-599.

[3]) The same effect is obtained as in the rotation of a single plane mirror viewed at a grazing angle.

the rotation of the prism. As already explained, however, the three reflections taking place in prism Z not only rotate, but also reverse the image. To compensate for this reversal, a so-called roof prism (*fig. 7*) is employed at P instead of a simple prism or mirror. Reflection then takes place not at a flat surface (which only exchanges L and R), but at two surfaces at right angles to each other, forming a kind of trough in which B and O are also exchanged. Such a prism produces an erect, legible image, as will be seen from fig. 7. In principle, it would also be possible to place the roof prism at S (fig. 2);

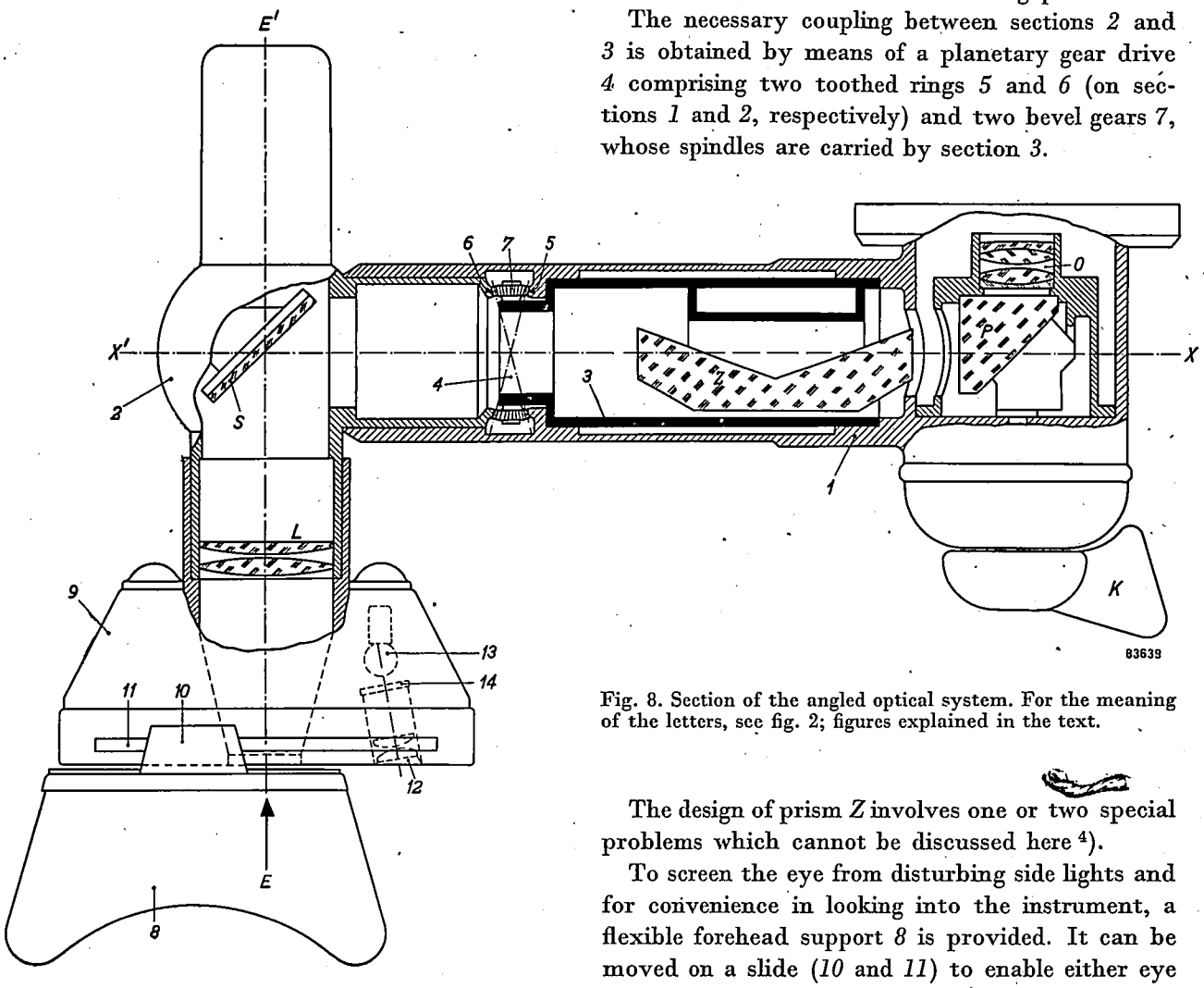


Fig. 8. Section of the angled optical system. For the meaning of the letters, see fig. 2; figures explained in the text.

however, point P is preferred as a location for this prism because in this case the image to be photographed is also the right way round (see below).

It will be evident from the above that prism Z is shown in fig. 2 rotated 90° from its true position; this is to give a clearer view of the situation.

Constructional details

It will be seen from the preceding explanation of the effect of the erecting prism Z, that when the eyepiece E is rotated about the axis of the periscope X-X', prism Z must rotate with it but at half its angular velocity. Accordingly, the viewing system comprises three individual sections (*fig. 8*) viz:

1) a section (*1*) fixed to the image intensifier tube; it contains the objective O and the roof prism P
2) a section (*2*) adjoining section *1* and rotating about axis X-X'; it contains magnifying lens L and mirror S;
3) a section (*3*) rotating about the same axis X-X', but at only half the angular velocity of section *2*; it contains the erecting prism Z.

The necessary coupling between sections *2* and *3* is obtained by means of a planetary gear drive *4* comprising two toothed rings *5* and *6* (on sections *1* and *2*, respectively) and two bevel gears *7*, whose spindles are carried by section *3*.

The design of prism Z involves one or two special problems which cannot be discussed here [4]).

To screen the eye from disturbing side lights and for convenience in looking into the instrument, a flexible forehead support *8* is provided. It can be moved on a slide (*10* and *11*) to enable either eye to be employed. Moreover, section *9* rotates about exis E-E', enabling the observer to tilt his head to any angle.

For the convenience of observers averse to keeping one eye closed, or alternatively staring into darkness, during an examination, small peep-holes (*12*) are provided, one on each side of the eyepiece aperture (E). A lamp (*13*) and a green filter (*14*) behind these peep-holes present to the eye not employed in the examination, a low-luminance

[4]) H. Jensen, Die Abmessungen von Abbe-Umkehr- und Aufrichteprismen, Optik **12**, 150-152, 1955 (No. 3).

field of the same colour as the image in the intensifier [5]).

Photographing the image is a very simple matter, as will be seen from fig. 1. Another tube at right angles to the axis of the periscope, contains the camera (shown on the right in fig. 2). When a photograph is required, prism P is simply turned towards the camera C by means of a knob K, thus deflecting the light from the intensifier through objective O and another lens D on to the film in the camera. Since P is a roof prism, this gives an unreversed image on the film.

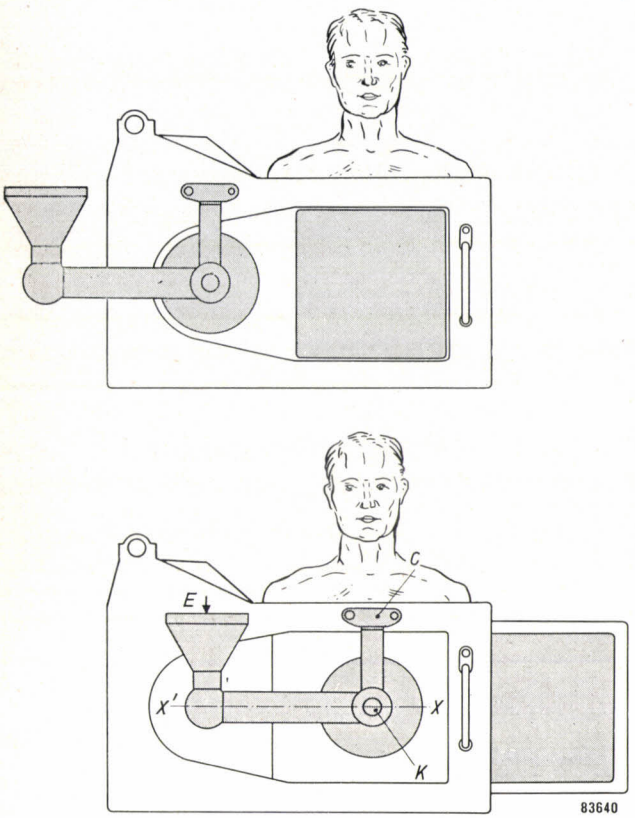


Fig. 9. Top: The apparatus as used with an ordinary fluorescent screen. Bottom: With the image intensifier positioned in front of the patient. For the meaning of the letters see fig. 2 and fig. 8.

[5]) This device was suggested by Prof. H. Schober.

The movement of the film and the opening of the shutter are likewise controlled by knob K; it also closes an electric contact to switch the X-ray apparatus from the screening voltage and intensity to the voltage and intensity required for the photograph. Correct exposure is ensured by an electronic time switch incorporated in the apparatus; 35-mm film with a picture-size of 24 × 24 mm is employed in the camera. The camera is equipped with a clockwork film-feed to move 55 frames of film automatically.

The image intensifier itself is mounted on a slide to enable it to be moved aside to introduce an ordinary fluorescent screen into the X-ray beam (*fig. 9*), or for the taking of direct radiographs with ordinary large-size film-casettes as employed in conventional X-ray diagnostic apparatus. The doctor employing the apparatus described here is thus provided at all times with all that he requires for the existing techniques, and with the means of employing image intensification; he can thus compare the two in practice.

V. MEDICAL ASPECTS OF THE IMAGE INTENSIFIER

by J. FEDDEMA *).

621.386.8:616-073.75:621.383.8

In both fluoroscopy and radiography, which may be considered to form the basis of modern X-ray diagnosis, the image intensifier will play a very important part. These two branches of X-ray diagnosis are discussed separately in the present article.

Fluoroscopy

In an X-ray examination the patient is exposed to radiation which is to some extent harmful — a fact which is, fortunately, now better appreciated by doctors employing X-radiation for diagnostic examinations. From the quantitative point of view, fluoroscopy (long exposure at low radiation intensity) produces a very much stronger effect in this respect than radiography (very short, but intense exposure). To reduce the dose required in fluoroscopy as far as possible, the radiologist usually adapts his eyes to a very low luminance level for at least 15 minutes, and to preclude any unnecessary loss of time in re-adapting of the eyes, it is nowadays customary to take all cases requiring fluoroscopic examination one after the other.

In ordinary chest fluoroscopy, involving such adaptation, an X-ray tube current of $2\frac{1}{2}$-5 mA at 60-70 kV is usually employed to ensure an

*) Medical Department of N.V. Philips, Eindhoven.

accurate appraisal of the lung structure. Using the image intensifier, the same result is obtained at the same voltage, but without preliminary adaptation and with a current which need not exceed 0.5 mA. Hence the dose is reduced by a factor of 10. A smaller, but nevertheless appreciable dose-reduction is obtained in the case of relatively thicker and more solid subjects (as in lateral fluoroscopy of the abdominal organs).

Experience has shown that such relatively thick subjects necessitate the use of a scatter grid.

Tests carried out on a "Philite" phantom, as designed by Burger [1]), for a subject thickness corresponding to that of the human thorax have shown that, at the same X-ray tube voltage, fluoroscopy with the image intensifier produces roughly the same contrast-detail perceptibility at 0.1 mA as ordinary fluoroscopy at 4 mA (see II). A tube current of 3 mA is enough to bring the perceptibility of contrast and detail in intensifier fluoroscopy quite a long way towards the standard attained in ordinary, full-size radiographs. Despite the intensifier, however, this standard will never be equalled in fluoroscopy, since the number of X-ray quanta released during the storage time of the eye (0.1-0.2 sec) is invariably very much smaller than the number effective in radiography (see II). Moreover, the high gamma of the photographic emulsion increases the contrast in the radiograph by a factor of $2\frac{1}{2}$ or 3. In fluoroscopy, the actual contrasts are the same with, as without the intensifier, but with it they are raised to a very much higher luminance level. Observers employing the intensifier for the first time are therefore often disappointed, having assumed from the fact that the fluorescent image is seen at virtually the same luminance level as a radiograph examined in front of an ordinary light box, that the two must produce very much the same impression (high contrast).

Photography

The image quality in photography with the image intensifier is limited not only by the blurring effect of the two fluorescent screens, but also by the grain of the film. Since the viewing screen to be photographed is very bright, fine-grain film of low sensitivity may be employed for single photographs. It is found that the X-ray intensity required to take an image intensifier photograph on Kodak "Micro-File" film using a tandem optical system, each component having an aperture ratio of 1 : 1.5, is only from a half to a quarter of that involved in the taking of an ordinary full-size radiograph of the same subject at the same tube-voltage. Also, the high contrast and very fine grain in such a 35-mm photograph results in almost the same information as a normal radiograph.

Comparing image intensifier photography with ordinary screen photography (fluorography) [2]), we find that it has a drawback, viz. the small field of view. On the other hand, it has one or two advantages which should not be underestimated, viz.; the apparatus is relatively small and easy to handle; adjustments prior to taking photographs are readily effected by viewing through the intensifier without any preliminary dark-adaptation; single photographs can be taken on fine-grain 35-mm film giving high contrast and therefore excellent picture-quality: all these advantages are procured with an X-ray dose at least a factor of 3 smaller than that required in ordinary radiography, instead of 3 or 4 times larger as in fluorography.

Cinematography

To all appearances, cinematography with the image intensifier will find a great deal of scope in the future. It is the only method of photographing transient processes in the human body without risk either to the subject (overdose) or to the X-ray tube (overloading). For this purpose, a more sensitive 35-mm film, e.g. Gevaert "Orthoscopix" or Agfa "Fluorapid", is employed. Experience has shown that the dose required to expose a single photograph on such film is less than 1/10 of the dose ordinarily employed in radiography: this at once implies the possibility of cinematography. Several films taken with the image intensifier are already available. Equipment for this type of cinematography is shown in *fig. 1*; for particulars see *Table I* [3]).

Table I. Examples of films made with the image intensifier.

1) Deglutition at larynx level, in lateral projection. 80 kV, 10 mA, 20 frames/sec, duration 15 seconds. 6 m of film, with a total dose of roughly 2 r.
2) Film in frontal projection of the bulbus (entrance of the duodenum). 120 kV, 11 mA, 8 frames/sec., duration 2 minutes. 20 m of film, with a total dose of 50 r.
3) Micturition. 125 kV, 20 mA, 8 frames/sec., duration 56 sec. 10 m of film, with a total dose of 45 r.
4) Cerebral angiography (examination of blood vessels in the brain) in lateral projection. 90 kV, 10 mA, 16 frames/sec, duration 20 sec. 7 m of film, with a total dose smaller than 5 r.

[1]) Philips tech. Rev. **11**, 291-298, 1949/50.

[2]) See for example, Philips tech. Rev. **13**, 269-281, 1951/52.

[3]) These experiments were carried out in collaboration with J. van der Wal and J. Proper of the Research Laboratories at Eindhoven.

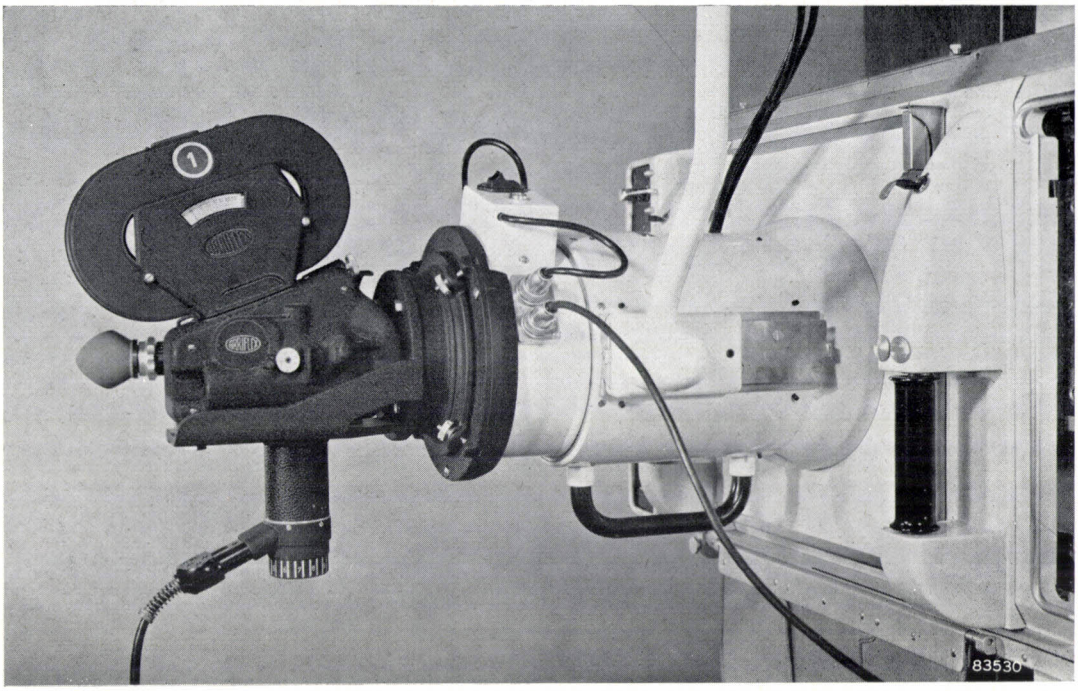


Fig. 1. Apparatus for cinematography with the image intensifier. Right: Viewing system on the stand of an ordinary diagnostic X-ray installation. The usual fluorescent screen is replaced by an image intensifier. Left: The film camera, with the film magazine on top; an eyepiece to view the image during the taking of the film is at extreme left, on the axis of the camera and image intensifier (see III).

Most of these films were made with an X-ray tube having a 0.3 mm focus, without overloading it. For each of them the distance from focus to screen was 90 cm, since a scatter grid designed for this distance was employed.

Table II. Dose to patient in different methods of gastric radiography (object thickness 21 cm). All exposures with 3-phase full-wave rectification, 120 kV, 1 mm Al-filter, conventional scatter grid, focus-screen distance 100 cm.

Method	Dose in r
Ordinary radiography	0.5
Miniature radiography with mirror camera	
a) Single exposures	
Gevaert "Ortho Scopix"	2.5
Same film 2 × magnification	10
b) Serial exposures on 70-mm film	
5 frames per sec for 1 sec	12.5
5 frames per sec for 10 sec	125
c) Cinematography on 35-mm film	
20 frames per sec for 1 sec	35
20 frames per sec for 1 minute	2000
Image intensifier photography with tandem lens system	
a) Single exposures	
Gevaert "Ortho Scopix"	0.05
Kodak "Micro-File"	0.75
b) Cinematography on 35-mm film Gevaert "Ortho Scopix" or Agfa "Fluorapid", 8 frames per sec for 2 min; overall length of film 20 m, total number of frames on film 1000.	
Total dose	50

Although it is already evident from this table that the dose itself is a more or less minor problem in X-ray cinematography with the image intensifier, this all the more evident from *Table II*, indicating the dose to the subject in different methods of gastric radiography.

Methods of diagnostic exmination

To show in how far it is practicable to employ the image intensifier in the various methods of diagnostic X-ray examination, a general survey of these methods will now be given.

Examination of the skeleton and the joints of the limbs

Because of the importance of minor changes in the bone structure in skeletal examinations, ordinary radiography is still the only effective method. Here, then, the image intensifier cannot supersede the radiograph, although it is very useful in exploratory examination before the actual exposure. Such visual examination facilitates the correct positioning of the object (say, for the projection of joint spaces and small calcifications). Image intensifier fluoroscopy is also useful for locating foreign bodies and for the examination of joints into which contrast medium, or air, has been injected.

Again, the image intensifier is ideal for follow-through examinations to ensure, for example, that the re-setting of bone fractures, dislocations, etc has been satisfactory. It is in examinations of this kind that the danger from radiation has so often been underestimated, and many a doctor has injured his fingers or hands in the course of them. The same applies to the pinning of fractures, that is, driving a stainless metal pin through the shaft of a broken bone. It is now possible to make sure, by a check examination in the operating theatre itself, that the pin has been driven into the precise position selected for it this eliminates the often enervating delay whilst radiographs are developed, and also cuts down the overall operating time.

Examination of the spinal column

Although it is now possible to obtain a fairly accurate impression of possible defects in the spinal column, especially in the region of the cervical vertebrae, with the aid of the image intensifier, the ordinary radiograph is still the obvious choice for such examinations, at any rate for the time being, because it enables the bone structure to be assessed correctly. However, there is scope for the image intensifier in myelography, in which a certain amount of contrast medium is injected into the canal of the spinal cord and any obstructions preventing the passage of this medium are located by examining the subject in different positions. Not only the possibility of working in daylight, but also the small X-ray dose to the patient is important in such examinations, since abnormalities of this kind usually occur in the lumbar region, where the proximity of the genitals, extremely sensitive to radiation, necessitates more than ordinarily careful dose control.

Examination of the skull

A simple X-ray apparatus may be very valuable in the consulting room of an ear, nose and throat specialist, say, as a means of examining an inflammation of the nasal sinuses, visible either by a swelling of the mucous membrane, or by an accumulation of fluid in the sinuses. The present field of view of the image intensifier, viz. 13½ cm, is ample to enable such a condition to be diagnosed at a glance (*fig. 2*). Also, X-ray cinematography of the jaw joint may well be useful to the specialist.

Again, cinematography may be employed in cerebral angiography, that is, studying the circulation in the blood vessels of the brain by injecting contrast medium into the carotid artery.

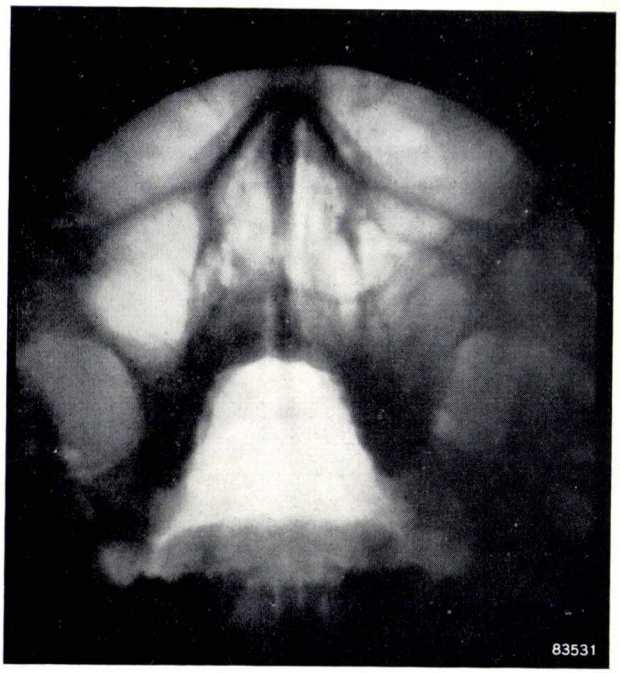


Fig. 2. Reproduction of a still photograph taken with the image intensifier. It shows the nasal sinuses in the upper jaw.

Examination of the digestive organs

Because food passes very quickly through the upper portion of the oesophagus, there has been for many years now a desire to record the movements of the larynx, etc. cinematographically. Holmgren accomplished this in Sweden as early as 1946 by means of screen cinematography. However, image intensifier cinematography offers very much better opportunities in this respect by enabling a higher frame frequency to be employed (*fig. 3*). Since the passage of food through the lower portion of the oesophagus, and through the stomach and the small intestine, is much slower, cinematography of these regions is not essential; however, image intensifier screening of these organs is very useful. Films showing the movements of the stomach and intestines are undoubtedly spectacular, and eminently suitable for purposes of instruction.

Examination of the kidneys and genitals

A clear view of the kidney and pelves (filled with contrast medium) and their contractions is readily obtained by fluoroscopy with the image intensifier. In retrograde pyelography, in which the contrast medium is injected through the urinary ducts, the above-mentioned movements can be easily followed.

Cystography usually involves taking single still photographs of the bladder during micturition. Image intensifier cinematography may be employed for this purpose in the future. Owing to the proxi-

mity of the genitals it is necessary to employ only a very small dose of radiation in this examination. This is also the case in hysterosalpingography in which contrast medium is injected into the uterus, usually to ascertain whether the oviducts are clear of obstruction. The insertion and manipulation of instruments can invariably be observed in daylight.

cluded, at any rate for the time being, by its small field of view. However, the intensifier may well be employed for exploratory examinations of individual subjects; it also deserves consideration as a means of examining local disorders, in view of the high contrast-detail perception in the image. Shadow-producing foreign bodies in the respiratory system

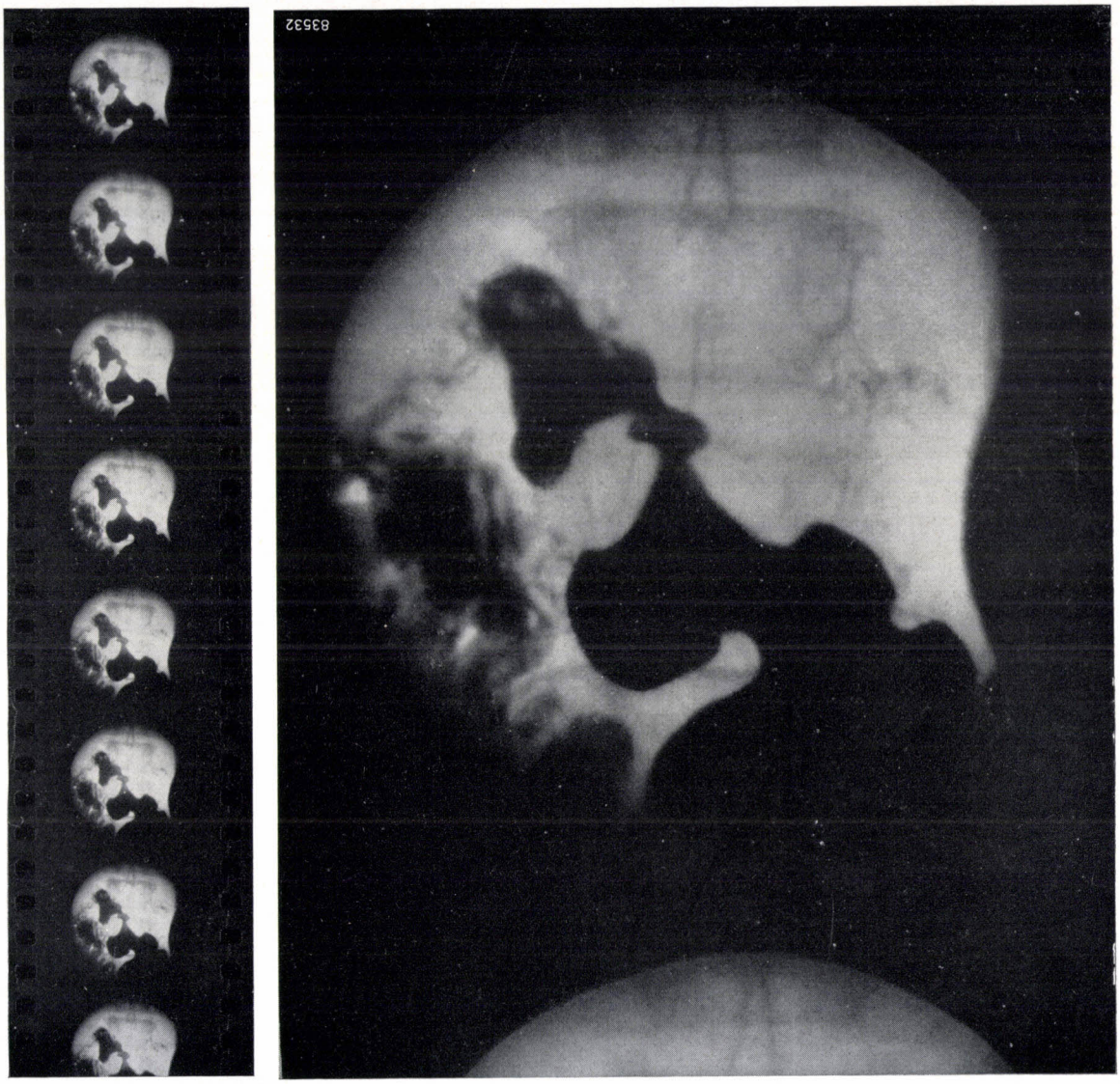


Fig. 3. Gastric cinematography. Left: Part of the film strip. Right: Enlargement of one frame in this strip.

X-ray examination during pregnancy should be kept to a minimum: it is necessary only in some cases during the final weeks, to show the position of the foetus or pelvic anomalies, etc. Here, ordinary radiography is still the best method.

Examination of the lungs and respiratory passages

In serial chest-fluoroscopy the time required for adaptation is virtually immaterial; the use of the image intensifier for this purpose is therefore pre-

are readily located; moreover, image intensifier fluoroscopy is ideal as a means of observing the position of such objects in relation to the instruments inserted to remove them.

Examination of the heart

In heart-catheterisation, the image intensifier enables the X-ray dose employed during the insertion of the catheter, which sometimes takes a very long time, to be reduced considerably.

Cinematography merits consideration as a means of examining highly localised disorders of the heart and adjacent blood vessels.

Whereas the field of view is not large enough to permit of a total examination of the heart in adults, the study of the heart function in children by means of the intensifier has already produced some remarkable results.

Although this summary is by no means complete, the examples given in it show clearly enough the potentialities of the X-ray image intensifier as an aid to medical diagnosis.

VI. INDUSTRIAL RADIOLOGY WITH THE IMAGE INTENSIFIER

by G. LANG *) and R. O. SCHUMACHER *). 620.179.1:621.386.8:621.303.8

Industrial radiology is now a widely used method for examining materials; in fact, it is one of the most important and reliable of the non-destructive methods of inspection. In the light metals industry, for example, X-rays are employed on a large scale for examining castings for defects fluoroscopically. The relatively low absorptive power of light metal alloys enables X-ray shadow pictures of adequate luminance to be obtained. With an X-ray tube of small focus (e.g. 0.4 × 0.4 mm) an enlarged shadow image can be projected [1]), giving unusually good detail-perception.

The situation with regard to the examination of steel, however, is less favourable. The welding of parts normally subjected to heavy loads, for example, parts of boilers, tanks, bridges, ships, etc., usually rests upon a reliable, non-destructive method of inspection enabling the quality of the welds to be properly assessed. Because of the high absorptive power of the iron, however, steel contructional elements cannot usually be examined fluoroscopically; such X-ray shadow pictures are very faint and can therefore be observed only in a darkened room after the eyes have been thoroughly dark-adapted. In many cases, where the particular constructional elements are either too large or too heavy to be conveyed to a screening room, fluoroscopic examination is out of the question for this reason alone. Moreover, fluorescent images of steel parts afford only limited perceptibility of detail; to ensure adequate screen luminance it is necessary to employ thick, coarse-grained fluorescent screens and also high-powered X-ray tubes which preclude all possibility of a small focus. In practice, then, steel is suitable for fluoroscopic examination only if not more than 6 or 8 mm thick. Since in most welded constructions the material to be examined is very much thicker, it has hitherto been impossible to make such X-ray examinations other than by the photographic method, which, however, is time-consuming and expensive.

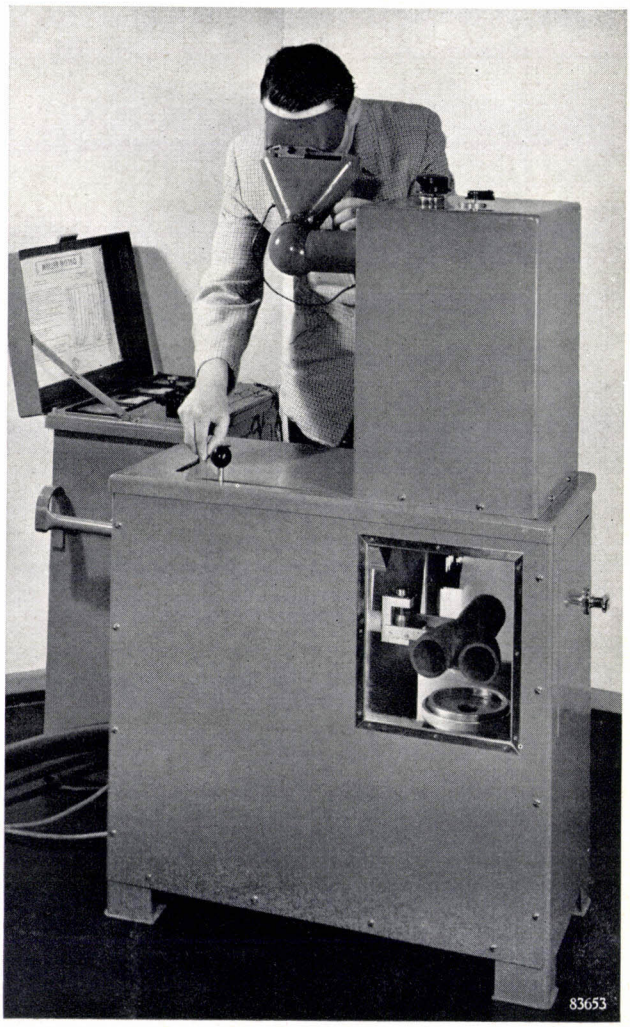


Fig. 1. Industrial radiology equipment using the image intensifier, for experiments and demonstrations. Note the beam exit aperture of the X-ray tube, radiating vertically upwards, behind the lead glass window. Above it is the work to be examined, in this case a welded Y-joint.. The image intensifier, mounted in a box on top of the apparatus, is provided with an angled optical system. The two handles seen on the left of the apparatus, are manipulated by the observer to locate the work as required.

*) C. H. F. Müller Aktiengesellschaft, Hamburg.

[1]) See G. C. E. Burger, B. Combée and J. H. van der Tuuk, X-ray fluoroscopy with enlarged image, Philips tech. Rev. **8**, 321-329, 1946.

It will be evident that the appreciable increase in luminance given by means of the image intensifier reduces or even eliminates the above-mentioned objections to the use of fluoroscopy for examining steel. X-ray examination as applied to industrial production has thus acquired new possibilities, which have been investigated in the application laboratory of C.H.F. Müller in Hamburg. A brief account of the results of this investigation will now be given [2]).

either direct from the viewing screen of the image intensifier by means of a camera attached to the viewing system, or in the usual way on film or X-ray paper in a casette exposed in front of the intensifier.

The experiments described here were carried out on steel plates and tubes of different thicknesses. A measure of the possible detail-perception was obtained by means of the DIN test objects employed in ordinary radiography; such a test object (pene-

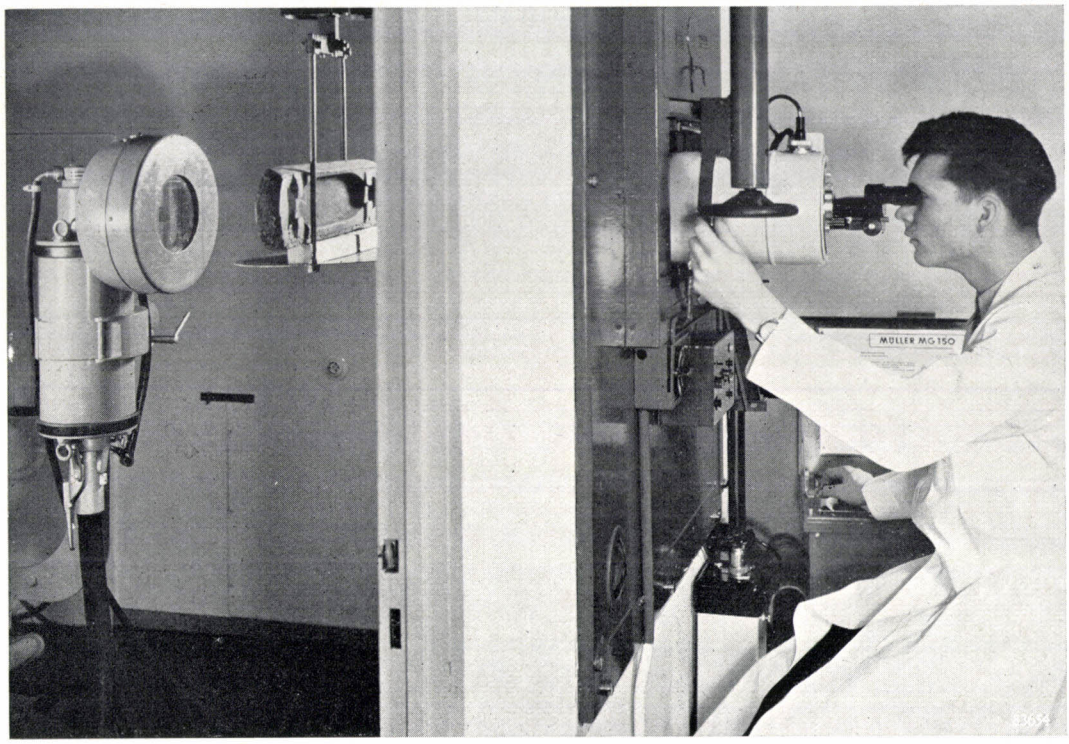


Fig. 2. Viewing equipment with built-in image intensifier. Here, the viewing optical system is a binocular microscope.

The experimental installations employed in the investigation are shown in *fig. 1* and *fig. 2;* the one incorporates an angled optical system as described in article IV of this series, and the other a binocular microscope (see article III). We consider a binocular system most suitable for the examination of materials because over long working periods it is less fatiguing to view with both eyes than with one. Defects revealed by fluoroscopy can be photographed for more accurate appraisal, or for test records,

trameter), secured to the side of the object facing the X-ray tube, contains a series of wires of progressively increasing thickness, of the same material as the work being examined. If the diameter of the thinnest wire just discernible by the observer is, say, 3% of the thickness of the material screened, then it is said that the *wire-sensitivity* is 3%, which is taken as an indication of the image quality [3]).

Curves showing the image quality so determined, plotted against the material thickness, are shown in *fig. 3*. X-ray tube voltages between 80 kV (for steel roughly 2 mm thick) and 150 kV (for steel

[2]) See R. Lang, Röntgendurchleuchtungseinrichtung mit Bildverstärker, Energie und Technik **5**, 163, 1954 (July). The arrangement described in this article, as shown in fig. 1, was demonstrated at various exhibitions last year.
Investigations into the examination of materials with the X-ray image intensifier have also been carried out in the laboratories of the Philips factories at Balham (England); see A. Nemet and W. F. Cox, Intensification of the X-ray image in industrial radiology, to be published in Proc. Instn. Electr. Engrs. **A 103**, 1956.

[3]) The wire-sensitivity is not an exact measure of the lateral dimensions of the smallest perceptible defect, since they are also governed by the shape and nature of the defect (gas occlusions, slag, cracks, and so on). Similarly, in comparing the fluoroscopic and photographic methods for image quality, the wire-resolution can be taken only as a rough indication, since conditions vary considerably between different observations.

roughly 20 mm thick) were employed in the observations; the viewing was carried out in daylight. A 0.4 × 0.4 mm focus was employed.

The increase in the smallest perceptible wire diameter with the thickness of the material screened is not proportional to the latter, but more gradual

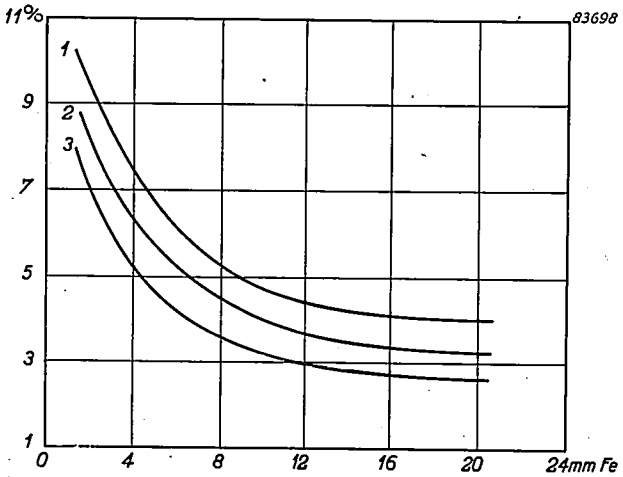


Fig. 3. Wire-sensitivity (in % of thickness of object) in the fluoroscopy of steel with an X-ray image intensifier, plotted against the thickness of the material (in mm). Tube voltage 80 to 150 kV, distance between focus and intensifier 50 cm, size of focus 0.4 × 0.4 mm. Enlarged X-ray shadow picture projection: the magnifications associated with curves *1*, *2* and *3* are 1.2, 1.6 and 2.2, respectively.

(in curve *1*, for example, this diameter is 0.3 mm for 4 mm Fe, and increases only to 0.8 mm for 20 mm Fe). The consequent improvement in image quality, expressed as a percentage, with increasing material thickness, as demonstrated in fig. 3, is analogous to what is found in ordinary fluoroscopy[4]). Experience has shown that for materials thicker than 10 mm, the most important in practice, the wire-sensitivity attainable with the image intensifier is 3%. This may be compared with direct examination with an ordinary fluorescent screen (which, as we have already seen, is possible only in the dark and with steel not thicker than 8 mm, and imposes a very much heavier voltage and power load on the X-ray tube), in which the wire-sensitivity is at best 6%.

The wire-sensitivity of 3% obtained in fluoroscopy with the image intensifier is ample for most purposes. Even better image quality, to roughly 1%, is obtained in photographic X-ray examination. In many cases, however, such a high degree of detail-perceptibility is unnecessary, and in such cases X-ray examination has hitherto been dispensed with in view of the above-mentioned disadvantages of the photographic method. It is probable that the simplicity and efficiency of fluoroscopy with the image intensifier will now cause this method to be adopted also in the steel industry. An additional feature of such fluoroscopy is that by virtue of the relatively low X-ray intensities employed, one or two simple precautions are sufficient to ensure that the operators are fully protected against radiation hazard.

The above-mentioned results are valid for a normal distance between focus and intensifier, e.g. 50 cm. This relatively long distance enables an enlarged X-ray shadow picture to be projected; the advantage of such projection, long employed in the radiology of light metals, is demonstrated by curves *2* and *3* of fig. 3. In the screening of relatively thicker materials, the quality of the image may be likewise improved by reducing the focus-to-intensifier distance, and so increasing the luminance of the screen (placing the intensifier near the object instead of projecting an enlarged image). Reducing this distance to roughly 20 cm (a closer approach to the focus is impossible owing to the shield round the X-ray tube) enables steel up to 30 mm thick to be screened, with an image quality of roughly 3%. Tests at the same distance with a larger focus and higher tube power have shown that it is even possible to screen steel 40 mm thick, although the image quality then deteriorates to some extent.

It is necessary to take steps to ensure that no unfiltered, primary radiation strikes the fluorescent screen of the image intensifier, since the dazzling brightness produced by such radiation at some points on the viewing screen and the associated afterglow impair the overall image quality.

When thick steel is examined a slightly unsteady image is obtained ("noise"), owing to the fact that by selective absorption in the steel, the X-rays are attenuated and only the shorter wavelengths remain; hence the image is built up from only a small number of large quanta (see article II). This effect does not seriously affect the image quality, however.

To conclude this brief article one or two cases from practical experience in the radiology of steel with the image intensifier will now be discussed.

To our knowledge, the image intensifier was first employed industrially to determine the level of liquid in steel bottles; with a tube voltage of 130 kV and a tube current of 6 mA, it is possible to observe the liquid level direct on the viewing

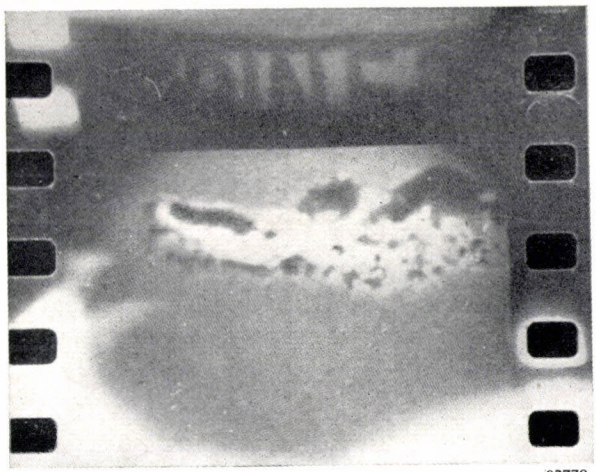
[4]) This effect is attributable to the fact that the geometrical blurring and the blurring of the image caused by the fluorescent screen, which impose a limit on the wire-resolution in the case of thin wires (thin walls), do so less noticeably with thick wires (thick walls). Although the amount of X-radiation scattered in the material increases at the same time, thus reducing the contrast, the first-mentioned factor evidently predominates.

screen of the image intensifier in daylight. The overall thickness of material thus viewed was 6 mm, but laboratory tests have shown that the same process can be applied to steel bottles with a wall-thickness of 6 mm, i.e. an overall steel thickness of 12 mm.

Another promising application of the image intensifier is in the inspection of bearings. *Fig. 4* shows a fluorograph (on miniature film) of a defect, detected visually with the aid of the image intensifier, in a bearing 13 mm thick. *Fig. 5* (likewise on miniature film) shows defects similarly detected in a welded seam in sheet steel 15 mm thick. In accordance with expectation (see article II of this series), the quality of these miniature photographs is better than that of the visual image referred to in fig. 3.

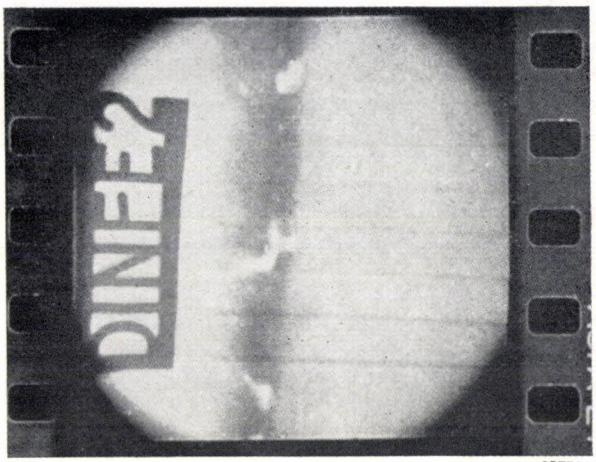
It is also possible to foresee important uses for fluoroscopy with the X-ray image intensifier in the inspection of pipes with welded longitudinal seams, as employed in long distance gas and water mains and oil pipe lines. In these mass-produced pipes we have a typical example of inspection being dispensed with altogether hitherto, because photographic X-ray examination was too expensive. A possible arrangement for screening such pipes, generally between 8 and 10 metres long, is as follows. The X-ray tube is secured to a long arm inserted into the pipe and radiates outwards through the welded seam towards the image intensifier. The pipe is moved parallel to its axis in such a way that the seam remains between the stationary intensifier and the X-ray tube. Employing such an arrangement it is possible to work comfortably with a distance of no more than 20 cm between focus and image intensifier. Provisional tests have shown that seams can be examined continuously at the rate of roughly 5 cm per second.

A similar method could be employed to examine longitudinally welded steel girders.



83770

Fig. 4. Photograph made with the image intensifier of one half of a bearing block: the radiation penetrated 10 mm of steel plus 3 mm of bearing bronze. The segregation of lead can be cleasly seen. Tube data: 150 kV, 3 mA; focus 0.4 × 0.4 mm. Exposure 20 seconds.



83771

Fig. 5. Welded seam in steel sheet 15 mm thick, photographed with the image intensifier. Note the faultsin the seam, and the shadow of the DIN test-object placed in front of the plate. X-ray tube data: 140 kV, 3 mA; focus 0.4 × 0.4 mm. Exposure 15 seconds.

Summary I-VI. The first of the articles (I) in this series on the application of the X-ray image intensifier deals with the actual operation of the intensifier tube. In this tube, the luminance of the X-ray shadow picture on the fluorescent screen is intensified between 800 and 1200 times. This luminance intensification may be employed both to give better viewing conditions for the eye and to reduce the required X-ray intensity. As explained in article II, however, it is not advisable to reduce the intensity too far, since undue reduction causes the fluctuations in the number of X-ray quanta to become visible, which affects the observation of detail. By a quantitative comparison of fluoroscopy, with and without the image intensifier, direct radiography, fluorography, and photography with the image intensifier, it is shown that under practical conditions, say, in fluoroscopy, the perception of detail is governed almost entirely by the fluctuations, whereas in direct radiography and image intensifier photography on fine-grain film this is not the case.

Article III discusses the various optical devices which can be employed to enlarge the reduced image on the screen of the image intensifier tube without any appreciable loss of luminance to the observer; optical systems for photography and cinematography are also discussed. Detailed descriptions are given of the following: a binocular microscope for fluoroscopy; an angled microscope with a large exit pupil readily located by means of a movable, frosted glass viewing plate; a tandem lens system comprising two conventional fast photographic objectives, for photographing the small viewing screen (15 mm in diameter) of the image intensifier on miniature film. Article IV describes an angled viewing optical system enabling the intensifier to be used for X-ray diagnosis with an ordinary, universal examination table. Here, a rotatable eyepiece enables the radiologist to observe the X-ray image in a normal position, and without any image-rotation when tilting the patient.

Finally, the practical application of the intensifier is discussed in the last two articles (V, VI). It is shown that the limited

field of view of the tube is not necessarily a handicap either in medical examinations or in industrial radiology. As regards medical applications, article V contains a survey of the parts of the body and organic functions the examination of which is facilitated by the use of the image intensifier. Points emphasized are the value of the intensifier in examinations involving the use of contrast media, and the development of X-ray cinematography, whose practical possibilities can be fully exploited only with the aid of the intensifier. The article contains one or two examples of X-ray films already made, including one of the duodenum; the subject dose employed in the taking of this 2-minute film (roughly 1000 frames) is only 50 r. An investigation into the industrial possibilities of the image intensifier (article VI) shows, amongst other things, that steel constructional elements 20 mm thick can be examined readily in daylight, the smallest perceptible detail then being 3% of the material thickness. Where the detail is larger, it is even possible to examine steelwork 40 mm thick.

A TECHNIQUE FOR MACHINING TUNGSTEN

by R. LEVI *).

669.276

Developed primarily for the manufacture of "dispenser" type cathodes, the technique for machining tungsten described in this article may well prove valuable for other applications of tungsten metal.

Tungsten metal plays an all-important part as filament material in incandescent lamps because of its very high melting point (about 3400 °C) and its low vapour pressure and high strength at elevated temperatures. For similar and other reasons, the metal is extensively used in X-ray tubes, both for anodes and cathodes, and for relay contacts, etc. Undoubtedly its physical properties would make tungsten ideally suited for many more applications in the laboratory and in industry, but its potential usefulness has been limited by its almost complete lack of machinability; pure tungsten is very hard and brittle at normal temperatures so that it is virtually impossible by normal methods to fabricate tungsten parts of intricate shapes and close tolerances.

A new technique, developed in the Philips Laboratories at Irvington during recent years and to be described in this article, has opened a new approach to this problem and holds good promise for future applications of tungsten.

When it was first attempted to use tungsten for incandescent lamp filaments, the very fact of its high melting point necessitated the application of unusual manufacturing methods (powder metallurgy). The present-day technique for making tungsten filaments may be briefly summarized as follows [1]). Tungsten powder of a suitable grain-size distribution and other characteristics, obtained by chemical processes, is pressed into bars at a pressure of 6-25 tons per square inch (1000-4000 kg/cm^2) and heated in an oven to a temperature of say 1100 °C. Under such treatment (pre-sintering) the metal grains are bonded together to a certain extent and the bars acquire sufficient strength to permit subsequent handling in the sintering process proper. This consists in heating the pre-sintered bars in a hydrogen atmosphere by an electric current to a temperature of about 3000 °C. Sintering of the tungsten metal grains under these conditions occurs to such an extent that the density of the material, which in the pre-sintered bars may have been about 55% of the value for solid tungsten, may rise to more than 90%. The density of the sintered bars is further increased by passing them a number of times in a hot state through a hammering or swaging machine. This process results in rods a few millimeters in diameter which, in the hot state, are sufficiently ductile for drawing into wire. Tungsten wire of diameter 1 mm down to 0.01 mm or even less is currently produced in this way, and coiled for filaments.

Apart from this highly developed and mechanized technique of drawing tungsten wire, machining possibilities for the very hard sintered tungsten ingots (or rods and sheets obtained from them) at normal temperatures are restricted to grinding and slicing by means of silicon carbide cut-off wheels [2]). Such a procedure can obviously be useful only for simple parts of convenient dimensions. Attempts to form the parts *before* the final sintering operation, either by pressing the tungsten powder in a die or by machining pre-sintered bars (density about 55%) have not been very successful. Machining of the

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

1) C. J. Smithells, Tungsten, Chapman & Hall, London 1945, 2nd ed.; see also J. D. Fast, The preparation of metals in a compact form by pressing and sintering, Philips tech. Rev. 4, 309-316, 1939.

2) For hot machining methods see D. White and J. J. Aust, Materials and Methods **27**, 81, 1948.

very porous pre-sintered bars does not yield smooth surfaces since the particles are torn out in clusters rather than cut. Moreover, considerable shrinkage and warping generally occur during the final sintering process, making it very difficult to attain the required shape and dimensions.

The technique to be dealt with in this article [3]) was specially developed for applications of tungsten in which a certain accurately controlled *porosity* of the metal is an essential condition, viz., the Philips "dispenser" cathodes (the L-cathode, described in this Review some years ago[4]), and its more recent "impregnated" version which will be described in these pages shortly [5])). Basically the technique consists in the following steps. A porous ingot of tungsten already sintered to the required degree is first *infiltrated* with a suitable molten metal which does not react with the tungsten in any way. The impregnated tungsten body can then be machined at normal temperatures with conventional carbide or high speed steel tools. Finally the infiltrant is removed by volatilization and a precisely machined pure tungsten part is thus obtained, with porosity restored to the exact value established during the sintering operation.

Fig. 1 illustrates the appearance of a tungsten part in various stages of the process. Copper was used as the infiltrant in this case. Owing to subsequent oxidation of the copper at the surface, the machined part will exhibit only a slight difference in colour before and after volatilization; these two stages therefore are not shown separately in the photograph.

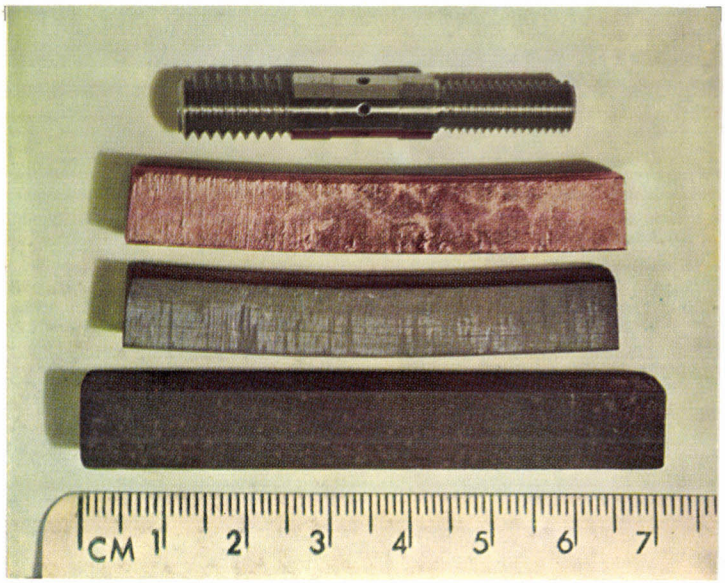


Fig. 1. Steps in machining a tungsten part according to the new method. From bottom to top: Pre-sintered tungsten bar (density 55% of the value for solid tungsten); tungsten bar sintered to the required density (83%); sintered tungsten bar after impregnation with copper; part machined by normal operations from the impregnated bar. The copper is volatilized after machining; this process does not affect the dimensions and restores the density precisely to its original value of 83%.

It should be pointed out that the infiltration of porous tungsten bodies with metals such as copper or silver has been known for many years and has been applied for making spot-welding electrodes and certain types of electrical relay contacts. Such contacts must primarily possess a good electrical conductivity, the imbued copper or silver contributing to this end, while the role of the tungsten is to prevent the contact from sticking or being welded together by the effect of arcing. The purpose of the infiltration in our case being quite different, the requirements to be met will obviously be quite different too. This is seen most clearly in the selection of the infiltrating metal. In both cases the molten infiltrant must satisfy the condition of

[3]) R. Levi, U.S. Patent No. 2 669 008, Feb. 16, 1954. See also R. Levi, The machining of tungsten and its application in the fabrication of Philips dispenser cathodes, Convention record of the I.R.E. 1954 National Convention, part 3, 70-73.

[4]) H. J. Lemmens, M. J. Jansen and R. Loosjes, A new thermionic cathode for heavy loads, Philips tech. Rev. **11**, 341-350, 1949/50.

[5]) R. Levi, J. appl. Phys. **24**, 233, 1953; Le Vide **9**, 284-289, Nov. 1954; J. appl. Phys. **26**, 639, 1955 (May).

wetting the tungsten and of penetrating the porous body by capillary action. For the machining technique, however, the additional requirement must be met that the infiltrant and tungsten should be mutually insoluble either below or above the melting point of the infiltrant. Moreover, the infiltrant should act not only as a "filler" but also as a *lubricant* during the machining operation, this combined action preventing the tearing out of particles as well as burnishing and high tool wear which would otherwise occur.

Gold, copper and alloys of the two in all proportions appear to conform best to the above requirements. The cost factor should not prevent the use of gold, since the latter when removed by volatilization can be recovered. Silver, on the contrary, is not a very satisfactory infiltrating material in our case since tungsten shows a slight solubility in molten silver. The reprecipitation of tungsten onto the larger grains which takes place upon cooling changes somewhat the value and the character of the porosity attained during the initial sintering operation.

When using copper, the impregnation is carried out at about 1350 °C [6]), for a period of not less than 10 minutes in the case of ingots $^3/_8'' \times {}^3/_8''$; larger ingots require a longer impregnating time. It is important to "fill" the ingot completely: if a small portion of it is not properly infiltrated, breakage of the tungsten or of the tool may result. In order to ensure proper filling, the ingot is placed on top of a weighed amount of copper (OFHC), slightly in excess of the amount which will be necessary; the weighed amount will be 8-10 % of the weight of the tungsten when the porosity is 83-84 %. The infiltration is carried out in a hydrogen atmosphere and the temperature is first slowly raised to a point below the melting point of the impregnant and held there a few minutes to permit the interior of the ingot to attain the same temperature as the surface. When the temperature is finally raised to 1350 °C the molten copper will penetrate the tungsten body from the bottom by capillary action, this process being facilitated by the fluxing action of the hydrogen.

No tungsten grains can be detected under microscopic examination of a freshly machined surface, since a thin copper film has been smeared over the entire area. If the copper film is chemically removed from the machined surface, the smoothness and flatness of the grains indicate that they have actually been cut by the tool and not merely torn out.

The volatilization of the copper is effected by heating the machined parts in a vacuum furnace at 1800-1900 °C for a sufficient time. The resulting parts under spectroscopic examination show only an extremely faint trace of copper. In order that the machined parts will retain dimensional stability and proper porosity during the volatilization (*fig. 2*), it is essential that the sintering of the tungsten frame prior to the infiltration was conducted for a sufficient length of time at a temperature considerably higher than that necessary for the subsequent evaporation of the copper (and, of course, higher than the temperatures at which the parts will further be treated or used). Since the proper sintering temperature depends to a large extent on the characteristics of the tungsten powder, on the pressure used in forming the bars and on the sintering atmosphere (e.g. its water vapour content), all these factors have to be carefully selected. An example of the technique as developed for the dispenser cathodes [5]), is the following. Tungsten powder

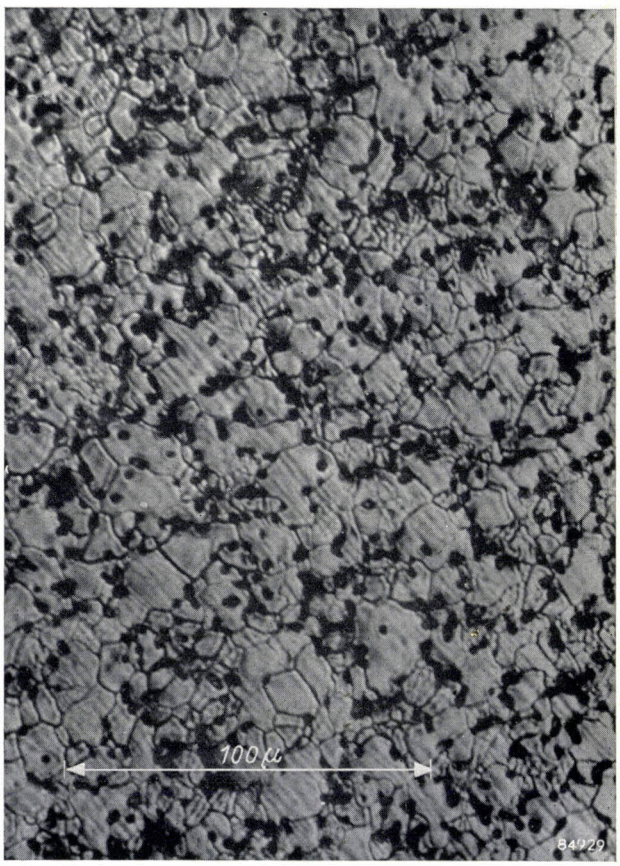


Fig. 2. Photomicrograph showing a polished section of the tungsten surface after the volatilization of the copper. The average pore size is of the order of a few microns, the pore separation varying between a few microns and a few tens of microns. (Such small pore distances are desirable for the surface of dispenser cathodes [7]).)

[6]) All temperatures indicated for our process are brightness temperatures measured by sighting on the tungsten.

[7]) This will be shown in a forthcoming publication by E. S. Rittner and R. H. Ahlert of the Irvington Laboratories.

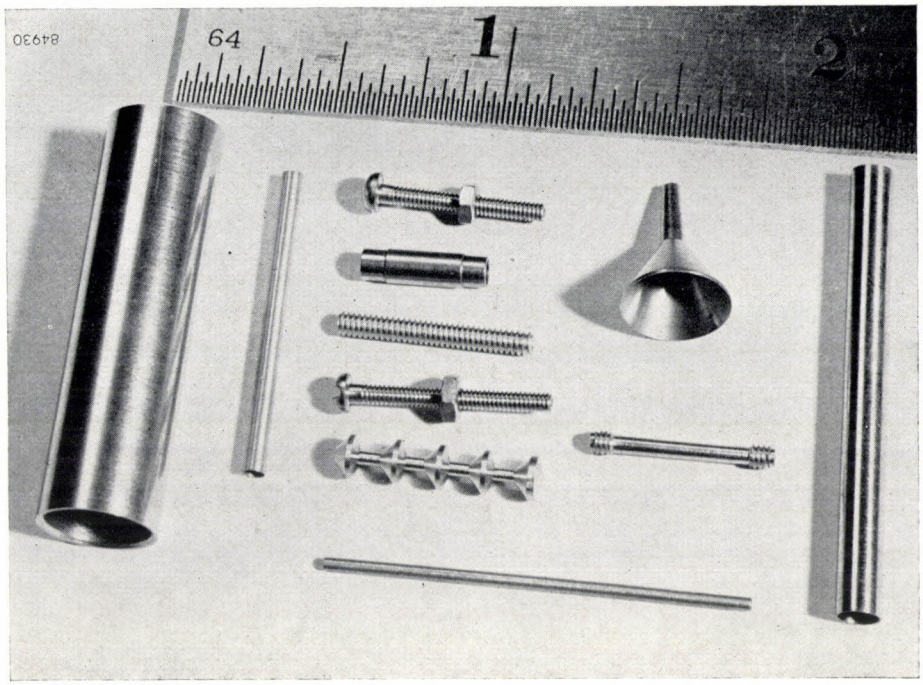


Fig. 3. A number of tungsten parts made by the new technique.

with the characteristics shown in *Table I* is pressed into bars at 2000 kg/cm². After pre-sintering for 20 minutes at 1150 °C, the ingots are sintered for 20 minutes more at 2400 °C in a water-free reducing atmosphere (cracked anhydrous ammonia). The density of the ingot then reaches a value of 83-84% of the solid tungsten value.

Normal machining operations on the infiltrated ingots can be carried out with relative ease at densities up to about this value. Increases in density above this value will make the machining progressively more difficult because of a rapid increase in the percentage of non-connecting pores which cannot be infiltrated. This will actually limit the application of the technique described to tungsten parts of density slightly less than 90%.

It has been mentioned that the application to dispenser cathodes depends on the very porosity of the tungsten (a high percentage of the pores must also interconnect in this case). A number of other applications of tungsten may be conceived for which the porosity of the metal does not matter, while the ease of making intricate forms (see *fig. 3*), the smoothness of the surface and the close dimensional tolerances achievable by the new technique are of importance. An additional asset of this technique is that it allows fabrication of extremely fine parts which — even when disregarding tolerances etc. — could not previously be made either by pressing in a die or by machining a pre-sintered ingot, because of the inherent weakness of the material in this stage.

Table I. Grain size distribution of tungsten powder which in the case described as an example was used for machining parts of 83% density. The distribution is determined by a standardized elutriation analysis, passing an elutrient (water) over the powder sample a through and number of widening vessels, in which fractions of the powder are deposited. In the first, narrow vessels, where flow is rapid, chiefly large grains are deposited; smaller particles settle in the subsequent, wider vessels.

Fraction No.	% settled	Equivalent particle radius
1	29	$> 6\ \mu$
2	25	
3	11	
4	13	
5	22	$< 2\ \mu$

Another check on the powder characteristics is obtained from the Scott test, in which the density of the powder is measured after shaking for some time in a vessel. In our case this Scott density is about 68.4 gram/cubic inch (4.1 g/cm³).

Summary. Pure tungsten metal obtained by the well-known sintering process is extremely hard and brittle at normal temperatures, so that machining possibilities are very limited. It has been found that porous tungsten, of density up to 90% of the value for solid tungsten, can be accurately machined by normal methods and at normal temperatures when the tungsten body is infiltrated with suitable metals, such as copper or gold. After machining, the filling metal, which also acts as a lubricant, is removed by evaporation. No shrinkage or warping of the machined parts is to be feared provided that the tungsten ingot prior to machining was sintered to the required degree at a temperature higher than that to which the tungsten part is subjected during evaporation of the filler or in subsequent use. Intricate parts of different sizes, including very small ones, can be made in this way to close dimensional tolerances.

VOL. 17 No. 4, pp. 101-128 OCTOBER 1955

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 RECORDING AND PRODUCTION OF GRAMOPHONE RECORDS

by J. L. OOMS *).

681.854:621.395.625.3

Philips' Phonographic Industries are a comparatively young branch of the Philips' concern. They are engaged in the manufacture of gramophone records, both of the old, fast-running type (78 r.p.m.) and of the modern, long-playing records. The latter type, particularly, have reached a very high quality of reproduction. An essential factor in the manufacturing process, as explained below, is magnetic sound-recording.

Three stages can be distinguished in the manufacture of gramophone records: the actual recording, the preparation of the matrix from the recording and, finally, the actual mass-production — the pressing of records. These three stages will be discussed in turn.

Recording

The studio

The quality of reproduction of a gramophone-record is very largely dependent upon the acoustical properties of the recording studio. Acoustical imperfections during a musical performance attract far more attention when there is no visual contact, as is the case with gramophone music, than when the music is listened to directly. As is well known from broadcasting and allied activities, it is very difficult — if not impossible — to take all the necessary measures to guarantee good acoustical properties during the actual construction of a studio. Added to this, the conditions vary with the extent to which the hall is occupied, so that in recording for gramophone records (when there is not, as a rule, an audience), the sound-pictures are often totally different from those produced during a performance before an audience. For example, in the absence of an audience the reverberation time of high sound-frequencies will be longer. The liberal use of curtains to replace the acoustical properties of the audience's clothing is customary. Measures must likewise be taken to suppress undesirable resonance phenomena of the studio, especially of the floor and platform. A sound-picture distorted by this kind of phenomena will interfere but little during direct listening to a performance, since it occurs only sporadically; however, in a gramophone record, finding its way to the public in many thousands of copies, this fault will be multiplied, as it were, by the number of copies sold and by the number of times that each record is played.

Clearly, precautions must be taken to make the acoustics during the recording as favourable as possible. In view of the fact that this may involve costly alterations to convert a hall into a recording studio, it is understandable that gramophone record manufacturing companies are ever on the look-out for halls which already possess most of the desired properties. This explains why different companies sometimes make their recordings in the same hall.

The above remarks apply to records of serious music. Dance music and other types of popular entertainment require quite different conditions. Usually a hall with a very short reverberation time (a so-called "dry" room) is employed for this kind of music, the desired reverberation being added electronically. Use is made here either of a "reverberation chamber" or a "reverberation channel", using magnetic recording. This brings us to the field of trick effects, which especially in recent years has found wide application in the recording of modern popular music. Electrical and acoustical "décor" are also included under this heading. With these

*) Philips Phonographic Industries, Baarn, Netherlands.

techniques it is possible to make records which will reproduce a sound-picture that the artistes themselves could never achieve in a live performance.

"Electrical décor" consists of a signal which is directly mixed by electrical means with the signal to be recorded. In "acoustical décor" the mixing is brought about indirectly, via an acoustic link (loudspeaker in the studio). The signals for both electrical and acoustical décor are usually derived from a recording which was prepared beforehand; acoustical décor, however, can be reproduced (often in a special studio) while the actual recording is in progress.

The microphones

The quest for ever-better quality has led to the almost exclusive adoption of condenser-microphones [1]). The latter may be classified according to their directional response into multi-directional, bi-directional and uni-directional types (with roughly circular, figure-eight shaped or cardioid polar diagrams respectively). These types are variously used to counter the acoustical shortcomings of the hall, in the attempt to give the listener the same complete impression of what is offered, although he has no visual contact with the performers. In other words, the sound-picture recorded on a gramophone record, should not be merely as faithful a replica as possible of the performance, but should accentuate certain points so that the listener may find the result satisfying.

For the same reason a number of microphones are sometimes used simultaneously: by mixing, it is sought to improve the sound-picture. There is also the danger of augmenting undesirable effects however; recording with two or more microphones is therefore entrusted only to experts with considerable experience in this field. It is also often necessary to correct the sound-picture with respect to the frequency, so that the most favourable final balance is obtained over the whole frequency range. In general these corrections consist of slow variations in amplification with frequency, but in recent years it has also been the practice to carry out corrections within a relatively narrow range of frequencies.

Magnetic recording

The sound-picture balanced in the manner described above, is nowadays recorded first on magnetic tape [2]). The prosperity of the gramophone record industry in latter years is founded upon the development of the long-playing record [3]), and it would be no exaggeration to say that without magnetic recording, its rapid fruition would have been impossible.

Earlier recordings were made by a process of direct engraving — first on a wax disc and later on a lacquer disc. The subsequent manipulations in the manufacturing process allowed only a relatively small number of pressings to be made, so that the record was doomed to removal from the catalogue after only a few years. The situation is completely different when a recording is made on magnetic tape. From a tape-recording a practically unlimited number of new matrices can be prepared as soon as those in use are in danger of showing signs of wear.

Magnetic recording has a further important advantage. Even the best musicians are unable in general to play faultlessly for longer than some 10 minutes; a period of 15 to 20 minutes would be very exceptional. That there should be faults during performances of longer works is something which concert audiences accept as perfectly normal; usually minor errors are forgotten at once. When recording for gramophone records, however, faults are not acceptable, on account of the enormous multiplication factor referred to above; once the gramophone listener has noticed a fault, it will annoy him more and more intensely each time he plays the record.

To obtain two acceptable lacquer discs by a direct engraving process (at least two discs are necessary, one being a reserve), it was often necessary for long passages of the music to be re-recorded several times, since none of the playing faults could be corrected on the disc. Magnetic recording has an enormous advantage over this method, in that it allows the recording to be *edited*, i.e. spliced together in sections. Advantage is taken of this for the repetition of any passage in which a fault has been made, the repetition being subsequently fitted into the recording. In this way, flawless, magnetic recordings with a playing time of about half an hour can be obtained, this being of the greatest significance for the manufacture of long playing records.

Because of the risk that the recording might be spoiled be coughing and other noises, the public are not usually admitted into the hall during the recording; in any case, the repetitions of parts of the performance would of course greatly reduce its entertainment value [4]).

[1]) See Philips tech. Rev. **9**, 330-338, 1947/48.

[2]) See, for example: D. A. Snel, Magnetic sound recording equipment, Philips tech. Rev. **14**, 181-190, 1952/53; W. K. Westmijze, The principle of the magnetic recording and reproduction of sound, Philips tech. Rev. **15**, 84-96, 1953/54; W. K. Westmijze, Studies on magnetic recording, Philips Res. Rep. **8**, 148-157, 161-183, 245-269, 343-366, 1953.

[3]) See for example, L. Alons, New developments in the gramophone world, Philips tech. Rev. **13**, 134-144, 1951/52.

[4]) The only exceptions are historical occasions or performances of unique occurrence, the recordings of which are principally of documentary value. The circumstances under which such a record has been made, are usually reported on the label of the record.

In view of the fact that a gramophone record is actually derived not from the sound-picture direct but from the magnetic recording of the sound-picture, the very highest quality is necessary for the magnetic recording. Linear and non-linear distortion must be reduced to the absolute minimum. The same applies to brief fluctuations in the speed of the tape, which would lead to annoying variations in pitch ("wow" and "flutter").

During recording, the tape is run across a playback head and the reproduction is studied in a separate room. In judging the recording, it is taken into account that the future gramophone record will generally be played in living-rooms and must therefore give the greatest possible satisfaction under the circumstances obtaining there.

In practice however these circumstances cannot be realised in the listening room. In addition a fairly high sound level is chosen, so that any playing faults are most certain to be observed; however owing to the physiology of the ear this will mean that the observed sound-picture is not wholly accurate. The recording engineer must therefore translate the observed sound-picture, as it were, into one of lower sound level, more suitable for the living-room. Remembering also that the electrical and electro-acoustical properties of the future record owner's equipment are unknown, it will be clear that accurate adjustment of the sound-picture requires great practice.

Dynamic range

When recording it is necessary to reduce the "dynamic range", i.e. the intensity ratio of the strongest to the weakest passages. The excellent properties of magnetic tape with high-frequency bias [2]) allow it to record almost the whole dynamic range of a large symphony orchestra (about 70 dB). However, the fact that the record will eventually be played under the conditions obtaining in a living-room, means that the dynamic range of the record itself must be appreciably limited, if the reproduction is not to be marred by too strong a non-linear distortion or drowned in the surface noise or the living-room sounds which can never be completely avoided.

A simple calculation will clarify these problems. Let us assume that the output of the average reproduction apparatus (radiogram or record player and radio set) amounts to 3 W and that this output is just reached in the strongest passages. For a dynamic range of 40 dB on the record, the pianissimo passages of the music would become so soft as to be completely dominated by all kinds of interfering sounds such as are normally present in a house: street noises, sounds from neighbouring apartments or houses, and even the faint hum of the gramophone motor itself. (It should be noted in this connection that the surface noise of modern plastic gramophone records is appreciably weaker than that of the older records and can be neglected here.)

In general therefore it is assumed that a gramophone record should have a dynamic range of not more than 30 dB. Were the dynamic range any greater, the listener would tend to increase the volume of reproduction somewhat during soft passages, with the result that the sound would be much too strong and, moreover, distorted during loud passages. The listener would then reduce the volume. This undesirable situation can be avoided by doing what the listener would otherwise have to do, during the actual recording, but in an expert manner with due consideration for the music. Only when the dynamic range of the reproduction has thus been reduced can it afford the listener undisturbed enjoyment.

From the magnetic recording to the matrix

Once a musical performance has been recorded magnetically, the next step is the transfer of the recording to a matrix which can then be employed to press out the gramophone records in a plastic material. This step is divided into a number of intermediate stages, the first being the production of a lacquer disc. The matrix is prepared in a number of chemical and plating processes from the lacquer disc.

Engraving the lacquer disc

The sounds recorded on the magnetic tape are transferred to the lacquer disc via a replay head, an amplifier and a groove-cutter (*fig. 1*). The lacquer disc turns at the same speed as the gramophone record which will be manufactured from it.

During this engraving process the necessity again arises for a very constant speed, both of the tape and of the disc. Moreover, exacting precautions must be taken to avoid vibrations (these largely determine the noise level of the unmodulated groove) and to ensure that the disc is flat. The latter is of especial importance for long-playing records; the profile of the micro-grooves (*fig. 2*) — of which there may be as many as 14 per millimeter — must be constant within very narrow limits. For this reason, the turntable of the cutting machine must be made very flat and must be accurately perpen-

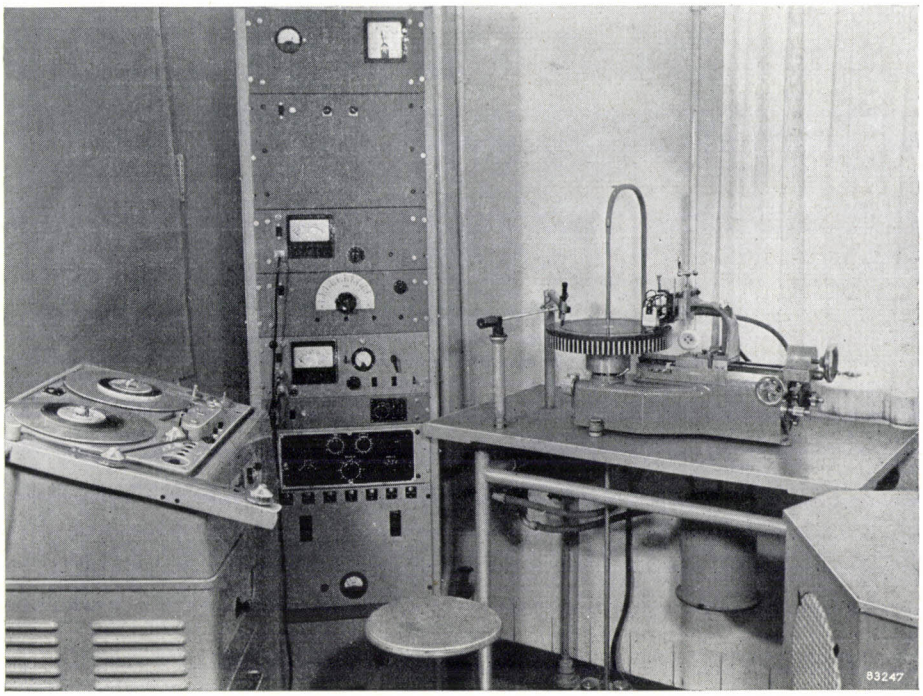


Fig. 1. Equipment for the transfer of magnetic recordings onto lacquer discs. From left to right: the machine in which the magnetic recording is played back, the amplifier rack (also containing other electronic apparatus), the machine on which the lacquer disc is engraved, and a monitoring loudspeaker.

dicular to the axis of rotation; the lacquer disc is drawn tight onto the turntable by suction.

The spectrum to be registered ranges from about 30 c/s to about 16 000 c/s. This wide range places high demands on the electronic apparatus and on the groove-cutter, especially because linear and non-linear distortion must be kept extremely low. (Linear distortion is here taken to mean deviation from a standardized frequency response curve).

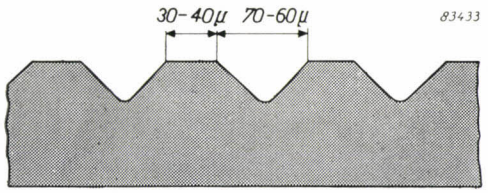


Fig. 2. Cross-section of unmodulated grooves in a long-playing record.

As regards this standardized frequency response curve, it should be noted that present-day gramophone record factories assume a *reproduction* channel with specific properties, which it is hoped will be established on an international basis. Two proposals for such a standard, one European and the other American (*fig. 3*), have been studied by the International Electrotechnical Commission, but no decision has yet been reached. When the lacquer disc is being engraved, a correction is applied which is the reciprocal of this standard, with a view to rendering the reproduction via the standard channel as far as possible independent of the frequency.

The groove-cutter is of the electrodynamic type. The problems involved in its development are of the same nature as those associated with the cutter in the Philips-Miller system [5]): the moving mass must be very small and the coil must have a high current-carrying capacity. On account of the extent of the frequency range, however, the requirements for the disc groove-cutter are still higher.

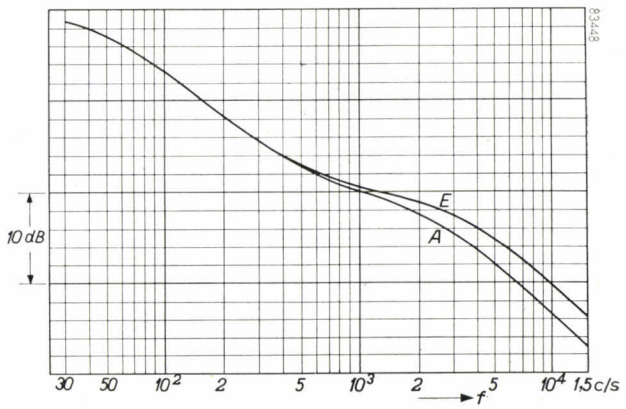


Fig. 3. Proposals for a standard reproduction channel for long-playing records, submitted to the I.E.C. for consideration. The speed v of the needle point for a constant sound intensity is plotted along the vertical axis. Along the horizontal axis the frequency f is plotted. Both are on a logarithmic scale. Curve A corresponds to the American proposal, curve E to the European proposal.

[5]) Philips tech. Rev. **1**, 135-141, 1936.

The following very approximate calculation will show that the coil is subjected to large currents and accelerations. For the sake of simplicity, only the moving mass of the copper wire of the coil is considered, so that the results will be somewhat on the low side.

If l is the length of wire on the coil, I the alternating current passing through it, and B the magnetic induction of the field in which the coil is located, then the force F which acts upon the coil (in Giorgi units) is:

$$F = IlB.$$

If the current density is denoted J, the total mass of the copper wire m and the density of copper d, then $Il = Jm/d$, and

$$F = \frac{JmB}{d}.$$

Moreover $F = ma$, when a is the acceleration; hence $a = JB/d$, or

$$J = \frac{ad}{B}. \quad \ldots\ldots\ldots \quad (1)$$

The acceleration a may be expressed as $\omega^2 A$, where ω is the angular frequency of the note being recorded and A the amplitude of the groove. To obtain a sufficiently high amplitude, the acceleration a should be as large as possible. An upper limit to the value of a is set by the tracing distortion (see p. 138 of the article cited in [3])), which occurs whenever the radius of curvature of the recorded sine curve becomes less than the radius of the spherical needle point with which the gramophone record is later traced during reproduction. The radius of curvature ϱ of the sine curve is smallest at the peaks; thus it is there that the limit is most likely to be exceeded. At the peaks we have approximately

$$\varrho = \frac{V^2\sqrt{2}}{a},$$

where V is the speed at which the groove passes beneath the needle. In the limiting case ϱ is thus equal to the radius r of the needle point, the minimum speed must be taken for V, i.e that of the innermost grove (V_i), and the maximum value for a, viz. a_{max}. Then:

$$r = V_i^2\sqrt{2}/a_{max}$$

or

$$a_{max} = \frac{V_i^2\sqrt{2}}{r}. \quad \ldots\ldots\ldots \quad (2)$$

Substituting the following values in (1) and (2):

d = 8900 kg/m³ (copper),
B = 1 Wb/m² (= 10 000 gauss),
V_i = 0.22 metre/sec (for a record with 33⅓ r.p.m.),
r = 25×10^{-6} m,

we find a maximum acceleration of:

$$a_{max} = 2740 \text{ metres/sec}^2,$$

or 280 × the acceleration due to gravity. The maximum current density is given by

$$J_{max} = 24.2 \text{ A/mm}^2,$$

neglecting, as stated earlier, moving masses other than that of the copper wire of the coil.

In the groove-cutter in its present form as produced by Fonofilm of Copenhagen the difficulties have been satisfactorily overcome, thanks to the following measures: new, very light construction materials of great strength have been used; the components have been very carefully finished, as a result of which undesirable air gaps have been avoided and consequently the magnetic induction can be raised to a high value. In addition, negative feedback is applied [6]): attached to the cutting tool there is a small coil located in a magnetic field. A voltage is set up in this coil which is proportional to the speed of the cutting stylus. This voltage serves as the negative feedback in the amplifier which feeds the groove-cutter. This reduces linear and non-linear distortion of the engraving process to a very low level; measurements have shown that the intermodulation is about 1%.

Another special feature is the electrical heating of the point of the cutting stylus (sapphire). Thanks to this heating a smoother groove is formed, resulting in less surface noise. Moreover the reproduction of high notes is improved. In its cold state the lacquer is not purely plastic but still somewhat elastic. Under these conditions the amplitude of the groove becomes smaller immediately after cutting, the reduction in size being more pronounced at smallest wavelengths. This phenomenon therefore occurs particularly at high notes and for the innermost grooves, where the wavelength is shortest. Heating the stylus renders the lacquer so soft that this undesirable effect is avoided. The lacquer regains its original hardness at a very short distance behind the stylus.

To lengthen the playing time of gramophone records, the system of *variable groove pitch* has found application in recent years. In this system the average distance between two consecutive grooves is made to depend on the amplitude and is kept as small as possible. This means that in soft passages more grooves are engraved per millimeter than in the usual system in which the groove separation is constant and based on the loudest passages [7]).

The variation of the groove separation with amplitude is, of course, done automatically. Before passing the replay head, the magnetic tape passes over an auxiliary replay head which gives a preindication of the strength of the signal about to be cut in the disc. The alternating voltage produced in the auxiliary replay head is rectified and the variable rectified voltage so obtained determines the frequency of a valve oscillator. The latter supplies a synchronous motor via an amplifier, and the

---

[6]) F. Schlegel, Einige Schallplattenaufnahmeprobleme, Acustica 4, 45-47, 1954 (No. 1).

[7]) The average extension in playing time thereby achieved is about 20 % for classical works.

number of revolutions of the motor is thereby made to depend on the strength of the signal about to be cut. The motor drives the mechanism which displaces the cutting-stylus radially in the required direction.

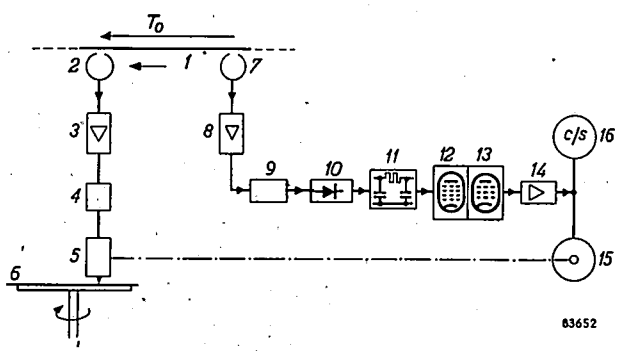


Fig. 4. Block diagram of the equipment for transferring the magnetic recording (*1*) to a lacquer disc (*6*) via replay head *2*, amplifier *3*, corrective network *4* and groove-cutter *5*. The equipment for variable groove pitch is on the right: *7* auxiliary replay head, situated about one revolution (T_0) of the disc ahead of head *2*; the signal from *7* passes through amplifier *8*, filter *9*, rectifier *10* and delay network *11* to the reactance tube *12*. The latter regulates the frequency of a valve oscillator *13*, which supplies a synchronous motor *15* via an amplifier *14*; the synchronous motor displaces the groove-cutter radially; *16* frequency meter.

Fig. 4 is a block diagram of the installation for variable groove pitch. Between the rectifier and the reactance tube (the tube which governs the frequency of the oscillator) there is a delay network, whose time constant is approximately equal to the time required for a certain point on the tape to cover the distance between the two replay heads; this time T_0 is selected roughly equal to the time for one revolution of the disc.

Fig. 5 shows the delay network in greater detail. The amplified and rectified signal V_s from the auxiliary replay head is applied to C_1. The capacitor C_2 gets a delayed charge, via a resistor R_1. As the signal V_s increases, the grid potential V_{g3} of the reactance tube gradually becomes less negative, and this increases the frequency of the oscillator; the speed of the synchronous motor and the groove separation s are proportional to this frequency.

Also indicated in the circuit are a threshold voltage V_d and a limiting voltage V_{lim}. The former is a direct voltage which opposes V_s; as a result the groove distance which had initially been adjusted to a very small value, only begins to increase when the amplitude exceeds a certain level. The limiter consists of a diode in series with a source of direct voltage V_{lim}; this ensures that the oscillator frequency and the groove distance do not increase indefinitely as a function of the signal

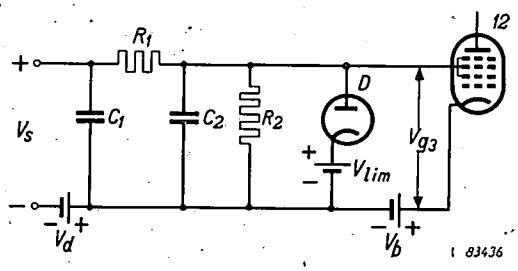


Fig. 5. Circuit of the delay network *11* of fig. 4. V_s input signal (varying direct voltage). V_d threshold voltage. C_1 capactitor from which capacitor C_2 is given a delayed charge via resistor R_1. R_2 discharge resistor. The diode D with bias V_{lim} forms the limiter. V_b grid bias. *12* reactance tube.

amplitude. *Fig. 6* shows the influence of the threshold voltage and the limiter. In *fig. 7* the voltage-time functions of various points in the circuit are shown.

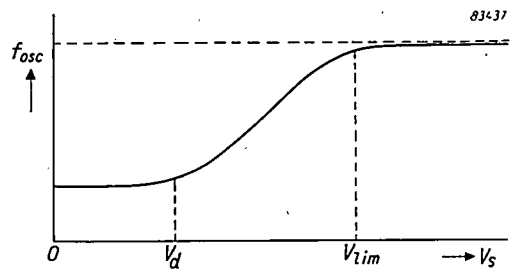


Fig. 6. The frequency f_{osc} of the oscillator (*13* in fig. 4) as a function of the signal voltage V_s. The threshold voltage is V_d, the limiting voltage V_{lim}.

The preparation of the matrix

The matrix is prepared from the lacquer disc by electroplating and press techniques. The lacquer disc must first be rendered conducting. The modern method is roughly the same as that by which silver mirrors are made, namely by precipitation of silver from an ammoniacal solution of silver nitrate. Processes which take place in vacuo, such as silver-

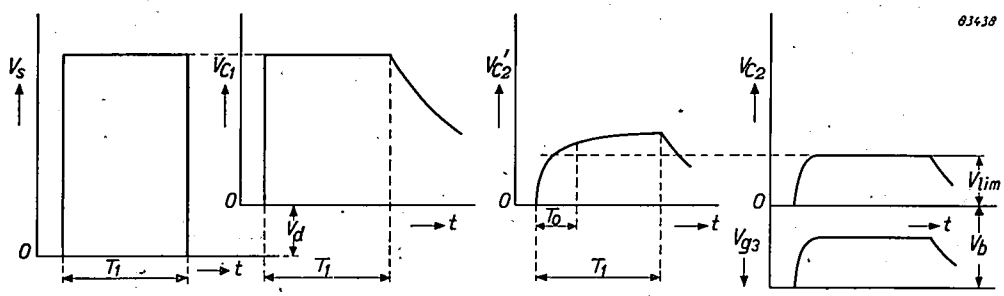


Fig. 7. Various voltages in the circuit of fig. 5, as functions of the time t, for a fortissimo passage (duration T_1; the revolution time of the record is T_0).
V_s = signal voltage. V_{C_1} = voltage across capacitor C_1. The voltage across C_2 would be as that shown for V_{C_2}' if no limiter were present; with the limiter the voltage is as shown for V_{C_2}. The voltage between grid and cathode of the reactance tube is V_{g3}; this voltage determines the frequency of the oscillator.

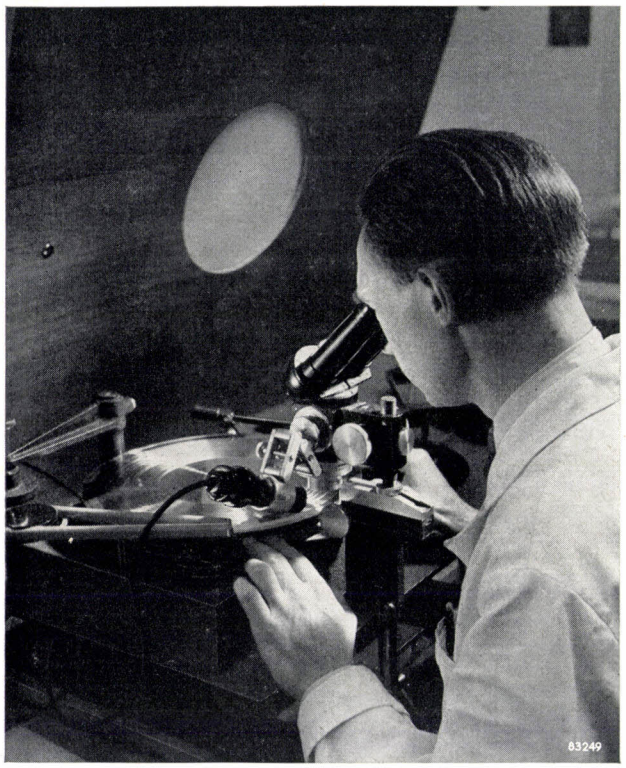


Fig. 8. Checking the positive for damage and dust particles.

ing by cathodic sputtering or evaporation onto the disc have for all practical purposes been abandoned, since both the vacuum and the temperature rise associated with these methods have an unfavourable influence on the lacquer disc.

When the lacquer plate has been covered with a layer of silver — only a fraction of a micron in thickness — a layer of nickel, followed by a layer of copper, are electro-deposited on the disc. The lacquer plate can now be separated from the metal. The thin coating of silver comes away with the metals which have been built up on it. A metallic copy of the lacquer disc is thus obtained, with the difference however that the copy bears ridges instead of a grooves; it is therefore called the negative.

In principle it is possible to use this negative as a matrix, for an impression of it in a plastic material would be an exact positive replica of the original, i.e. the lacquer disc. This procedure is followed, however, only if the record is to be issued in small numbers. The method is not followed for larger issues because, should the negative be damaged during pressing there would be no "reserve"; it would then be necessary to make a new lacquer disc engraving and to prepare a new negative from it (the latter operation can generally be performed only once on a lacquer disc), and the manufacture of this particular record would be held up during this time. To avoid this, a metal positive is first prepared from the negative, by an electro-forming operation on the negative, again of nickel and copper. (To avoid fusion of the positive and the negative, the latter is first covered with a layer of a separating agent.) In this way a positive is obtained which is in fact a metal gramophone record (with grooves on only one side), which it is possible to play. This is in fact done in order to check the positive for irregularities which might have arisen during the plating processes; these irregularities are if necessary corrected. With the aid of very fine engraving tools and a microscope, particles of dust and other extraneous matter can be removed from the groove (*fig. 8*). A metallic negative is now prepared from the positive in the same way as the positive was derived from the negative. This second negative is the pressing matrix or stamper (*fig. 9*). The copper back of the matrix is rendered absolutely plane parallel with the front side on a special lathe. The centre is then determined by an optical method (*fig. 10*), this being the point with respect to which the groove forms an equiangular spiral, and the centre hole punched (on a disc with variable groove distance there is always a part with constant groove distance, which can be used for determining the centre). The outer edge is then turned to size. To improve the durability, the matrix is finally plated with a layer of chromium, a few microns thick.

Fig. 9. Separation of the positive and the electroformed stamper.

Fig. 10. Centring a matrix. The latter is loosely champed to a rotatable table. The position is altered until on turning the table, the grooves, viewed through a microscope, are seen to move in one direction only (either inwards or outwards). The lever is then pulled down, punching out the centre hole.

Should the matrix be damaged during manufacture a new one can be made without occasioning a long delay, from the metal positive. The latter, like the metal negative from which it was derived, suffers no damage whatsoever from the electroforming copying process. Should the positive also be unsuitable for use when a number of matrices have been made from it, there is a further reserve in the form of the metallic negative.

The mass-production of gramophone records

Before discussing the pressing, i.e. the mass-production of the actual gramophone-records, let us first say a few words on the materials from which the discs are pressed.

Pressing materials

Two kinds of material are employed nowadays for the manufacture of gramophone records. The older is a shellac material, the more modern a synthetic resin, which is actually a copolymer of vinyl chloride and vinyl acetate. Apart from giving a very smooth surface (thus an almost noiseless groove), the synthetic material has the added advantage of being unbreakable [3]).

The methods of preparation of these materials differ considerably. The shellac material consists principally of shellac, copal resin, a pigment and a filler. These constituents, having been very finely ground, are mixed while still dry. The mixture is then plasticized and kneaded in a rolling plant with one heated roller; the stiff mass is then fed to a special type of calender which cuts the material into tablets of the size required for one gramophone record. In the preparation of the synthetic resin the constituents — principally copolymer powder — are mixed with a pigment and a stabilizer, the latter to ensure that the copolymer does not decompose during the manufacturing process, giving off chlorine, which is harmful. The material then passes through an extrusion press where it is heated and plasticized prior to extrustion; it then passes through a cutting machine. It leaves the latter in the forms of grains about 5 mm in diameter. This is the basic material for long-playing records. The material is examined for the presence of undesirable particles of metal by means of a metal detector [8]).

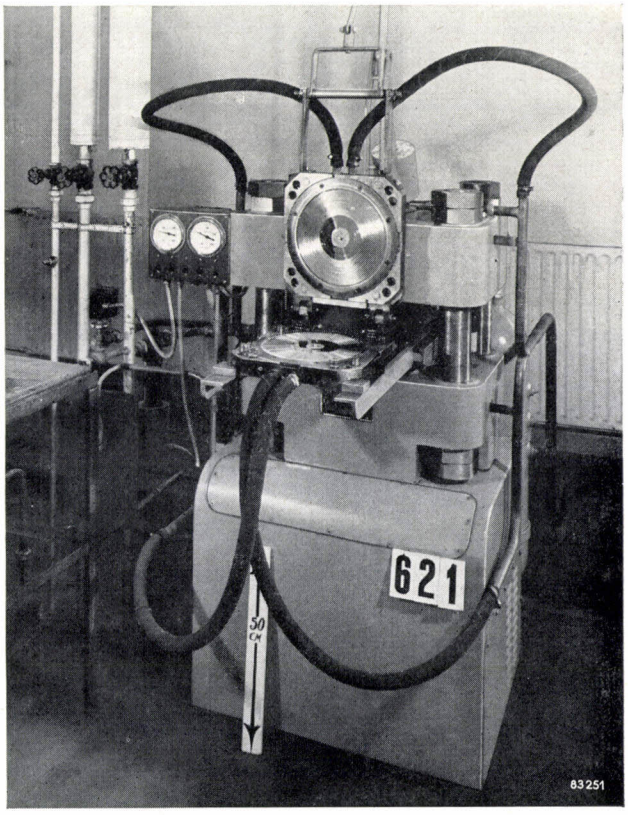


Fig. 11. Gramophone record press shown open.

[8]) E. Blasberg and A. de Groot, Philips tech. Rev. **15**, 97-104, 1953/54 (fig. 8).

The pressing

Gramophone records are pressed in hydraulic presses (*fig. 11*) of the order of 100 tons capacity. The presses are equipped with moulds to which the matrices are fixed. The material to be pressed is pre-heated, the shellac tablets on a steam-heated table, the synthetic resin grains in an electric oven. The pre-heating is necessary to accelerate the pressing cycle and more especially to save the matrices, which would wear rapidly against the hard, cold material.

Prior to pressing, the matrices and the corresponding labels are attached to the moulds and the pre-heated material (either a shellac tablet or a suitable amount of granular resin) is fed in. During these operations the moulds are steam heated, the heating being continued during the pressing operation itself. The hot material spreads out rapidly over the whole surface, so that it completely fills the mould; the steam is then shut off and cold water is fed in to cool the moulds and matrix rapidly to a temperature at which the record has become sufficiently solid to be removed from the press.

The edge of the record is freed from extruded material and carefully finished on a polishing machine. This is followed by the visual inspection of each record, packing in covers and delivery to the store. At regular intervals a check is conducted in which records are selected at random and tested for musical quality.

Summary. The chain of operations in the manufacture of gramophone records is described. The music is recorded magnetically on tape, and subsequently transferred onto a lacquer disc by means of a groove-cutter. A metal negative is prepared by an electroplating on the lacquer disc; from this negative a positive and a further negative are made by electroforming. The last negative is the matrix (stamper), with which the gramophone records are pressed out in a hydraulic press. Some records are still manufactured from shellac, but the modern synthetic resin material — a copolymer of vinyl chloride and vinyl acetate — is becoming increasingly used for long-playing records. Various phases in the manufacturing process are discussed in some detail.

A GENERATOR FOR FAST NEUTRONS

621.384.6:539.185.442

In the last few years interest has developed in the effects produced by fast neutrons, i.e. neutrons with a kinetic energy of the order of 10^5 or 10^6 electron volts. Such neutrons, like other types of corpuscular radiation, possess a strong ionizing action especially in media containing hydrogen and can therefore have a considerable effect on living matter and on substances consisting of large organic molecules.

While intense beams of slow neutrons can be obtained from atomic piles, the generation of high energy neutrons depends on particular nuclear reactions which take place in collisions of particles which have been accelerated to high energies. In the Philips Laboratory in Eindhoven, an apparatus has been developed which produces fast neutrons by means of the so-called D-D reaction:

$$_1D^2 + {}_1D^2 \rightarrow {}_2He^3 + {}_0n^1 + 3.28 \text{ MeV}.$$

For this reaction, deuterium ions $_1D^2$ (heavy hydrogen nuclei) are accelerated by a high voltage and directed onto a target plate composed of a material rich in deuterium. In our neutron generator the latter consists of frozen D_2O, i.e. "heavy ice".

The accelerating apparatus (*fig. 1*) is based on a cascade generator which delivers a direct voltage which can be varied continuously between zero and 1 million volts. This generator, which is composed of a series of condensors and high voltage rectifiers, departs from the conventional design and is built as three columns around the porcelain tube in which the ions are accelerated. An ion source is situated at the upper end of the tube and connected to the high potential pole of the generator so that it is at a high positive potential with respect to earth.

The ions are produced here by an auxiliary discharge in deuterium gas at low pressure. The discharge is maintained by a high frequency alternating voltage. The power needed for the source is supplied by a small dynamo, driven via an insulating belt by a motor mounted by the earthed foot of the cascade generator (see fig. 1). A special feature of the cascade generator is that instead of thermionic valves, selenium rectifiers are used (nominal peak inverse voltage 250 kV each); these require no heater current, so that the well-known problem of supplying heater current to valves which are at a high potential does not arise. The accelerating tube

is built up of five accelerating sections, the electrodes of which are of such a form that they focus the beam of downward directed ions. The ion current can be regulated between zero and a maximum of 1200 μA.

The target plate consists of a hollow copper disc, the edge of which is kept covered with a layer of heavy ice. The layer need not be more than about 20 μ thick, i.e. about the penetration depth of

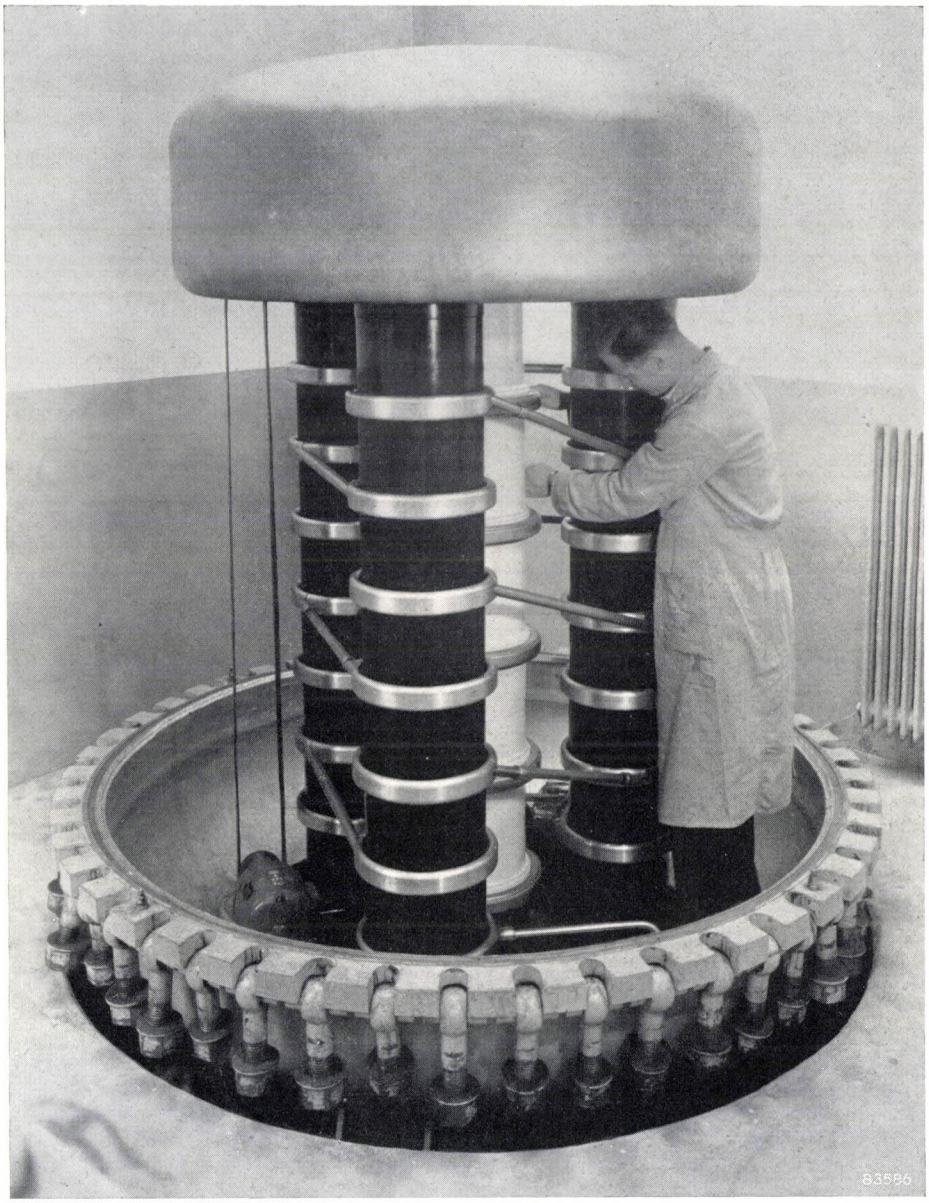


Fig. 1. 1 MeV cascade generator, arranged in three columns around the porcelain acceleration tube. Left, insulating belt which drives the dynamo for supplying the ion source and other auxiliary apparatus.

The cascade generator and the accelerating tube take up comparatively little space, since they are placed in a steel drum, three metres high, in which nitrogen at a pressure of 10 atm. provides the insulation. The pumps which maintain the high vacuum in the acceleration tube, are mounted at the bottom of the drum (*fig. 2*). This photograph is taken one floor lower than the room in which the generator stands, and shows also the earthed end of the acceleration tube and the target.

deuterium nuclei of 1 MeV energy in heavy ice. The temperature of the ice surface may not rise above about — 100 °C, since the vapour pressure of the heavy water would then be too high for proper functioning of the acceleration tube. The hollow target is therefore filled partly with liquid nitrogen, which continuously evaporates and so provides the necessary cooling. To avoid local overheating of the ice by the narrow ion beam, the beam is made to trace Lissajous figures on the ice surface (tracing

speed a few hundred m/sec) and the target is rotated (peripheral speed 8 m/sec) about a shaft introduced through the wall of the vacuum vessel and driven by a small electric motor. The shaft is hollow and through it the evaporated nitrogen escapes from the target and liquid nitrogen is supplied (*fig. 3*). These measures make it possible for about 500 W to be continuously dissipated on the layer of ice. At a voltage of 1 MeV, therefore, the ion current can be raised to about 500 μA.

It has been shown that with this neutron generator a continuous flux of 10^{10} neutrons per second can be obtained for many hours at a stretch. The neutrons leave the target in all directions through the walls of the target chamber. The neutron intensity is greatest in the direction of the ion beam. The neutrons also have the highest energy in this direction, viz. at the maximum ion acceleration voltage, about 4 MeV.

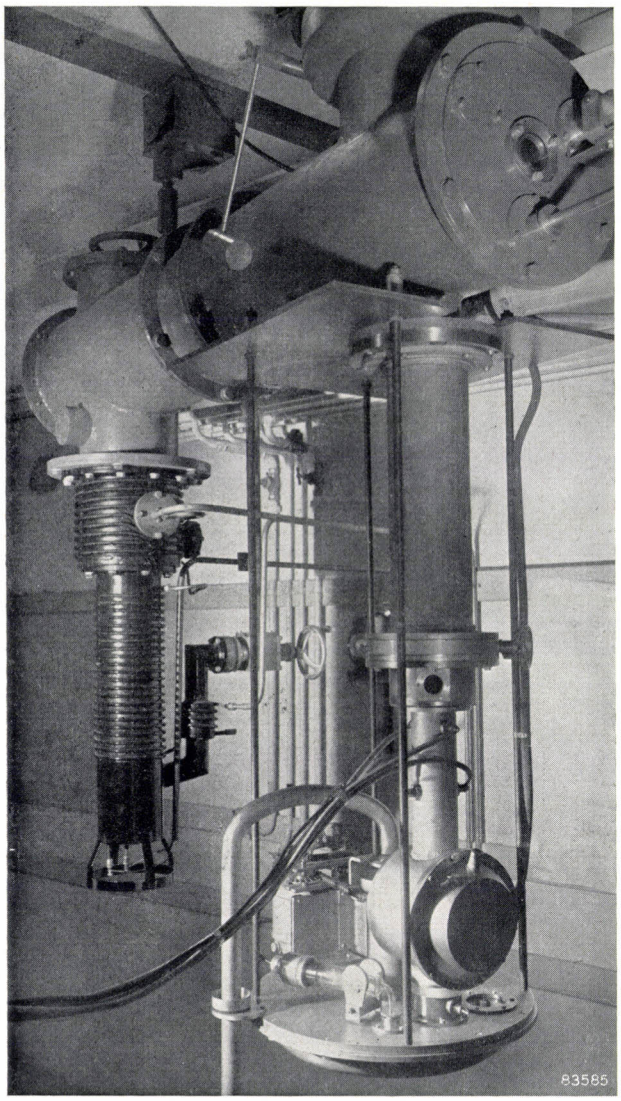


Fig. 2. The earthed lower end of the acceleration tube (a floor lower than in fig. 1). In the background is the pump installation which maintains the high vacuum in the acceleration tube.

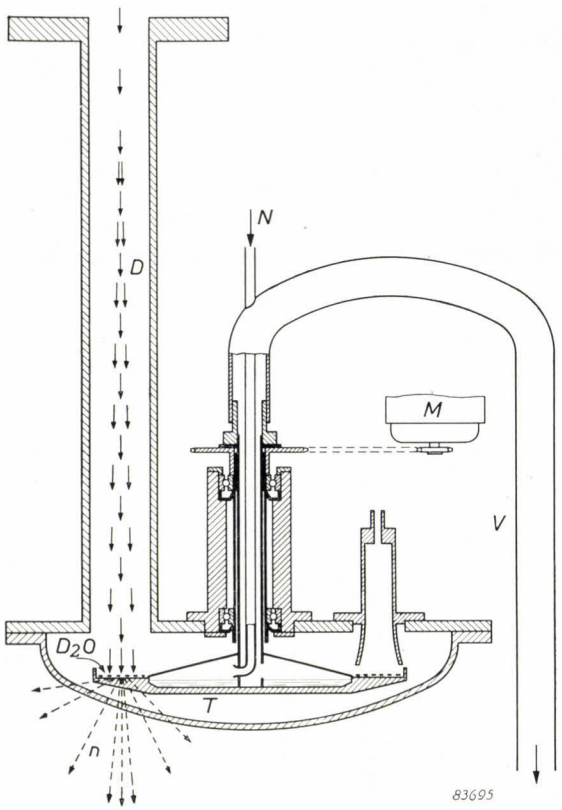


Fig. 3. Cross-section of the lower end of the acceleration tube, with hollow target *T* and its hollow axis of rotation. *D* deuteron ion beam, D_2O layer of heavy ice, *n* neutrons produced, *M* electric motor, *N* supply of liquid nitrogen, *V* chimney for extracting the evaporated nitrogen.

A. C. van DORSTEN.

THE "SYMMETRIX", A UNIVERSAL TABLE AND STAND FOR X-RAY DIAGNOSIS

by J. J. C. HARDENBERG and H. W. DUMBRILL *).

616-073.755.1

Many articles on the medical application of X-ray have been published in this Review, mostly concerned with physical and electrical problems: the production of the required X-rays, the X-ray tube supply, the quality of the X-ray image, the effect of irradiation upon human tissue, and so on. However, the actual radiological work also involves interesting mechanical problems, and the ultimate success of diagnosis or treatment depends quite considerably upon the ingenuity of the designer in solving these problems.

Introduction

In general, an X-ray diagnostic apparatus comprises five essential components: the X-ray tube; the H.T. generator; the control desk for the voltage and current of the tube, and the exposure[1]); a fluorescent screen or a film holder; and finally, the mechanical equipment, usually a universal tilting table, enabling the radiologist to position the patient as required. This equipment has been the subject of study for many years by mechanical engineers. This is a consequence of the development of many different methods of radiological examination, now in regular use in hospitals: clearly the equipment has to be designed to facilitate these procedures, which involve the examination of every part of the human body. The ease with which the radiologist can perform his task will evidently depend very largely upon the thought and precision embodied in the design of this equipment.

The requirements for the equipment are that the X-ray tube, fluorescent screen, film holder etc. must be movable to enable them to be positioned correctly in relation to the patient. It may also be convenient to be able to move the patient himself, the most common example being movement of the patient from a vertical to a horizontal position, which, by further rotation may then be converted into the so-called Trendelenburg position. In general, the X-ray tube, fluorescent screen, etc. should move with the patient so as to maintain the same relative position. This enables the radiologist to examine the patient not only in the vertical, horizontal, or intermediate positions, but also during the actual movement of the patient.

This technique is not employed in all countries. In Sweden, for example, separate examinations in the vertical and horizontal positions are usually preferred, and separate installations are employed for these examinations. However, in most countries so-called universal tables, suitable for the above-mentioned technique, are employed and are undergoing steady development.

Most of these equipments have a table turning about a horizontal axis. *Fig. 1* shows such a table in the horizontal position. The patient lies on top of the table; beneath it is the X-ray tube, which can be moved parallel to the table both longitudinally and laterally. Above the subject is the viewing equipment, that is, a fluorescent screen, usually combined with a serial cassette changer (described later). The viewing system is coupled to the X-ray tube in such a way as to follow its movement parallel to the table top. Moreover, the viewer is adjustable in a direction perpendicular to the table

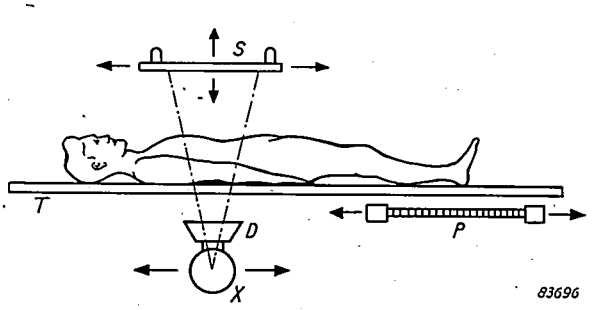


Fig. 1. Essential parts of a table for X-ray diagnosis. *T* table to carry the patient; *X* X-ray tube (with adjustable lead diaphragm, *D*, to limit the X-ray beam); *S* viewing system comprising a fluorescent screen and film-cassette; *P* Potter-Bucky diaphragm, and film-cassette employed in conjunction with an alternative separately mounted X-ray tube. The tube (*X*) can be moved parallel to the table, that is, in the directions indicated by arrows, and also at right angles to the plane of the drawing; the viewing system (*S*) moves with the tube, and can also be moved perpendicular to the surface of the table. The table top is made of a material of low X-ray absorption.

*) Philips Balham Works, London.

1) Previous articles dealing with these components are: H. A. G. Hazeu and J. M. Ledeboer, A universal apparatus for X-ray diagnosis, Philips tech. Rev. **6**, 12-20, 1941; A. Nemet, W. A. Bayfield and M. Berindei, A diagnostic X-ray apparatus with exposure technique indication and overload protection, Philips tech. Rev. **10**, 37-45, 1949/1949.

top, to enable the radiologist to lower it as required to the patient (usually positioned with his back to the table) or even to employ the so-called compression technique (see later).

For some examination procedures, the table is used in conjunction with a second X-ray tube, mounted separately. Use is then often made of a so-called Potter-Bucky diaphragm to reduce X-ray scatter; this is fitted underneath the table as shown in fig. 1 [2]). The Potter-Bucky diaphragm must move or oscillate in a certain manner during the exposure. In addition, it must be adjustable to various positions beneath the table to enable different parts of the body to be radiographed. Accordingly, the designer must provide for all these movements.

Apart from speed and reliability, the various movements of the equipment must also be such that once the adjustment has been made, the observation or radiograph may be made immediately. For example, individual parts of the equipment must not vibrate at the moment that the exposure is made and so affect the sharpness of the picture. Again, an efficient braking system is required, to arrest the equipment at the instant that the desired position is obtained. Finally, all the movements of table, tube, film-holder, patient, etc. must be effected without undue effort on the part of the radiologist.

Principle of the "Symmetrix" design

It will be evident from the introduction that so many requirements are put on a universal diagnostic equipment that it is difficult to satisfy them all. Moreover, developments during the past 10 or 20 years have shown that whenever a new design appears on the market, users tend to add the new facilities provided by it to their list of requirements.

This was particularly noticeable after the second world war, when it was found that the design of mountings on the continent of Europe had followed somewhat different lines from those in England and America. In view of this, Philips set out to find a design to satisfy, as far as possible, the requirements of both schools of thought. This project led to the development of the "Symmetrix" mounting, now approved by radiologists in many countries. [3]) Essentially, this design is based on two ideas: *a*) the table movement comprises consecutive rotation about two separate axes. *b*) the two rotations are symmetrical with respect to the horizontal position; the table can be tilted from the horizontal to the vertical either to the left or to the right. With this arrangement is is possible to satisfy a large number of requirements.

The idea mentioned in *a*) is relatively new. To explain it, we shall now consider a conventional type of universal diagnostic table, which is adjusted to the different positions by rotation about a single axis. Such a mounting is, for example, the Philips "Diagnost". This has a centrally located base containing the drive, the table pivoting about a horizontal axis at the top of the base (*fig. 2*). The axis is so positioned as to leave enough space between it and the table top for the movement of the scatter grid mentioned above, and to enable to table to be adjusted to Trendelenburg positions up to roughly 10°.

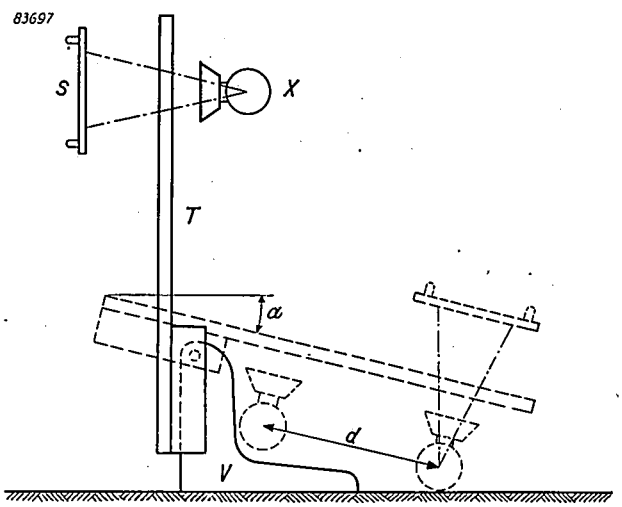


Fig. 2. Diagram of the Philips "Diagnost". The table (*T*) pivots on a spindle at the top of the central base (*V*). When the table is swung from the vertical to the horizontal position and beyond it (Trendelenburg position), there is a risk of the X-ray tube coming in contact with the floor; hence it is necessary, even for relatively small Trendelenburg angles (*a*), to fix the axis of rotation relatively high above the floor, Moreover, owing to the presence of the central base containing the spindle, the X-ray tube can be moved only a short distance (*d*) along the table.

Although satisfactory in many respects, the design of this stand satisfies the requirements only to a limited extent. Firstly, the central base limits the longitudinal movement of the X-ray tube under the table; secondly, the adjustment of the table to the Trendelenburg position involves a risk that the tube will touch the floor, and to avoid this it was necessary, even with the relatively small angle of 10°, to mount the table at a level somewhat higher than that which has been shown to be the most convenient.

In the "Symmetrix", this limitation is avoided by providing the table with two axes of rotation.

---

[2]) See W. J. Oosterkamp, Eliminating scattered radiation in medical X-ray photographs, Philips tech. Rev. 8, 183-192, 1946.

[3]) The design of the "Symmetrix" is based on a tilting table developed and produced by the Philips factories at Balham.

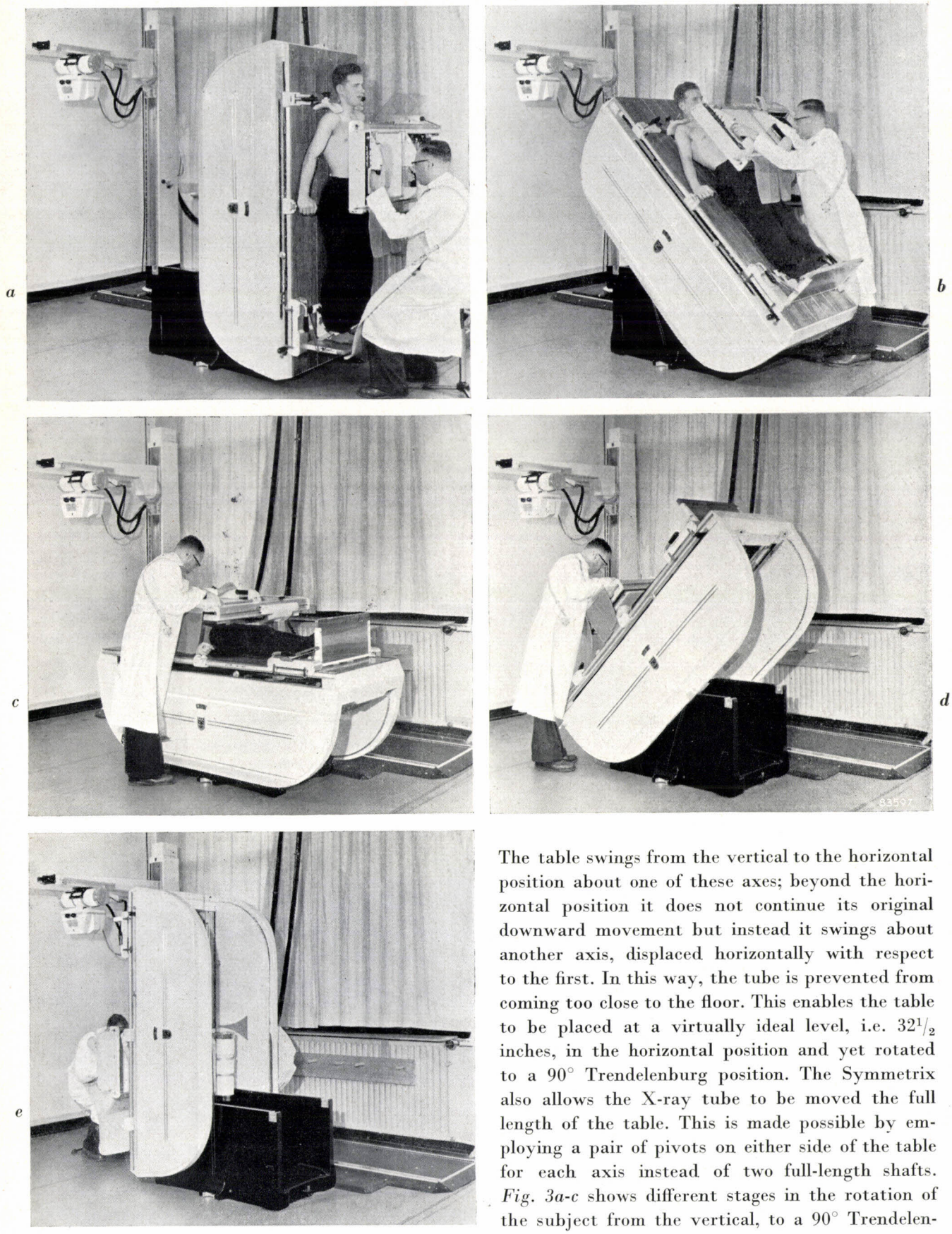


Fig. 3. The "Symmetrix" examination table can be rotated from the one vertical position (*a*) to the horizontal position (*c*) and thence to the opposite vertical position (*e*). In spite of this, the surface of the table in the horizontal position is only $32^1/_2$ inches above the floor.

The table swings from the vertical to the horizontal position about one of these axes; beyond the horizontal position it does not continue its original downward movement but instead it swings about another axis, displaced horizontally with respect to the first. In this way, the tube is prevented from coming too close to the floor. This enables the table to be placed at a virtually ideal level, i.e. $32^1/_2$ inches, in the horizontal position and yet rotated to a 90° Trendelenburg position. The Symmetrix also allows the X-ray tube to be moved the full length of the table. This is made possible by employing a pair of pivots on either side of the table for each axis instead of two full-length shafts. *Fig. 3a-c* shows different stages in the rotation of the subject from the vertical, to a 90° Trendelenburg position, in which the two symmetrically places axes of rotation successively come into operation.

The symmetrical location of the two axes of rotation is the logical outcome of the demand for deeper and deeper Trendelenburg positions (30°, 45°, 60°, etc.). Such positions can, of course, also be obtained in other ways: some of the existing systems combine two rotary movements, others one rotary movement and a translation. However, we consider that our system offers a considerable advantage as compared with these possibilities, in that the left-hand and right-hand vertical positions are equivalent and so enable the radiologist to choose either of them as the normal position for handling the patient.

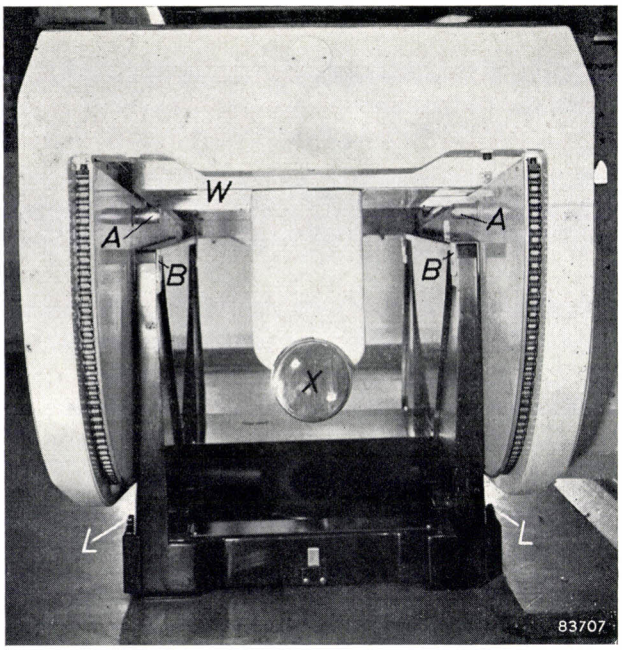


Fig. 4. End view underneath the "Symmetrix" with the table tilted slightly about the axis of rotation at the far end of the table. Both alternative axes of rotation consist of a pair of short spindles which turn in open bearing bushes at the top of the base walls. The spindles (*A*) at the near end of the table have just left their bushes (*B*). The use of short spindles in place of long shafts leaves the space under the table entirely free from obstructions and therefore enables the carriage (*W*) carrying the X-ray tube (*X*) to be moved the full length of the table underneath the patient. Note the two sprocket wheels (*L*, see fig. 6), one on either side of the base

It will now be clear that the principal feature of the "Symmetrix" table is that its axis of rotation changes from one pivot to the other when the table passes through the horizontal position. This is made possible by employing bearing bushes open at the top, enabling the one pair of pivoting lugs attached to the table to leave its bearings when the table begins to rotate on the other pair (*fig. 4*). The weight of the table is sufficient to prevent the spindles from leaving the bearings at the wrong moment.

Details of the "Symmetrix" table mechanism

Fig. 5 shows a cross-section of the "Symmetrix" table in the horizontal position. It is seen that the general profile of the base is U-shaped, and that the table fits over it as another, inverted, U.

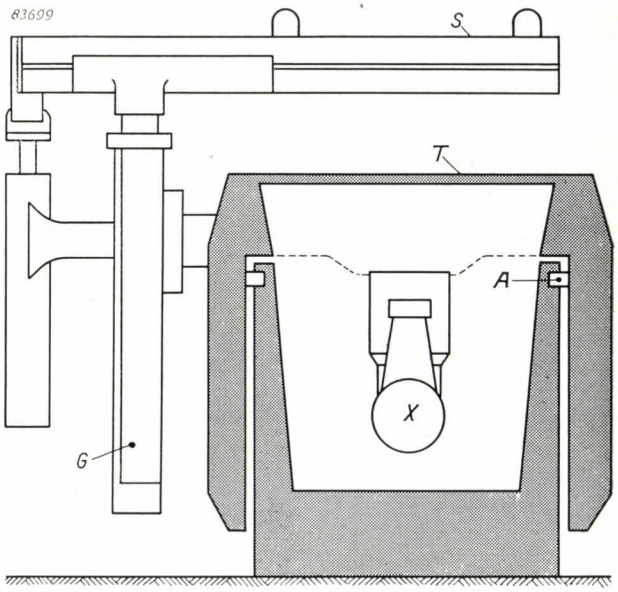


Fig. 5. Cross-section of the "Symmetrix" table. The base forms a U, over which the table, with its two sidewalls, fits as an inverted U. *G* is the mounting of the viewing system *S*. For the other letters see fig. 1 and fig. 3.

The mechanism to tilt the table is shown diagrammatically in *fig. 6*. Two driving spindles rotating in bearings in the base are positioned almost perpendicularly below the two axes of rotation of the table. A single electric motor placed at the centre of the base operates both driving spindles simultaneously at the same speed and in the same direction. Gear wheels at both ends of the two driving spindles, outside the base, engage toothed segments attached

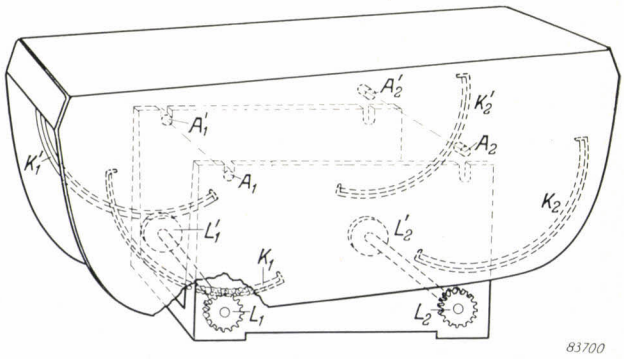


Fig. 6. Schematic drawing of the driving mechanism of the table. The table carries two heavy sprockets acting as toothed segments, K_1—K_1' and K_2—K_2', centring upon the two axes of rotation, A_1—A_1' and A_2—A_2', respectively, These sprockets engage with two pairs of sprocket wheels, L_1—L_1' and L_2—L_2', which are driven (at the same speed and in the same direction) by a centrally mounted motor (see fig. 9).

to the table and centring upon the respective axes of rotation. Except in the horizontal position, only one pair of toothed segments is engaged with the gear wheels. When the table passes the horizontal position, the pair of segments hitherto disengaged takes over from the other pair, the system being so designed that this transfer takes place without any interruption of the movement. In reality, ease of production and cost considerations have lead to the use of sprocket wheels instead of gears, and sprockets stretched round the quadrant-shaped ends of the table (see fig. 3), instead of the toothed segments.

The change-over of axis when the table is swung past the horizontal position can be explained more fully with the aid of *fig. 7*. The diagram shows the mechanism schematically, that is, the teeth of the wheels and segments are omitted in the drawing. Now, point P (fig. 7*a*), which will be the unique point of meshing at the moment of transition from the one pivot to the other, rotates first about axis A_1 and later, with the same angular velocity ω, about axis A_2. The peripheral velocity of rotation about the latter axis is $v_2 = h\omega$ (fig. 7*b*); the original peripheral velocity $v_1 = l\omega$ includes a component p in the v_2 direction. From the geometry of the figure we obtain

$$p = v_2 \, (\cos^2\varepsilon - \tfrac{1}{2} \sin 2\varepsilon/\tan \alpha),$$

or, if the angle ε is small,

$$p/v_2 \approx 1 - \varepsilon/\tan \alpha.$$

It can be seen that p would be precisely equal to v_2 if $\varepsilon = 0$, that is, if each gear wheel operates in bearings vertically below its associated axis of rotation of the table. This then, is the condition to ensure continuous movement of the table in the theoretical case when the teeth of the wheels are ignored.

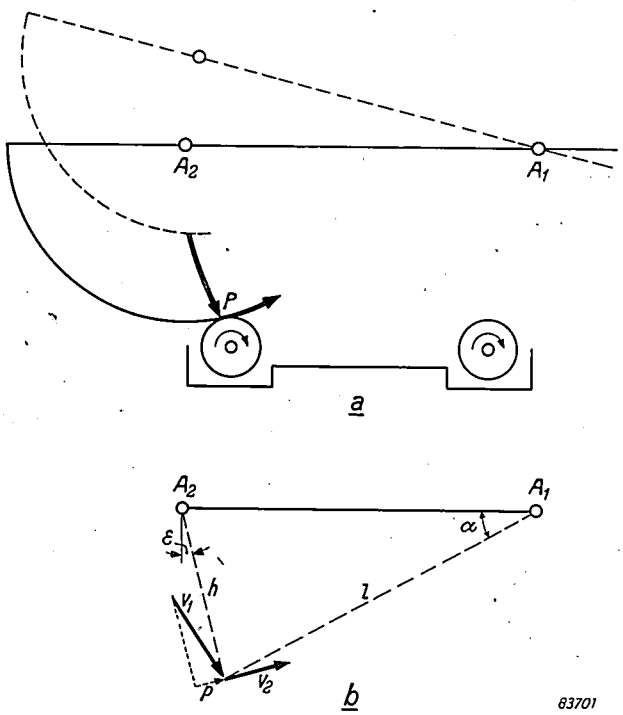


Fig. 7. Diagram illustrating the transition from rotation about axis A_1 to rotation about axis A_2.

In practice however, the teeth cannot be ignored. The changing of the pivot is then also governed by another important condition, that is, that the rollers of the sprocket engaging with the rotating sprocket wheel at point P must not scrape along the leading or trailing edges of the teeth on the sprocket wheel. To satisfy this condition it is necessary to make ε a small positive angle. In the "Symmetrix" $\varepsilon = 3°$. From *fig. 8*,

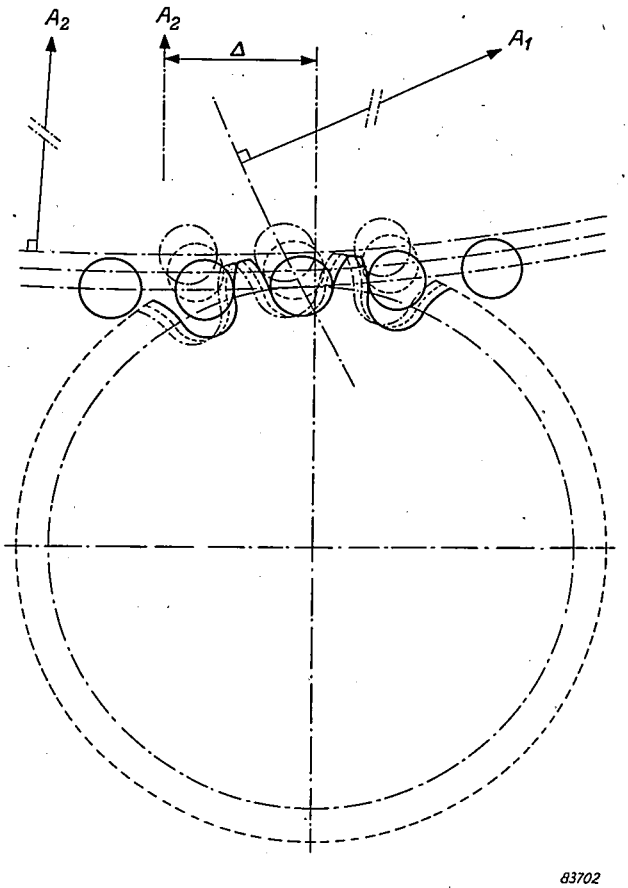


Fig. 8. The teeth of one of the rotating sprocket wheels and the rollers of the sprocket engaging with it (rotation about axis A_1), in three consecutive positions. Instead of being vertically below axis A_2, the sprocket wheel axis is displaced a certain distance (Δ) inwards.

illustrating the mating of the sprocket rollers and sprocket wheel teeth in three consecutive positions, it can be seen that in fact all the rollers clear the sides of the teeth by a small margin. This still occurs if the point of engagement of sprocket and wheel lies somewhat to the left or to the right of the position shown, as may well occur due to slight variations in the position in which the sprocket is mounted.

A single-phase, two speed, reversible electric motor is employed, enabling the radiologist to swing the table either quickly or slowly in the desired direction by operating a foot switch, or a conveniently placed hand switch. The higher speed, employed, for example, to adjust the table when unoccupied, brings the table from the vertical to the horizontal position in 20 seconds; with the lower speed, used in certain examinations involving continuous viewing of the image of a moving subject, the table takes 40 seconds for this movement. To

avoid an unduly perceptible jolt as the table swings past the horizontal position, owing to the fact that the vertical component of the motion of part of the table then reverses direction, feeler contacts are fitted in the bearing bushes to switch the motor automatically to the lower speed whenever one of the pairs of table pivots approaches its bearings (*fig. 9*).

To make the table top readily accessible for the positioning of the patient, whether in the vertical or the horizontal position, the mounting for the viewing unit runs on one side only of the Symmetrix (fig. 5). This mounting is constructed of light alloy and is fixed rigidly to the carriage carrying the X-ray tube. The whole assembly, guided by rollers, can be moved along the table. The viewing system, comprising a fluorescent screen and a serial film-cassette, is fixed on slides so that it can be moved both perpendicular to the table surface and across it. For the lateral movement, this system is normally coupled to the X-ray tube carriage by a telescopic connecting rod. By disconnecting this rod it is possible to push the serial cassette back far enough to all but clear the table surface. This is an advantage in that it facilitates the taking of radiographs with the separate X-ray tube with the Potter-Bucky diaphragm and cassette below the table (see next section).

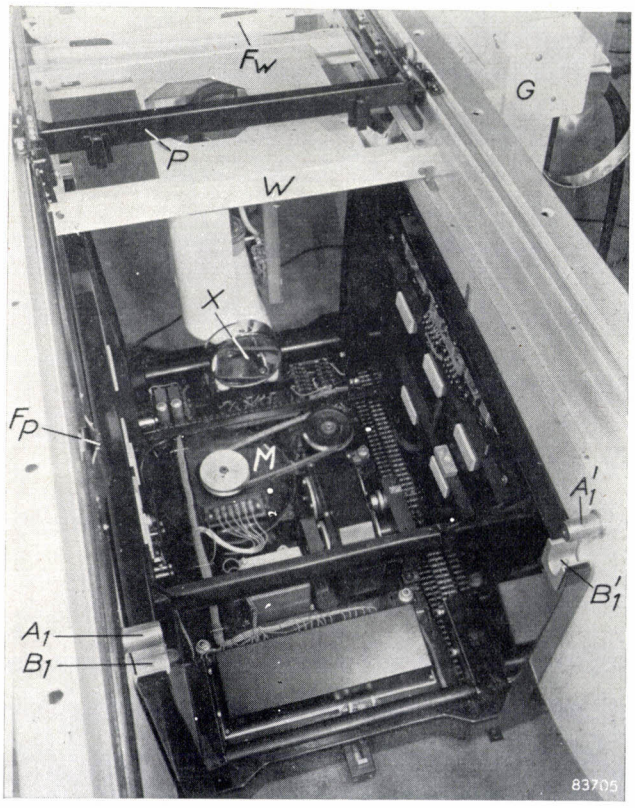


Fig. 9. Base of the "Symmetrix" table seen during assembly. M centrally mounted electric motor to drive the table: A_1, A_1' spindles and B_1, B_1' their associated open bushes. In B_1' there is a feeler switch which causes the table to rotate at the lower speed when passing throught the horizontal position; the switch operates through a relay system which is visible in the side wall of the base. *X* X-ray tube; *W* tube carriage with "concertina" cable (F_w); *G* mounting for viewing system; *P* frame of Potter-Bucky diaphragm, with concertina cable (F_p).

The X-ray tube and the viewing system coupled together may be adjusted a distance of 130 cm along the table, 30 cm across it and 35 cm in the perpendicular direction. Perpendicular movement of the viewing system when the table is in the horizontal position, is facilitated by a counterweight in the mounting. Movement of the viewing system and its counterweight and the tube carriage along the table when the latter is in the vertical position is also balanced by counterweights running along the side walls of the table (*fig. 10*).

Although this movement therefore involves virtually no work against the force of gravity, the setting in motion and stopping of such large masses would nevertheless mean that the radiologist had to overcome considerable forces if inertia. Accordingly, the longitudinal movement of tube and viewing system is driven, via sprockets and a friction and gear-wheel transmission by an auxiliary motor mounted at one end of the table. By pressing one of the lightly spring-loaded handles on the series cassette in the desired direction of travel, the radiologist closes a contact to operate the auxiliary motor in the appropriate direction. The motor thus drives the X-ray tube and viewing system in the same direction as the radiologist moves his hand. The torque-versus-speed characteristic of the motor is such as to ensure adequate initial acceleration of the masses to be moved: as the motor accelerates, the torque decreases until it is just enough to maintain the speed of movement usually required in practical examinations. This gives the radiologist the sensation that he is pushing the fluorescent screen and the cassette along the table by hand, with no more effort than is required to push them across the table. The latter movement is performed without the aid of a motor, since no counterweights have to be moved and therefore only relatively small inertia forces are involved.

The movable Potter-Bucky diaphragm under the table, referred to earlier in this article, is likewise provided with counterweights in the side-walls of the table to facilitate its movement along the latter.

Instead of balancing the viewing unit by means of counterweights, it would be possible to obtain the same effect by means of external springs; this would involve relatively smaller forces of inertia. However, we chose the above-mentioned method because it places less restriction on the positioning of the whole installation and does not impair the all-round accessibility of the table.

The flexible H.T. supply cables of the X-ray tube are carried along by the tube carriage as it moves along the table. A lateral guide passes them through the viewing system mounting to be drawn along by the tube as it moves across the table. table and the Potter-Bucky diaphragm, which is movable in relation to the viewing unit mounting. Finally the tilting table and the fixed base are connected by a cable passing over spring-loaded rollers in the side-walls of the table which pay out the cable

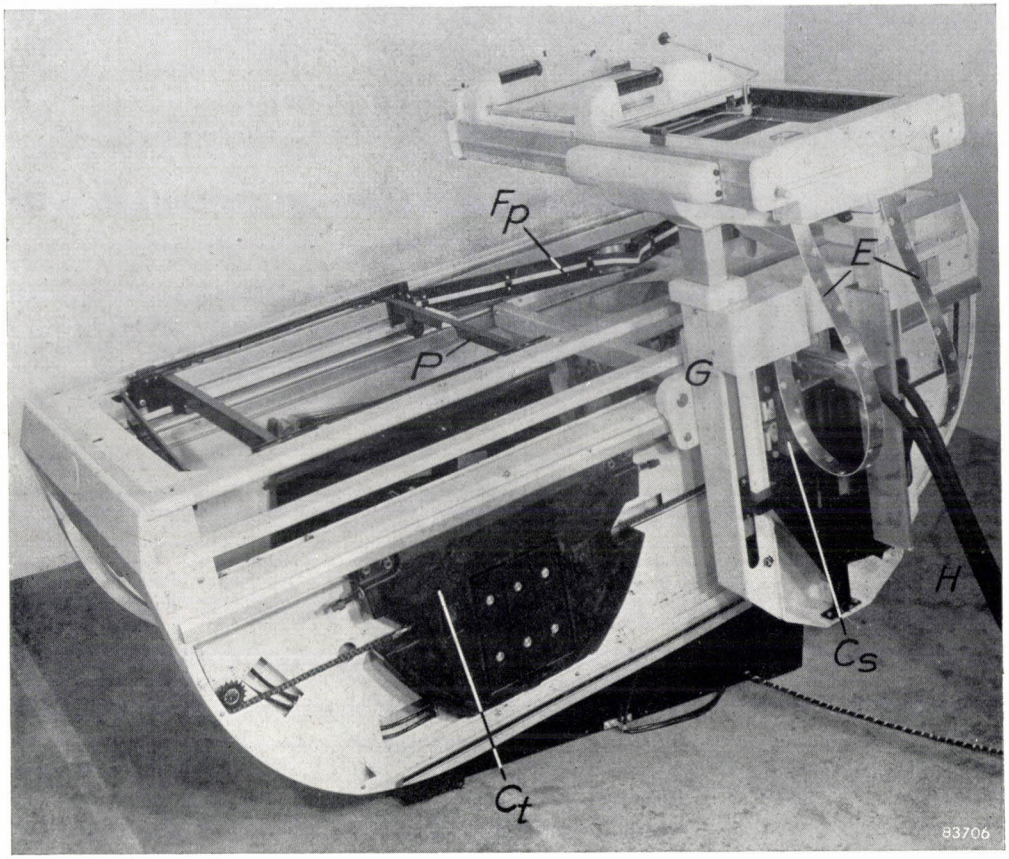


Fig. 10. The "Symmetrix" table during assembly. The table top, the side plates of the table and the cover plates of the viewing unit mounting G are not in position. C_s is the counterweight to to balance the perpendicalar movement of the viewing unit. One of the two heavier counterweights C_t employed to balance the movement of viewing unit and X-ray tube carriage along the table can be seen in the side wall of the table. H high-tension cables, passing through a lateral guide to the X-ray tube under the table; E flat multicore cables connecting the controls on the viewing unit to the mounting; F_p concertina cable to connect the scatter grid P to the table (only the frame of P is shown).

The "Symmetrix" table countains a complex system of wiring to energize the motors and to connect motors, X-ray tube, etc. to operating contacts on the viewing system, in the base of the equipment and in a hand switch. The connecting cables are designed to follow the different parts of the table as they move in relation to one another. Since the movements of the viewing system and the tube carriage are only short, i.e. 35 cm perpendicular to, and 30 cm parallel to, the table surface, they are connected by two looped flat multi-core cables (E in fig. 10). The viewing system and the tube carriage are linked to the table by a cable attached to one or two hinged strips, following in the manner of a concertina (F_w in fig. 9); a similar system is employed for connections between the as the table tilts from the horizontal to the vertical position, and take up the slack as it returns to the horizontal position. The need for such a system arises, of course, from the fact that the "Symmetrix" is not provided with one permanent axis of rotation; otherwise, it would be possible to connect the table to the base by means of a straight, flexible cable close to the axis.

Mounting of the second X-ray tube

As stated in the introduction, there are certain methods of examination for which the X-ray tube incorporated in the table is not suitable. Important examples are chest radiography, in which the tube is placed quite a long way from the subject (say, 1.5 m), and the radiography of internal organs with

the aid of the Potter-Bucky diaphragm, for which the subject must lie flat on the table, either prone or face upwards, and be irradiated vertically from above. Another example is planigraphy, a technique necessitating an altogether different arrangement of tube, subject and film.

In the case of the "Symmetrix", the second X-ray tube required for these and other techniques is mounted on a separate column, running on floor rails parallel to the table (*fig. 11*). In other designs ceiling rails are sometimes employed, either in conjunction with floor rails or alone. Although this enables the column to be dispensed with as an unnecessary obstruction, the many degrees of freedom and the corresponding counterweights which make up a rather heavy installation, would impose a severe strain on the ceiling. The floor rail system was therefore adopted, the rails being bedded in a low platform with sloping sides to prevent it from being a dangerous obstruction even in the dark.

The X-ray tube is mounted on a horizontal arm which can be moved up and down the column and rotated about its longitudinal axis. Since the column likewise rotates on its axis, and the X-ray tube can be moved to and fro along the arm, and also rotated about its own axis, the tube can easily be adjusted in relation to the "Symmetrix" table to obtain the required examination of the patient without the inconvenience of moving him. Moreover, the column-mounted X-ray tube can also be used to examine other patients on tables or stretchers arranged either in line with, or parallel to, the "Symmetrix" table. This combination of column and table therefore has some claim to be described as a universal mounting.

To facilitate the vertical adjustment of the tube, the arm and tube are balanced by counterweights sliding up and down in the column. To be within easy reach, the operating handles controlling the various movements are located at the tip of the arm. The X-ray tube is so mounted that the focus, from which the X-rays proceed, is on the geometrical axis of rotation of the arm. This is important in planigraphy In this technique, the X-ray focus and the cassette are moved in opposite directions parallel to the surface of the table, in such a way that the central ray of the X-ray beam always passes through a particular point inside the patient [4]). To satisfy this requirement it is necessary to turn the X-ray tube about the axis of the arm, and the above-mentioned arrangement enables it to be rotated without affecting the parallel movement of the focus.

Fully automatic planigraphy with the subject horizontal (in which the pillar must move 1 metre along the rails in roughly 1 second) or upright (in which the arm must move 1 metre vertically in roughly 1 second) can be carried out by means of a separate auxiliary unit.

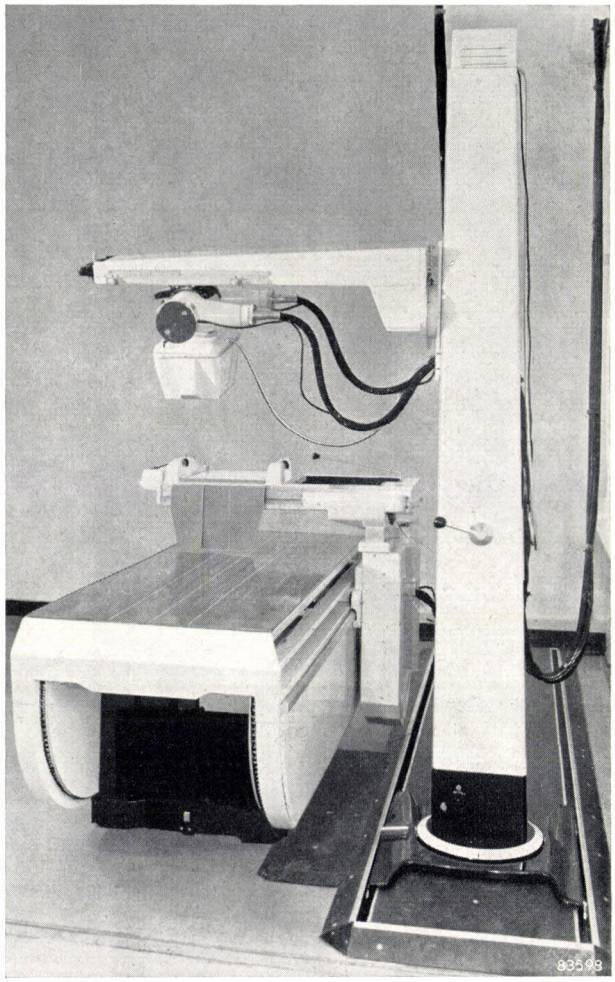


Fig. 11. Column carrying the alternative X-ray tube, for use with the "Symmetrix" stand. The column runs on double floor rails and rotates on its axis. The X-ray tube and its diaphragm are attached to an arm and can be moved to and fro along it. The arm rotates on its axis and can be raised and lowered on the column.

Practical application of the "Symmetrix"

Although it is beyond the scope of this article to go into all the operations involved in the various possible examinations with a universal diagnostic apparatus it may be useful to devote our concluding paragraph to a description of the procedure in carrying out a particular examination. Accordingly, the examination of the stomach by means of the serial cassette will now be considered.

The "Symmetrix" table having been adjusted to the vertical position, the subject walks up to it without having to stoop to avoid obstructions in the

[4]) See A. Bouwers, The localization of objects in the human body, Philips tech. Rev. **5**, 309-314, 1940.

dark, stands against it (if necessary on a foot rest) and may be given barium meal to swallow. The radiologist, sitting in front of the fluorescent screen, studies the X-ray image. He is protected against direct X-rays by a sheet of thick lead-glass, and from radiation scattered in the patient by a lead cover and sheets of lead-rubber hanging from the screen (fig. 4). He also wears a lead-rubber apron; lead-rubber gloves are sometimes used to protect his hands. Two motors mounted on the shield of the X-ray tube and controlled by push-buttons beside the screen enable the radiologist to vary the rectangular aperture of a lead diaphragm situated in the X-ray beam, and so mask the image, as required. (The image should not be larger than is strictly necessary for the particular examination, in order to minimize the decrease in contrast owing to scattered radiation; see the article referred to in note [2])). Gripping the handles on either side of the screen, the radiologist moves it over the subject to find the best position for a radiograph; he does this without effort, since the vertical component of the movement, involving the movement of the viewing system, tube carriage and counterweights, is carried out for him by the auxiliary motor. The radiologist may palpate the subject by hand or with a suitable instrument, or push the screen forward to "compress" the subject by a special pad of low X-ray absorption attached to the other side of the screen (the adipose and other tissues are thus pushed aside as far as possible, so that they do not absorb and scatter X-radiation unnecessarily). When once the screen is suitably positioned and the proper phase in the assimilation of the contrast medium is reached, the radiologist clamps all the moving parts of the apparatus and takes the radiograph. All that he has to do to clamp the apparatus is to turn either of the two screen grips slightly on its spindle, thus applying electromagnetic brakes to all the moving parts. To take the radiograph, he merely presses a button; the tube voltage and current and the exposure, are preselected, usually by an assistant operating the control desk.

The pressing of the button releases a spring-loaded slide to carry the cassette into position in front of the screen; as it moves, this slide closes a contact to prepare the X-ray tube for the load to be applied to it a moment later (the heater current then increases and the anode begins to rotate). If necessary, the same contact may be used to switch on an electric motor which oscillates a thin Potter-Bucky grid behind the screen. On reaching the end of its track, the slide closes another contact to switch on the tube H.T. for the pre-selected period. This completes the taking of the radiograph; the radiologist then pushes the slide back to its starting position, thus enabling the whole process to be repeated to record another phase of the stomach movements. The serial cassette is so designed that it automatically makes a series of exposures on the one standard-size film sheet; it may be adjusted to expose the film either all at once, or in 2, 4 or 6 equal parts.

A switch is provided on the serial cassette to enable the radiologist to tilt the table to certain angles indicated on a scale; in stomach radiography this facility is employed only in special cases, but it is important in certain other examinations.

Summary. The mounting of a diagnostic X-ray installation usually includes a movable examination table upon which the subject is placed, with an X-ray tube and viewing system (fluorescent screen, film-holder) mounted so as to enable the radiologist to position them as required in relation to one another. A certain widely used examination procedure involves turning the patient about a horizontal axis (at right angles to the X-ray beam) from the vertical to the horizontal position, and in some cases beyond it. During this movement the X-ray beam must tilt with the patient. The "Symmetrix" table described in the present article enables the subject to be tilted through 180° from the one vertical position to the other without limiting the movement of the X-ray tube along the table, and without the table having to be mounted too high above the floor. The essential feature of the design is that the table has two alternative axes of rotation, equidistant from the centre, which come into operation one after the other. Each axis comprises two short spindles turning in bushes which are open at the top, so that the pair of spindles not in operation may leave to permit rotation about the other pair. Two sets of counterweights are employed, the one to balance the viewing system as it moves perpendicular to the table, and the other in the side-walls of the table, to balance this system, together with its counterweights and the X-ray tube, when they are moved along the table as one unit. An auxiliary motor is employed to move the larger masses. The "Symmetrix" installation also includes a column, running on rails in the floor, carrying a second X-ray tube. The latter is provided with various degrees of freedom for use in certain important examination procedures.

ERRATUM

AN X-RAY TUBE FOR MICRORADIOGRAPHY

In *figs 3* and *4* of the above mentioned article (Philips tech. Rev. **17**, p. 46, Aug. 1955) the scale marks in the photographs should read 100 μ and 200 μ respectively instead of 10 and 20 μ.

A SIMPLE AND RELIABLE IONIZATION MANOMETER

by E. BOUWMEESTER and N. WARMOLTZ.

531.788.74

Since 1916, when Buckley constructed and described the first ionization manometer, a large variety of designs both of the manometer tube and of the associated circuits, have been devised. By a careful evaluation of the merits of the various systems, the authors have arrived at a combination of circuits and details which, though not in themselves original, form a reliable and accurate manometer.

Principle of the ionization manometer

Among the many methods for measuring low gas pressures [1]), within the range 10^{-2} to 10^{-8} mm Hg, the ionization manometer is still one of the most accurate instruments, and for its accuracy, the most convenient. Let us briefly recapitulate the principle.

The manometer head is mounted in a glass tube which is fused to the apparatus in which the pressure is to be measured. The tube contains as a rule three electrodes (*fig. 1*) viz. a cathode *f*, and

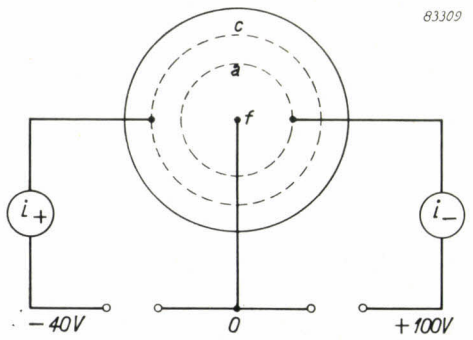


Fig. 1. Principle of the ionization manometer. The electrons emitted by the cathode *f* are accelerated by the anode *a* (which has the form of a grid), at which they finally arrive; they produce the current *i*—. The ions formed arrive at the ion collector *c* and produce the current *i*+.

two coaxially arranged electrodes *a* and *c*. The grid-like electrode *a* (the anode) is given a positive potential of, say, 100 V relative to the cathode, while the electrode *c* is given a negative potential of about — 40 V. The electrons are accelerated by the electric field between *f* and *a*. A number of them shoot through the anode and, in general, oscillate about it before impinging upon *a*. They therefore describe a path of considerable length, in the course of which they ionize some of the gas molecules within the envelope. The positive ions formed in the region between *a* and *c*, will impinge upon *c*, which is called the "ion collector" [2]).

The number of ions is proportional to the gas density and hence to the gas pressure, as well as to the ionizing electron current. If the latter is constant, then the current flowing from *c* is a measure of the gas pressure.

In designing an ionization manometer two essential problems present themselves: the adjustment and stabilization of the electron current, and the measuring of the ion current, which is generally very small (often of the order of 10^{-11} A). In this article we shall describe the design of an ionization manometer which has been used in the Philips Research Laboratories in Eindhoven and elsewhere for a considerable time (*fig. 2*) [3]).

This will be followed by a discussion of the calibration curves and of some of the measures taken to improve the accuracy at low pressures (down to 10^{-8} mm Hg).

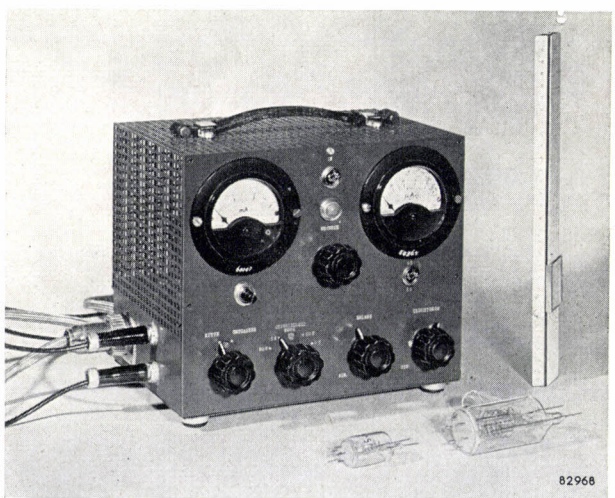


Fig. 2. The ionization manometer, showing two manometer tubes, one of a large and one of a small type.

[1]) A systematic survey of these methods is given e.g. in F. M. Penning, Manometers for low gas pressures, Philips tech. Rev. **2**, 201-208, 1937.

[2]) Sometimes *c* is given the function of anode and *a* that of collector.

[3]) A short description of the instrument in an earlier stage of development was given by: N. Warmoltz and E .Bouwmeester, An easily degassable ionisation gauge, Appl. sci. Res. **B2**, 273-276, 1951.

Adjusting and stabilizing the electron current

If the anode potential is given a sufficiently high value, so that the saturation current of the cathode is used as ionizing electron current, then fluctuations of the anode potential do not influence the calibration. This, however, requires a circuit for keeping this saturation current constant at the desired value by automatic control of the filament current. A control of this kind as a rule requires an amplifying tube and two regulator tubes and is thus rather complicated. If the anode current is limited by the space charge, then the current is virtually independent of the cathode temperature (provided only amounts to 10-15% of the mean value without feedback, is reduced to 1-2% by the application of feedback.

The manometer has been calibrated for two values of the electron current, viz. at 0.5 and at 2.5 mA. The latter value is applied for low pressures, since then the sensitivity is approximately five times greater. These values, which can be read from meter M_1 (fig. 3), are adjustable by varying the cathode resistance by means of R_3. The configuration of the electrodes in the manometer has been so selected, that at these emission current values, the cathode f becomes slightly positive with respect to g.

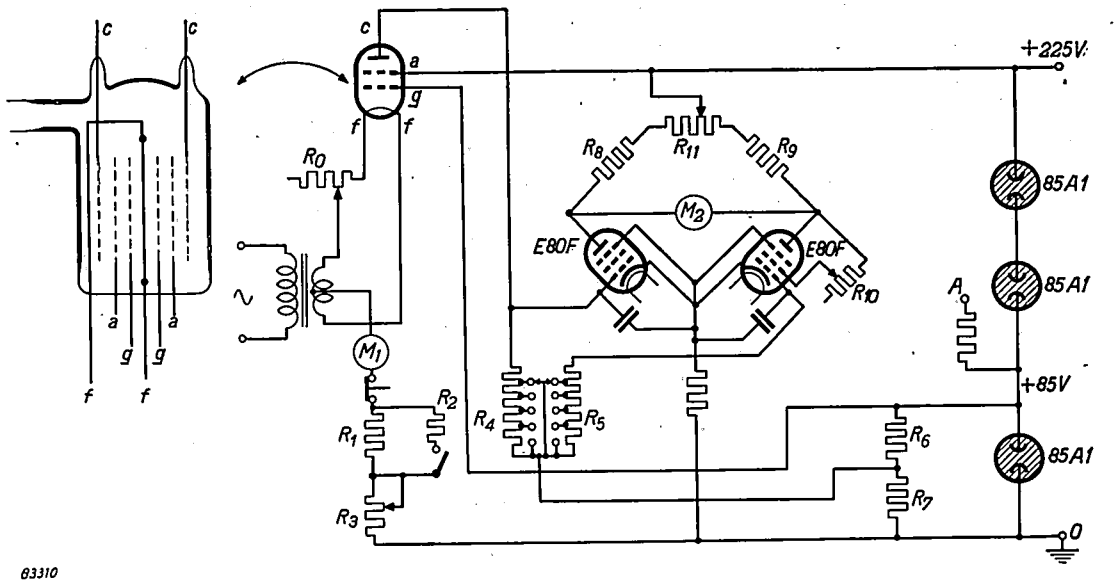


Fig. 3. Circuit diagram of the ionization manometer. The manometer tube is shown as a normal tetrode. The drawing on the left shows the tube more in its actual form. The glass wall is covered with an earthed conducting layer, except where drawn with thin lines. The meter M_1 is used for adjusting the ionizing electron current; the pressure is read from M_2.

that the latter does not drop so far that the current is again limited to the saturation current). Following C. G. and D. D. Montgomery [4]), we have applied an improved variation of the latter method. A grid g is fitted between cathode f and anode a (*fig. 3*). Together they form a triode, the anode current of which can be stabilized very simply by negative feedback via a suitable resistance in the cathode current. The potentials of control grid and anode are both stabilized by means of voltage-stabilizing tubes 85 A1, to + 85 and + 255 V respectively relative to earth. The ripple on the anode current, caused by the potential distribution along the directly heated cathode varying with the mains frequency, which

The cathode saturation emission has to be adjusted by means of the filament current to a value slightly (e.g. 0.5 mA) in excess of the actual value of the emitted electron current. The cathode temperature required for this emission and, therefore, also the necessary filament current, however, depends on the work function of the cathode metal, which in its turn is considerably affected by the type of gas inside the manometer. Oxygen, for instance, has the effect of greatly increasing the work function (poisoning of the cathode); with the presence of this gas, therefore, the cathode temperature should be far higher than when the tube is filled, say, with pure argon. The necessary filament current, however, is not determined by the required cathode temperature alone. The loss of heat from the cathode to its surroundings, which depends both on the gas pressure

[4]) C. G. and D. D. Montgomery, A grid-controlled ionisation gauge, Rev. sci. Instr. 9, 58, 1928.

and on the type of gas, also has an effect. It is thus obvious that the filament current necessary to give the required emission, depends on many variable factors. In order to avoid the risk of making the cathode temperature unnecessarily high and thereby shortening the working life of the cathode, it is desirable that the saturation emission can be directly indicated and adjusted. As shown in *fig. 4* this can be realized by temporarily reducing the cathode

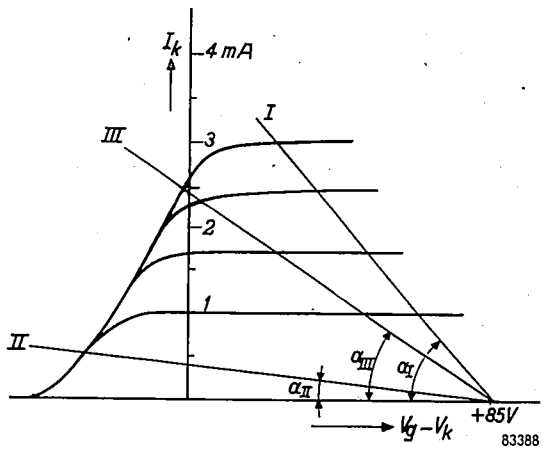


Fig. 4. The current I_k through the cathode circuit of the ionization manometer shown as a function of the voltage $V_g - V_k$ between the control grid and the cathode, for various values of the saturation emission of the cathode, and at constant control-grid and anode potential (respectively + 85 V and + 255 V relative to earth). The working point of the triode so formed is given in the usual way by the point of intersection of the relevant characteristic with a straight line (the load line), passing through the point + 85 V on the abscissa. (This latter point is in reality considerably further to the right than shown in this schematic diagram.). The load line makes an angle α with the abscissa; cot α is a measure of the resistance between cathode and earth. If this resistance is low (R_2 connected in parallel to R_1, and R_3 completely short-circuited, see fig. 3), then the saturation current flows through the tube (load line *I*). If R_2 is switched off, cot $\alpha_{II} = R_1 + R_3$; the cathode current can now be accurately adjusted at 0.5 mA by means of R_3, (load line *II*). If R_2 is switched out, the load line can be placed in position *III* by means of R_3, provided the saturation emission is sufficient; the anode current now amounts to 2.5 mA.

resistance, for this purpose the resistance R_2 can be shunted with R_1 (see fig. 3). The meter M_1 then indicates the saturation emission current. The filament current is subsequently adjusted by means of R_0, until the saturation emission has the desired value.

Measuring of the ion current

The ion collector *c* is connected, via the resistor R_4 (fig. 3), to a point having a potential of approximately + 42.5 V relative to earth ($R_6 = R_7$). The collector is, therefore, 212.5 V negative with respect to the anode and therefore attracts the positive ions formed beyond the anode.

The voltage produced across the resistor R_4 is applied to the control grid of a pentode E 80 F, and governs the current through this tube. The resistor R_4 is adjustable to five values, each differing by a factor of 10. For each value of the electron emission current, therefore, five measuring ranges are available.

The current through the pentode, however, is not directly used to indicate the ion current. The change in anode current due to change in the gas pressure is very slight and fluctuations of this current, e.g. due to insufficient stabilization of the supply voltages, would cause grave errors in the indication. This is prevented by incorporating the pentode in one of the branches of a Wheatstone bridge. Another pentode of the same type and the resistors R_8 and R_9 form the other arms of the bridge. The current through a microammeter M_2 (100 μA f.s.d.) connected between the anodes of the two tubes is a measure of the ion current and the corresponding pressure. Any fluctuations of the supply voltages have practically the same effect upon the two tubes and do not disturb the balance of the bridge. This method of measuring the ion current is due to Laferti [5]). Provisions are made to enable adjustment of the bridge, such that: 1) M_2 indicates zero (equilibrium of the bridge) if no ions are formed, which is the case when the ionizing electron current is interrupted by the switch under M_1; 2) the sensitivity has such a value that the indication of M_2 corresponds to the calibration curve (current through M_2 as a function of the gas pressure). The latter is effected by ensuring that with a given voltage on the control grid of the first pentode (obtained by connecting the calibration point *A* to this control grid), M_2 shows a certain deflection. In order to satisfy both the above conditions at the same time, two mutually independent adjustments are required, which are provided in this case by the variable resistors R_{10} and R_{11}. By means of R_{10} the screen-grid voltage of the second tube (and hence its characteristic) is adjusted, whilst R_{11} makes it possible to control the ratio of the two resistor-branches of the bridge. It has been found necessary to make sure that the grid resistors R_4 and R_5 of the two tubes have always the same value. Otherwise the voltages across the grid resistors caused by grid currents, mainly originating from positive ions from residual gas in both pentodes would disturb the adjustment after the instrument is switched over to another measuring range.

---

[5]) S. Dushman, Scientific foundations of vacuum technique, J. Wiley, New York 1949, p. 361.

The calibration curves

The manometer tube is made in two types, one of 50 cc and one of 10 cc capacity (see fig. 2). The smaller type has been specially developed for measurements on vessels of small volume. *Fig. 5* shows the calibration curve for both types, for an ionizing electron current of 0.5 mA and for the gases argon and hydrogen [6]). The current measured increases somewhat more rapidly than proportional to the pressure. In the range of the higher pressures

Improvement of the accuracy at low pressures

During the development of this manometer one of the foremost aims was a sturdy and reliable instrument. We have, therefore, avoided sacrificing these properties for an extension of the measuring range to pressures below 10^{-8} mm Hg, which is the normal limit for this type of gauge. Some provisions have been made, however, to increase the accuracy and reliability of the indication in the lowest range: 1) The wall of the glass tube containing the electrode

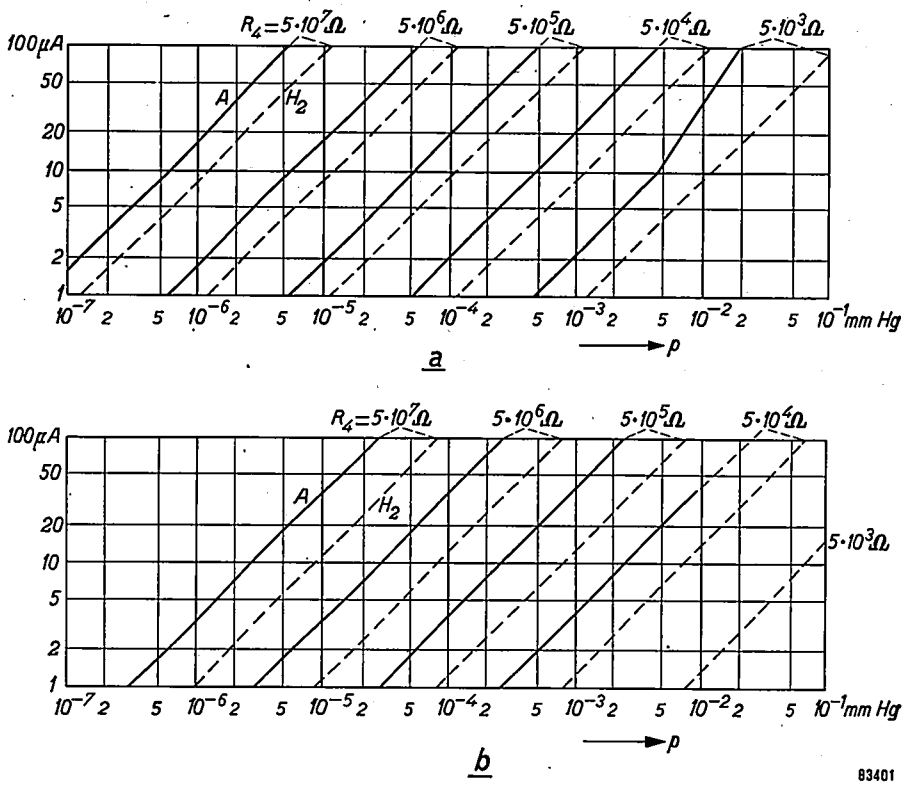


Fig. 5. Calibration curves for the large type of manometer tube (*a*) and for the smaller type (*b*). The full lines apply to argon (A), the broken lines to hydrogen (H_2). The ionizing electron current was 0.5 mA. The value of the saturation emission has some influence on the calibration at pressures higher than 5×10^{-3} mm Hg. In the present case the saturation emission was 1 mA. Both the pressure p and the current read from M_2 (see fig. 3) (not the ion current itself!) have been plotted logarithmically. Only a straight line with an slope of 45° then represents a linear relationship between pressure and current, so that with this instrument the current increases slightly more than proportional to the pressure. The values of the resistor R_4, corresponding to the various parts of the graphs, are included in the diagram. If the ionizing electron current is adjusted to 2.5 mA, pressures down to about 10^{-8} mm Hg can be measured. The ion current itself may be evaluated from the bridge sensitivity (760 μA/V) and the value of R_4.

(a few times 10^{-3} mm Hg) the large type deviates rather more from the proportionality in the case of argon. This is probably caused by an incipient avalanche-effect. It does not however impair the usefulness of the large type in this range.

The calibration curves as far down as 10^{-4} mm Hg have been obtained by comparison with an accurate McLeod gauge. For low pressures these measurements have been extrapolated.

[6]) As is known (cf. the article referred to in footnote [1])), the calibration curve depends on the type of gas.

system is made internally conducting and earthed with the exception of the parts in the vicinity of the electrode lead-ins. Positive ions formed on the wall — where many adsorbed gas atoms are present — due to light from the incandescent cathode cannot now reach the collector since the latter is positive to earth. Moreover, the earthed wall precludes the accumulation of any wall charges, which might indeterminately distort the potential distribution inside the manometer tube, a most undesirable effect, particularly at low pressures.

2) The ion collector is lead in through two protuberances on the head of the tube, designed to reduce leakage current between the collector and the other electrodes.

3) The ion collector takes the form of wide helix of fairly thin wire rather than a cylindrical plate, in view of the fact that at low pressures a disturbing effect is caused by the X-rays created by the electrons impinging on the anode. Part of this radiation reaches the ion collector and liberates electrons from it. In this way a current flows through the resistor R_4 which would be wrongly interpreted as an ion current. When the collector has the form of a helix the greater part of the radiation passes through the collector. The fairly thick lead-in wires of the collector are enamel-covered, to prevent their being hit by the X-rays.

This X-ray effect is one of the most important factors putting a lower limit to the measuring range of an ionization manometer. Alpert [7]), who was the first to adopt measures for suppressing this effect, used a very small and carefully situated collector and was thus able to measure pressures as low as 10^{-10} mm Hg.

4) Adequate de-gassing is particularly important when low pressures are to be measured. The three concentric electrodes (control grid, anode and collector), are mounted without supports, but each have lead-in wires at either end so that for de-gassing they can be heated by a current. For this purpose they can be connected together to a special winding of the supply transformer via a multiple switch (not shown in fig. 3). The various wire thicknesses and lengths of the grids are so chosen that during this process they all assume about the same temperature. They are made of molybdenum, which remains sufficiently rigid at high temperatures to prevent sagging of the grids. During the process of de-gassing, the filament may also be heated, and the envelope may be surrounded by a small furnace or heated in a gas flame in order to liberate the gas from the glass wall.

In order to obtain an idea of the influence of the above-mentioned X-ray effect, we have combined the ionization manometer with a mass spectrometer into a closed unit. A mass spectrometer can in principle be regarded as a sort of ionization manometer (*fig. 6*): here, too, the ion current impinging upon the collector is a measure of the gas pressure. In a mass spectrometer, however, the collector is so situated that it cannot be reached by X-rays from the anode.

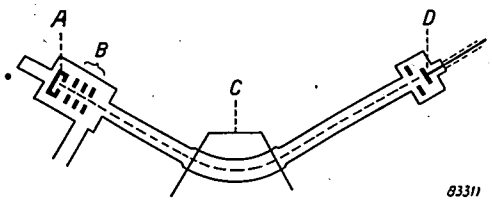


Fig. 6. Principle of a mass spectrometer. The gas to be analyzed, under low pressure, fills the entire spectrometer and is, therefore, also present in the ionization chamber *A*, where an electron current perpendicular to the plane of the drawing ionize the gas molecules. The ions thus formed are accelerated by the accelerating and focussing electrodes *B*. The magnet *C* produces a field perpendicular to the plane of the drawing, and is so adjusted that ions of the desired component of the gas impinge upon the collector *D*.

The mass spectrometer was first calibrated as a manometer in the range 10^{-4}-10^{-7} mm Hg by comparison with the ionization manometer, the system being filled with argon. (It was known that this mass spectrometer was capable of measuring pressures down to 10^{-12} mm Hg.) By prolonged pumping of the system, the pressure measured with the mass spectrometer reached about 10^{-9} mm; nevertheless the ionization manometer did not indicate lower than about 10^{-8} mm Hg. This, therefore, constitutes the limit of the measuring range for this type of manometer.

Summary. An ionization manometer with four electrodes is described. The ionizing electron current is stabilized by means of a grid at constant potential placed near to the cathode and a high resistance in the cathode circuit. The ion current is measured with a Wheatstone-bridge, two of whose branches consist of pentodes (E 80 F), the other two branches being resistors. The ion current passes through the control-grid circuit of one of the pentodes, thus unbalancing the bridge. A microammeter serves as the indicating instrument. Some measures for improving the accuracy at low pressures are discussed. The measuring range extends from about 10^{-2} mm Hg to approximately 10^{-8} mm Hg. The lower limit is restricted by the fact that under the influence of X-rays from the anode, electrons are liberated from the collector.

7) D. Alpert, New developments in the production and measurement of ultra high vacuum. J. appl. Phys. **24**, 860-876, 1953.

ABSTRACTS OF RECENT SCIENTIFIC PUBLICATIONS OF 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 Philips Research Laboratory, Eindhoven, Netherlands.

2184: Y. Haven and J. H. van Santen: Electrically and elastically active relaxation modes (J. chem. Phys. **22**, 1146-1147, 1954 (No. 6)).

An assembly of particles, each of which can move about in a system of potential wells separated from each other by energy barriers, proceeds toward the equilibrium distribution with a velocity given by a sum of exponential functions of time, i.e. there are a number of relaxation modes. The number of the relaxation modes and their symmetry properties are briefly discussed. For charged particles, some relaxation modes correspond to a change of electric moment (electrically active modes) while other modes correspond to redistributions of the particles that can be brought about by mechanical means (e.g. hydrostatic pressure, stretching or shear stresses). Some examples are given.

2185: J. M. Stevels: Iron in glass (Proc. Int. Comm. Glass **1**, 68-76, 1154).

Agreements and discrepancies are outlined concerning the explanations of the colour of iron-containing glasses. Proposals are made to introduce a "language" to be used in glass technology to describe the problems in a more general way, without reference to pre-conceived theories. The situation in a glass may probably be described by three parameters, namely: a structural parameter, α, which describes approximately how the rest of the network influences the position of the iron in the network; a reduction parameter, β, which indicates to what extent the glass is oxidized or reduced; and an electronic energy parameter, γ, which defines the ease with which electrons move within the medium consisting of oxygen ions.

2186: P. Baeyens: Electroplating with modulated current (to be published in Trans. Inst. Met. Finishing **31**, 1954).

A review of the many methods of electroplating with "modulated" current. The adjective "modulated" is proposed in this connection for a plating current that undergoes periodic variations in value or in sign. An extensive survey of the literature is given (excluding anodic oxidation and the formation of conversion coatings).

2187: H. P. J. Wijn: Quelques propriétés physiques des ferrites (Onde électrique **34**, 418-424, 1954, No. 326). (Some physical properties of ferrites; in French.)

After a brief survey of the chemical and physical properties of the ferromagnetic ferrites used in h.f. techniques, some special properties of ferrites are reviewed. The use of Ferroxcube as cores for high Q coils and for transformer cores is discussed. It is also shown that powdered ferrites can in principle be used at higher frequencies then the same material in solid form. Measurements of the magnetization curve of ferrites are given, from which conclusions may be drawn regarding the distortion caused by ferrite cores in circuits. Finally, a glass-enclosed ferrite ring having a rectangular hysteresis loop is described.

2188: J. S. C. Wessels: A possible function of Vitamin K in photosynthesis (Rec. Trav. chim. Pays-Bas **73**, 529-536, 1954 (No. 7)).

It is postulated that vitamin K plays a part as energy acceptor and hydrogen donor in photosynsthesis. Indications are given which may support this hypothesis. A possible function of Vitamin K in the process of oxidative phosphorylation is suggested.

2189: W. Elenbaas: Gasentladungslampen (Vakuum-Technik **2**, 36-42, 1954, No. 2). (Gas discharge lamps; in German).

Review of modern gas-discharge lamps, against the background of the physical processes involved. Low-pressure fluorescent lamps are dealt with first, with a short account of the concept of colour-rendering and colour-temperature. The conditions for the attainment of optimum efficiency in these lamps are also described. Increasing the vapour-pressure leads to increased efficiency of the discharge itself (i.e. without regard for the fluorescent powder coating). Lamps of this type (high and super-high pressure lamps) necessitate changes in construction and manufacturing processes, due to the increased loading of the discharge-tube walls and the resulting higher wall-temperature. A summary is given of the more common types of high and super-high pressure

lamps being manufactured at the present time, together with their applications.

Of the remaining gas-discharge lamp types, sodium lamps are described, followed by a short discussion of neon lamps, and, finally, of neon sign lamps.

R 241: J. S. C. Wessels: Investigations on photosythesis; the Hill reaction, Part II (Philips Res. Rep. **9**, 161-196, 1954, No. 3).

Continuation of *R 240*. The possibility of the formation of free radicals in the Hill reaction is discussed. Neither polymerisation of acrylonitrile nor formation of phenol in the presence of benzene could be demonstrated. From this it is concluded that free radicals are either not present, or only present in very low concentrations. It appears that the standard potential of the compound concerned chiefly determines its suitability to serve as a Hill oxidant. The maximal reducing power of chloroplasts in vitro is highly dependent on the oxygen concentration. In an oxygen atmosphere, final potentials lower than about 230 mV were never attained, whereas the lowest final potential observed upon exclusion of oxygen was about zero. Redox systems with a negative standard potential, the reduction of which requires more energy than corresponds to one light quantum, are hardly reduced at all by chloroplasts. The theoretical and biochemical aspects of these results are discussed at length and a simplified scheme for the Hill reaction is proposed. The reaction mechanism can be interpreted satisfactorily by this scheme. The influence of a number of inhibitors upon the Hill reaction was investigated and some discrepancies occurring in the literature cleared up. The fact that the reaction rate is not influenced by *p*-chloromercuribenzoate and *p*-aminophenyldichlorarsine indicates that free SH groups are not essential for the activity of the chloroplasts.

R 242: K. F. Niessen: Spontaneous magnetization as a function of temperature for mixed crystals of ferrites with several Curie temperatures (Philips Res. Rep. **9**, 197-208, 1954, No. 3).

For mixed crystals of ferrites with more than one Curie temperature it is shown by an example how the spontaneous magnetization can be calculated as a function of temperature.

R 243: J. B. de Boer and A. Oostrijck: Reflection properties of dry and wet road surfaces and a simple method for their measurement (Philips Res. Rep. **9**, 209-224 1954, No. 3).

During the last 20 years or so, much work has been devoted to the study of the reflection properties of road surfaces. However, for the purpose of road lighting, the results published up to now are either incomplete or insufficiently accurate. Methods and equipment which have been applied up to now for the measurements of the reflection properties of road surfaces are rather complicated. The very simple means used by the authors for making measurements in the laboratory and in the field are briefly described. In this paper quantitative data are given on the reflection properties of a number of road surfaces common in the Netherlands. These data have been given for those derections of light incidence and observation which are of importance in public lighting practice. They have been presented in a special diagram in order to enable the public lighting engineer to make calculations on the luminance of road surfaces in a quick and simple manner.

R 244: F. van der Maesen and J. A. Brenkman: The solid solubility and the diffusion of nickel in germanium (Philips Res. Rep. **9**, 225-230, 1954, No. 3).

Nickel produces rapidly diffusing acceptors in germanium, just as copper. Hall and resistivity measurements show the existence of a Ni acceptor level lying 0.23 eV above the valence band. On the basis of this picture, the solid solubility between 700 and 900 °C is derived from resistivity measurements. From these values and the liquidus curve of the phase diagram Ge-Ni, the distribution coefficients (k) at various temperatures are calculated. The distribution coefficient (k^*) of Ni at the melting point of Ge is calculated to be 1.8×10^{-6} according to a method of Thurmond and Struthers. The diffusion coefficient of Ni in Ge is measured between 700 and 850 °C; the activation energy of diffusion is found to be 21 kcal/mole. Annealing of a Ni-saturated Ge sample restores the original resistivity.

R 245: M. E. Wise: Converting a number distribution of particle size into one for volume or surface area (Philips Res. Rep. **9**, 231-237, 1954, No. 3).

A number distribution of equivalent radii of particles in a powder can be accurately converted into a distribution by volume or surface area, even if the data are highly and/or non-uniformly grouped. A formula is derived to do this and is applied to a microscopic analysis into a frequency distribution of radii between 1, $\sqrt{2}$, 2, $2\sqrt{2}$ 4,.....units.

R 246: C. G. J. Jansen, R. Loosjes and K. Compaan: The velocity distribution of electrons of thermionic emitters under pulsed operation, Part II. Experimental results and a tentative theoretical explanation (Philips Res. Rep. **9**, 241-258, 1954, No. 4).

The velocity distribution among electrons emitted by metallic and oxide-coated cathodes is investigated by means of the apparatus described in a previous paper (*R 207*). Both under square-pulsed and DC loads metallic cathodes emit electrons with a single well-defined velocity. BaSrO and SrO coated cathodes, on the other hand, show under pulsed operation a dispersion in electron velocities up to several hundreds of volts with spectra consisting of a set of more or less discrete lines. Under a DC load the same oxide coatings emit electrons with nearly uniform velocities, the spectrum consisting of one rather sharp line. When DC is added to a square-pulsed load the line spectrum gradually contracts into a single line when the proportion DC is increased. Coatings of BaO, ThO_2, and of a mixture of (BaSr)O and nickel powder give a one-line spectrum, although this line is less sharp than with metallic cathodes. The velocity spectra of (BaSr)O and SrO are explained by a high resistance in the surface layer of the coating which arises owing to the combined effect of the conduction through the solid material and through the pores. The occurence of discrete lines is tentatively attributed to the combined effect to the potential distribution in the surface layer, of secondary emission, and of the geometry of the pores. The effect of a DC load is ascribed to polarization in the oxide layer.

R 247: L. M. Nijland: Some investigations on the electrical properties of hexagonal selenium (Philips Res. Rep. **9**, 259-294, 1954, No. 4).

A short survey of the recent literature on selenium is given in the first section. The resistivity of polycrystalline, hexagonal selenium can be lowered by halogens and can be raised by thallium. No reliable data on fundamental quantities like hole densities and mobilities are found in the literature. Investigations on these subjects are described in the next sections. A method to purify selenium by evaporation near its melting point is developed High-frequency measurements of the conductivity of pure and of thallium-doped selenium samples are consistent with the assumption that polycrystalline selenium consists of rather well-conducting crystals embedded in badly conducting layers of more or less amorphous selenium. Thallium increases the resistance of the layers and does not or only slightly affect the resistance of the crystals. Measurements of the Hall effect and of the shunt resistivity vs. frequency of pure and of bromine-containing selenium samples point to the same layer structure. Bromine does not affect the resistivity of the crystals but lowers that of the layers. Single crystals prepared in a bromine-containing atmosphere have the same resistivities as pure crystals. The last section contains a discussion of the experimental results and points to a model with both amorphous and crystallized parts in the same selenium chain.

R 248: E. W. Gorter: Saturation magnetization and crystal chemistry of ferrimagnetic oxides, Part I (Philips Res. Rep. **9**, 295-320, 1954, No. 4).

Measurements of the saturation magnetization (σ) against temperature are carried out for a number of mixed crystal oxides with spinel structure. The results are in agreement with Néel's theory of ferrimagnetism: the resultant magnetic moment m is the difference of the moments of the tetrahedral (A) and octahedral (B) sublattices, either (a) with complete parallelism of the ionic moments inside each sublattice, or (b) with angles between the ionic moments inside one of the sublattices. The spinel structure is described in section 1.1; experimental and theoretical data from literature on cation distribution are summarized in section 1.2.

VOL. 17 No. 5, pp. 129-160 NOVEMBER 1955

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

CAPACITOR MATERIALS WITH HIGH DIELECTRIC CONSTANT

by G. H. JONKER.

537.226:621.315.612:621.319.4

In the construction of capacitors in recent years, use has often been made of ceramic materials such as $BaTiO_3$, which exhibits a very high value of the dielectric constant within a limited temperature range. It is desirable to reduce the temperature variation of the dielectric constant as much as possible, while retaining its high value. This may be done in a number of ways, such as by using mixed crystals of $BaTiO_3$ with other related compounds and by "diluting" the material with substances forming a second phase. A few examples are described here.

Introduction

Barium meta-titanate, $BaTiO_3$ (hereafter referred to simply as barium titanate), is a crystalline solid exhibiting the same crystal structure as the mineral perovskite ($CaTiO_3$). For rather more than ten years it has been increasingly used because of its unusual dielectric properties. It was already known that titanium dioxide (TiO_2) had a large dielectric constant, viz. about $\varepsilon = 114$, averaged for all crystal directions. For some compounds of titanium dioxide with other metal oxides, even larger values of ε were found. Of the simple compounds, barium titanate exhibits the greatest value. This material, however, exhibits a remarkable peak in its dielectric constant-temperature curve [1]). As can be seen in *fig. 1*, the dielectric constant of barium titanate shows a very high maximum ($\varepsilon \approx 10{,}000$) at $T = 123$ °C and a slight peak at $T \approx 10$ °C. These peaks correspond to crystallographic transition points. Above 123 °C, barium titanate forms cubic crystals, between 10 °C and 123 °C, tetragonal, and below 10 °C, rhombic [2]).

Below 123 °C, $BaTiO_3$ has a number of other remarkable properties which can be combined under the name "ferro-electric" properties. These are: *a*) the dielectric polarization exhibits the phenomenon of hysteresis, so that a residual polarization or remanence can occur; *b*) the dielectric constant is dependent on the amplitude of the alternating voltage used for the measurements and on the direct biassing voltage which may be applied.

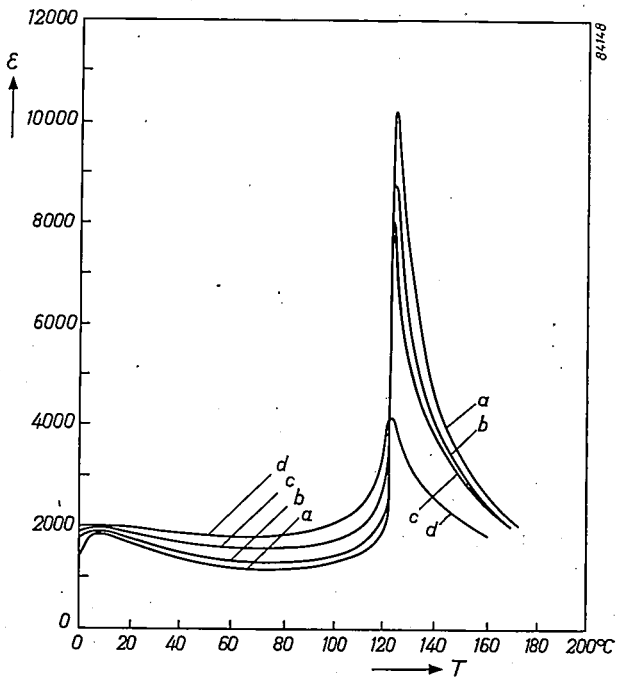


Fig. 1. The dielectric constant ε of $BaTiO_3$ as a function of temperature T. The curves *a*, *b*, *c*, *d* have been measured for specimens sintered at 1400, 1350, 1300 and 1280 °C respectively. In the same order the size of the crystals decreases from about 100 μ to 5 μ. The maximum at 123 °C corresponds to the transition from cubic to tetragonal structure, the maximum at about 10 °C to the transition from the tetragonal to the rhombic structure.

[1]) G. H. Jonker and J. H. van Santen. The ferro-electricity of titanates. Philips tech. Rev. **11**, 183-192, 1949/50; further referred to as I.

[2]) Below −70 °C, $BaTiO_3$ is rhombohedral. The transition point, which also gives rise to a low maximum of ε (see I, fig. 4), is not shown in fig. 1.

These special properties form an obstacle to the use of barium titanate in capacitors for use in oscillators. Apart from small temperature dependence of ε and low dielectric losses ($\tan\delta < 10^{-3}$), an essential requirement in this case is that the dielectric should not possess property (*b*).

The conditions are less rigid for coupling and by-pass capacitors in amplifiers and for smoothing capacitors and capacitors for interference suppression. We shall direct our attention in this article to this sort of application, for which the principal requirement is that the capacity is large in comparison with the dimensions. Variations of 10 % in both directions, resulting from temperature changes and from a voltage bias, are permissible. The dielectric losses may also be much greater ($\tan\delta < 0.03$).

Even these moderate demands cannot be met by $BaTiO_3$. In particular, the strong dependence of ε on the temperature raises objections. This material only shows large values of ε in a limited temperature range which, furthermore, is unfavourably situated. Fortunately it is possible to modify the properties of barium titanate in various ways. In the first place, by careful choice of the firing process, the ε peak can be considerably broadened. Further, one can work, not with $BaTiO_3$, but with mixed crystals of this substance with related compounds, obtained by replacing either the Ba-ion, or the Ti-ion, or both, by analogous ions. In this way the ε peak can be shifted to a more suitable temperature range, independent of the methods employed to make the peak broader. One can also "dilute" the material with another substance, either of related composition or not, which can behave as a second phase between the crystals of the original material. If the material chosen for this purpose has a smaller temperature dependence of ε, then the dependence on temperature of the mixture will be decreased. Finally, the properties can be favourably influenced, not only through the firing process, but by the admixture of small amounts of substances of different structure, which may dissolve in the material or settle along the crystal boundaries.

Combinations of these measures now make it possible to prepare a series of materials related to $BaTiO_3$, which do meet the requirements described above and which have been responsible for an increase in the application possibilities for materials with a large dielectric constant. We shall now consider these possibilities in more detail.

Influencing the transition points of $BaTiO_3$

The sharp peak in the ε-T curve for $BaTiO_3$ at $T = 123$ °C appears to be dependent on the manner of preparation of the material. For densely sintered, coarse crystalline material, with crystals of dimensions greater than 100 μ, a very high maximum is found, with a sharp drop on the lower temperature side. The graph of $\varepsilon = f(T)$ (see fig. 1, curve *a*) corresponds closely to that of a single crystal [3]).

If the material is prepared so that the crystals are much smaller, the maximum becomes lower and wider (fig. 1, curve *d* [4])). From a thorough investigation by means of X-ray diffraction photographs and from measurements of the specific heat (carried out in co-operation with J. Volger in the Research Laboratory at Eindhoven) it appears that in the ceramic material the crystallographic transition, which for a single crystal is sharply defined and takes place at a particular temperature, can be spread out over a large temperature range, e.g. between 100 °C and 170 °C. In this range one finds a mixture of two crystal forms, viz. the cubic, which is stable above 123 °C for a monocrystal, and the tetragonal, which occurs below 123 °C for a monocrystal.

That individual crystals of $BaTiO_3$ can have such divergent transition points is attributable to the fact that the transition point of $BaTiO_3$ is very sensitive to outside influences. This appears from the following thermodynamic considerations.

We start from the formula for the Gibbs free energy G:

$$G = U - TS + pV, \quad \ldots\ldots \quad (1)$$

where U is the energy, S the entropy, p the pressure and V the molecular volume [5]). Since $T\mathrm{d}S = \mathrm{d}U + p\mathrm{d}V$ it follows that:

$$\mathrm{d}G = -S\mathrm{d}T + V\mathrm{d}p. \quad \ldots\ldots \quad (2)$$

In *fig. 2* the variation of the Gibbs free energy of the cubic phase (G_c) and of the tetragonal phase (G_t) are schematically represented in the neighbourhood of the transition point. Since

$$\left(\frac{\partial G}{\partial T}\right)_p = -S,$$

the slope of these lines is determined by the entropy S of the phases; the angle between the lines is determined by the difference in entropy $\Delta S = S_c - S_t$, and thus by the heat of transition

$$\Delta H = T_{ct}\,\Delta S,$$

[3]) W. Merz, Phys. Rev. 75, 687, 1949.
[4]) cf. H. Kniepkamp and W. Heywang, Z. ang. Physik 6, 385-390, 1954 (No. 9).
[5]) E.g. see J. D. Fast, Entropy in science and technology, Philips tech. Rev. 16, 258-269 (esp. page 267), 298-308 and 321-332, 1954/55 (Nos. 9, 10 and 11).

where T_{ct} is the transition temperature between the tetragonal and cubic phase.

If, by some means or another, the values of G_t and G_c are changed by amounts dG_t and dG_c, which are not equal:

$$dG_c - dG_t = \Delta dG \neq 0,$$

it follows from simple geometrical considerations that the transition temperature is shifted according to the formula:

$$dT_{ct} = \frac{\Delta dG}{\Delta S} = T_{ct}\frac{\Delta dG}{\Delta H}. \quad . \quad . \quad . \quad (3)$$

For $T_{ct} = 396$ °K and $\Delta H = 220$ joules/mol (according to the measurements of J. Volger) we find:

$$dT = 1.8\,\Delta dG\,\frac{°C}{J/mol}. \quad . \quad . \quad . \quad . \quad (4)$$

For some external influences ΔdG is simple to calculate. Consider, for example, the effect of external pressure. If the external pressure increases by an amount dp, at constant temperature T,

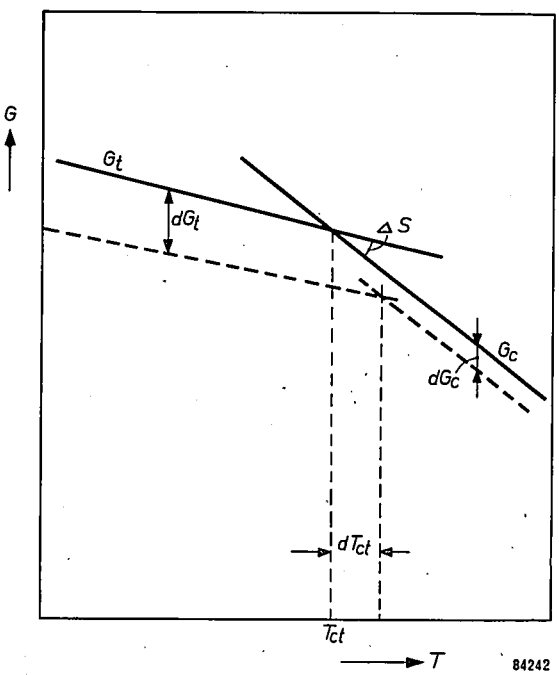


Fig. 2. Variation of the Gibbs free energy G with the temperature T (schematic) for cubic $BaTiO_3$ (G_c) and for tetragonal $BaTiO_3$ (G_t). The point of intersection of these lines gives the transition temperature T_{ct}. This point is displaced if, for any reason, the lines G_c and G_t are moved over different distances dG_c and dG_t.

equation (2) becomes $dG = V dp$, and we obtain, by substitution in eq. (3), the well-known equation of Clapeyron:

$$dT_{ct} = T_{ct}\frac{\Delta V\,dp}{\Delta H}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

The most important component of ΔV is the spontaneous change of volume at the transition point, $V_c - V_t$.

Table I. Dimensions of the elementary cell of $BaTiO_3$ at the transition point ($T = 123$ °C).

	a in Å	b in Å	c in Å	v in Å^3	V in cm^3
Cubic (c)	4.006	4.006	4.006	64.288	38.701
Tetragonal (t)	4.000	4.000	4.021	64.336	38.730
Difference (c − t)	0.006	0.006	−0.015	−0.048	−0.029

In *Table I*, ΔV is calculated from the formula $V = N_{Av}\,v$, where v is the volume of the elementary cell ($v = abc$) [6]) and N_{Av} is Avogadro's constant (6.02×10^{23}/mole). With the help of these figures we find for a uniform pressure p:

$$\Delta T_{ct} = -5.2 \times 10^{-8}\,p \text{ °C per newton/m}^2. \quad . \quad . \quad (6)$$

or (since 1 atm = 101,325 newtons/m^2), −5.3 °C per 1000 atm, which agrees fairly well with the value found experimentally by Merz [7]) for single crystals, viz. −5.8 °C per 1000 atm.

If the external pressure does not act isotropically, account must be taken of the fact that in the transition from the cubic to the tretagonal phase, the orientation of the tetragonal crystals is preferentially influenced by the direction of the external force. Equation (2), which only takes account of pressure acting in all directions, is not valid in such a case. Under uni-directional pressure a tetragonal crystal occurs with the c-axis perpendicular to the direction of the pressure. In such a case T_{ct} increases as a result of the pressure so that:

$$\Delta T_{ct} = +11.2 \times 10^{-8}\,p \text{ °C per newton/m}^2. \quad (7)$$

If the crystal be subjected to a tension, the tetragonal form is orientated so that the c-axis lies in the direction of the force, and T_{ct} also increases [8]):

$$\Delta T_{ct} = +27.7 \times 10^{-8}\,p \text{ °C per newton/m}^2. \quad . \quad . \quad (8)$$

For a force of 1000 kg/cm^2 (approx. 10^8 newtons/m^2), we find ΔT to be 11 and 27 °C respectively.

It is thus clear that internal stresses, which may

---

[6]) For this purpose use was made of unpublished measurements by M. G. Harwood, Material Research Lab., Philips Electrical Ltd., Mitcham, England.

[7]) W. Merz, Phys. Rev. **78**, 52-54, 1950. Compare equation (6) with that for the equilibrium of ice and water, for which $\Delta T = -8.8 \times 10^{-8}\,p$ °C/(N/m^2). Despite the fact that ΔV is much larger (viz. 2 cm^3 compared with about 0.03 for $BaTiO_3$) $\Delta T/p$ is of the same order of magnitude because ΔH is much greater for water-ice(about 6000 J/mol against 220 J/mol for BaTiO3).

[8]) See also P. W. Forsbergh, Phys. Rev. **93**, 686-692, 1954, who has found experimentally a quadratic effect for a pressure applied in two directions.

arise from non-uniform cooling and which may be of the order of 1000 kg/cm², can easily result in a spread of some tens of degrees in the transition temperature. It has been shown experimentally that, particularly in fine crystalline material, the spread may range from 100 °C to 170 °C.

The Gibbs free energy and with it, the transition point can be influenced not only by internal stresses but also by other causes, which do not appear in equation (2). As an example of such effects, we will consider the difference in free surface energy between the tetragonal and cubic phases. It is obvious that this effect will be most strongly felt in fine crystalline material, where the total internal surface is greatest. Such material is generally obtained by sintering at as low a temperature as possible. It is then probable that the crystal growth is so incomplete that many faults occur in the lattice, which may also make a contribution to the Gibbs free energy.

From the above it is clear that by means of a combination of several measures it is possible to alter radically the $\varepsilon - T$ curve of the pure material. Further variation in the form of this curve can be obtained by making use of mixed crystals, which we shall now discuss further.

Mixed crystals of $BaTiO_3$ with related compounds

Barium titanate forms mixed crystals with many other substances which, like itself, crystallize in the perovskite structure. Even quite small quantities of these substances affect the ε-T curve appreciably. The effect of $SrTiO_3$ has been thoroughly studied (see I). This substance causes all three transition points to shift to lower temperatures. By the addition of $SrTiO_3$, the peak can be displaced from 123 °C to a temperature in the range in which the capacitor is to be used. For mixed crystals containing 75% $BaTiO_3$ and 25% $SrTiO_3$ for example, the main peak lies at about 50 °C.

The same is found when the Ti-ion, instead of the Ba-ion is replaced, for example, by mixing with $BaZrO_3$ or $BaSnO_3$. The zirconates and stannates shift the main peak to lower temperatures.

In such mixed crystals of $BaTiO_3$ with zirconates and stannates and other titanates, the volume change at the transition point is smaller and hence the effect of mechanical stresses is smaller. The spread at the transition point is therefore smaller and the peak in the ε-T curve sharper[9]), at least in the case of well sintered, somewhat coarse crystalline specimens.

[9]) G. A. Smolenski, M. A. Karamyshev and K. I. Rozgachev, Doklady Akademii Nauk. S.S.R. **79**, 53-56, 1951.

The desired widening and lowering of the peak must therefore be obtained by other means, such as the effect of crystal size (small crystals are favourable) and of lattice faults and inhomogeneities.

Sometimes, in the case of mixed crystals, special circumstances help in achieving the desired widening of the peak. The second transition point, which for pure $BaTiO_3$ lies at 10 °C, is shifted to higher temperatures by the addition of $BaZrO_3$ or $BaSnO_3$. For mixtures in certain proportions the first and second transition points coincide (see *fig. 3* and *4*). Mixed crystals for which the two transition points lie close together, show a broader, lower peak in the ε-T curve.

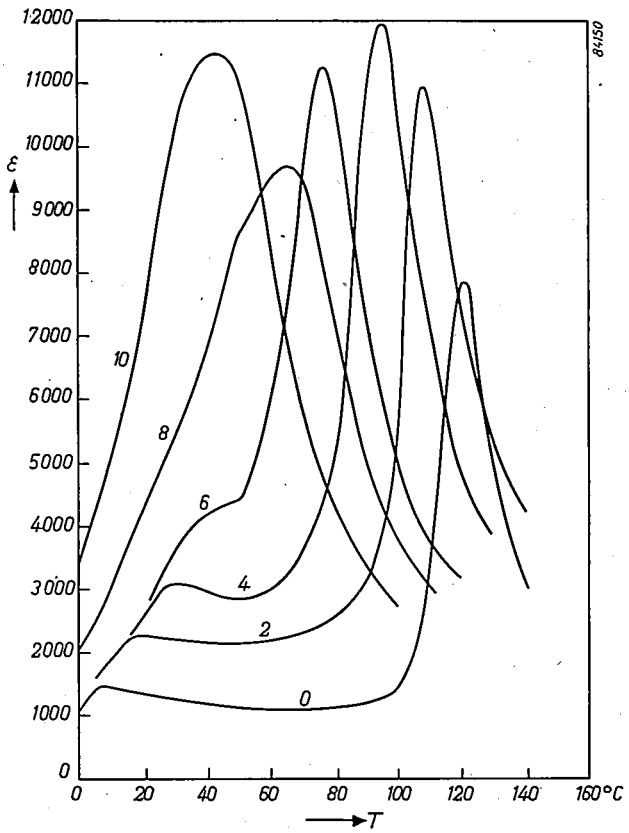


Fig. 3. ε-T curves for mixed crystals of $BaTiO_3$ and $BaSnO_3$. The figures beside the lines give the percentage of $BaSnO_3$ present. The mixed crystals have a higher ε maximum than pure $BaTiO_3$. For specimens where the first and second maxima lie close together (see fig. 4) a disturbance occurs which makes this effect less obvious.

Often when making dielectric material consisting of mixed crystals, a wide maximum in the ε-T curve is obtained without taking any special measures to achieve this. In manufacturing these products, the basic ingredients are powdered carbonates and oxides, such as $BaCO_3$, $SrCO_3$, TiO_2 and ZrO_2. These materials have to react with each other in the solid state, for which purpose diffusion of the ions must

take place. At low firing temperatures both the chemical reaction and the diffusion take place slowly, so that the formation of crystals is incomplete, resulting in small crystals, inhomogeneities and consequent widening of the ε-peak. This applies particularly to the case where ZrO_2 is present.

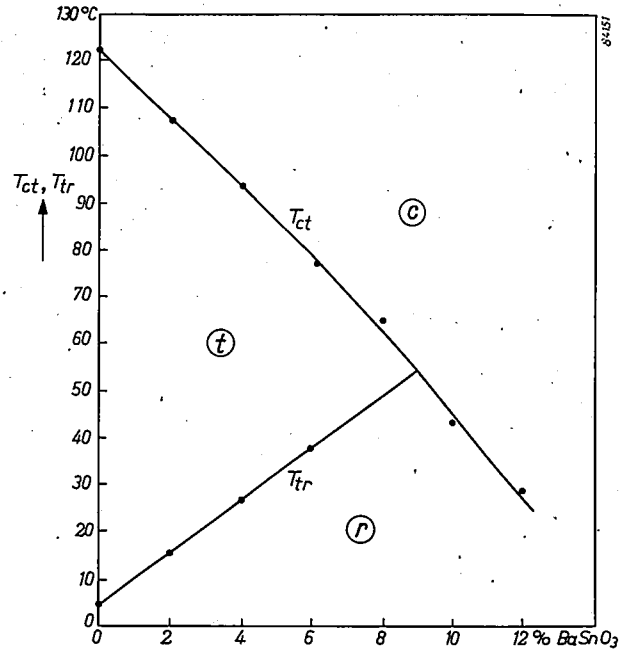


Fig. 4. The effect of $BaSnO_3$ on the transition points of $BaTiO_3$ (an almost similar figure applies for additions of $BaZrO_3$) c = region of cubic structure, t = tetragonal, r = rhombic.

Influence of a second phase

Another possible means of reducing the temperature dependence of the dielectric constant of the materials under discussion, is to dilute them with a substance which shows little or no variation of the dielectric constant with the temperature. A mixture of ceramic dielectrics behaves roughly according to the Lichtenecker rule of mixtures [10]):

$$\log \varepsilon = v_1 \log \varepsilon_1 + v_2 \log \varepsilon_2, \quad . \quad . \quad . \quad . \quad (9)$$

where v_1 and v_2 are the volume fractions of components with dielectric constants ε_1 and ε_2.

If the dielectric constant of the first material has the values ε_{1a} and ε_{1b} at temperatures T_a and T_b respectively, and that of the second ε_{2a} and ε_{2b}, we then find from the rule:

$$\log \varepsilon_a = v_1 \log \varepsilon_{1a} + v_2 \log \varepsilon_{2a},$$
$$\log \varepsilon_b = v_1 \log \varepsilon_{1b} + v_2 \log \varepsilon_{2b},$$
$$\log (\varepsilon_a/\varepsilon_b) = v_1 \log (\varepsilon_{1a}/\varepsilon_{1b}) + v_2 \log (\varepsilon_{2a}/\varepsilon_{2b}). \quad (10)$$

[10]) K. Lichtenecker, Phys. Z. 27, 115-158, 1926.

Now if ε for the second phase shows only a small dependence on T, so that $\varepsilon_{2a}/\varepsilon_{2b}$ differs only slightly from unity, we see that the mixing has resulted in the logarithm of $\varepsilon_a/\varepsilon_b$ being multiplied by a factor v_1 (<1), which is independent the value of ε_2.

It is seen from (9) that in order to achieve a large final value of ε, it is desirable to choose as the second phase some material which itself has a large value of ε. From this point of view it is fortunate that $CaTiO_3$ and $BaTiO_3$, although both crystallizing in the perovskite structure, differ so greatly in the dimensions of their elementary cells that they do not form mixed crystals in all proportions; in certain particular proportions the mixture splits into two phases, one rich and one poor in calcium. Now $CaTiO_3$ itself is a material with a fairly large value of ε ($\varepsilon = 130$ at room temperature); the ε-T curve exhibits no peaks. The temperature coefficient is negative and constant over a large range of temperatures; its value is given by:

$$\frac{1}{\varepsilon}\frac{d\varepsilon}{dT} = -1500 \times 10^{-6} (°K)^{-1}.$$

The Ca-rich phase of mixed crystals of Ca- and Ba-titanate also has a constant negative temperature coefficient. The Ca-rich phase is therefore the ideal material to improve the ε-T curve of the Ba-rich phase. This is still valid if a third substance of perovskite structure is added to shift the peak to a lower temperature. Furthermore, the remaining peak can, of course, be made wider and lower by firing at low temperatures to encourage inhomogeneity. An added advantage here is that the miscibility of Ca- and Ba-titanates decreases with decreasing temperature. During the cooling process after firing, therefore, a segregation takes place which does not lead to an equilibrium state. This produces extra inhomogeneity and internal stresses, which make the maximum of the ε-T curve still broader.

We shall now discuss two systems in which the above-mentioned segregation plays a part.

1) The system $BaTiO_3$-$SrTiO_3$-$CaTiO_3$. This system with five components (Ba, Sr, Ca, Ti, O) is in fact a ternary system; it is restricted to compounds $MTiO_3$ with perovskite structure (M = Ba, Sr or Ca); all compositions can thus be represented graphically in a triangle with the three titanates at the corners (*fig. 5*) and each point is defined by two coordinates.

$SrTiO_3$ forms mixed crystals in all proportions with both $BaTiO_3$ and $CaTiO_3$. It crystallizes cubically with a lattice constant (3.90 Å) lying between

the average values for all crystal axes of $CaTiO_3$ (3.82 Å) and $BaTiO_3$ (4.00 Å), both of which deviate slightly from cubic crystal structure. In such a case the miscibility of the original components (Ca and Ba here) usually increases as more of a third component (here Sr) is added. This in fact occurs in the present case, as may be seen from fig. 5, which shows the segregation area (two-phase area). In this region a number of straight lines have been drawn approximately parallel to the $BaTiO_3$-$CaTiO_3$ side. These are the so-called co-existence lines or connodes. The ends of each connode, lying on the boundary of the two-phase area, give the compositions of the two-phases into which a mixture, represented by a point on the connode, separates out.

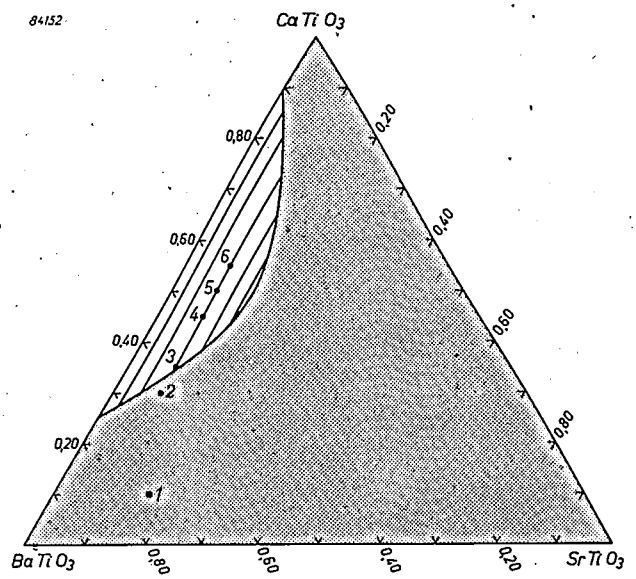


Fig. 5. Phase diagram at 1350 °C of the combination $BaTiO_3$-$SrTiO_3$-$CaTiO_3$. $BaTiO_3$ and $CaTiO_3$ are only partly miscible: random mixtures of these two materials separate out into two equilibrium phases, viz. mixed crystals of 26% $CaTiO_3$ + 74% $BaTiO_3$ and mixed crystals of 89% $CaTiO_3$ + 11% $BaTiO_3$. The miscibility is improved by the addition of $SrTiO_3$. In the segregation region, the straight lines (connodes) join the mixed crystals which are in equilibrium with each other.

The quantities of the two phases are inversely proportional to the distances of the point from the two ends of the line. A number of the mixtures studied are plotted in fig. 5, and in *fig. 6* the corresponding $\varepsilon - T$ curves are shown.

2) The system $BaTiO_3$-$CaTiO_3$-$BaZrO_3$-$CaZrO_3$. This can also be represented in a plane diagram, because only compounds with the formula ABO_3 occur, all of which crystallize in the perovskite structure. The number of independent variables is thus once more two: the ratio Zr/Ti can be expressed by the fraction $x/(1-x)$ and the ratio Ca/Ba by the fraction $y/1-y)$. If x and y are plotted as perpendicular coordinates, then all possible com-

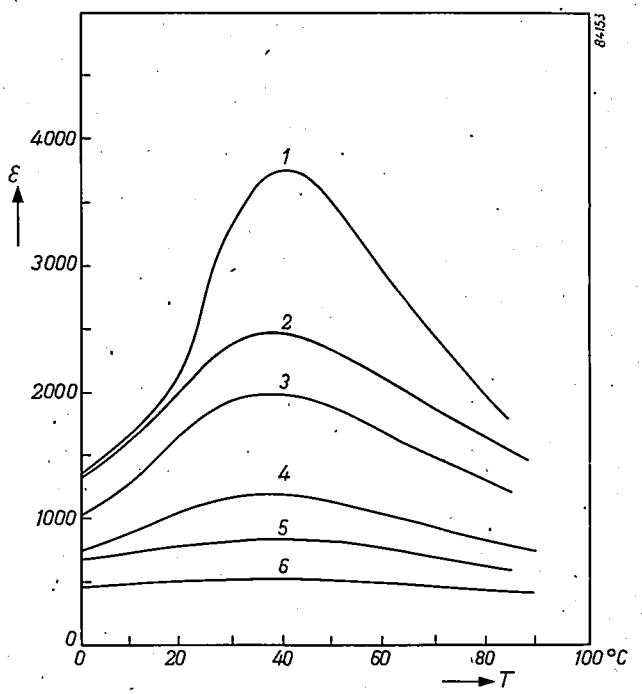


Fig. 6. ε-T curves for sintered specimens, composed of $BaTiO_3$, $SrTiO_3$ and $CaTiO_3$, corresponding to the compositions *1* to *6* in fig. 5. Compositions *3* to *6* lie in the two-phase region and in this order contain increasing amounts of the phase rich in $CaTiO_3$, which makes the ε-T curves increasingly flat.

positions can be found within the square $0<x<1$, $0<y<1$. The position of the ε-peak is mainly determined by the ratio Zr/Ti, thus by the parameter x. *Fig. 7* shows the segregation area with the connodes and *fig. 8* the ε-T curves for the four specimens whose compositions are defined in fig 7.

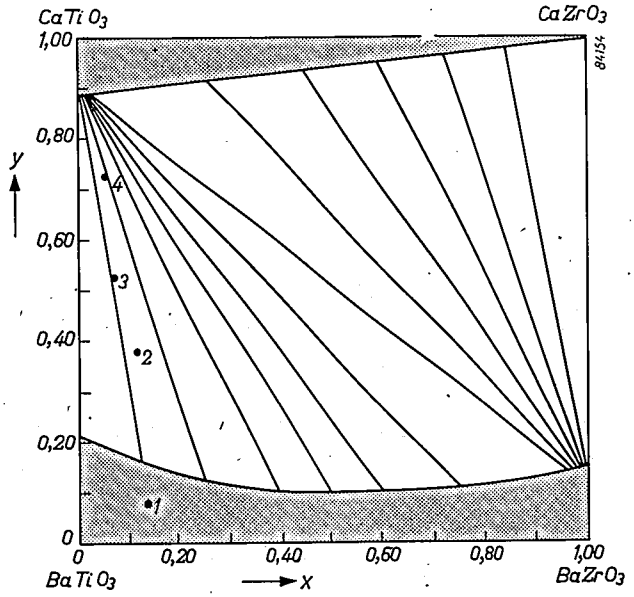


Fig. 7. Phase diagram at 1300 °C of the system $BaTiO_3$-$CaTiO_3$-$CaZrO_3$-$BaZrO_3$ with two perovskite phases (shaded areas). The ratio Zr/Ti is $x/(1-x)$, the ratio Ca/Ba is $y/(1-y)$. The straight lines in the segregation region are the connodes. From the direction of the connodes it can be seen that the combination $BaZrO_3$ + $CaTiO_3$ is more stable than the combination $BaTiO_3$ + $CaZrO_3$.

This system has also been described by Mc Quarrie and Behnke [11]). Our results differ somewhat from their's, especially in the area in the neighbourhood of the $BaTiO_3$ corner, which is so important for practical dielectrics and which we therefore studied more particularly.

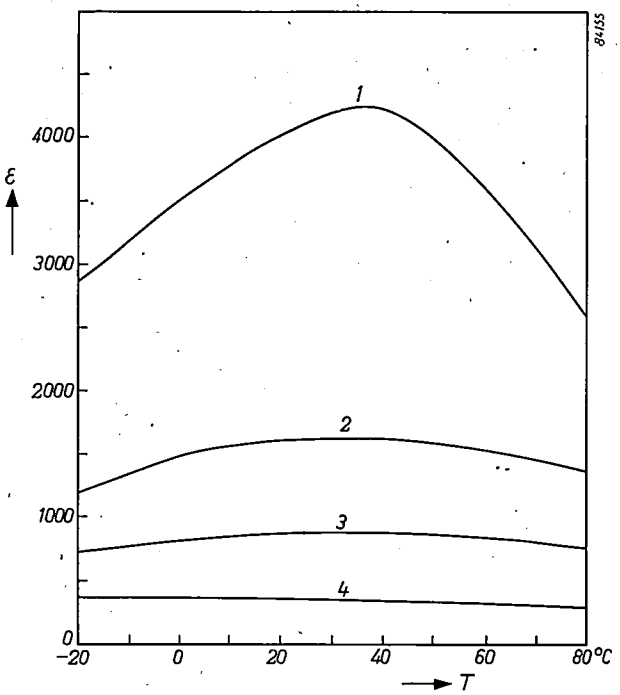


Fig. 8. ε-T curves for sintered specimens prepared from $BaCO_3$, $CaCO_3$, TiO_2 and ZrO_2, corresponding to the compositions *1* to *4* in fig. 7. Here too, the strong flattening action of the increasing proportions of the phase rich in $CaTiO_3$ is demonstrated.

As was evident from the work of the above-mentioned authors and can be seen from fig. 7, the connodes lie in such a way that the system practically falls into two ternary systems, viz. the systems $CaTiO_3$-$BaTiO_3$-$BaZrO_3$ and $CaTiO_3$-$CaZrO_3$-$BaZrO_3$.

The behaviour expressed in fig. 7 can be more easily understood if the system (Ba, Ca) (Ti, Zr)O_3 is compared with a simpler system which behaves in very much the same way, viz. that of Fe-Ni-FeO-NiO. This system is also determined by two parameters, viz. the ratio Ni/Fe and the ratio oxide/metal. For the sake of simplicity, one can assume that no metal dissolves in the oxide and no oxide in the metal. The one-phase areas thus shrink to the sides Fe-Ni ($y=0$) and FeO-NiO ($y = 1$) (see *fig. 9*). If a mixture of Fe and Ni is partially oxidized, mainly FeO will be formed because of the greater affinity of Fe for O. The two-phase lines (connodes) thus become straight lines running from the FeO corner ($x \approx 0$, $y = 1$) to points on the FeNi line ($y = 0$). On the other hand, if a mixed oxide crystal (Fe,Ni)O is partially reduced, mainly Ni will be formed and thus the metal in the diagram is represented by a point which coincides nearly with the Ni corner ($y = 0$, $x \approx 1$). The rest of the diagram is filled with lines joining the Ni corner to points on the NiO,FeO line ($y = 1$). If we turn our attention to the centre of the square ($x = y = 0.5$), we see that for this net composition, the system separates out almost entirely into FeO + Ni and not into the combination NiO + Fe, which corresponds to the other diagonal. This is related to the fact that the reaction $NiO + Fe \rightleftarrows FeO + Ni + Q$ has a positive heat of reaction Q i.e. it is exothermic. The heat of reaction mainly determines the difference in Gibbs free energy because the differences in entropy in this case can be neglected, and this makes the Gibbs free energy of FeO + Ni smaller than that of NiO + Fe, so that the former is more stable.

In the same way the diagonal in fig. 7, running approximately from $CaTiO_3$ to $BaZrO_3$ and dividing the two "ternary" systems, indicates that the reaction

$$BaTiO_3 + CaZrO_3 \rightleftarrows CaTiO_3 + BaZrO_3 + Q$$

is exothermic ($Q > 0$), so that the right-hand combination is the more stable.

A similar behaviour to that described above for (Ba,Ca) (Ti,Zr)O_3 is found for the system (Ba,Ca) (Ti,Sn)O_3.

If one tries to prepare substances with the composition (Ba,Ca) (Ti,Ce)O_3, a two-phase (even partially three-phase) system arises, in which the phase consisting chiefly of $CaCeO_3$ has a small Gibbs free energy. This is to be seen in *fig. 10*, where the diagonal which divides the two nearly ternary systems, now runs from $BaTiO_3$ to $CaCeO_3$, in contrast to fig. 7.

In fig. 10 the substance $CaCeO_3$ is given as $CaO.CeO_2$. The replacement of the Ti-ion by another random tetravalent ion does not, in fact, always lead to compounds of the perovskite type. If one attempts

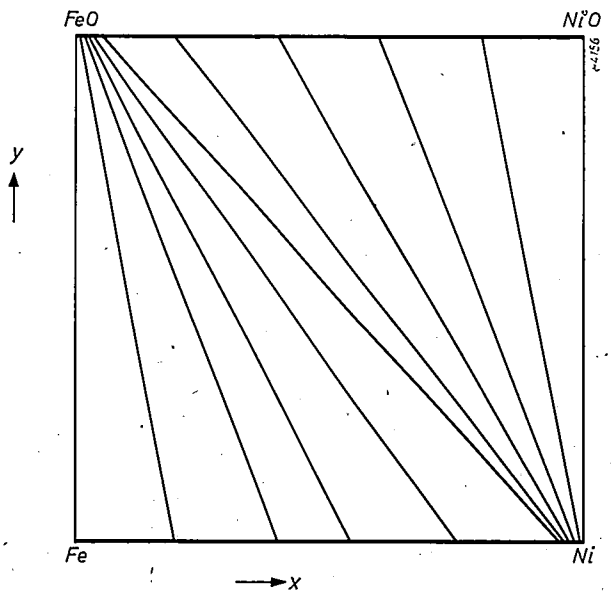


Fig. 9. Phase diagram for the system Fe-Ni-FeO-NiO (schematic). This diagram serves to elucidate fig. 7. The connodes have been drawn in order to demonstrate that the combination Ni + FeO is more stable than the combination Fe + NiO.

[11]) M. McQuarrie and F. W. Behnke, J. Amer Ceram. Soc. **37**, 539-543, 1954.

to prepare $CaCeO_3$, it appears that this does not crystallize in the perovskite structure, but produces a solution of CaO in CeO_2, which, like CeO_2 itself, shows the same structure as fluorite [12]). Furthermore, the cerates, including $CaO.CeO_2$, have a fairly small value of ε ($\varepsilon \approx 35$). Thus the addition of Ce to a (Ba,Ca) titanate has a less favourable result than that of Zr or Sn.

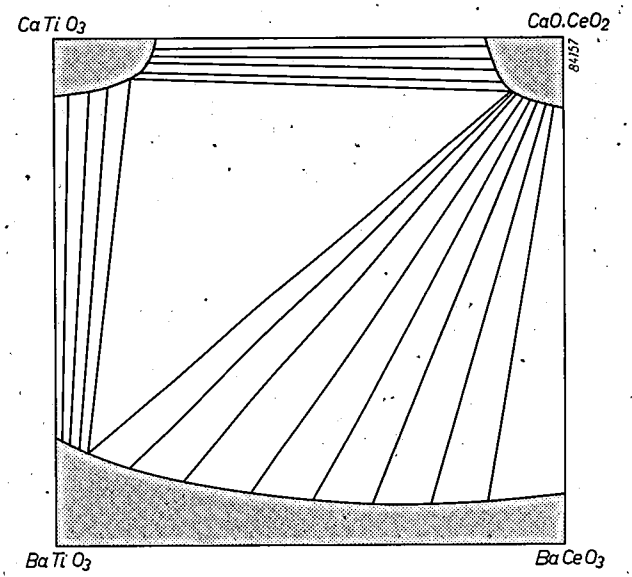


Fig. 10. Phase diagram for the system $BaTiO_3$-$CaTiO_3$-CaO.CeO_2-$BaCeO_3$ (schematic). In this diagram, besides the two perovskite phases, a third phase with fluorite structure, occurs in the $CaO.CeO_2$ corner. The Gibbs free energy is so small there that the disposition of the connodes in this diagram differs from that in fig. 7.

The stability of materials with perovskite structure

The above considerations bring us to the question of the stability of perovskite compounds. The value of the Gibbs free energy (which determines the stability) can be related to the degree to which the ions do or do not fit into the perovskite lattice.

The most important contribution to the Gibbs free energy is, in fact, made by the electrostatic component of the lattice energy (Coulomb energy). If, however, one compares these contributions for the combination $BaTiO_3 + CaZrO_3$ with those for the combination $CaTiO_3 + BaZrO_3$, it appears that when calculated for perfect perovskite structure, they are almost equal. The chief contribution te the lattice energy is A/a, where A is the Madelung constant and a the cell dimension for a given component. For the above substances a has the values 4.00, 4.00, 3.92 and 4.18 Å respectively and we find for the first-mentioned combination

$$\frac{1}{4.00} + \frac{1}{4.00} = 0.500,$$

and for the second combination,

$$\frac{1}{3.82} + \frac{1}{4.18} = 0.501.$$

[12]) H. Gränicher, Helv. Phys. Acta 24, 619-622, 1951.

Under these circumstances, secondary influences, such as the fit of the ions in the lattice, become of importance in determining the difference in Gibbs free energy.

A measure of the fit of the ions in the lattice is Goldschmidt's tolerance factor t (see I):

$$t = \frac{R_A + R_O}{(R_B + R_O)\sqrt{2}}. \quad \ldots\ldots (11)$$

Here R_A, R_B and R_O represent the radii of the large metal ions (A), small metal ions (B) and oxygen ions (O), all imagined to be spherical. Compositions which fit well, like $SrTiO_3$, have a value $t \approx 1$. Ba compounds also fit fairly well. Ca compounds show greater deviations, as can be seen from *Table II*.

Table II. Tolerance factor t for various perovskite compounds, calculated with the help of ion radii corrected according to Zachariasen.

$CaTiO_3$	0.94	$BaTiO_3$	1.07
$CaSnO_3$	0.91	$BaSnO_3$	1.025
$CaZrO_3$	0.87	$BaZrO_3$	0.98
"$CaCeO_3$"	0.82	$BaCeO_3$	0.92

If one takes the value of $|t - 1|$ as a measure of the Gibbs free energy, then the most stable combination in the (Ba, Ca) (Ti, Zr)O_3 system will be found to be $BaZrO_3 + CaTiO_3$. In the (Ba,Ca) (Ti, Ce)O_3 system discussed above, the combination $BaCeO_3 + CaTiO_3$ would be the more stable according to the table. But the "compound" $CaO.CeO_2$ which, as stated above, possesses a fluorite structure, has a so much smaller Gibbs free energy than the hypothetical $CaCeO_3$ that as a result the other combination, $BaTiO_3 + CaO.CeO_2$, is the more stable.

Effect of foreign admixtures

Up to now we have spoken mainly about influencing the behaviour of $BaTiO_3$ by formation of mixed crystals with substances which also have a perovskite structure. It has also become apparent that many other substances will dissolve in small quantities in $BaTiO_3$ and sometimes have a marked effect on the form of the $\varepsilon - T$ curve. It is not yet known with certainty whether their influence is due to their hinderance of the crystal growth or to some specific influence in the material itself. In either case, the transition temperatures are not shifted (or are shifted only very slightly); the added quantities are too small for that. Probably the principal action is upon the crystal growth during the sintering process and in hindering the transition from the cubic to the tetragonal phases. This change in crystal structure occurs by a whole row of ions of one kind moving with respect to a row of another kind. If the crystal structure is now disturbed by

solutions of foreign materials which do not fit into the perovskite structure, the change can be wholly or partially hindered. A favourable additional effect is that foreign admixtures cause the dielectric hysteresis loop to be considerably flattened, so that the non-linear properties which pure $BaTiO_3$ exhibits strongly below 123 °C partly disappear.

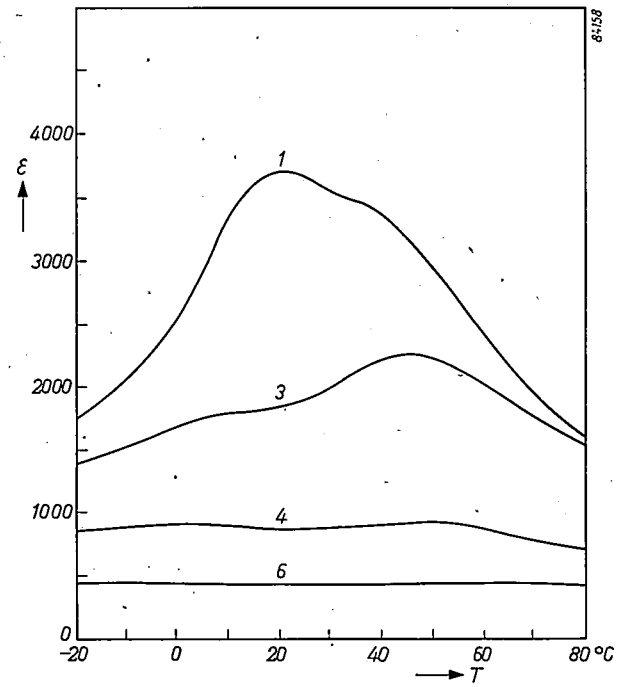


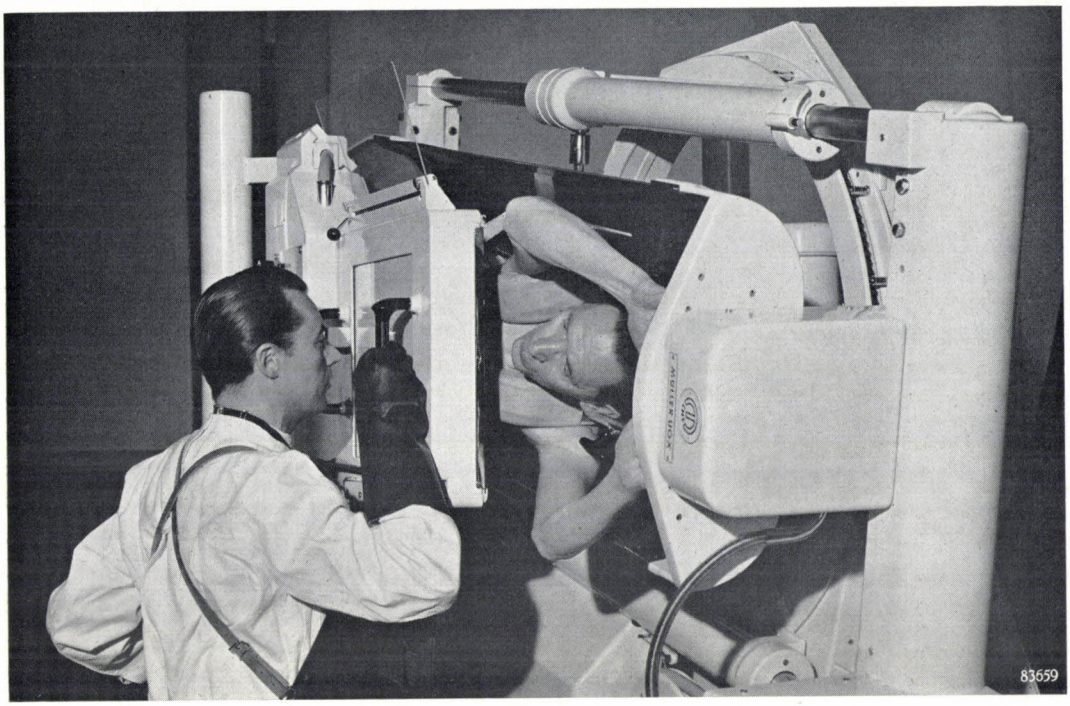
Fig. 11. ε-T curves for a series of specimens in the $BaTiO_3$-$SrTiO_3$-$CaTiO_3$ system to which about 10% Mg_2TiO_4 has been added. The compositions (calculated without the Mg_2TiO_4) of 1,3,4 and 6 correspond approximately with the similarly numbered points in fig. 5. The effect of the additive is to make the curves flatter than those in fig. 6.

An objection, however, is that some impurities only temporarily hinder the crystallographic transition, i.e. they only slow it down. This results in prolonged after-effects, which are observed as a prolonged continuous decrease in the dielectric constant in material which has been cooled from a temperature above the transition point to room temperature.

One substance which has been successfully used to improve the ε-T curve is magnesium oxide, which is added to the $(Ba,Sr,Ca)TiO_3$ system in the form of Mg_2TiO_4. Some ε-T curves for these specimens are shown in *fig. 11*.

The work discussed here has been carried on in co-operation with V.I. Middel of the Material Research Laboratory, Philips Electrical Ltd, Mitcham (England), J. Möllers of the Deutsche Philips Ges.m.b.H. Keramische Werke, Hamburg and G. H. Weber and N. W. Smit of the Ceramics Laboratory at Eindhoven.

Summary. $BaTiO_3$ and related compounds of the perovskite structure show very high values of the dielectric constant. In certain temperature ranges, ε shows peaks corresponding to crystallographic transition points. Since the values of the transition temperatures are sensitive to all kinds of influences (internal stresses, surface phenomena, lattice imperfections) the manner in which the material is fired has a great effect upon the shape of the ε-T curve. This curve can be further influenced by the formation of mixed crystals (displacement of the peaks), by diluting the material with a second phase which has a flatter ε-T curve, and also by the addition of small quantities of foreign substances (such as Mg_2TiO_4) which do not crystallize themselves in perovskite structure but still manage to enter the material either as a solution or between the crystals. Such admixtures help, at the same time, to suppress the dielectric hysteresis.



A NOVEL TYPE OF DIAGNOSTIC X-RAY UNIT

by H. VERSE *) and K. WEIGEL *).

616-073.755.1

In general, the designers of practical medical equipment must be guided by medical opinion. This applies particularly to X-ray diagnostic apparatus, whose special feature is an effective combination of mechanical, electrical and optical components. The apparatus described in the present article is novel in that, to improve the above-mentioned combination, the designers have ventured to suggest to the radiologist an examination procedure which is in some respects new and should be complementary to the older and more familiar methods of examination.

Principles of the apparatus

The problem of mechanical aids for medical X-ray examination was discussed in a recent issue of this Review, in connection with a description of the "Symmetrix" [1]). The "Symmetrix" belongs to a class of apparatus now employed for routine diagnostic examinations in almost all countries. The essential feature of such apparatus is that it enables the examination table to be rotated about a horizontal axis at right angles to the X-ray beam, from the vertical to the horizontal position and beyond it (Trendelenburg position); with the most advanced equipment of this type, say, the "Symmetrix", the table can be rotated even far enough to invert the subject. As the table carrying the patient tilts, the X-ray tube and the viewing system (fluorescent screen and serial cassette) move with it. Moreover, the X-ray tube and viewing system can be moved, within certain limits, parallel to the table and the viewing system can be shifted at right angles to the table surface. The balancing of the movable parts of the "Symmetrix" is described in the article referred to in note [1]).

The "Symmetrix" is a development on conventional lines, based on examination procedures to which radiologists are already fully accustomed. The diagnostic unit to be described here, however, is based on quite different ideas. In this apparatus (type UGX) designed in Hamburg, the X-ray beam remains horizontal and at eye-level (*fig. 1*) [2]). The patient is strapped in a special cradle (which is transparent to X-rays) situated between

*) C. H. F. Müller Aktiengesellschaft, Hamburg.

[1]) J. J. C. Hardenberg and H. W. Dumbrill, The "Symmetrix", a universal table and stand for X-ray diagnosis, Philips tech. Rev. **17**, 112-120, 1955/56 (No. 4).

[2]) H. Verse and K. Weigel, Eine gerätetechnische Betrachtung zur Röntgendiagnostik, Fortschr. Röntgenstrahlen **80**, 520-524, 1954 (No. 4).

the X-ray tube and the fluorescent screen. This cradle, driven by motors and controlled from the position of the radiologist, enables the subject to be shifted and rotated so as to position the required part of the body in the X-ray beam at an angle corresponding to the desired projection [3]), enabling different positions of the internal organs relative to the direction of gravity to be obtained.

The development of this apparatus is based on the general premise that with the advance of X-ray technique it will become more and more difficult to satisfy the requirements imposed by the conventional examination procedures referred to at the beginning of this article. Viewing systems are becoming more and more complex; for example, there is a tendency towards higher tube voltages, necessitating heavier shields on the screen as protection against scattered radiation; an X-ray image intensifier with its viewing system may be used instead of the ordinary fluorescent screen; again, the viewing system may be provided with special attachments, e.g. a camera mount for miniature radiography or an X-ray image intensifier and ciné camera. Owing to the size and weight of such attachments it is no simple matter to enable them to be moved with the table, and the problem of balancing becomes almost insuperable when the possibility of making them interchangeable is also considered. Moreover, there is the important question of the physical strain imposed on the radiologist during an examination. Despite motors to enable the table to be tilted, and the viewing system and associated counterweights to be shifted, almost without effort, the mere act of following the movements of the fluorescent screen or image intensifier imposes a considerable strain on the radiologist.

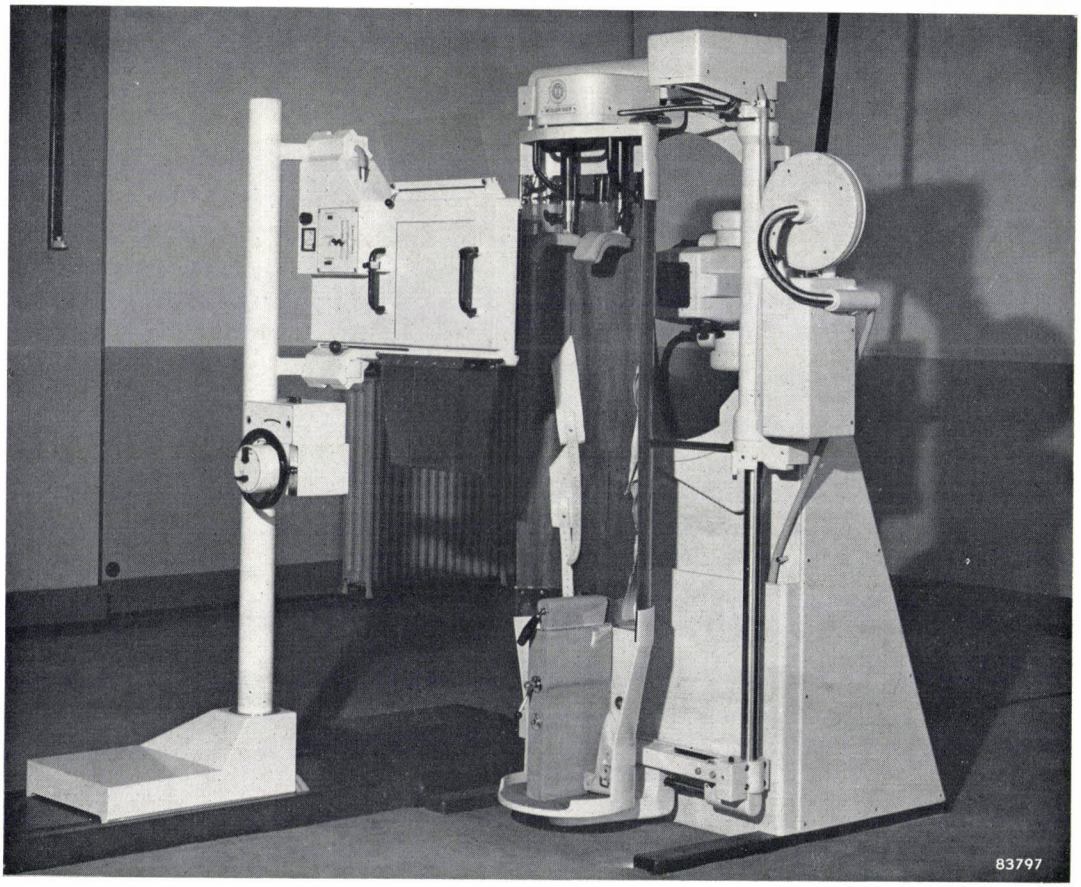
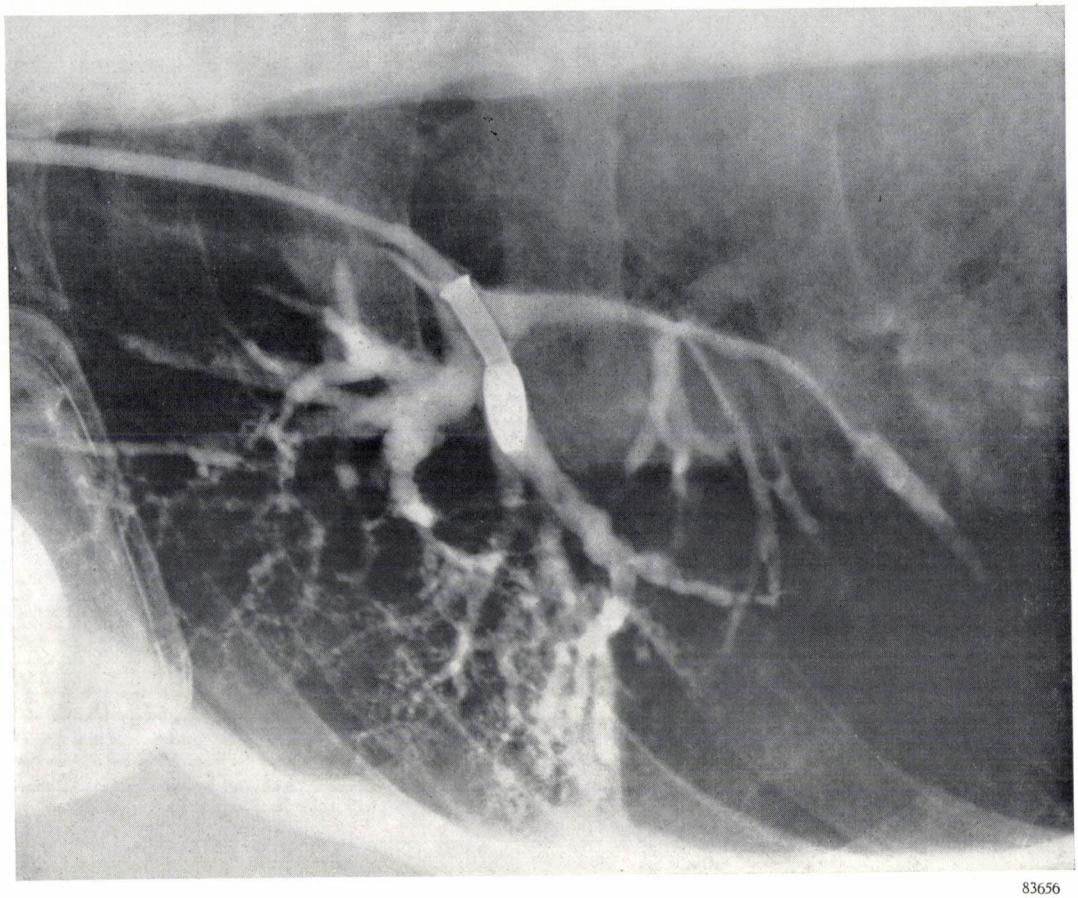


Fig. 1. The "Müller" UGX diagnostic X-ray unit. The column seen on the left of the picture, running on floor rails, carries the viewing system and the control box; a cradle of thin wood, to carry the patient, supported by a frame mounted on the base is seen in the centre of the picture; on the right of the cradle and base is the X-ray tube, on a separate stand, also on floor rails.

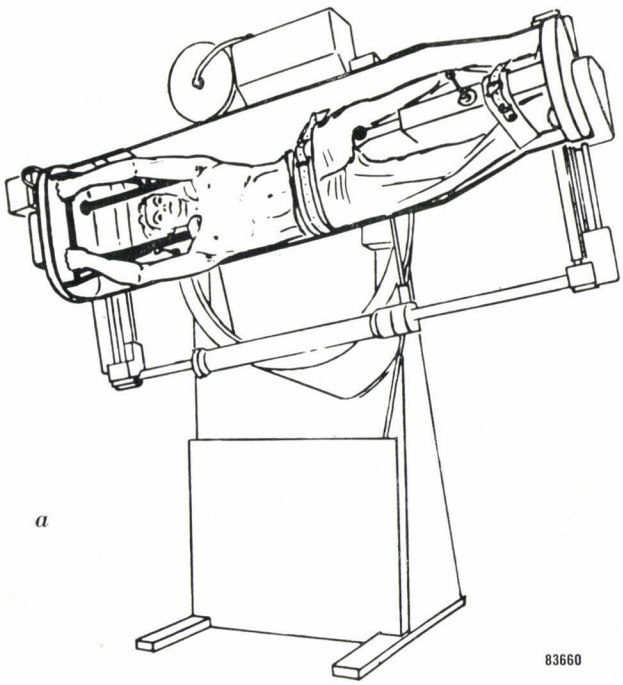
None of these objections apply to the new unit shown in fig. 1. It enables the radiologist to carry out any examination whilst standing comfortably erect and looking straight in front of him. The only physical effort involved is that required to move the viewing system horizontally either towards or away

[3]) The unit considered here is employed mainly for the examination of organs in the trunk of the patient.



b

83656



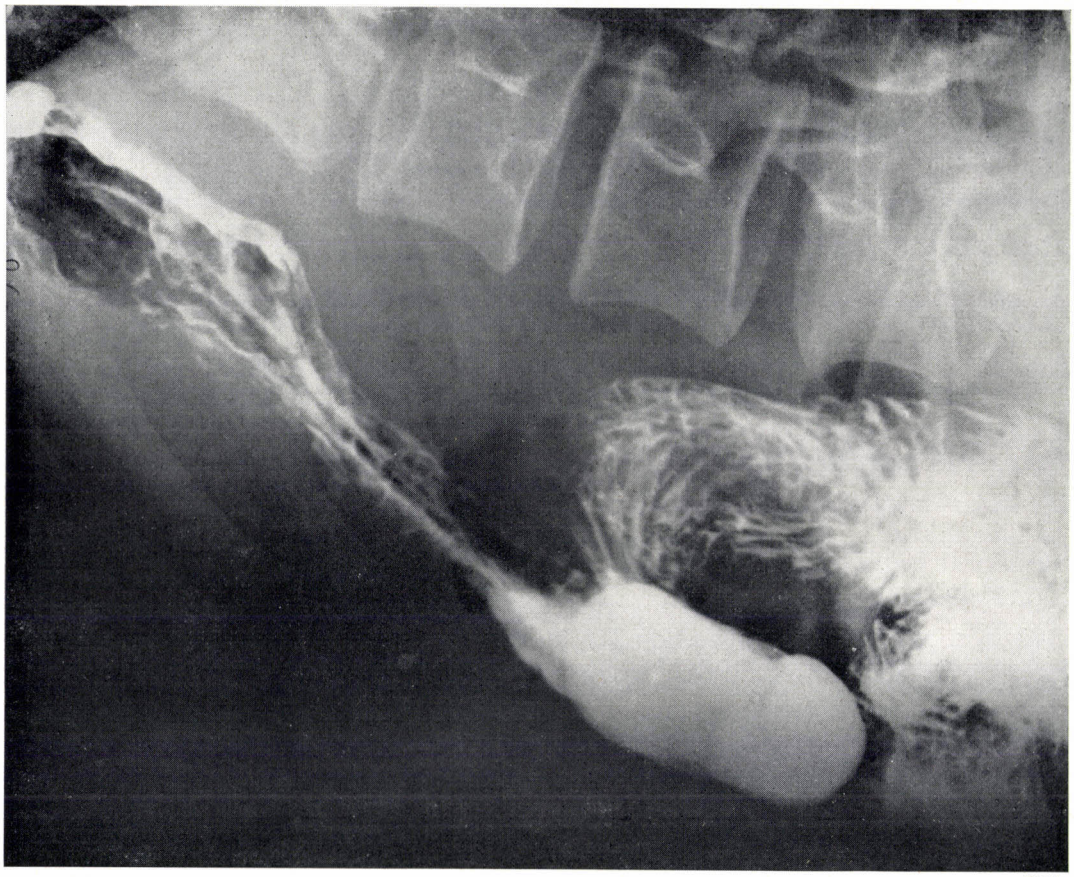
a

83660

Fig. 2. Example of bronchography with the "Müller" UGX. The subject lies on his right side, tilted slightly with his head downwards (*a*). Contrast medium is injected into the bronchia by means of a cannula. The X-ray picture (*b*) shows the nozzle of the cannula and the branch of the trachea leading to the centre lobe of the right lung. Most of the branches of the top lobe, seen on the left of the picture, are filled with contrast medium, enabling their variations to be readily appraised. The branches of the bottom lobe are only partly filled.

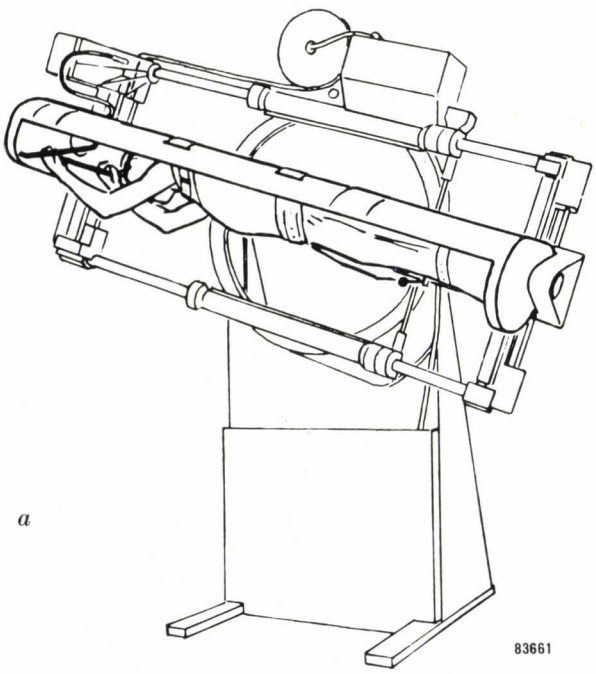
from the patient, a movement which needs no balancing against the force of gravity. Moreover, apparatus so designed eliminates the above-mentioned problems associated with the interchanging of the different attachments for viewing and recording the X-ray image. Also, the radiologist can be given very much better protection against scattered radiation, with less discomfort to himself, because this apparatus is operated almost from one position. On conventional equipment the radiologist must move about for the examination of the patient in various positions and is obliged to carry a considerable weight of lead on his person in order to be roughly equally protected in all positions.

Considering the diagnostic possibilities of the new unit, it will be noted that the possibility of tilting the patient together with the X-ray beam in the conventional manner has been deliberately abandoned. Accordingly, this apparatus cannot be employed instead of the existing equipment to diagnose cases for which the above-mentioned method of examination is indispensable. On the other hand, it gives scope for a considerable extension of the existing possibilities through a series of entirely new patient-positions and projections. Experience



b

83657



a

83661

Fig. 3. Example of stomach radiography with the "Müller" UGX. The subject is virtually prone, but tilted feet downwards at roughly 20° (*a*). This novel position gives an unusually clear view of the front and rear walls of the stomach and of the position of the duodenum. The entrance to the stomach, with the air bubble, always present in the stomach, rising into it, is seen at the top left-hand corner of the radiograph (*b*). Stretching diagonally across the picture towards the bottom right-hand corner is the long middle tract of the stomach. The contrast medium has settled in the lower end. The stomach exit is concealed behind the completely filled tract, but the duodenum, joining this exit, is seen freely projected between the filled tract and the spinal column.

with the apparatus so far has shown that these new facilities may well constitute a very useful adjunct to the conventional examination procedures. Two of the new patient-positions, together with the associated X-ray pictures, are shown in *figures 2* and *3*.

Another point worth mentioning is that the apparatus provides many of the practical facilities of so-called universal X-ray diagnostic units. For example, it can be adjusted for taking pictures at a distance, or to project an enlarged shadow picture, merely by moving the X-ray tube or viewing system horizontally. Similarly, in stomach examinations, a precisely regulated compression of the subject can be effected (see for example [1])). The apparatus is also suitable for planigraphy, as will be explained later.

Design of the unit

The essential feature of the new unit is the cradle and drive, enabling the radiologist to move the subject in the directions indicated by arrows in *fig. 4*. The design of the cradle and drive may

now be explained with the aid of *fig. 5*. The principal components are the cradle itself, in which the patient is placed (or laid), a frame to carry the cradle and enable it to be rotated and translated, and a sturdy, bottom-heavy base carrying a large ring ball-bearing. This bearing — in principle of the same type as conventional ball bearings — comprises two rings, one fixed to the base and the other, driven by an electric motor (in the base) carrying two cylindrical slides for the longitudinal tubes of the frame. The latter, which carries the cradle, can thus be moved across the bearing. This enables the patient to be moved up to 90 cm longitudinally. Power for this movement is supplied by an electric motor, mounted on the rotating ring of the large bearing and geared through an intermediate spindle to two pinions in the end pieces of the cylindrical slides; these pinions engage with racks along the backs of the long tubes of the frame.

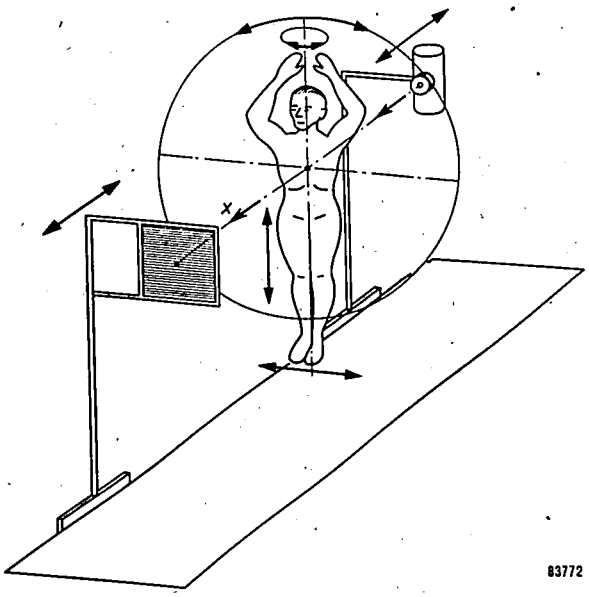


Fig. 4. Degrees of freedom of the patient, the X-ray tube and the viewing system in the "Müller" UGX diagnostic unit. The centre-line of the X-ray beam remains in a fixed horizontal position throughout these movements.

The lateral tubes at the top and bottom of the frame carry rails for head and foot members with bearings to support the cradle. Inside these tubes are threaded spindles, coupled together by a spindle in one of the long tubes and driven by an electric motor attached to the lateral tube at the head of the frame. These spindles engage with threaded bushes on the head and foot members of the cradle, thus driving them at the same speed along the rails. In this way, the patient can be moved 18 cm in

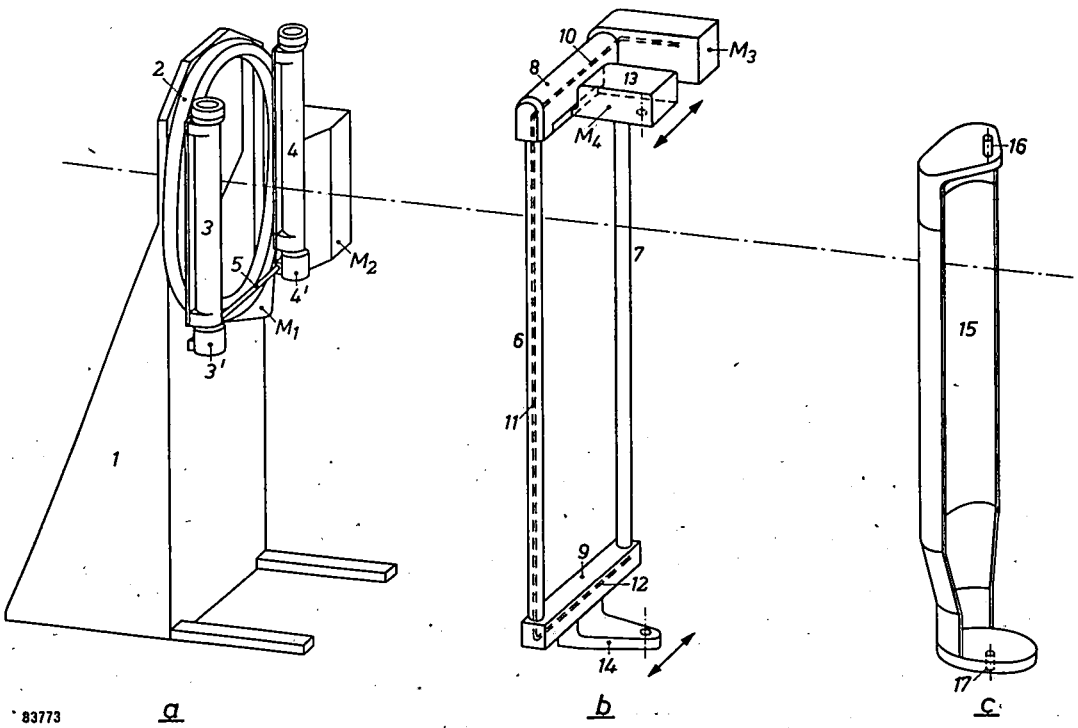


Fig. 5. The various parts of the cradle and drive. *a*) Base *1* of the diagnostic unit, with electric motor M_1 to drive the rotating ring of the large ball bearing *2*; cylindrical slides *3* and *4*, and electric motor M_2 to drive the pinions in the end-pieces *3'* and *4'*; *5* coupling spindle. *b*) Frame, with long tubes *6* and *7* fitted with racks and supported by slides *3* and *4*; *8* and *9* lateral tubes; M_3 electric motor to drive the head and foot members *13* and *14* through spindles *10*, *11* and *12* and thus move the cradle and patient laterally. *c*) Cradle *15*, with short spindles *16* and *17* fitting into bearings in the head and foot members *13* and *14*; spindle *16* is driven by another electric motor (M_4, in *b*) to rotate the patient about his longitudinal axis.

either direction at right angles to his longitudinal axis.

Short spindles at both ends of the cradle fit into thrust bearings in the head and foot slides. An electric motor housed in the head slide drives the head spindle through a reduction gear to rotate the cradle about its longitudinal axis, enabling the patient to be turned a full 360° in either direction. Hand grips and shoulder supports at the head, and a leg support at the foot of the cradle, together with various belts and buckles, hold the patient firmly and give him a feeling of complete safety, even in positions such as those shown in figures 2 and 3.

It is seen that only one of the motors, that is, the one driving the rotary ring of the large bearing, is secured to the base. The motors for the longitudinal and lateral movements, and that to rotate the cradle, are mounted on moving parts of the apparatus. Moreover, most of the limit contacts and change-over switches, employed (see below) to control the different movements, are likewise attached to moving parts. Accordingly, it was necessary to design a special cabling system to connect them. Owing to the number of connections involved it was not practicable to employ slip rings; hence another system, shown in *fig. 6*, was adopted (see also fig. 1); all the leads are enclosed in one, flexible, multi-core cable, running to the furthest terminal on the head slide and held taut, without sagging loops, in every position of the cradle. This cable is divided into three sections (*a*, *b* and *c*). The first section runs from a fixed terminal (*I*) in the base of the apparatus to a point (*II*) near the electric motor on the rotating ring of the large bearing; the second section connects this point to another (*III*) on the lateral tube at the head of the frame, and the third one connects the frame to the head slide of the cradle (*IV*). Since the distance to the head-slide to be covered by the last length of cable (*c*) varies only slightly, it is taken in a simple loop, held taut by a wire spiral fitted round the cable. On the other hand, sections *a* and *b* must span very considerable ranges of movement. The cable slack required for this purpose is taken up by two suitably positioned cable drums (*A* and *B*). Both these sections of cable run from their fixed points (*I* and *II*) to the correspondingd rum, entering through the hollow drum-spindle (*I'* and *II'*). They then cross to a guide on the inner face of the drum and thence span the latter over whatever angle is necessary for the position of the cradle (see fig. 6). Leaving the drum tangentially, they pass to the subsequent fixed points (*II* and *III*). The drums are spring-loaded and so keep the unsupported lengths of cable taut. The cable itself is extra flexible, enabling the cores to move freely

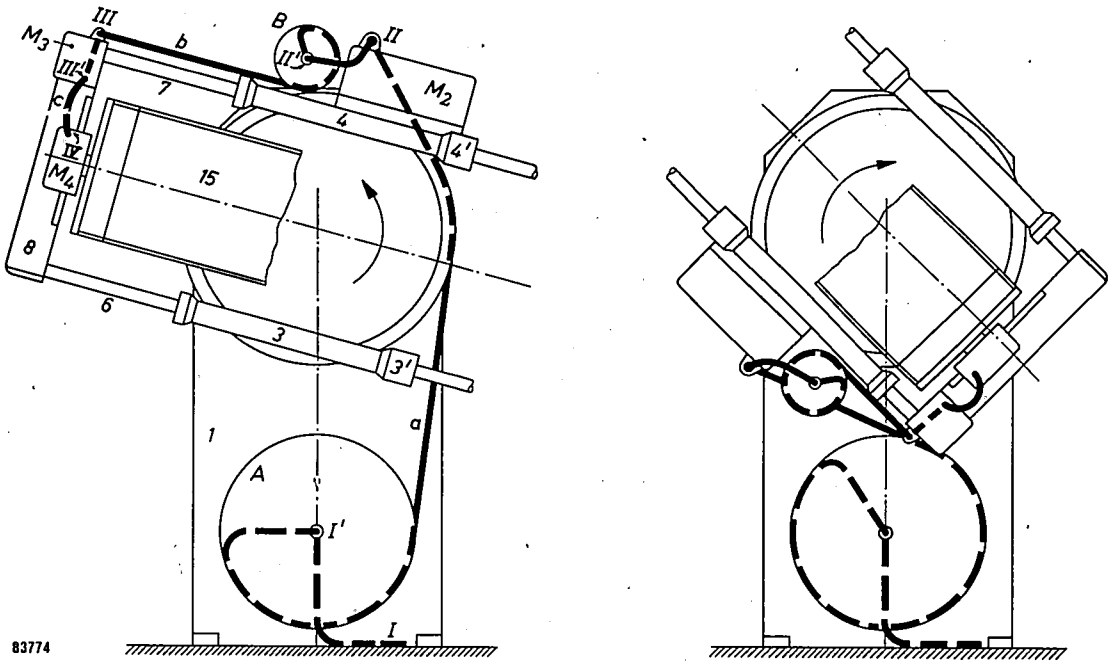


Fig. 6. Connecting cables on the equipment. During all movements, cable sections *a* and *b* are held taut between terminal points *I*, *II* and *III* by spring-loaded cable-drums *A* and *B*. Section *c* is prevented from sagging by a wire spiral fitted round it. Other components numbered as in fig. 5. The diagram on the right shows the change in the configuration of the cable sections when the patient is adjusted to another position.

in relation to one another and to the cable sheath, and so absorb the torsion occurring at the entries to the drums.

Control of cradle and drive

As we have seen, the radiologist controls the cradle from his position near the viewing system. For this purpose, a control box is mounted on the latter, within easy reach of his left hand (fig. 1 and *fig.* 7). On the front of the box is a handwheel to control the rotation of the subject, in either direction, about the axis of the horizontal X-ray beam; in the hub of this wheel are two switches to control the longitudinal and lateral movements of the subject. Another switch, to control the rotation of the subject about his longitudinal axis, is located on top of the control box.

To enable the apparatus to be operated without difficulty in a darkened room, the hub of the handwheel, with the two switches to move the subject longitudinally and laterally, is made to rotate with the subject, that is, in the same direction and with the same angular velocity, by means of a motor housed in the control box. This ensures that the movement of the hand manipulating the control switches produces a corresponding movement of the subject, regardless of how he is tilted (*fig. 8*). All the various movements have limit contacts in suitable positions, which switch off the associated driving motor when the limit of the movement is reached. In this way rotation about the horizontal X-ray beam is limited to 180° clockwise and 180° anti-clockwise; the patient can thus be tilted to the inverted vertical position in either direction.

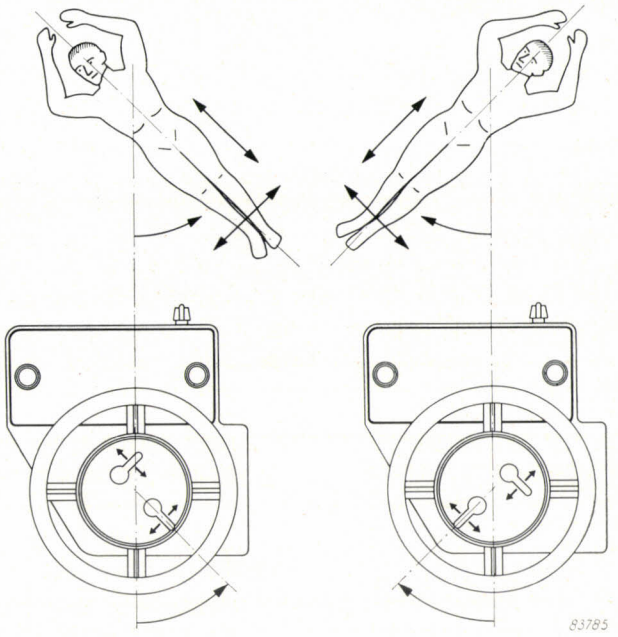


Fig. 8. The hub of the hand wheel driven by a small motor housed in the control box, rotates with the patient so as to maintain the switches in the same positions relative to the patient-movements controlled by them.

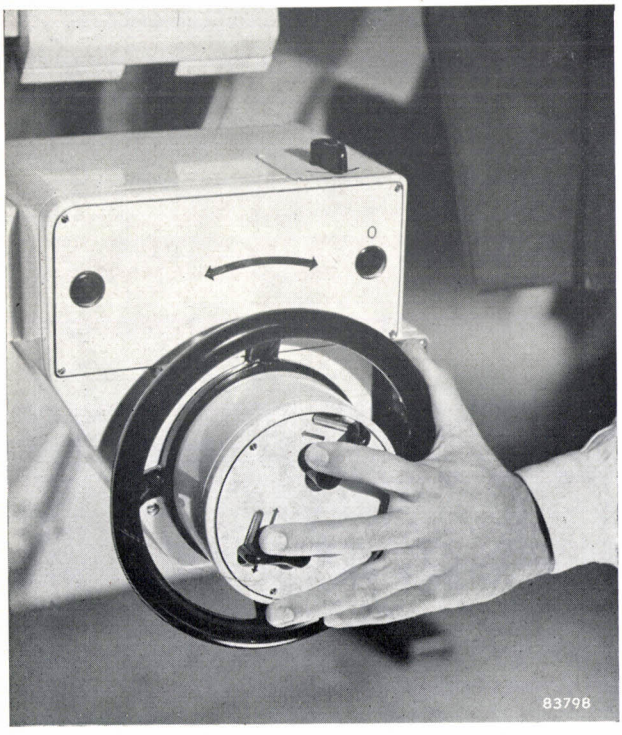


Fig. 7. The control box. The hand wheel controls the rotation of the patient about the horizontal X-ray beam. The two switches on the hub of this wheel control the longitudinal and lateral movements of the patient, and the switch on top of the box controls the rotation of the patient about his longitudinal axis. When the button *0* on the right is pressed, all the moving parts of the apparatus return to their normal positions. The button on the left enables the whole apparatus to be stopped immediately in an emergency.

To return the subject to the normal position after an examination, the doctor merely presses a button at the front of the control box (*0* in fig. 7). The initial movement of any of the parts of the apparatus away from the normal position, operates one of two toggle switches, according to the direction of the movement. This in turn energizes a relay which ensures that each motor rotates in the proper direction when the return-button is pressed. On reaching their normal positions, the various moving parts operate other limit contacts which switch of the driving motors automatically. The rotating part of the control box, carrying the two switches, is similarly controlled. The cradle and the rotating hub of the control box thus return to the same relative positions after each examination; hence, although the motors are asynchronous, cradle and hub can never be out of phase with each other, since any slight variations that may occur are not cumulative.

The relay system for the operation of these controls is housed in the base of the apparatus (*fig. 9*).

Planigraphy

There may be some doubt as to why it is necessary for the main bearing to have so large an inner diameter, since in principle the X-ray beam would have plenty of room to pass through a very much smaller one. However, this large bearing was chosen to make possible the use of one or two accessories which enable the apparatus to be employed for planigraphy. This secondary function involves swinging the X-ray beam, fixed for all other purposes, through a given angle about a particular point in the body of the patient. *Fig. 10* shows a model demonstrating how this may be done. In the model the X-ray tube is swung in such a way that the central ray of its beam remains in the vertical plane of symmetry of the apparatus. The film cassette must move with the central ray in planigraphic exposures; hence the viewing system is shifted away from the normal operating position to make room for a column carrying a movable film cassette (and if necessary, a Potter-Bucky diaphragm).

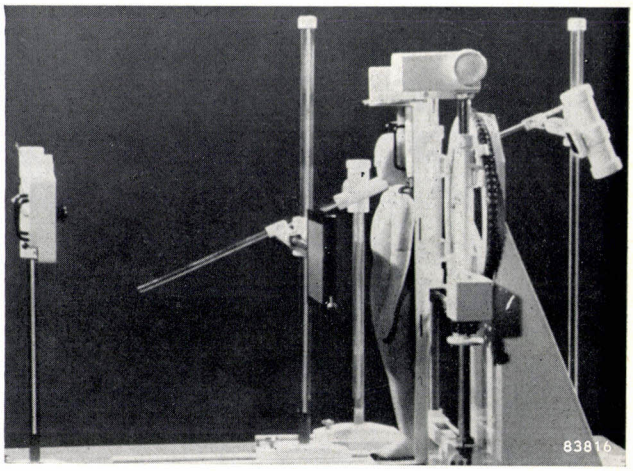


Fig. 10. Model of the UGX diagnostic unit as set-up for planigraphy. The fluorescent screen is not required; it is moved away to make room for a film cassette mounted on a separate stand and connected to the X-ray tube by a coupling rod.

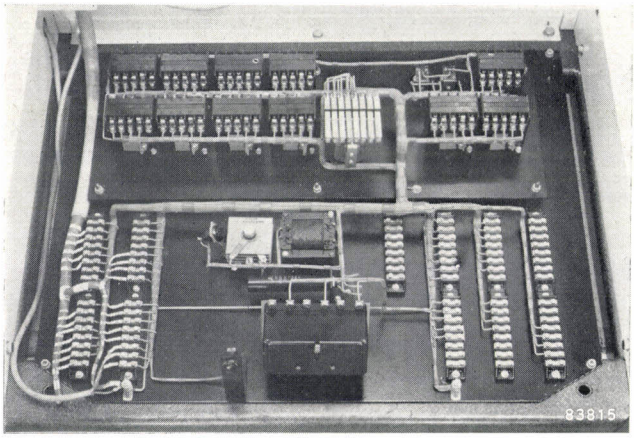


Fig. 9. Relay system, housed in the base of the apparatus, through which the limit contacts, "memory" switches, etc., control the various movements.

Summary. One of the characteristic examination procedures carried out with so-called universal X-ray diagnostic equipments consists in tilting the patient, together with the X-ray tube and the viewing system, about a horizontal axis at right angles to the X-ray beam. With the "Müller" UGX diagnostic unit, however, this procedure, and in particular the associated vertical examination of patients in the horizontal position, is dispensed with. The X-ray beam of this new apparatus remains in a fixed horizontal position. A movable cradle between the X-ray tube and the fluorescent screen enables the patient to be rotated about an axis coinciding with the horizontal X-ray beam; moreover, the patient can be moved longitudinally and laterally, and rotated about his longitudinal axis. The radiologist, standing in front of the viewing system, controls all the motor-driven parts of the apparatus with the aid of a conveniently placed control box. The principal advantage of this arrangement is that it facilitates the use of modern diagnostic aids such as the X-ray image intensifier and the ciné camera in examinations, and minimizes the physical strain on the radiologist, even when using very hard X-rays which require heavy protective screening. Instead of the conventional examination procedure dispensed with, the apparatus permits several entirely new patient positions and unusual projections, one or two examples of which are given.

RESEARCH ON THE CONTROL OF ANIMAL PESTS

by J. MELTZER *).

632.9:632.6/.7

*Agriculture and horticulture must conduct a perpetual battle against plant diseases and pests. The dangers that threaten when adequate quantities of pest control agents are not kept at hand, were brought home only last year by the locust plague in Morocco, where in a short space of time enormous damage was inflicted. At Boekesteyn *) and elsewhere an intensive search is in progress to extend the arsenal of chemical combating agents. Research at Boekesteyn has led, inter alia, to a new agent for combating the spider mite.*

Cultivated plants are continuously exposed to attack, and even to destruction, by various diseases and pests. A review of such injurious influences, together with the chemical means at our disposal for combating them appeared recently in this journal [1]). As explained in that article (hereafter referred to as I), pests can be divided into two types: vegetable (fungi, weeds) and animal. We shall be concerned here only with the latter class.

Although the animal kingdom embraces the most divergent creatures, from the primitive microscopic, non-cellular, animalcules (Protozoa) to the largest mammals, the insects form far and away the greatest class. According to Imms [2]), 70% of all the known species of animals belong to the insects, i.e. 700 000 species, while it is conjectured that this is less than a fifth part of the number of species of insects which actually exists. In this group is found the majority of injurious animals. It should not go unmentioned that there are also many useful insects, useful either because they bring about fertilization of plants, or because they are the natural enemies of injurious animals. The existence side by side of harmful and useful insects renders the control of the former a complex problem.

After the insects, the most important harmful animals are the rodents (rats, mice), snails, worms (especially certain eelworms) and the mites, related to the insects, which belong to the spider family (*Arachnida*). At Boekesteyn various species of experimental animals are cultured on a large scale, including flies (see fig. 2 of article I), grain-weevils, flour-moths, flour beetles (*Tribolium*), colorado beetles, thrips (*Thysanoptera*), spider mites, etc.

*) Agrobiological laboratory Boekesteyn, N.V. Philips-Roxane, 's-Graveland (Holland).

1) R. van der Veen, "Boekesteyn", the agrobiological laboratory of N.V. Philips-Roxane, Philips tech. Rev. **16**, 353-359, 1954/55 No. 12).

2) A. D. Imms, A general textbook of entomology, Methuen, London 1946.

Stomach poisons and contact poisons

Before going into the work carried out at Boekesteyn, it will perhaps be profitable to dwell for a moment on the historical development of insecticides, for this will give us at the same time an idea of the action of these substances.

According to its mode of action, a poison is known either as a stomach poison or a contact poison. In the last century only general (as opposed to specific) stomach poisons were available for combating insects, e.g. lead arsenate. These poisons, however, were not only active against insects, they were also dangerous to all animals and to man. Another disadvantage of stomach poisons is that they are only lethal to insects if the latter eat a relatively large amount of leaf. Sucking insects and spider mites on the other hand, which do not eat the leaves, but pierce their proboscis into the plant to take up the plant juices, are not killed by a stomach poison which covers the surface of the plant. In combating these pests, use must be made of a contact poison.

Mineral and vegetable contact poisons

At the end of the 19th century, paraffin and tar oil came into use as contact poisons against scale insects (*Coccidea*). Later, special mineral and tar oils were developed for combating the winter eggs of various insects and spider mites, which pass the winter at the egg stage on fruit trees.

We cannot here go into the mode of action of these oils. It is sufficient for our discussion to know that they function not as stomach poisons but as contact poisons. Their applicability, however, is seriously limited by the fact that they can only be used when the trees are bare of leaves, since otherwise they would do more harm than good.

The first contact poison that was used against insects was of vegetable origin: nicotine, which at the end of the 17th century was applied on a small scale in the form of tobacco juice. But this poison,

too, is extremely dangerous to man and is rapidly absorbed through the human skin. When, around 1925, ground Derris root came onto the market, it was a great advance, for this product was shown to contain a contact poison (rotenone) which, although very strongly active against insects, is but slightly toxic to man. It was, moreover, far cheaper than nicotine and the pyrethrins (prepared from *Pyrethrum*) which had been discovered in the meantime and which are still used in the home against insects.

Synthetic organic contact poisons

With the discovery of DDT [6] [3]), the first synthetic organic contact insecticide came onto the market. It would take us too far from our subject to describe the revolutionary consequences which this has had for the combating of insects. Sufficient be it to say that this discovery gave the chemical industry the great stimulus to explore this field further.

One of the reasons for this was mentioned briefly in article I: caution must be exercised to ensure that the natural enemies of the insect species to be exterminated are not destroyed. The ideal would require the availability of a number of compounds, each of which was active against one harmful insect species, without being poisonous to other animals — a series, that is, of highly specific (selective) compounds.

Another reason for the intensive research is the noteworthy fact that, however effective an insecticide may be against a few or even a large number of pests, there are always some insect species which in some way or another escape the lethal action. It will suffice to mention only one of the causes: many insect species are naturally protected by their structure, mode of life, or physiology against the action of certain poisons. Some for example live *within* the plant (in wood, bark, leaf or stem) and by reason of this are difficultly accessible from without; such pests can only be combated from within the plant.

Emphytic (systemic) agents

All the above mentioned insecticides are either sprayed or dusted onto the plant. They might therefore be called *ectophytic*, i.e. *outside* the plant. An ectophytic contact poison kills both friend and foe among insects, for neither parasites nor predatory enemies of the insect to be exterminated nor even bees are spared. Another drawback with ectophytic insecticides is that the shoots which the plant puts forward after treatment, do not share in the protection. Moreover, the applied layer is exposed to sun, rain and wind, so that under unfavourable weather conditions it may rapidly lose activity.

These disadvantages do not hold, or hold to a lesser degree, for poisons which are taken up *into* the plant and are dispersed throughout the whole plant in the sap streams. Young shoots which are put out after the treatment are protected from the very beginning; the weather no longer has a direct influence; insects living in the plant, of which something was said in the previous section, come within the reach of the poison, as well as sucking insects and mites, while useful insects which neither feed on parts of the plant nor suck up its juices, run no danger.

Combating agents which are assimilated by the plant are now often called systemic insecticides. Following Prof. Kuenen we prefer to replace the less accurate term systemic by *emphytic*, a term which implies both *in* and *throughout* the plant and forms a clear antonym to ectophytic. The emphytic agents so far known, for example schradan [12], are all very toxic to man (which is likewise one reason why the ectophytic agents, in spite of the disadvantages summarized, have by no means seen their day).

Artificial resistance

This survey would be incomplete without some mention of artificial resistance, i.e. a resistance which is built-up in certain insects by some insecticides. Thus, flies are known which have become resistant to DDT, and there are mites which have even acquired resistance to highly poisonous parathion [10].

Due account must be taken during research of the possibility that resistance may be set up. At Boekesteyn resistant flies are regularly included in the investigations, as we shall see presently.

Research at Boekesteyn

From the foregoing it will hardly be surprising that in the search for means of controlling animal pests our interest is mainly in the field of contact poisons; a stomach poison is only of real interest when it appears to have an emphytic mode of action. A further line of investigation is dictated by the need for specific agents, which do not attack the enemies of the insect to be exterminated, are only mildly toxic to man and domestic animals and evoke no, or only a very slight degree of resistance.

---

[3]) The numbers in square brackets refer to the structural formulae in the appendix.

In addition it is required that a spray or dust residue on the plant is able to kill insects alighting upon it. Use is only very seldom made of methods in which the insects themselves are dusted or sprayed.

Part of the research at Boekesteyn is carried out "under glass" and part on the living plants. This is followed — as was outlined in article I — by a continued investigation of those compounds which have passed the primary test.

Investigations under glass

For testing the action of compounds against grain-weevils and flies, petri dishes are employed, covered on the inside with a coating of the compound under investigation. The coating is obtained by dissolving the compound in a volatile solvent (usually acetone) to give a solution of a definite concentration and pouring a known amount of this solution into the dish; the compound is uniformly distributed during the evaporation of the solvent by rotating and shaking (*fig. 1*). For tests on grain-weevils the bottom, walls and lid are treated, for flies only the lid. When the flies are attacked by the agent, they fall onto the untreated bottom; the dose which they have received from the lid must in itself be sufficient to kill (*fig. 2*).

Resistance tests on flies

As already mentioned, flies are now known that have become resistant to one or more insecticides; thus, wherever there has been intensive spraying with DDT, flies are found which are resistant to DDT. If such flies are treated with lindane — the gamma isomer of hexachlorocyclohexane (BHC) [7] — they very quickly die, but if the flies are treated with lindane for many generations without interruption, they likewise become resistant to this agent. In latter years we have cultured various resistant strains by regular spraying with a certain agent, so as to include them in our investigations.

Fig. 1. Irrigation of petri dishes with a solution of the compound to be investigated for tests on flies and grain-weevils. The solvent (usually acetone) is allowed to evaporate, leaving behind a known quantity of the compound in a thin layer.

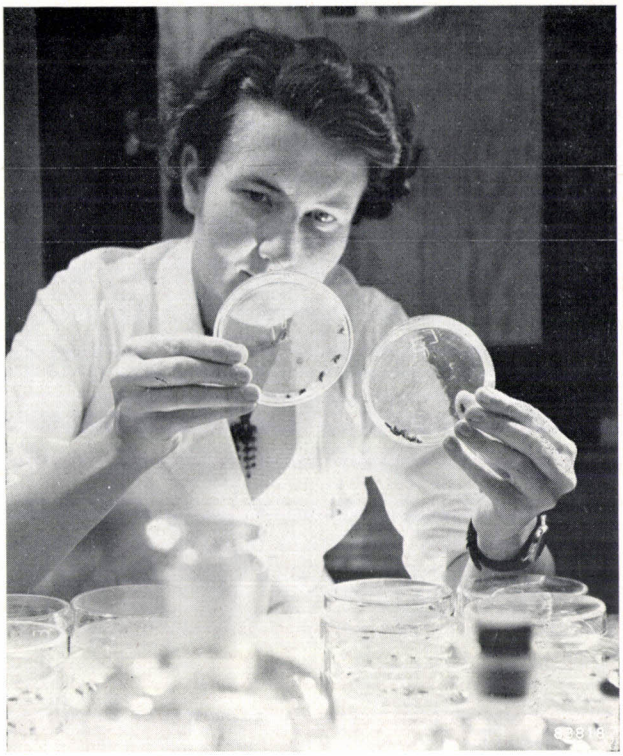


Fig. 2. Tests on flies in petri dishes that have been coated with lindane. The lindane has killed the strain of normal sensitivity (right) but has had no effect on the resistant strain (left).

As to the nature of resistance we are groping in the dark. Some investigators assume that resistance is based upon dechlorination of the compound in the body of the insect. This, however, has not yet been satisfactorily demonstrated.

In practice it is found that resistance by no means occurs in all species of insects; the behaviour of flies in this respect is the exception rather than the rule. Nor can flies be rendered resistant to all poisons. Thus, it has not so far been possible to impart resistance against phenyl N,N-dimethyl carbaminate [19] — provisionally designated as S 17 — although it is possible to render flies resistant to its chlorinated derivatives.

S 17 is not the only compound against which flies only become resistant with difficulty; their sensitivity to parathion and the pyrethrins, mentioned early in this article, does not depreciate significantly. Nevertheless none of these compounds are very suitable for controlling flies: parathion is far too toxic for use in houses and cow-sheds, and S 17 and pyrethrins have only a short active life. Latterly good results have been obtained against resistant flies with "Diazinone", a phosphorus compound of low toxicity.

For the time being, resistance can only be avoided by not using the same agent for too long a period at a stretch. Our laboratory experience has shown that it is desirable, even after only one spraying with DDT against flies in houses and cow-sheds, to transfer to another insecticide, since DDT will give rise to an appreciable resistance even in the first generation. Insecticides of the BHC group, such as lindane, are much more satisfactory; in their case resistance only occurs after several generations.

Tests on the potato root eelworm

The dreaded "potato sickness" is caused by the potato root eelworm, a small worm about 0.5 mm in length (*fig. 3*). It lives in the roots of the potato plant and gives rise to a severe festering of some of the roots which develop gall-like growths. The latter interrupt the supply of food through the root, as a result of which the parts of the plant above ground wilt.

During the development of the female eelworm the abdomen greatly swells as it fills with eggs, and finally bursts out through the skin of the root. The spherical females can be seen as small knobbles on affected roots. When the female dies its body wall forms a tough envelope — the cyst, *fig. 4* — around the eggs and larvae. The larvae only make their appearance from the cyst when they come into contact with excretion product(s) of the potato root; that is, they are enticed from their lair only by one or more specific emergency stimulating substances. Should this not happen, the cysts remain virile in the ground for many years and thus constitute a serious threat to potato cultivation.

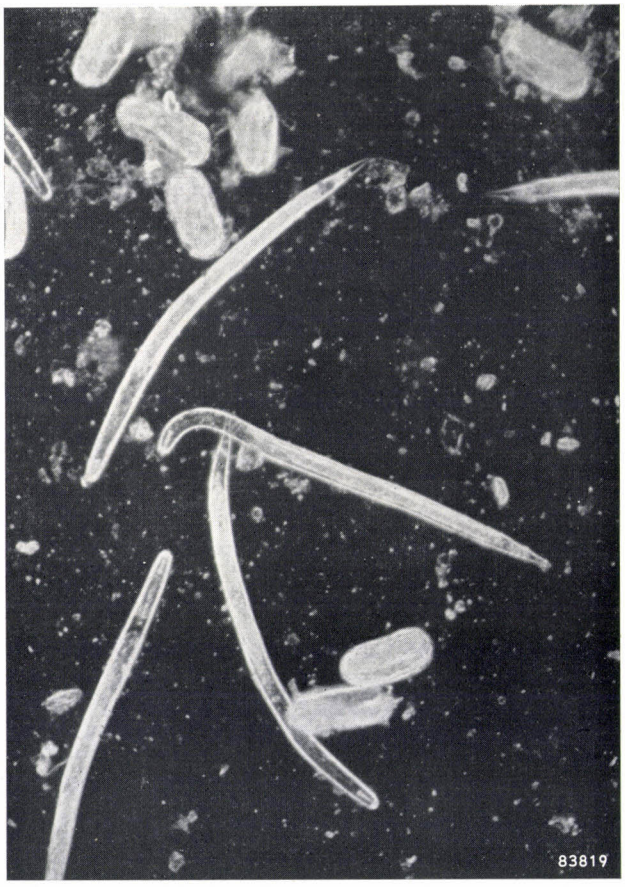


Fig. 3. Larvae and eggs of the potato root eelworm (*Heterodera-rostochiensis*, Wollenw.), responsible for potato sickness. The larvae are about 0.5 mm long.

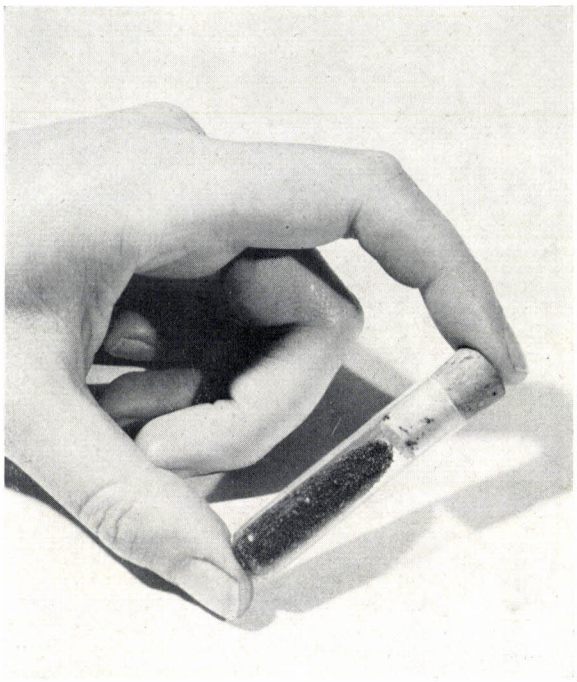


Fig. 4. Tube containing 10 000 cysts of the potato root eelworm (fig. 3).

In our tests, the cysts are exposed for a short while to the compounds to be investigated, washed and placed in plastic dishes. They are then put into water with which pots containing sand cultures of potatoes have been washed; this water contains the stimulating substance and activates the emergence of the larvae. The more active the combating agent the fewer the larvae leaving the cysts. The method is illustrated in *fig. 5*.

Tests on living plants

Tests on leaf-devouring caterpillars and beetles, and on sap-sucking greenflies, thrips, spider mites etc., can only be carried out on living plants (*fig. 6*).

As previously remarked, it is the action of the residue left behind on the treated plant which especially interests us in the ectophytic agents. In our laboratory tests the treatment consists in immersing the plants in a solution, emulsion or suspension of the compound under review. When dry the plants are infected with the experimental animals.

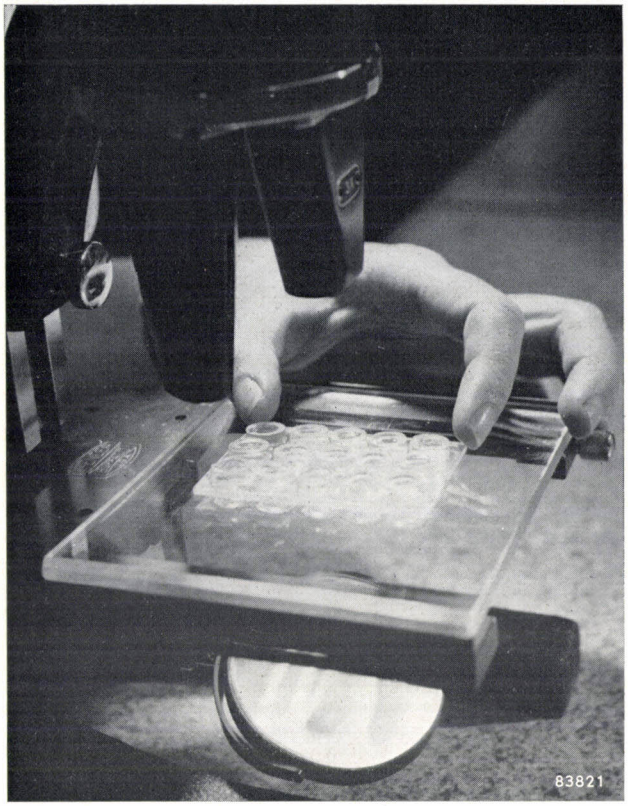


Fig. 5. With the aid of a binocular microscope it is ascertained whether a certain agent has been active against the eggs and the larvae of the potato root eelworm.

The normal arrangement for tests on caterpillars, beetles and greenflies may be seen in *fig. 7*. The plants infected with the insects are placed in glass cylinders, covered with gauze. Very small animals, such as thrips and spider mites, are applied to the leaf in small plastic boxes (*fig. 8*), so that they do not get lost.

The same arrangements are used in emphytic research; the only difference is that now the earth in the pot and not the plant is treated with the compound to be investigated. To exclude any possible action by the vapour from the compound the lower opening of the glass cylinders is carefully sealed with a plastic plate.

The rate at which the compound is transported through the plant can be studied by attaching the insect boxes (fig. 8) at various levels, and measuring how much later the one group is affected than the other. It can also be ascertained whether there is any transport from one leaf to the other or from the upper to the lower parts; for this purpose one leaf is treated with the emphytic agent and the boxes are attached at other suitably chosen positions.

Follow-up investigations

When a material has passed the first test — i.e. has shown a lethal action towards the animal to be combated — the second phase of the research begins. This includes an accurate determination of the degree of activity on as great a number of animal species as possible, and an investigation of the behaviour of the compound under various conditions in which it is compared with known combating agents.

Of the many methods which can be applied here, we will name only those that give an insight into selectivity. When insects are injected with DDT

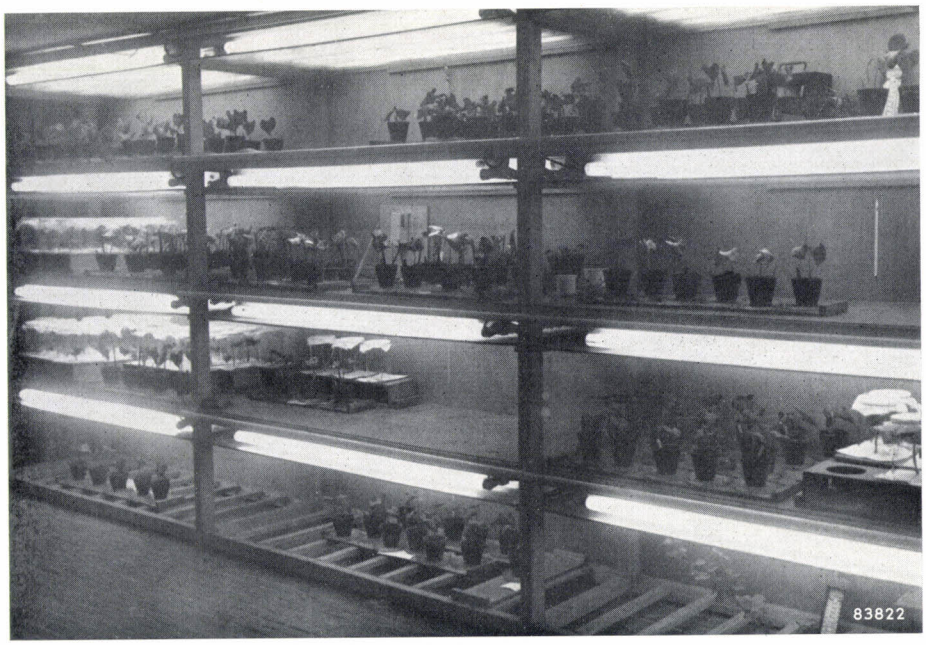


Fig. 6. Insect tests on living plants in an air-conditioned room.

Fig. 7. Testing colorado beetle larvae on potato foliage in glass cylinders.

(*fig. 9*) and warm-blooded animals are given DDT by intravenous injection, the toxic doses per kilogram of living weight are found to be the same. But if DDT is brought into contact with the skin it is found that warm-blooded animals are relatively insensitive, compared with those insects that are sensitive to DDT. It is almost self-evident that for insects themselves there is likewise a difference in reaction time on injection or on contact with the skin. The two values thus found measure the toxicity and the rate of penetration respectively. The injection tests supply a more general value for the toxicity and from this value in combination with the contact toxicity an idea of the insecticidal value can be

Fig. 8. Bean plants infected with spider mites. The latter are enclosed in a small plastic box.

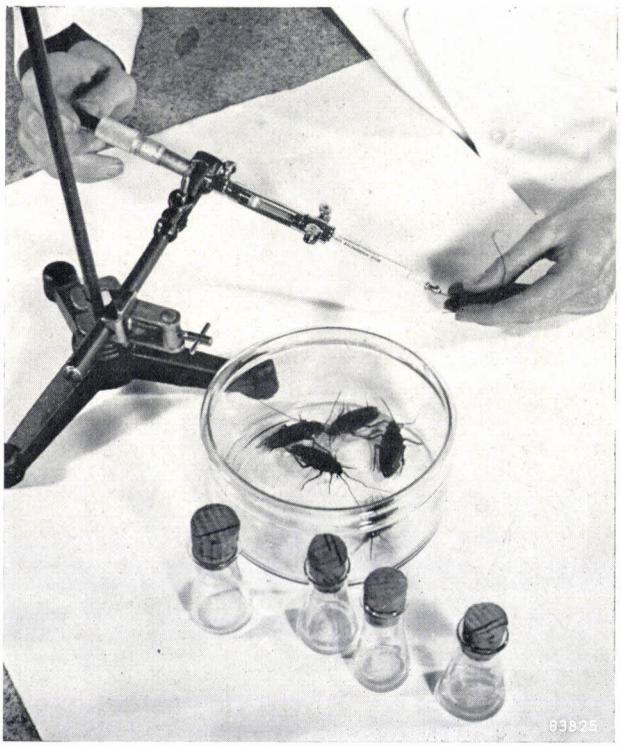


Fig. 9. Cockroaches injected with an insecticide by means of a micro-syringe.

obtained. Some idea can also be gleaned about the rate of penetration, but it should be remembered that little is known of the fate of insecticides within the body of insects. Once the second phase has passed off satisfactorily, there follows a third, in which the compound is tested in the field after its possible toxicity to warm-blooded animals has been determined. This will form the subject of a subsequent article.

A new agent for combating the spider mite ("Tedion V 18")

Around the turn of the century, "fruit tree carbolineum" was used very extensively in fruit growing against insect eggs. It is not active against spider mite eggs. As a consequence the spider mite became an ever more serious pest ("fruit-tree red spider"). No adequate means was known for combating the spider mite. At first it was thought that a suitable agent had been found in parathion, but in practice it was not always satisfactory. True, the emphytic agent schradan [12] is most effective, but is unattractive because of its toxicity. In later years specific agents have been developed against the spider mite (acaricides) which are practically non-toxic to man. The most important are p-chlorophenylbenzene sulphonate [20], known as CPBS or PCPBS, and p-chlorobenzyl sulphide [21], known as chloroparacide.

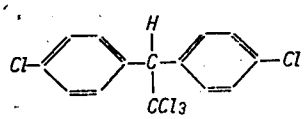
Some years ago we were investigating a series of sulphones. They caused no mortality worth noting either with insects, spider mites or the eggs of the flour-moth. When later we examined these compounds for their action on spider mite eggs it turned out, much to our surprise, that they were active. The most active was 2,4,5,4′-tetrachlorodiphenyl sulphone [22], which was provisionally designated as V 18. It is remarkable that only the growing stages of the mites (eggs, larvae and nymphae [4])) are killed; V 18 has no more effect on mature mites than it has on insects. It is poisonous neither to plants nor to warm-blooded animals: for mice 5 grams of V 18 per kilogram of living weight produces not a single symptom of poisoning. Assuming that mouse and man are about equally sensitive, this would mean that a man weighing 70 kilograms could safely consume 350 grams of V 18. When it is considered that in our experiments a solution of only one part of V 18 in a million parts of spraying solution is often able to bring about total extermination of spider mite eggs and larvae, it will be seen that it is not only remarkably selective but also extremely active. It is in fact more active than the compounds just mentioned ([20] and [21]) and has the additional advantage that the risk of damaging the foliage is appreciably smaller.

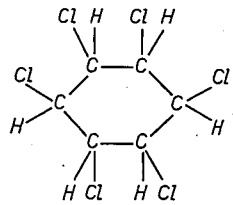
The manufacture of V 18 will begin this year. It will be marketed under the name "Tedion V 18".

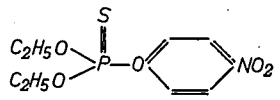
[4]) A nymph is a stage intermediate between the larva and the adult and in appearance has already assumed great similarity to the latter.

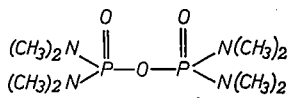
APPENDIX: STRUCTURAL FORMULAE

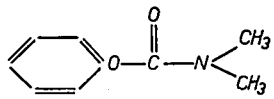
Below are given the structural formulae of the compounds mentioned in the text.

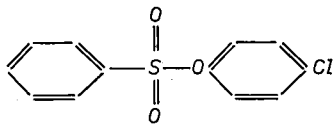
[6] DDT . . . . . 

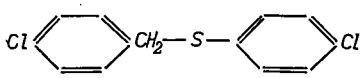
[7] HCH . . . . . 

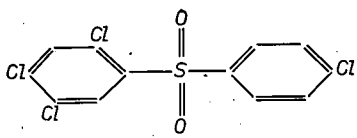
[10] Parathion . . . . . 

[12] Schradan . . . . . 

]19] S 17. . . . . 

[20] CPBS . . . . . 

[21] Chloroparacide . . 

[22] „Tedion V 18" . . 

Summary. Pests injurious to cultivated plants are primarily of the insects species. After a historical survey of the development of insecticides — classifiable as stomach poisons and contact poisions — there is some discussion of "emphytic" (systemic) agents which render the plant itself poisonous to certain insects and consequently have many advantages over "ectophytic" agents.

From the work of the Boekesteyn laboratory in the search for pest control agents for animal pests, reports are given of: 1) resistance tests on flies, 2) tests on the potato root eelworm responsible for potato sickness, and 3) a new, highly specific agent against spider mite, harmless to man and domestic animals. This acaricide will be marketed under the name "Tedion V 18". In follow-up investigations, both contact tests and injection tests are made with the investigated compounds. The structural formulae of some of the compounds discussed are given in an appendix.

THE GLOSS POINT OF GLAZES

by R. W. P. de VRIES.

666.291.22:536.421.2.08

With the development of glass manufacture and glass processing from an art into a science, objective methods of characterizing and identifying the behaviour of glass have gained ground in this branch of industry. Such developments have perhaps lagged somewhat behind as regards glazes and enamels.

In glass working, just as in metal working, specific temperatures should be strictly adhered to, in order that the glass will have the physical properties required. As regards the types of glass from which the envelopes of lamps, radio tubes, X-ray tubes etc. are made, it is above all the viscosity that determines the requisite temperature for each of the different stages in the production process; it also frequently sets an upper limit to the temperature at which the product can safely be used. To characterize such a type of glass, therefore, not only are the coefficient of expansion and certain other important properties indicated, but also the temperature at which the viscosity possesses the specific values required for the various phases of manufacture; see *fig. 1*.

Away from the glasses used for blowing and working into bulbs, tubes, etc., there is another group, about which much less is known, viz. the glazes and enamels, for coating ceramics, glass and metals. The object may be to form a joint, or to provide a protective or decorative finish. The coating is effected by applying to the object a thin layer of a suspension of finely-ground glaze in water [1]). The layer is dried and the remaining grainy coating is fired in a furnace until fusion takes place. As soon as the object reaches a certain temperature during the firing process, the surface undergoes a fairly sudden transition from dull to bright. This temperature is termed the gloss point.

For the comparison of glazes, the viscosity curve discussed above is of little use. In the first place, the phenomenon in question, namely the smoothing out of the surface is not a unique function of the viscosity. The surface tension also has an effect and the time factor too plays a part. In the second place, the viscosity values of glaze glasses at the temperatures of firing are relatively low and for this reason they are more difficult to determine.

The appropriate method therefore is to ascertain the gloss point of the glaze direct from a firing test. This can be done during firing by observing the glaze through a viewing-aperture in the furnace wall and by noting the temperature at which the surface of the glaze is seen to become reflective by the light radiated from the walls inside the furnace, or (at temperatures below 700 °C) by the light supplied by an incandescent lamp.

When visually determining the gloss point in this way, it was found that there was sometimes an unaccountable spread in the results. To trace the cause of this, an arrangement was set up for

[1]) Physico-chemically there is no difference between glazes and enamels. For convenience we shall therefore speak henceforth only of glazes. A detailed account of these substances and of enamelling technique is given by L. Vielhaber, Emailtechnik, V.D.I., Düsseldorf 1953.

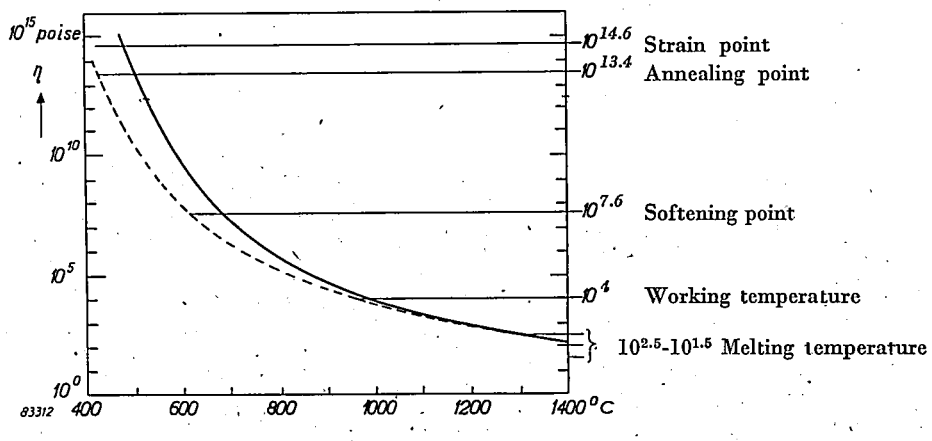


Fig. 1. The viscosity η (in poise) of two types of glass as a function of the temperature. The full curve applies to soda glass, the dashed curve to lead glass. For processing, certain values of η are required, which thus determine the processing temperatures.

ascertaining the gloss point more objectively.

The basic layout is shown diagrammatically in *fig. 2*. A sample of the glaze is placed on a slide and heated in a small furnace. A beam of light is directed upon the sample through an opening in the furnace,

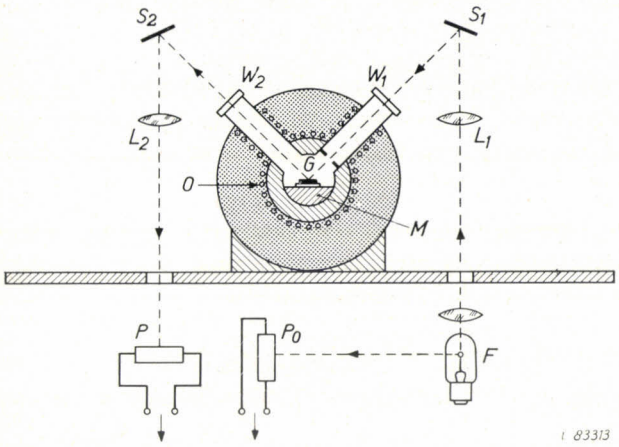


Fig. 2. Basic layout for the objective determination of the gloss point. In the furnace O the sample G is placed on a chrome-iron block M. By means of a small incandescent lamp F, lens L_1 and mirror S_1, a narrow beam of light is directed through window W_1 on to the glaze. Specularly reflected light leaves the furnace via window W_2 and is directed upon photo-electric cell P by mirror S_2 and lens L_2. Direct light from the light source falls upon the photo-electric cell P_0. The ratio of the photo-electric currents from P and P_0 is recorded as a function of the furnace temperature.

and the light reflected from the sample emerges through a second opening and falls on a photo-electric cell. Initially, the sample reflects diffusely and the photo-electric cell therefore receives only a faint gleam. However, as soon as the gloss point is reached, the sample reflects almost as a mirror surface. The configuration of source, sample and photo-cell is so chosen that the reflected beam of light falls on the cell, which therefore now indicates a much higher luminous intensity. If the furnace temperature is gradually raised and the photo-electric current is plotted as a function of the temperature on an X-Y recorder, the resultant curve will show a sharp upward step when the gloss point is reached. The special furnace is shown in *fig. 3* and the complete apparatus in *fig. 4*.

The measuring procedure is subject to a number of difficulties. The luminous intensity of the light source may vary during the recording of a curve. These variations are compensated by allowing a direct beam of light from the source to fall on a second, similar photo-electric cell and recording the ratio of the two photo-electric currents (reflected light/direct light) on the X-Y recorder. A more serious difficulty is the radiation from the furnace which causes heating of the photo-electric cells. To minimize this effect, a type of selenium-cell with a small temperature effect is used and the cells are positioned lower than the furnace by deflecting the path of the light as shown in fig. 2. Moreover the furnace is so constructed that it works with a minimum of power, and its radiation is therefore kept to a minimum. The slide that carries the glaze sample must contribute as little reflected light as possible. At temperatures above approx. 600 °C the measurement becomes inevitable unreliable owing to the amount of light radiated by the slide and the inner walls of the furnace. This interfering effect can be avoided by using an intermittent light source and measuring only the synchronously varying part of the light received by the photo-electric cell; at present, however, our tests have not gone beyond 600 °C. It is of course important for the temperature of the sample at every instant to be defined and known with sufficient accuracy. This is not so easy to realize as it seems; a considerable loss of heat occurs through the opening near the sample, and it is consequently difficult to obtain a zone around the sample of sufficiently uniform temperature. Again, while the temperature is being raised, the sample temperature may wholly or in part lag a little behind that at the point of measurement (the location of the thermo-element used). In view of the importance of the rate of heating, as mentioned earlier, it is necessary to use a standardized heating curve in order to arrive at reproducible results. This is best done by automatic

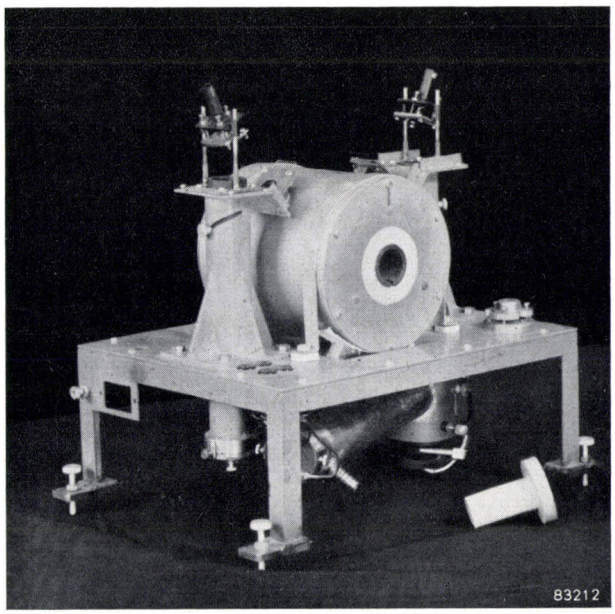


Fig. 3. Photograph of the apparatus sketched in fig. 2. The components marked in fig. 2 are easily identified. A spirit level can be seen on the right of the base plate of the furnace; the slide carrying the glaze sample must lie horizontally so that the glaze will not move out of the beam of light during the test.

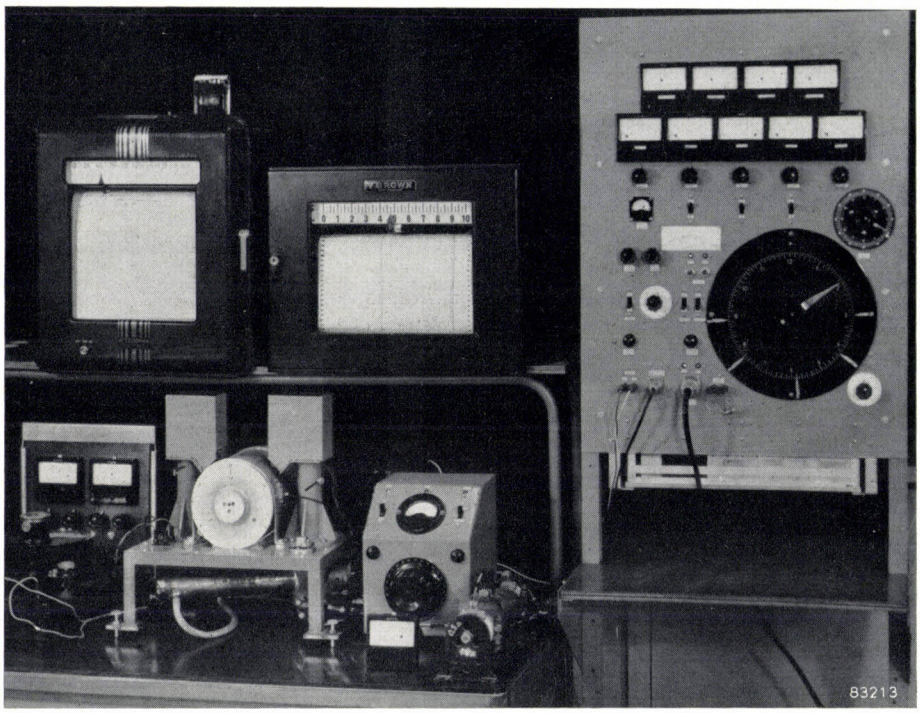


Fig. 4. Complete arrangement for the objective determination of the gloss point. Top left is the X-Y recorder, and beside it an instrument which records the temperature as a function of time (for checking the heating-rate of the furnace; see fig. 7). Below the recorders is the furnace, with meters on either side giving direct indications of the separate photo-electric currents and the temperature and for adjusting the furnace power. At the right is the programming apparatus for controlling the furnace temperature.

programme control of the furnace. Finally, the preparation of the sample should be standardized too, since variations in grain size, grain distribution and packing can also influence the measurements.

So much attention is normally paid to the latter factors, which are of course equally important in the visual determination of the gloss point, that the cause of the spread in results mentioned above must be sought elsewhere. The explanation of the spread indeed soon became apparent from the curves recorded with the apparatus described. We shall now discuss some of these curves.

The curve in *fig. 5* was obtained for a jointing glaze. With its curious dip at the beginning of fusion, i.e. just below the gloss point, it was found very readily reproducible. Because of this reproducibility, it was possible to break off the firing of different samples at various points in the critical temperature range in order to find out what appearance of the glaze surface corresponded to the various points on the curve. Some photographs of such samples are shown in *fig. 6;* they correspond to points I-V in fig. 5. At point I the sintering has only just started (the grains can no longer be blown from the slide). At point IV the surface has not yet become fully bright, whereas at point V the gloss point has evidently been well exceeded. The visually determined gloss point was found to lie at point g; with this glaze the visual determinations were fairly consistent.

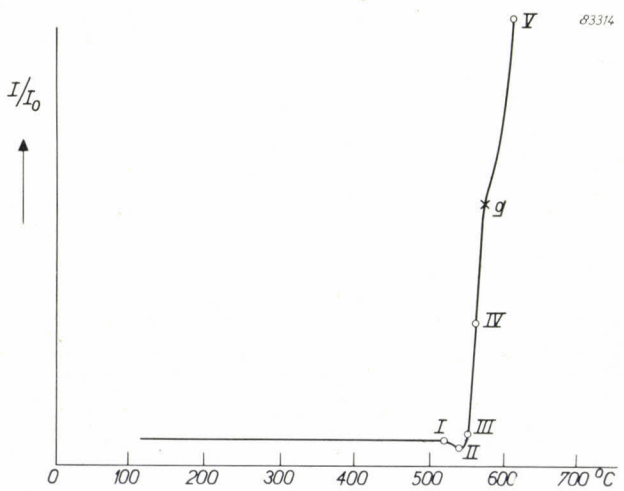


Fig. 5. Recorded gloss curve of a jointing glaze. The ratio I/I_0 of specularly-reflected light to incident light is recorded as a function of the temperature. The test, carried out on a number of samples, was broken off at various points of the curve (I-V), by quickly removing the sample from the furnace. The sudden increase in viscosity, due to rapid cooling, "freezes" the surface appearance of the sample, so that it can be inspected at room temperature. Point g can be regarded as the gloss point.

Comparing sample *I* and sample *II* under the microscope, it was found that at point *II* of the curve the grains were just becoming rounded off at the beginning of fusion (grains sticking together) whereas this had not yet started at point *I*. The ducible result — but the mere reproducibility of this particular result is of little consequence; the important thing is to recognize the deglazing effects, since it is upon these that the glaze will have to be rejected for most applications.

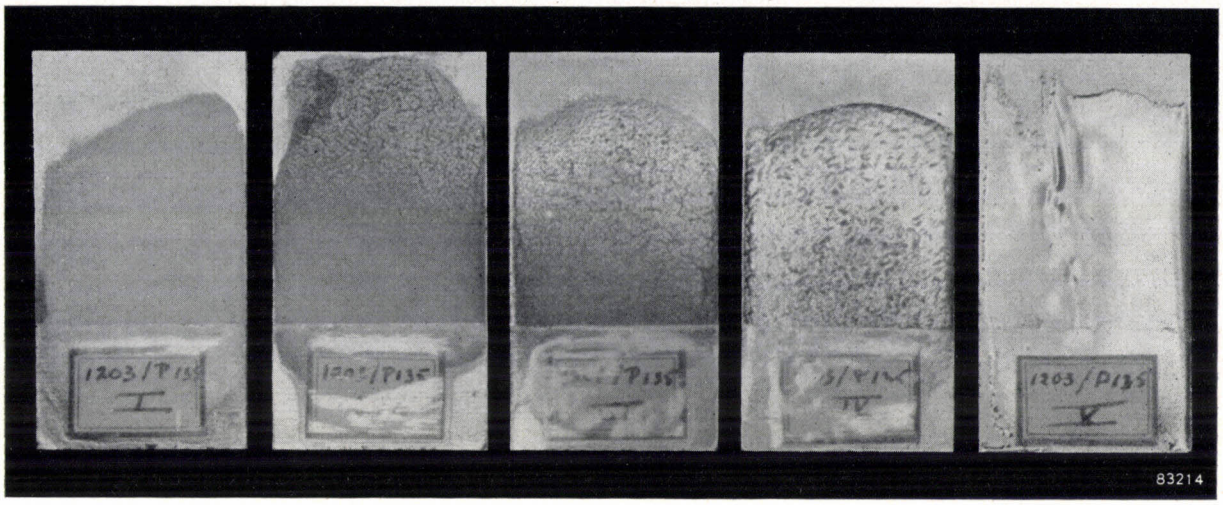


Fig. 6. Photographs of glaze samples, heated to points *I-V* indicated in fig. 5.
I: The grains of glaze powder have only just started to adhere to the underlayer: very first signs of sintering.
II: Grains rounded off, beginning of fusion.
III: First signs of specular reflection in the hollows between the grains.
IV: Whole surface not yet completely bright.
V: Completely bright surface; the gloss point (*g*) has been well exceeded.

dip at point *II* can now be explained by the fact that after entering a grain the light has to travel further in the glaze before leaving it again due to scattering. Owing to absorption in the glaze it will have become attenuated, so that the measured reflection which is still diffuse, is somewhat reduced.

Fig. 7 shows the much more complicated curve plotted for an experimental glaze, consisting mainly of boron oxide (10%) and lead oxide (85%). Here too the dip is obtained just before fusion. The further great fluctuations in the reflection from the sample can be related to the physico-chemical behaviour as follows: after the glaze has fused to a glossy layer, it begins, at a somewhat higher temperature, to deglaze (crystallize) and therefore to assume a dull appearance again. When the temperature is raised still higher, the minute crystals that have been formed either fuse again or dissolve in the residual glass mass. A moment later, the mass deglazes once more, after which final and real fusion takes place. It is hardly surprising that this chain of events gives conflicting results in the visual determination of the gloss point. An observer who conscientiously records the very first signs of brightening has the best chance of arriving at a repro-

Again, the curve in fig. 7 was found to be readily reproducible. It was therefore again possible to break off the firing process at certain points of the curve in order to confirm the deglazing effects by means of X-ray diffraction analysis. Analyses were carried out for points *IV*, *V* and *VI* of the curve in fig. 7.

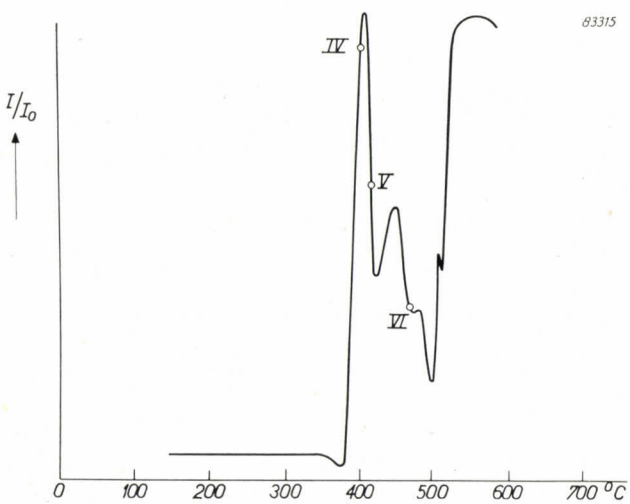


Fig. 7. Recorded gloss curve of an experimental glaze prepared in the laboratory. The heating-rate was 2 °C per minute. As the heating-rate increased, the curve was displaced as a whole towards somewhat higher temperatures without appreciable changes in the various features of the curve.

For point *IV* only a glass phase was found (no sharp diffraction lines); for point *V* the preparation was found to have, in addition to the glass phase, a crystalline phase of β-$Pb_4B_2O_7$; in point *VI* this crystalline phase had disappeared and in its place were identified crystals of α-PbB_2O_4. In this way confirmation was found for the deglazing as such, and for the difference between the two deglazing processes.

As the foregoing clearly shows, the objective determination of the gloss point can be a valuable means of investigating and, eventually perhaps, of mastering the composite physico-chemical processes that take place during the firing of glaze powders.

Summary. When firing a layer of glaze powder on a slide in a small furnace it is found that at a certain temperature the surface of the layer undergoes a sudden transition from dull to bright. The conventional visual determination of this "gloss point" on some glazes prepared and tested in the laboratory produced widely varying results. For this reason an objective method of determining the gloss point was developed, whereby two photo-electric cells are used in conjunction with an *X-Y* recorder, which plots the degree of specular reflection obtained against the temperature of the glaze. The resulting curves were found to be readily reproducible. In some cases they are of complicated form, which makes the non-reproducibility of the visual determinations understandable; in the course of further investigations it was possible to relate various features of the curves to deglazing effects.

ABSTRACTS OF RECENT SCIENTIFIC PUBLICATIONS OF 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 Philips Research Laboratory, Eindhoven, Netherlands.

2190: W. J. Oosterkamp and T. Tol: Prinzipielle und praktische Grenzen der Detailerkennbarkeit bei verschiedenen röntgenologischen Beobachtungsmethoden insbesondere bei Verwendung der Bildverstärkerröhre (Fort. Röntgenstr. Röntgenpraxis **81**, 381-392, 1954, No. 3). (Theoretical and practical limits of etdila perception in various radiological techniques, in particular, using the image intensifier: in German).

The theoretical limit of the detail perceptibility in fluoroscopy with and without image intensifier, in full size normal radiography, miniature radiography and in intensifier radiography, is determined by the fluctuations of the number of quanta or particles conveying the information during the different stages of the radiological process. This theoretical limit is calculated on the basis of a measurement of these numbers of quanta. Every X-ray quantum absorbed in the primary fluorescent screen contributes to the visual impression when the image intensifier is used. This is not the case in normal fluoroscopy and miniature radiography, where only 1% of the absorbed X-ray quanta are effective.

The observed detail perceptibility is limited by the fluctuations of the X-ray quanta for intensifier fluoroscopy at low X-ray intensities and for intensifier radiography with very sensitive emulsions and large apertures; X-ray quanta fluctuations and instrumental deficiencies (lack of definition and loss of contrast) have approximately the same influences in the case of intensifier radiography on fine grain emulsions; the instrumental deficiencies are predominant in miniature radiography and in full size radiography (with intensifier screens).

The X-ray dose is 70-300 times smaller in intensifier radiography as compared with miniature radiography if the same emulsion is used, the detail perceptibility being approximately the same. The detail perceptibility of the intensifier radiograph on a fine grain 35 mm film approaches that of the full-size radiograph; only half the dose is necessary in the former case.

2191: H. Bremmer: The extension of Sommerfeld's formula for the propagation of radio waves over a flat earth, to different conductivities of the soil (Physica **20**, 441-460, 1954).

The influence of non-homogeneous soil conditions on the propagation of radio waves over a flat earth can be investigated with the aid of an integral equation based on Green's theorem. This equation applies to all types of distributions (also continuous) of the conductivity and of the dielectric constant of the earth; it is comparable with the integral equation considered by Hufford for the propagation over irregular terrain. The special solution for two adjacent regions of homogeneous electrical constants can be treated numerically with the aid of two different expansions for the field near the

separating boundary and for the field far beyond this boundary; the rigorous solution of the integral equation proves to be identical with the corresponding expression derived in a very different way by Clemmow. The solution for three adjacent homogeneous regions can be obtained by solving the general eqation by a similar method.

2192: C. J. Bouwkamp and H. B. G. Casimir: On multipole expansions in the theory of electromagnetic radiation (Physica **20**, 539-554, 1954).

A new method is developed for expanding the electromagnetic field of radiating charges and currents in multipole components. Outside a sphere enclosing all sources the field is represented in terms of two scalar wave functions (Debye potentials) which are expanded in Legendre and spherical Bessel functions. The various terms of these series expansions are shown to be closely related to the radial components of the electric and magnetic vectors and hence determine at once the various multipole terms of the radiation field. It is emphasized that an alternative method based on the conventional vector-potential presentation has several disadvantages. The authors avoid the use of group theory and tensor calculus and apply only ordinary vector calculus. Attention is drawn to some simple source representations of the radial field components.

2193: T. Tol and W. J. Oosterkamp: Visibility of X-ray quantum fluctuations on image intensifier radiography (Nature **174**, 397-398, 1954).

The gain in brightness in radiography with an X-ray image intensifier is accompanied by a visual "noise" effect, especially at low X-ray intensities and high intensification factor. Its origin lies in quantum fluctuations and the spatial distribution of the grain clusters of the photographic emulsion, each effective photon being responsible for the appearance, after development, of a cluster. Photographs illustrating the effect are shown. See also Philips tech. Rev. **17**, 75-76, 1955/56.

2194: F. J. Mulder: Simplified isolation of the non-saponifiable matter in the analysis of oils (Rec. Trav. chim. Pays-Bas **73**, 626-628, 1954, No. 8).

A simplified method is given for the analytical isolation of non-saponifiable matter from oils. After saponification in alcohol the alcoholic soap solution is completely dissolved in a measured quantity of benzene and the benzene solution is washed with water. The concentration in the benzene solution is determined after drying. By this method the tedious operations of extraction and quantitative transfer of the extracts are avoided. This method is suitable for the determination of the vitamins A and D.

2195: F. de Boer and W. F. Niklas: Application of a tracer to cathode-gettering and gas-adsorption problems (Brit. J. appl. Phys. **5**, 341-342, 1954, No. 9).

The residual gas pressure in a cathode-ray tube decreases when the cathode is heated. It is conjunctured that absorption by the barium-strontium oxide layer on the cathode is responsible. This absorption has been demonstrated by admitting into a cathode-ray tube some CO_2 containing a trace of C^{14}, and subsequently measuring the radioactivity of the cathode. In addition, it has been established that the gas absorption of other components of the electron gun is about 500 $\times$ less than that of the oxide layer on the cathode.

2196: F. de Jager: Les limites théoriques de la transmission en cas de niveau de bruit élévé, pour differents systèmes de modulation continue et de modulation codée (Onde electrique **34**, 675-682, 1954, Nos. 329-330).

The signal-to-noise ratio is examined for systems of frequency modulation, pulse position modulation, pulse code modulation (binary code) and delta modulation, in particular at high noise levels. In choosing the value of the optimum passband, a theoretical limit for the signal-to-noise ratio is found which is roughly equal for all the systems. To obtain, for example, a signal-to-noise ratio of 30 dB, the minimum power for each system must be 17 dB higher than the power of the noise found in a passband corresponding to the original signal. To get a signal-to-noise ratio of 40 dB, the corresponding figure is 20 dB. Finally the theoretical limits of the signal-to-noise ratio are examined for transmission in a chain of 50 links.

2197: H. G. Beljers, L. van der Kint and J. S. van Wieringen: Overhauser effect in a free radical (Phys. Rev. **95**, 1683, 1954, No. 6).

If a paramagnetic substance is placed in a magnetic field and electron spin resonance is excited by a microwave field, the thermal equilibrium of the electron spin system is disturbed. According to Overhauser, nuclear spins, if present, are aligned by relaxation processes which try to restore equili-

brium. The above experiment has been performed using the free radical diphenyl picryl hydrazyl (DPPH) in a magnetic field of 3300 Oe. A microwave field of 3.2 cm wavelength gave resonance absorption approaching saturation. Alignment of the protons in the DPPH is observed by a considerable increase of the proton resonance at 14 Mc/s.

2198*: F. A. Kröger and H. J. Vink: Die Regulierung von elektrischen und optischen Eigenschaften von polaren kristallen (Contribution to Halbleiterprobleme I, edited by W. Schottky, Vieweg, Braunschweig 1954). (The control of electric and optical properties of polar crystals; in German).

This article contains the more important results of an investigation which is reported in detail in Z. phys. Chem. **203**, 1-72, 1954, Nos 1-2 (see these abstracts No. **2180**).

R 249: E. W. Gorter: Saturation magnetization and crystal chemistry of ferrimagnetic oxides, Part II (Philips Res. Rep. **9**, 321-365, 1954, No. 5).

Continuation of *R 248*. Néel's theory is reviewed in section 2.2.1, with Yafet and Kittel's modification (section 2.2.2). When the moments of the two sublattices are approximately equal, Néel predicts a number of anomalous σ—T curves. All of these should occur in a series of mixed crystals of type (a) in which the resultant moment m changes sign. Our experimental methods are given in section 3. The measurements on the single ferrites $Me^{II}Fe_2{}^{III}O_4$, with $Me^{II} = Mn^{2+}$, Fe^{2+}, Co^{2+}, Ni^{2+}, Cu^{2+}, Mg^{2+} or $(0.5\ Li^{+} + 0{\cdot}5\ Fe^{3+})$ show that these belong to group (a); the mixed crystals $Me_{1-a}Zn_aFe_2O_4$ with $a >$ appr. 0.4 belong to group (b) (section 4). The moment of $Ca_{0.35}\ Zn_{0.65}Fe_2O_4$ is higher than that of any MgZn ferrite, perhaps because the angle A-O-B is increased by the presence of the greater part of the large Ca ions in the B sites, thus increasing the AB interaction; this behaviour would be in agreement with Anderson's theory (section 5.1; Anderson's theory is reviewed in section 2.3.2). The moments of ferrimagnetic oxides with other crystal structures may be predicted from the angles (metal ion)-(O^{2-} ion)-(metal ion); cf. $BaFe^{III}{}_{12}O_{19}$ and $KFe^{III}{}_{11}O_{17}$ (section 5.2).

R 250: G. Brouwer: A general asymptotic solution of reaction equations common in solid-state chemistry (Philips Res. Rep. **9**, 366-376, 1954, No. 5).

A theory of the chemistry of semiconductors, given recently by Kröger, Vink and Van den Boomgaard, is based on a number of reaction equations and certain additional conditions, including electroneutrality. A simple way to find an analytical or graphical approximate solution of this set of equations is presented in this report. Two examples are given; the method, however, is not restricted to these systems. The accuracy of the method is quite sufficient to serve its purpose.

R 251: H. A. Klasens: Temperature dependence of the luminescence and chemical stability of basic magnesium arsenate activated with tetravalent manganese (Philips Res. Rep. **9**, 377-390, 1954, No. 5).

The effect of activator concentration and partial substitution of the lattice constituents by other ions on the temperature dependence of the luminescence of 6 $MgO{\cdot}As_2O_5{\cdot}Mn$ is studied. The effect of the activator concentration is in agreement with the theory of Johnson and Williams. Incorporation of lithium improves the temperature dependence considerably. Measurements of the stability of various arsenates in reducing atmospheres show that phosphors containing lithium are less easily reduced. Incorporation of fluorine has the opposite effect. A theory for this induced variation of stability is given.

R 252: J. L. H. Jonker: The similarity law of secondary emission (Philips Res. Rep. **9**, 391-402, 1954, No. 5).

In the literature, the curves of the secondary-emission coefficient δ measured as a function of the energy εV_p of the primary electrons often show large deviations from the experimental universal curve, i.e. do not satisfy the similarity law. It is shown that some of these deviations may be caused by the fact that the target is partly or completely covered with foreign materials. Many materials have been re-measured with extra precautions to ensure a clean surface, and tables are given for a number of materials whose measured δ curve agreed with the experimental universal curve. In that case the indication of the values of δ and εV_p at the maximum of the $\delta = f(\varepsilon V_p)$ curve is sufficient; other values of δ can then be deduced from the experimental universal curve, for which a formula is given.

R 253: E. W. Gorter: Saturation magnetization and crystal chemistry of ferrimagnetic oxides, Part III (Philips Res. Rep. **9**, 403-443, 1954, No. 6).

Continuation of R 248 and R 249. In a number of mixed-crystal series, anomalous σ-T curves have

been looked for: these are not found in the systems $Ni_{1+a}Fe^{III}_{2-2a}Ti_aO_4$ and $Ni_{1.5-a}Zn_aFe^{III}Ti_{0.5}O_4$ because of the unexpected presence of Ti^{4+} ions in tetrahedral sites, proved for $Ni_{1.5}Fe^{III}Ti_{0.5}O_4$ by measurements of the effective g-factor (section 6.1.2). In the system $Li_{0.5}Fe^{III}_{2.5-a}Cr_aO_4$ ($0<a<2{\cdot}0$), the distribution of the Li^+ and Fe^{3+} ions is anomalous as a result of short-range order. The resultant moment remains positive, and only one type of anomalous σ-T curve, viz. that for which the spontaneous magnetization changes sign with temperature, occurs in a wide range of compositions. For $a \leqslant 1.25$ the materials belong to group (a) (section 7). A change of sign of the resultant moment does occur in the system $NiFe^{III}_{2-a}Al_aO_4$. Anomalous σ-T curves are here found in a narrow range of compositions, but not all types predicted by Néel: the reasons are discussed (section 8). The presence of Mn^{2+} ions apparently promotes the formation of angles between the ionic moments in B sites: the complete system $MnFe^{III}_{2-a}Cr_aO_4$ belongs to group (b) (section 9), as well as $MnFe_2O_4$ prepared by other authors (section 4) and part of the system $Ni_{1.5-a}Mn_aFe^{III}Ti_{0.5}O_4$ (section 6.3). In the first-named system m probably changes sign only as a result of these angles.

R 254: P. Cornelius: Vorschläge betreffend die Einheiten der Fläche, der dielektrischen Verschiebung und der magnetischen Feldstärke (Philips Res. Rep. **9**, 444-457, 1954, No. 6).

Suggestions with regard to the definition of the units of dielectric displacement and magnetic field strength, to remove certains difficulties in connection with the retionalization of units.

R 255: A. Versnel and J. L. H. Jonker: A magnetless magnetron (Philips Res. Rep. **9**, 458-459, 1954, No. 6).

A description is given of an oscillator tube that behaves like a magnetron. A feature of the new tube is that no magnetic field is needed. Measurements are discussed also.

R 256: J. Smit and H. G. van Bueren: Elastic after-effect in a-iron in relation to the elastic constants of cubic metals (Philips Res. Rep. **9**, 460-468, 1954, No. 6).

The elastic after-effect in α-iron, containing carbon or nitrogen in solid solution, can be interpreted as causing a change of the elastic moduli of the material. This interpretation renders it possible to estimate with a reasonable degree of accuracy the magnitude of the after-effect in polycrystalline materials. The estimate is based on a general relation which is shown to exist between the elastic constants of single crystals and of polycrystalline specimens of the cubic metals.

R 257: Th. P. J. Botden, C. Z. van Doorn and Y. Haven: Luminescence of F-centres in alkali-halide crystals (Philips Res. Rep. **9**, 469-477, 1954, No. 6).

Crystals of KCl, KBr, KI, NaCl and RbCl coloured additively have been irradiated in the F-absorption band at liquid-hydrogen and liquid-nitrogen temperatures. In all crystals an infra-red luminescence has been observed between about 1 and 1·5 μ. The luminescence is quenched at temperatures above 100-150 °K. The quantum efficiency is very low and the luminescence appears only when the crystals are kept at low temperatures after the formation of the F-centres and quenching in liquid nitrogen. Different results are obtained when the crystals are kept at room temperature in the dark or during irradiation. It is believed that the infra-red emission is to be attributed to F-centres.

VOL. 17 No. 6, pp. 161-188 DECEMBER 1955

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 TELEVISION RECEIVER SUITABLE FOR FOUR STANDARDS

by H. L. BERKHOUT.

621.397.62

Designing a television receiver suitable for four standards — 625 or 819 lines, positive or negative picture modulation, frequency modulation or amplitude modulation of the sound, and differing synchronizing signals — would seem at first sight to be hardly a promising undertaking. Nevertheless, it has been possible to design and produce such a receiver (e.g. type no. 17 TX 100A-70), and thus to meet a need which is felt particularly in Belgium and the adjacent areas.

The fact that various television standards are in use in Europe, faces receiver designers with the problem of constructing sets suitable for more than one standard. This problem is particularly pointed in Belgium, which has two standards of its own (for the Walloon and Flemish broadcasts), both of which differ from the standards in neighbouring countries. In the North and East of Belgium the Dutch and German broadcasts can be received, working on the "Gerber standard" [1]), and in the South good reception is possible from Lille, a transmitter using the French standard. Conversely, in the areas adjoining Belgium, reception of the Belgian stations is possible. A demand has therefore arisen for sets suitable for two, three or even four standards. The principal data concerning these standards are given in *Table I*.

[1]) So called after the chairman of the sub-committee of the C.C.I.R. which established this standard.

In this article some of the problems involved in designing a receiver for four standards will be discussed.

The intermediate frequency amplifiers

The receiver in question (*fig. 1*) is fitted with a knob for selecting the desired standard. The knob operates a number of switches (*fig. 2*) which constitute the standard selector. Wholly independent of this is the usual channel selector, with which the H.F. section of the receiver can be tuned to a number of television channels. This is done by changing coils, which are mounted in a rotating drum for this purpose (*fig. 3*). The channel selector in the four-standard receiver is fitted with components which give a choice of ten channels with the Belgian standards and the Gerber standard and an eleventh channel, viz. that for Lille (or, if desired, for another transmitter, e.g. Saarbrücken or Strasbourg).

The channel selector, of course, changes not only

Table I. Principal data for the television standards in use on the continent of Western-Europe. N = number of lines. f_s = frequency sound carrier. f_v = frequency vision carrier. AM = amplitude modulation. FM = frequency modulation.

Standard	N	Channel width Mc/s	$f_s - f_v$ Mc/s	Picture mod.	Sound mod.	Line pulse duration µs	Frame pulses		Equalizing pulses
							Number	Duration µs	
Belgium { Flemish	625	7	5.5	pos.	AM	5	6	25	present
Belgium { Walloon	819	7	5.5	pos.	AM	5	6	25	present
"Gerber"	625	7	5.5	neg.	FM	5	6	25	present
France	819	13.15	± 11.15	pos.	AM	2.5	1	20	absent

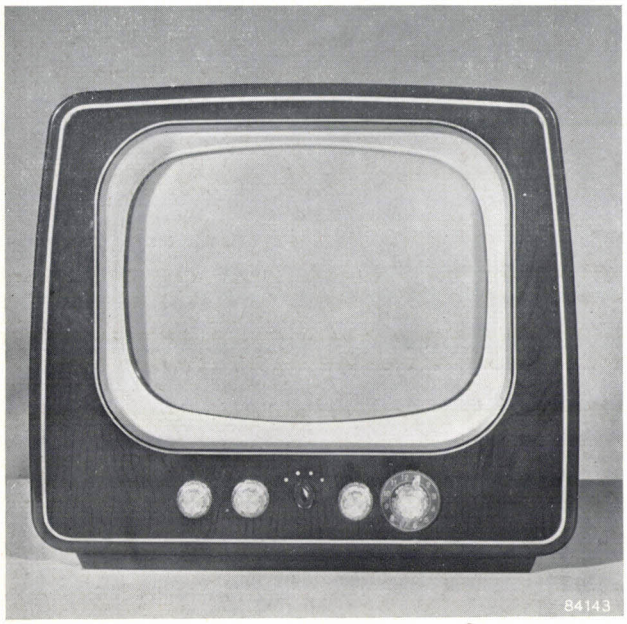


Fig. 1. Television receiver type 17 TX 100A-70 for four standards (the two Belgian, the French and the Gerber standard). In the centre is the standard selector, extreme right the channel selector.

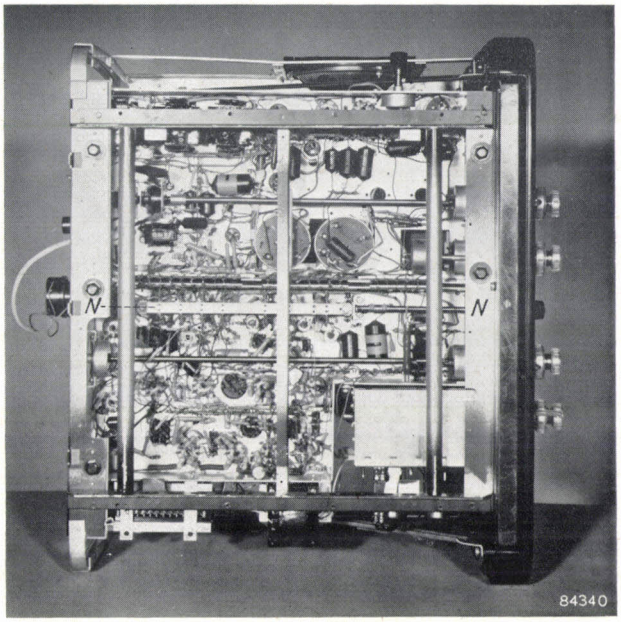


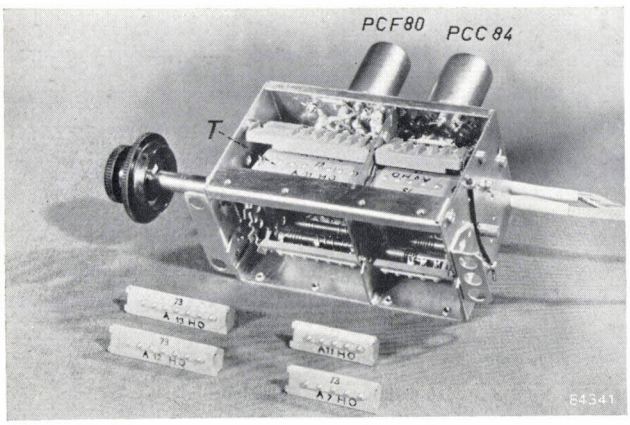
Fig. 2. Chassis of the set illustrated in fig. 1, seen from below. *NN* is the spindle of the standard selector which runs right through the chassis.

the tuning of the receiving circuits, but also that of the local oscillator which, in combination with the received signal, produces the I.F. vision and sound signals.

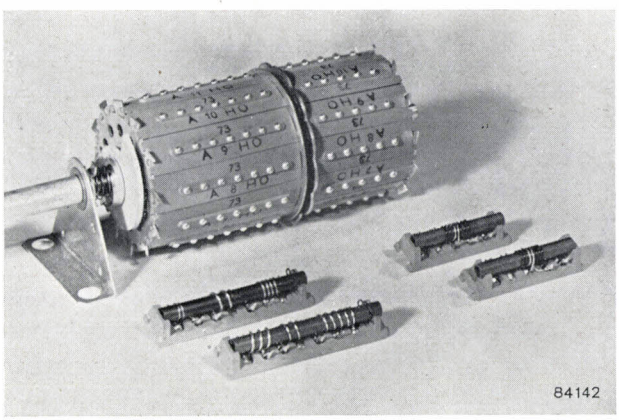
Picture channel

At first sight it appears to be impracticable to use one vision I.F. amplifier unchanged for the four standards, because the French standard works with about twice the bandwidth of the other three (about 11 Mc/s compared with about 5.5 Mc/s). It is, of course, possible in principle to construct an I.F. amplifier which can be adjusted at will to one bandwidth or the other, but this would make the switching very complicated. Furthermore, it would then be necessary to introduce switches between points with an I.F. potential difference, which would influence the stability unfavourably.

From a large number of observations it has become apparent to us that with a bandwidth limited to about 4 Mc/s, the French picture is acceptable in every respect. Without harming the quality of the image seriously, one vision I.F. amplifier, with a bandwidth of about 4 Mc/s, will therefore suffice



a



b

Fig. 3. *a*) High frequency stage incorporating channel selector (opened) as used in Philips television receivers. PCC 84 H.F. tube, PCF 80 first oscillator and frequency changer (both in screening cans). *T* rotating drum — shown separately in (*b*) — and the alternative coils. Each set of coils is mounted on an insulated block with contacts. In the foregrounds some of the coils are shown separately; in (*a*) from the contact side, in (*b*) from the coil side.

for all four standards. Thus we have to deal with only one vision I.F.

The question now arises as to how the desired I.F. is to be obtained: should the frequency of the local oscillator be higher or lower than that of the vision carrier? As with the design of ordinary radio sets, the possibility of interference (e.g. between harmonics of the oscillator and harmonics of the desired or another transmitter) is here decisive. For the Belgian standards and the Gerber standard, the chance of interference is least when the oscillator frequency is higher.

For transmitters with these three standards and for about half the French transmitters, the sound carrier frequency is also higher than the vision carrier frequency; for the other French transmitters — including Lille in channel 8a, the very one of importance for Belgium — it is lower (see *Table II*). If the same picture I.F. amplifier is to be used, it is therefore necessary to choose the oscillator frequency lower than the vision carrier frequency in order to receive a French transmitter in the latter group. This occurs automatically with the channel selector: In choosing the channel, the oscillator coil is inserted which gives the correct frequency.

The next question is: for what I.F. must the vision amplifier be designed? From interference considerations, it appears that, for the Belgian and Gerber standards, the most favourable I.F. for the picture is 38.9 Mc/s. For the sound, this amounts to 38.9 — 5.5 = 33.4 Mc/s for these three standards (*fig. 4a*) and to 38.9 — 11.15 = 27.75 Mc/s for the French standard (fig. 4*b*).

A filter in every stage of the vision amplifier (*fig. 5*) ensures suppression of the sound signal, which is attenuated by a factor 500, and of signals from neighbouring channels. The bandwidth of approximately 4 Mc/s is obtained by somewhat staggering the circuits of successive stages. *Fig. 6* shows the selectivity characteristics. That for the French standard shows a slightly greater bandwidth: this is because one of the filters is tuned to 27.75 Mc/s, and not to 33.4, so that in the neighbourhood of 34 - 36 Mc/s the suppression is less and the band therefore wider.

Table II. Carrier wave frequencies for the Belgian, Dutch, German and French television transmitters. f_v = vision carrier frequency, f_s sound-carrier frequency.

	Gerber standard and Belgian standards			French standard		
Freq. band	Channel	f_v Mc/s	f_s Mc/s	Channel	f_v Mc/s	f_s Mc/s
I (low)	2	48.25	53.75	2	52.40	41.25
	3	55.25	60.75	3	56.15	67.30
	4	62.25	67.75	4	65.55	54.40
III (high)	5	175.25	180.75	5	164.00	175.15
	6	182.25	187.75	6	173.40	162.25
	7	189.25	194.75	7	177.15	188.30
	8	196.25	201.75	8	186.55	175.40
	9	203.25	208.75	8a	185.25	174.10
	10	210.25	215.75	9	190.30	201.45
	11	217.25	222.75	10	199.70	188.25
				11	203.45	214.15
				12	212.85	201.70

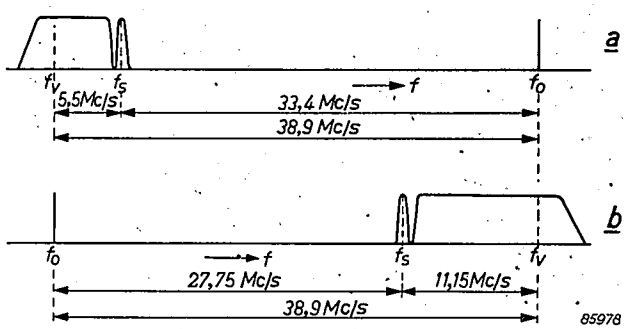


Fig. 4. *a*) Frequency spectrum for one of the Belgian standards and the Gerber standard. The sound carrier frequency f_s is 5.5 Mc/s higher than the vision carrier frequency f_v. A favourable value for the frequency f_o of the local oscillator is $f_v + 38.9$ Mc/s $= f_s + 33.4$ Mc/s.
b) The same for a French television transmitter such as Lille. Here $f_v - f_s = 11.15$ Mc/s. For $f_o = f_v - 38.9$ Mc/s, $f_s - f_o = 27.75$ Mc/s.

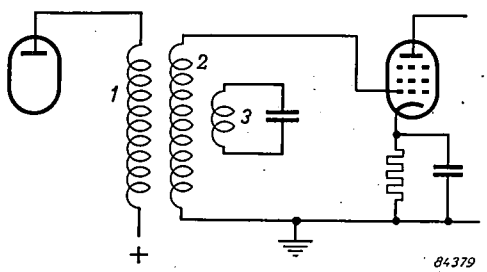


Fig. 5. Coupling between two stages of the vision I.F. amplifier. *1* primary, *2* secondary coil of the coupling transformer. *3* filter tuned to the frequency of an interfering signal.

Sound channel

We have just seen that in the sound channel we have to deal with two I.F. values, viz. 33.4 and 27.75 Mc/s. In amplifiers for frequencies of this magnitude the amplification per stage is limited by the grid-to-anode capacitance to a very small value. To obtain a considerable amplification, it is therefore necessary to have many stages. One is therefore not anxious to include two I.F. amplifiers for the sound channel in the set. One of these could be eliminated by transforming one I.F. signal to the frequency of the other. This, however, leaves another problem unsolved, viz. the design of a frequency detector (for the Gerber standard) which, at this high I.F.

not only suppresses amplitude variations well, but is also suitable for mass production. (In receivers for the Gerber standard alone, this difficulty is overcome by using the inter-carrier sound system [2]).

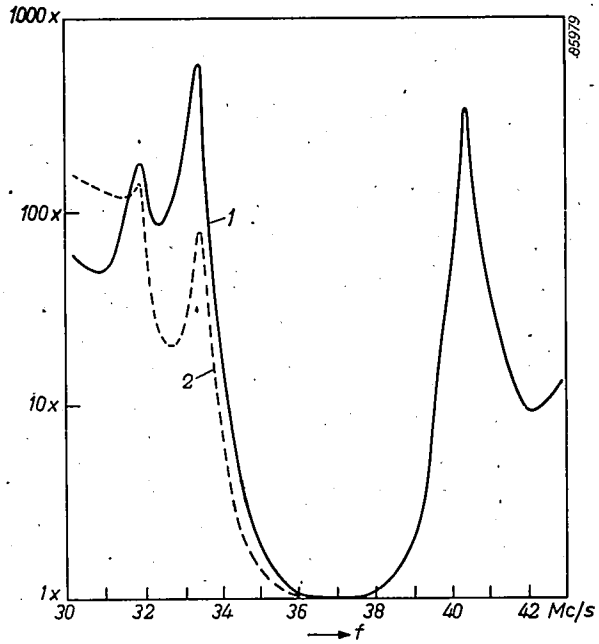


Fig. 6. Selectivity characteristic of the vision I.F. amplifier (MF_1 in fig. 7), *1* for the Belgian standards and the Gerber standard, *2* for the French standard.

Actually there is no reason why we should not use a second frequency transformation to bring both sound intermediate frequencies down to a much lower I.F. This second I.F. must, however, be very carefully chosen because the chance of interference is here very considerable. An analysis has shown that 7 Mc/s is the most suitable second I.F. — a value at which the frequency detector presents no particular problems.

To transform 33.4 Mc/s to 7 Mc/s, one can choose an oscillator frequency of either $33.4 + 7 = 40.4$ Mc/s or $33.4 - 7 = 26.4$ Mc/s. 40.4 Mc/s happens to be the more favourable. This frequency coincides with the sound carrier frequency of the higher adjacent channel (38.9 + 1.5 Mc/s); filters tuned to this frequency ensure that the amplification is reduced by a factor 200, which is sufficient to prevent interference. Of the harmonics, only the fifth, with the frequency $5 \times 40.4 = 202$ Mc/s, can interfere, but this lies on the sound side of channel 8 (table II) and can thus produce no interference in the picture. It is of importance that the frequency of the second oscillator be very stable, as otherwise multiples of it might assume values that lie in the pass-band of the vision I.F. amplifier.

[2]) See e.g. W. Werner, Philips tech. Rev. **16**, 195-200, 1954/55 (No. 7), in particular pp. 198-199.

To transform the sound I.F. of 27.75 Mc/s, which occurs in reception of a French transmitter, to 7 Mc/s, there is the choice of oscillator frequency between $27.75 + 7 = 34.75$ Mc/s and $27.75 - 7 = 20.75$ Mc/s. The former does not come into consideration because it falls in the band of the I.F. picture signal (see fig. 4*b*). Nine times the other value ($9 \times 20.75 = 186.75$ Mc/s) differs 1.5 Mc/s from the vision carrier frequency of Lille (185.25 Mc/s, see Table II), but in this case, too, filters in the amplifier suppress adequately.

Table III gives a summary of the chosen I.F. values and second oscillator frequencies.

Table III. Intermediate frequencies and second-oscillator frequencies in the four-standard receiver.

	Belgian standards and Gerber standard	French standard
I.F. picture	38.9 Mc/s	38.9 Mc/s
1st I.F. sound	33.4 Mc/s	27.75 Mc/s
Frequency 2nd oscillator	40.4 Mc/s	20.75 Mc/s
2nd I.F. sound	7 Mc/s	7 Mc/s

The fact that the receiver must be suitable for standards with an AM sound signal makes it necessary, 1) that the sound signal be tapped off immediately after the channel selector, and 2) that the sound signal be strongly suppressed in the vision channel, since the picture signal, which is always amplitude modulated, can easily suffer interference from a sound signal which is similarly modulated. This latter point has been given particular attention.

As regards the first item mentioned above, viz. the choice of the point at which the sound signal is tapped off, it would be preferable to place the second frequency-changer tube immediately after the channel selector; the switching in the sound I.F. section is then reduced as much as possible. An objection to this, however, is that the coupling between the second oscillator and the picture I.F. amplifier would then be so close that in the latter the 40.4 Mc/s signal could not be sufficiently suppressed and that an interference of $40.4 - 38.9 = 1.5$ Mc/s would remain visible. This difficulty is overcome by inserting an amplifier tube (MF_2, *fig.* 7) between the channel selector and the second frequency-changer tube, so that the coupling is considerably weakened. This extra tube must amplify signals of either of the sound intermediate frequencies (33.4 and 27.75 Mc/s). Switches operated by the standard selector tune the grid and anode circuits to one of these frequencies.

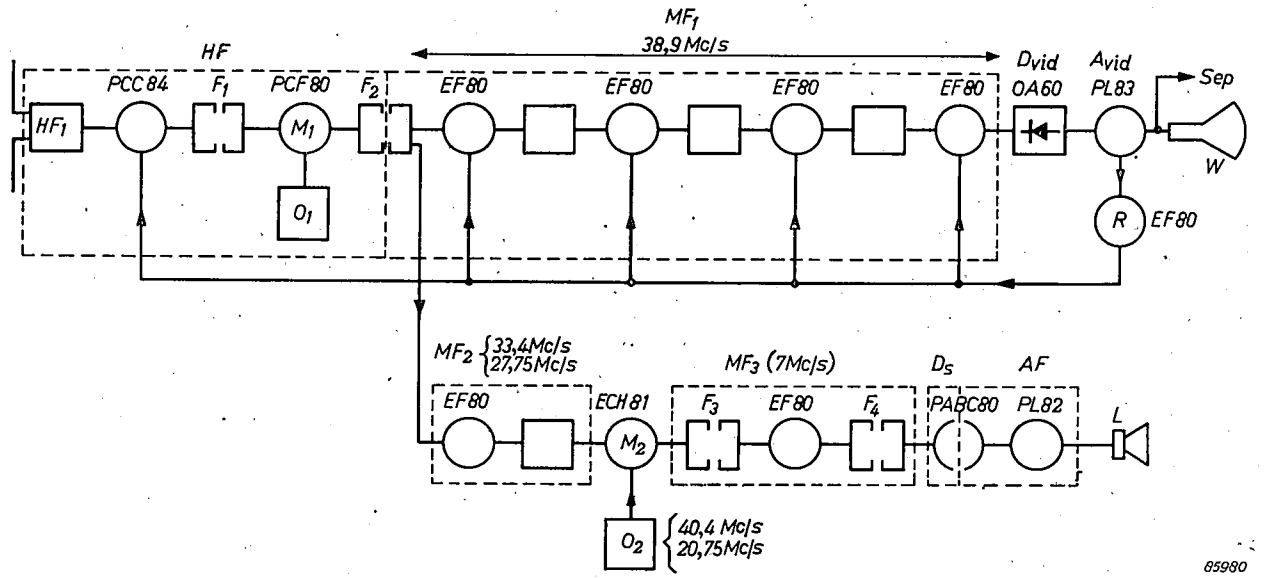


Fig. 7. Block diagram of the television receiver for four standards. *HF* high frequency section (incorporating the channel selector, see fig. 3), with high frequency stage HF_1, high frequency band filter F_1, first oscillator O_1, first frequency-changer M_1 and part of the I.F. band filter F_2.
Vision channel: MF_1 intermediate frequency amplifier with four tubes EF 80, D_{vid} video detector, A_{vid} video amplifier, R control tube (see further), W picture tube, *Sep* separation of the synchronizing signals.
Sound channel: MF_2 first I.F. amplifier, O_2 second local oscillator, M_2 second frequency-changer, MF_3 second I.F. amplifier (with band filters F_3 and F_4), D_S sound detector (with two diodes in the tube PABC 80), AF audio frequency amplifier (with the triode in the tube PABC 80 and a PL 82), L loudspeaker.

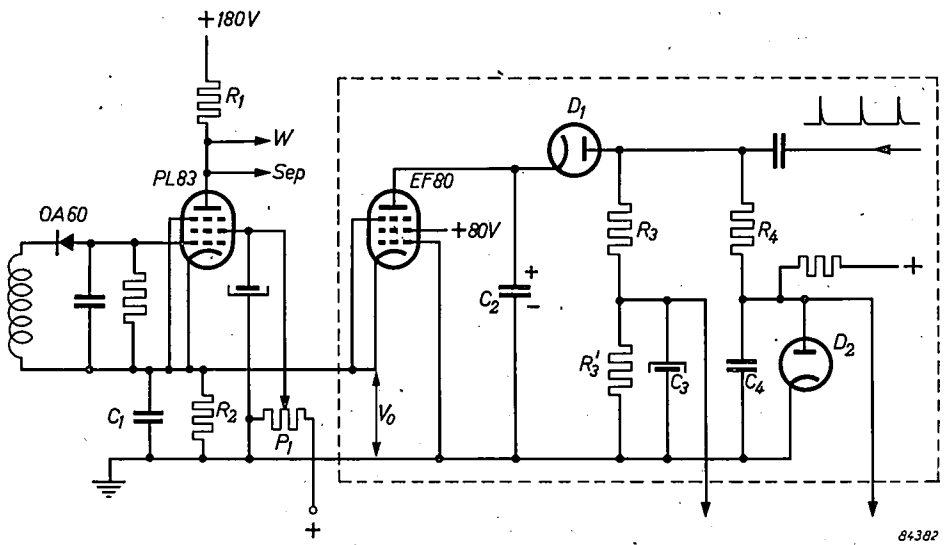


Fig. 8. Diagram of the video stage. Video detector is the germanium diode OA 60, video amplifier the pentode PL 83. The signal at the anode of the PL 83 controls the picture tube (W) and the separator (*Sep*), which separates out the synchronizing signals. P_1 contrast control.
Inside the dotted line: automatic gain control of the image channel. The control tube EF 80 receives a control voltage V_0, which appears across the cathode resistor R_2 (for I.F. current, short-circuited by the capacitor C_1) of the PL 83 Pulses originating in the line transformer (horizontal deflection) are rectified by the diode D_1 and supply (across the smoothing capacitor C_2) the anode voltage for the control tube. The control voltage for the I.F. tubes appears across the smoothing capacitor C_3; and the control voltage for the H.F. tube appears across C_4, when D_2 is rendered non-conducting by a strong signal.

To keep the switch-over simple, the coupling elements in the first sound I.F. amplifier are single tuned circuits. The second I.F. amplifier (7 Mc/s), which contains no switches, includes band filters.

The video section and the automatic gain control

The video signal is obtained by detection from the I.F. picture signal. The detector is a germanium diode OA 60 (*fig. 8*). The video signal is applied without bias to the control grid of a video amplifier tube, a pentode PL 83, a type which in this circuit with an anode supply of 180 V, can operate at zero grid voltage, as can occur in the absence of a signal. The tube is fed through an anode resistor; the amplified video signal thus appears with reversed polarity at the anode. This signal controls the picture tube.

According as the picture modulation is negative

(Gerber standard) or positive (the other three standards), the control grid and anode voltages of the video tube vary as shown in *fig. 9a* and *9b* respectively. It will be obvious that in order to avoid a negative picture on the screen, the signal must be inverted at some point in the video section when changing over from one modulation direction to another. In principle this can be done in the video detector, the video amplifier or the picture tube.

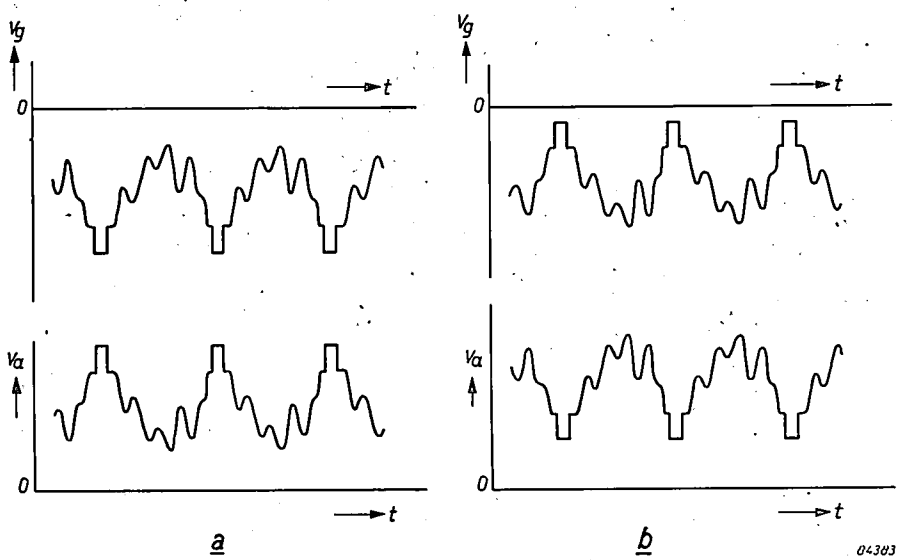


Fig. 9. Control grid voltage v_g and anode voltage v_a of the video tube PL 83 (fig. 8) as a function of time t, *a*) with negative modulation, *b*) with positive modulation.

Reversal at the picture tube has the advantage that no I.F. voltages are present there, and this, then, is the solution chosen. The circuit (*fig. 10*) is so arranged that for positive modulation the video signal is applied to the control grid and for negative modulation to the cathode of the picture tube, while in both cases, the brightness of the picture increases when the knob of the potentiometer P_2, which alters the bias of the picture tube, is turned to the right.

To avoid necessity of hand adjustment, it is desirable that the video signal automatically remains of the same magnitude when switching over from one transmitter to another which supplies a stronger or a weaker input signal. This purpose is served by the automatic gain control which, depending on the video signal strength, supplies the H.F. and I.F. tubes with a larger or smaller bias (control voltage) and so controls the amplification. We must now consider separately the cases of positive and negative modulation.

Negative picture modulation

The video signal on the control grid of the video tube PL 83, as represented in *fig. 11a* and *c*, results in a voltage across the cathode resistor R_2 (see fig. 8); this, in turn, acts upon the cathode of the control tube EF 80, whose grid is earthed. In the absence of a video signal, a large current flows through the video tube, producing across R_2 a direct voltage V_0 which lies far below the cut-off point of the control tube (fig. 11*b* and *d*). Furthermore, the circuit is such that when a video signal of the desired magnitude is present, only the synchronization pulses produce anode current pulses in the control tube. The A.C. component of these pulses passes through the smoothing capacitor C_2, while the D.C. component across the resistors R_3 - R_3' (fig. 8) gives rise to a direct potential across the smoothing capacitor C_3, which can be used as control potential for the I.F. tubes. Interference peaks (*i*, fig. 11*c*), which appear as black spots in the picture and can disturb the synchronization, are clipped by the video tube whenever they exceed the amplitude of the synchronization pulses.

The output voltage of the video tube can be varied simply by adjustment of the screen grid voltage of this tube. This is done by the potentiometer P_1 (fig. 8), which thus regulates the contrast. Adjustment of P_1 alters the cathode voltage V_0 which occurs in the control tube in the absence of a signal, and at the same time the threshold which the signal must exceed before the automatic gain control comes into action.

The H.F. tube PCC 84 (fig. 7) derives its control voltage from the smoothing condenser C_4 in fig. 8. Normally this voltage is zero (and the amplification thus a maximum, which is advantageous from considerations of noise) because the

diode D_2 passes current. Only when the input signal is very strong does there occur via R_4 and C_4 a negative control voltage which makes the diode non-conducting and causes the H.F. amplification to decrease.

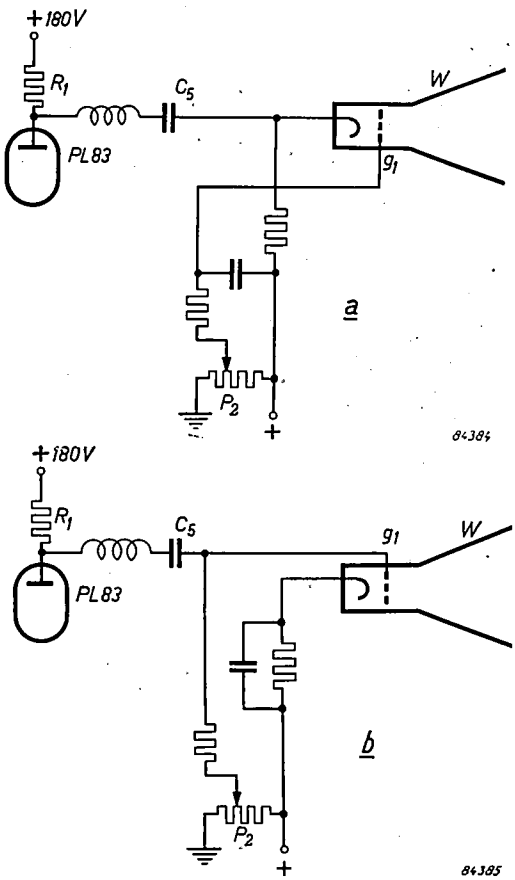


Fig. 10. Control of the picture tube W (with control grid, g_1), a) with negative modulation, b) with positive modulation. A double-pole switch (not shown) makes the change from one to the other. R_1 anode resistor of the video tube PL 83. C_5 coupling condenser. P_2 brightness control.

Positive picture modulation

With negative modulation there is in the peak of the synchronization signal a level which is independent of the picture information: it is thus convenient to use this level as the basis for the automatic gain control, which, as we have just seen, has been done. With positive modulation, however, the only level in the picture signal independent of the picture information is the black level, which is present as a "front porch" and "back porch" at 30% of the maximum white level. Circuits do exist to derive the control voltage from this level, but they are complicated and do not lend themselves well to switching-over into a circuit suitable for negative modulation.

However, we have established that the circuit discussed above can also serve very well for positive modulation. As we have seen, with this circuit the highest peaks of the video signal lie close to the cut-off point of the video tube (fig. 11c). With negative modulation, the peaks are those of the synchronization signal; with positive modulation they are the peaks which correspond to the whitest part of the picture.

This white is not, of course, constant from picture to picture and it may therefore seem strange to base a control system on a variable peak. From a great number of tests, however, it has been shown that, taken over not too short a time, the white level occuring in the various scenes does not vary appreciably. Only in an extreme case of a transmitter sending out a completely black "picture", would this be observed on the screen as white (at least, if the picture tube were fed with the D.C. component of the video signal; in fact this does not happen — see fig. 10 — and the black "picture" would be seen as grey).

The situation with positive modulation is illustrated in fig. 11e (video tube) and 11f (control tube). Interference peaks, which can be particularly tiresome with positive modulation since they appear as very bright, large white spots (see p. 196 of the article referred to in [2])), are clipped whenever they exceed the level of the brightest white in the picture.

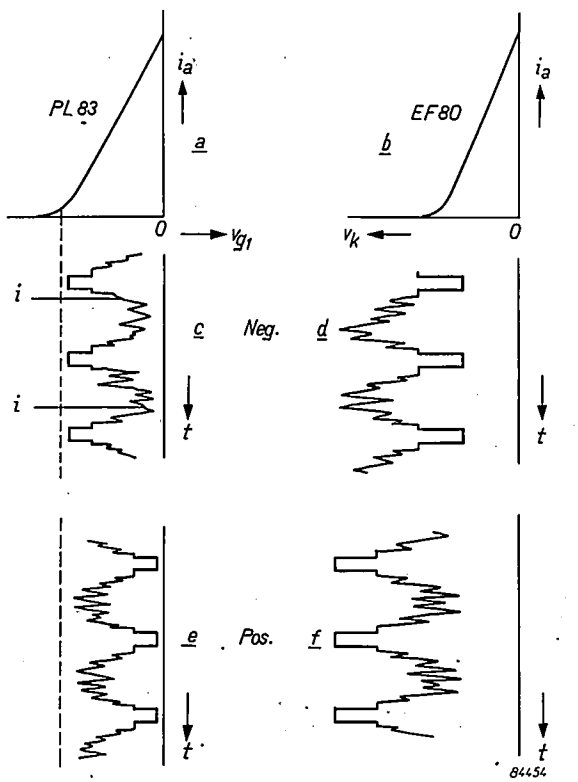


Fig. 11. a) Anode current i_a as a function of the control grid voltage v_{g1}, for the video tube PL 83. b) i_a as function of the cathode voltage v_k (with earthed grid), for the control tube EF 80. c) Control grid voltage of the PL 83, and d) cathode voltage of the EF 80 as a function of the time t, with negative picture-modulation; e) and f) the same for positive picture-modulation.

The sound detector

For the detection of the I.F. sound signal, we have to deal with the fact that of the four standards, three prescribe amplitude modulation (AM) and one frequency modulation (FM). Starting with a conventional FM detector, the so-called ratio detector (*fig. 12a*), and adding to it only a few

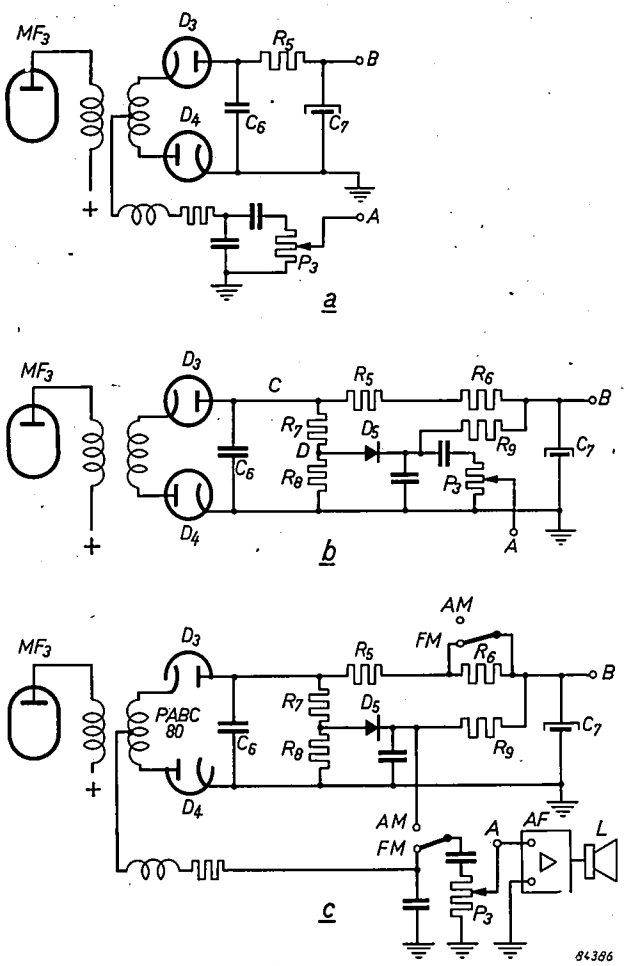


Fig. 12. *a*) Circuit of the ratio detector for FM. *b*) Circuit *a* modified to form an AM detector. *c*) Combination of both circuits, with two single-pole switches FM-AM.
MF_3 intermediate frequency amplifier. D_3 and D_4 are diode sections of a PABC 80, the triode section of which serves as first stage of the A.F. amplifier *AF*. D_5 germanium diode, which acts as interference limiter in the AM detector. P_3 volume control. The control voltage for the automatic gain control is taken off at *B*.

circuit elements and two switches, it has been extended to give AM detection as well. In both cases, besides the audio frequency signal, the detector delivers a direct voltage which is directly related to the amplitude of the I.F. sound signal and which can therefore be used as a control voltage for the I.F. amplifier in the sound channel.

Another well-known FM detector is that of Foster and Seeley [3]), which is preceeded by a limiting device to suppress amplitude variations and interference peaks; the limiting device is usually a pentode with strong input signal [4]). This method is less suitable for television than the ratio detector, however, since the limiter produces harmonics whose frequency may fall in the television channel and cause interference.

The ratio detector itself acts as a limiter, and for strong signals, the preceeding tube as well.

Fig. 12*b* shows the circuit of fig. 12*a*, modified to give AM detection. In both cases the detected signal appears across the capacitor C_6 and the control voltage across the electrolytic capacitor C_7; the potentiometer P_3 is the volume control. At the same time care has been taken that with AM, strong interference peaks are limited. In fig. 12*c* can be seen how the change-over is effected from the one circuit to the other; the two single-pole switches needed for this purpose are included in the standard selector.

The interference limiter of the AM detector works as follows. The anode of a germanium diode D_5 (fig. 12*b*) is connected to the tapping point *D* of the voltage divider R_7-R_8, and the cathode via the resistor R_9 to the point *B*. The detected AM signal appears across R_7 and R_8, i.e. an audio frequency fluctuating direct voltage. The potentials of the points *C* and *D* follow the curves *C* and *D* respectively in *fig. 13*. The direct voltage component of the detected signal occurs across the capacitor C_7 (the straight line *B* in fig. 13). It is seen that in the absence of interference, the point *D* is always positive with respect to the point *B*, so that the diode D_5 always passes current and the A.F. signal really does appear on the potentiometer P_3. Interference peaks such as *i*, however, which make *D* negative with respect to *B*, make the diode momentarily non-conducting and are thus limited to a specified amplitude.

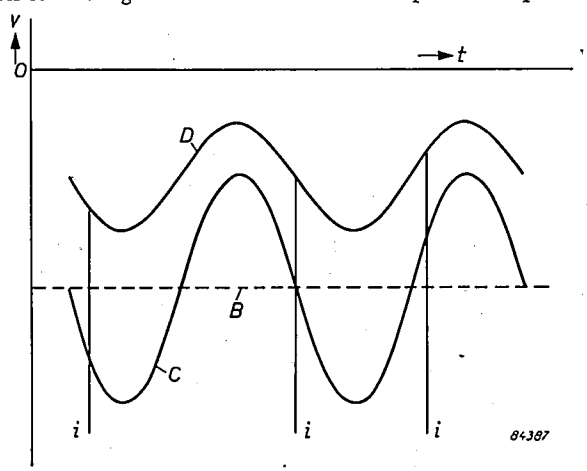


Fig. 13. The straight line *B* and curves *C* and *D* give the voltage variations at the points *B*, *C* and *D* respectively in fig. 12*b* with respect to earth (*0*). Interference pulses are clipped at the level *B*.

The deflection circuits

From the anode of the video tube (PL 83, fig. 8) the signal travels not only to the picture tube, but also to the circuit which separates out the synchroni-

[3]) See e.g. Philips tech. Rev. **8**, 48, 1946, fig. 10, or **11**, 4, 1949/50, fig. 5.

[4]) See e.g. Philips tech. Rev. **11**, 3, 1949/50, fig. 4*a*.

zation pulses. With negative picture modulation the synchronization pulses in the video signal are in the positive direction and the separation can thus be achieved by means of a tube which is so adjusted that the picture signal lies wholly in the cut-off area and only the synchronization pulses produce anode current pulses. With positive picture modulation, the video signal and synchronization pulses are in the negative direction; in this case an inverter tube is inserted between the video tube and the separator tube to reverse the polarity. The inverter stage only requires a small bandwidth, since only the synchronization pulses are of importance here.

parameters are such that the pentode conducts momentarily at intervals. While the pentode is non-conducting, the capacitor C_8 is charged via the anode resistor R_{10} and a resistor R_{11}. When the pentode conducts, C_8 discharges itself rapidly through this section of the tube and the resistor R_{11} ($< R_{10}$). The voltage across C_8 - R_{11} thus follow the form shown in fig. 14 on the right hand side. This variation is necessary for the control voltage of the final tube in the deflection circuit: the upward sloping section generates a linearly increasing current in the deflection coils and the large negative peak (from -100 to -150 V) cuts off the final

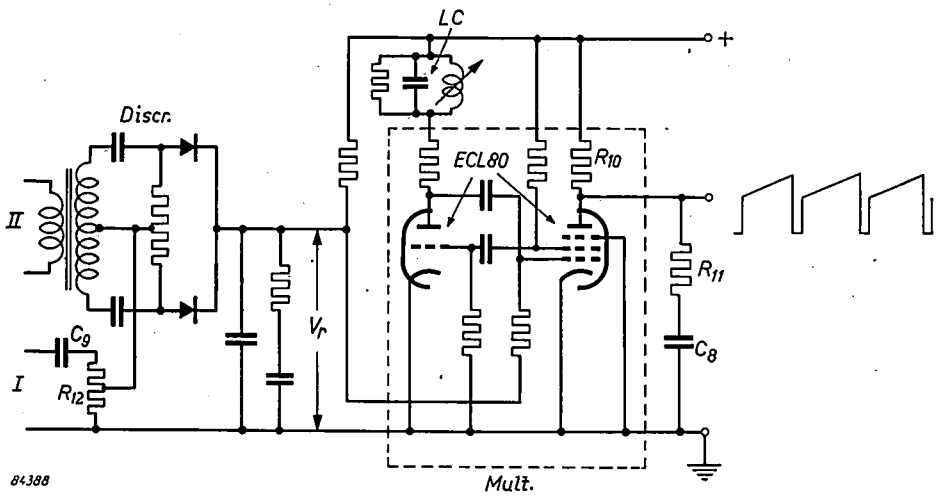


Fig. 14. Generation of the saw-tooth voltage for controlling the final stage for the horizontal deflection. *Mult* multivibrator, whose output voltage (of the form drawn on the right) appears across R_{11} - C_8. *LC* flywheel circuit, *Discr* phase discriminator, producing the control voltage V_r which determines the frequency of the multivibrator. *I* is fed with a voltage of the frequency of the multivibrator, *II* with the synchronizing pulses. C_9 - R_{12} differentiating network.

The frame synchronization pulse is derived in the usual way from the synchronization signal. Despite the fact that the French frame synchronization signal differs from the others (see table I), both forms of signal have proved to give good synchronization and interlacing. The vertical deflection in the four-standard receiver thus presents no further problems and we shall discuss it no further.

The horizontal deflection must be able to work at two line frequencies: $25 \times 625 = 15\,625$ c/s and $25 \times 819 = 20\,475$ c/s. The choice of the type of saw-tooth generator was influenced by the wish to reduce the switching to a minimum. A multivibrator meets this requirement best.

The multivibrator circuit used (*fig. 14*) works with two triodes. One is formed by the triode section of a triode-pentode ECL 80, the other by the cathode, first and second grids of the pentode section of the same tube. The circuit tube during the retrace, when its anode voltage is very highly positive.

For the horizontal deflection with negative picture modulation, so-called flywheel synchronization [5]) is necessary. With this modulation direction, the interference has the same direction as the synchronization, pulses and thus, with direct synchronization, could cause the saw-tooth generator to change over at the wrong moment. With positive modulation, this danger does not exist, because the interference is in the white direction, but here too, flywheel synchronization is of advantage. When the signal-to-noise ratio is small, the noise, superimposed upon the (not infinitely steep) sides of the synchronization pulse, could give rise to slight phase changes in the change-over of the saw-tooth

[5]) See e.g. P. A. Neeteson, Flywheel synchronization of saw-tooth generators in television receivers, Philips tech. Rev. **13**, 312-322, 1951/52; Television receiver design, Monograph 2: Flywheel synchronization of saw-tooth generators, Philips Technical Library, 1953.

generator, resulting in a "raggedness" of vertical lines. A flywheel system can improve the situation considerably. In the receiver under discussion this system has therefore been used for all four standards.

The flywheel system works, briefly, as follows. By means of a phase discriminator, a control voltage is derived from phase differences between the line saw-tooth (or pulses derived from it) and the line synchronizing pulses. This control voltage is then caused to act upon the saw-tooth generator in such a way that the frequency difference is reduced, even down to zero, owing to the fact that the phase discriminator reacts to the phase difference, which is the time integral of the frequency deviation ("integrating control"). The "flywheel" consists of an LC circuit tuned to the line frequency f_l or an RC combination with a time constant $1/f_l$. This has the effect that interference pulses remain ineffective (except in the very unlikely case that a long series of pulses occurs with the frequency f_l).

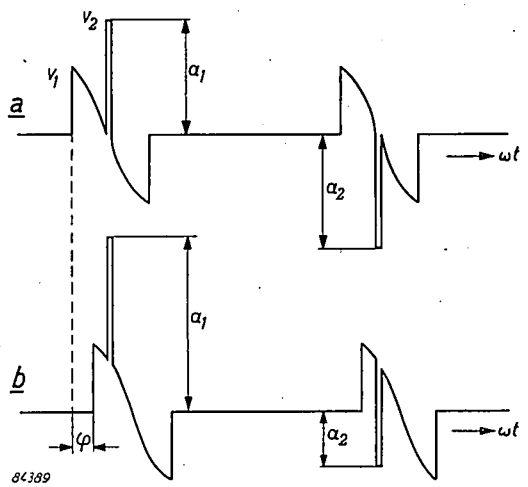


Fig. 15. Production of the control voltage by the phase discriminator. v_1 pulses of the frequency of the multivibrator, v_2 synchronizing pulses. In (a) the phase difference between the two series of pulses is zero; since amplitude a_1 = amplitude a_2, the control voltage is also zero. In (b) a frequency drift of the multivibrator has caused a phase difference of φ between the pulses v_1 and v_2, so that the amplitudes a_1 and a_2 are no longer of equal magnitude. This results in a particular value of the control voltage V_r which is approximately proportional to φ. The control voltage brings the frequency of the multivibrator back to the correct value.

In the four-standard receiver the flywheel is an oscillatory circuit (LC in fig. 14) which, by means of a tapping in the coil, can be tuned at will to either 15 625 c/s or 20 475 c/s (furthermore in the multivibrator one resistor and one capacitor must be changed). The synchronizing knob on the set changes the distance between a copper ring and the coil and so adjusts to synchronization.

The phase discriminator is to be seen at the left in fig. 14. It is fed at I with a voltage of the frequency of the multivibrator (derived from the line transformer), and at II with the synchronizing pulses. A differentiating network transforms the former voltage into a voltage v_1 of the form shown in *fig. 15*. The discriminator comprises two diodes. These are so connected that they deliver together the control voltage V_r, which is zero when the mid-point of the sloping side of v_1 coincides with a synchronization pulse (v_2) and assumes a positive or negative value when coincidence does not occur. This control voltage, as a bias on the control grid of the pentode, influences the frequency of the multivibrator and thus reduces the frequency deviation to zero.

Summary. The need in Belgium and the surrounding areas for television receivers suitable for two or more standards, has led to the design of a four-standard receiver, which has been brought on to the market in various types (e.g. type 17 TX 100A-70). Apart from the usual channel selector, the set is fitted with a standard selector, which gives a choice between the two Belgian standards, the French and the "Gerber" standard.

The vision channel contains for all four standards one and the same I.F. amplifier (38.9 Mc/s, bandwidth 4 Mc/s). Since both positive and negative picture modulation occurs, the video signal must be inverted somewhere; this is done at the picture tube. For the automatic gain control, a control tube derives a control voltage from the peaks of the synchronization pulses with negative modulation and from the peaks of the white in the picture with positive modulation.

The choice of 38.9 Mc/s as the I.F. for the vision amplifier leads to the following I.F.s in the sound channel: 27.75 Mc/s for the French standard and 33.4 Mc/s for the other standards. A second frequency transformation is used to bring these values down to a much lower one, viz. 7 Mc/s. The detector for FM is a conventional ratio detector, modified in such a way that it can be easily switched over to an AM detector with interference limiter.

The horizontal deflection is synchronized by means of automatic phase control by a flywheel circuit and the vertical deflection in the usual manner by means of an integrating circuit.

A COMPARISON BETWEEN REPRODUCED AND "LIVE" MUSIC

by R. VERMEULEN.

534.76:534.86:681.84.087.7

Endeavours to realize fidelity in the transmission of music are of long standing. The subject aroused interest as long ago as 1881, when Parisians were given the opportunity of listening to telephone transmissions from the Grand Opera via an installation designed by Ader. Although still very imperfect, the installation had already one modern refinement: it was equipped for "binaural" hearing. From these beginnings stereophonic reproduction has been developed, nowadays generally recognized as essential to the natural reproduction of music.

The prerequisites for the natural reproduction of music have recently been re-examined in this laboratory. From alternate performances of "live" music and music reproduced with the most modern equipment in the same hall, it was found that in many cases the listeners could hear no difference between them and sometimes systematically confused the two.

Why is it that in spite of all technical progress in electro-acoustics, it is still possible to distinguish between the music heard from a loudspeaker and that heard in the concert hall? Many will have their answer ready to this question. They will point out that the music is distorted in the various links of the reproduction channel: microphones, amplifiers, tape recorder or gramophone, and above all by the loudspeakers. These devices fall short in the reproduction of very low and very high notes, and, moreover, they introduce alien sounds. The dynamic range is restricted on the one hand, by hum and noise, and, on the other, by combination tones, which arise from overloading.

It cannot be denied that even in the best reproductions of to-day these distortions are still present. Nevertheless we doubt whether our question can be answered by simply laying the blame upon the shortcomings of electro-acoustical apparatus. The possibility of measuring certain shortcomings objectively and quantitatively (such as the lack of high overtones) has been a great stimulus for improvement. But the uncertainty regarding the permissible magnitude of these imperfections makes it very tempting to regard them as the only cause of the musically not entirely satisfactory result. The danger then is that the electrical engineer will treat the electro-acoustical instrument as a link between a signal generator and a voltmeter, and impose requirements upon it which, from a musical point of view, are meaningless, or even erroneous. A typical example of such a misconception was the tendency, very prevalent among technicians at the time but now discarded, to regard hiss as a criterion for a good reproduction of high tones, and thus to consider a high noise level as a favourable rather than as an adverse characteristic.

The "hole in the wall"

It may therefore be wise to look for other answers to the above question. Are we sure, for example, that a reproduction channel with no measurable technical defects will be able to create the illusion that an orchestra is playing actually in the room? Might there not be other aspects, so far neglected, which impair musical appreciation more than minor technical imperfections?

Some investigation into the problem shows that it is possible to give a satisfactory reproduction of a single, small, spatially concentrated sound source, such as a human voice or one small instrument, such as a clarinet, on condition that it is reproduced at the original volume. With a small ensemble and especially with a large orchestra, however, something is lacking in the reproduction as heard from the loudspeaker. The reason is that even a perfect loudspeaker can do no more than imitate the vibrations picked up by the microphone, and thus the best result will only be equivalent to a hole in the wall of the concert hall. The sound that such an opening transmits is absolutely free from all electrical and mechanical distortion, and should therefore certainly be designated as "super-highest fidelity". Nevertheless, the concert-goer who has arrived too late and who has had to listen to the beginning of the concert through a chink in the door, is relieved when he can enter the concert hall. Thus, irrespective of technical imperfections, there is evidently something missing with music, originating from an extensive sound-source, that reaches our ears via a small opening.

It is now generally known that this lack of auditory perspective can be rectified by means of stereophonic reproduction. A number of articles on this subject have appeared in this Review, in 1939

and in later years. Before describing some comparative tests made with stereophonically reproduced music and "live" music, we may usefully recapitulate the principles of stereophonic reproduction. We shall start with its predecessor: binaural reproduction.

Binaural hearing

To improve loudspeaker reproduction it is firstly necessary to overcome the "hole-in-the-wall" effect. This can be done by using two microphones instead of one. The microphones may be placed as ears on an "artificial head" (*fig. 1*) and connected to a pair of headphones in such a way that the left ear hears the sounds picked up by the left-hand microphone and the right ear the sounds picked up by the right-hand microphone. In this way "binaural" hearing, i.e. with two ears (*fig. 2*), is restored. True, there are only slight differences between the sound heard in the left earpiece and that in the right, but they are quite sufficient to give the listener the impression that he is seated at the place where the artificial head is set up. He may, for instance, easily be given the sensation that he hears persons moving and talking behind him, although there is no one there. The impression may be so realistic that the listener has to turn round to convince himself that there is really no one there. But as soon as he does so, the shortcomings of this system become evident, viz: the whole acoustical world rotates with his head. To avoid this, one might arrange for the artificial head to turn with the listener's head, and tests have confirmed that this does in fact overcome the deficiency [1]); moreover, the listener can then distinguish between sound-sources in front of him and behind him, which he cannot do if he keeps his head still. The disadvantages of this otherwise ideal solution are, however, obvious: headphones are in themselves a nuisance — and the coupling of the

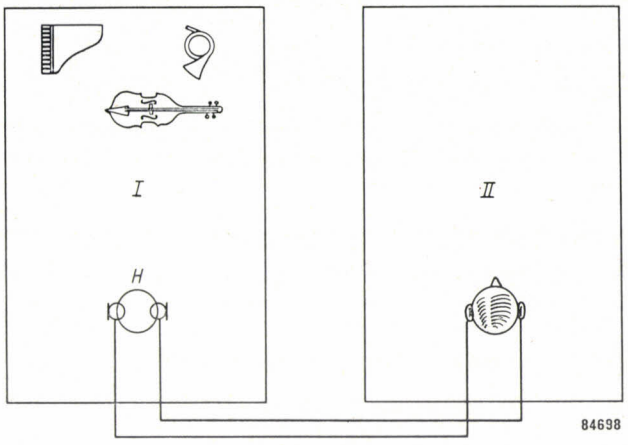


Fig. 2. Binaural hearing. The music played in room *I* is heard by the listener in *II* via headphones, each ear-piece of which is connected to a microphone. He thus receives an impression of auditory perspective, The microphones should preferably be fitted on an artificial head (*H*).

[1]) K. de Boer and A. Th. v. Urk, Philips tech. Rev. **6**, 359-364, 1941, in particular, pp. 360, 361.

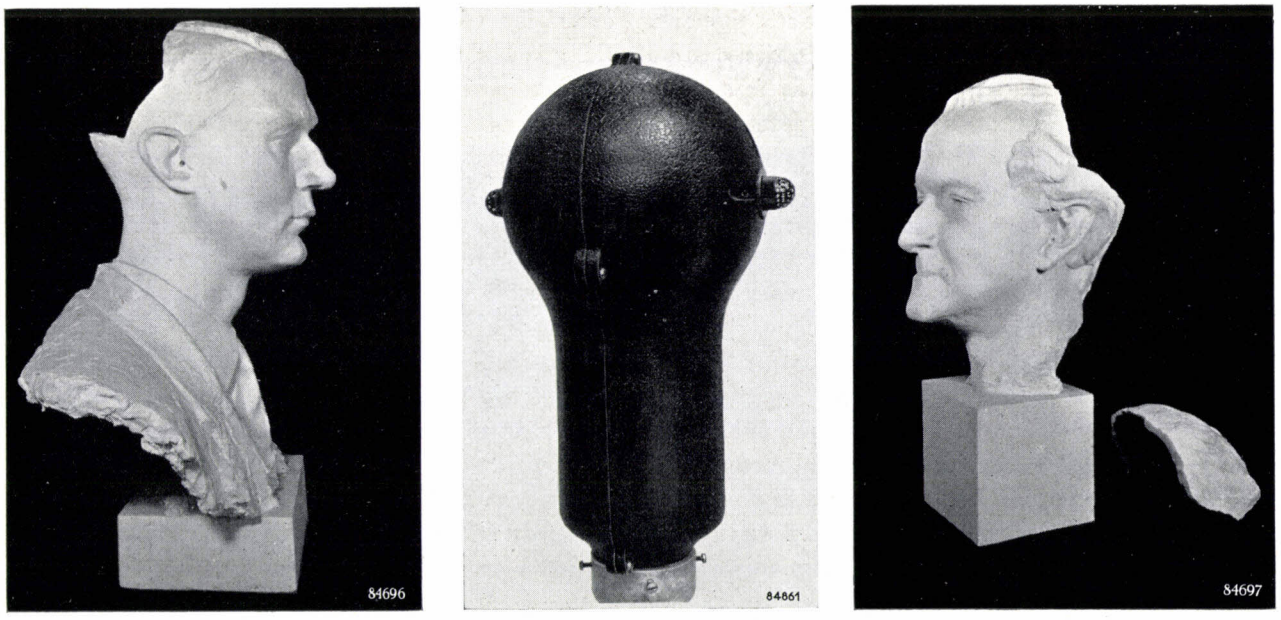


Fig. 1. Plaster heads, dating from the early days of stereophony and intended as experimental artificial heads. (They represent two workers who have contributed to the development of stereophony: on the left, K. de Boer; on the right, A. T. van Urk.) It was soon found that a sphere constitutes an adequate approximation to the human head. The photograph in the middle shows a modern artificial head.

artificial head with the listener's head would be quite impracticable. Nevertheless, these experiments were very instructive, showing as they did how essential is our binaural hearing to the sound impression received.

On the manner in which binaural hearing enables us to determine the direction of the sound-source, there have long been differences of opinion. It has been demonstrated by experiments that a time difference between the signals reaching the left and those reaching the right ear produces a sensation of direction, but others have shown that this is likewise the case with a difference in intensity. K. de Boer [2]), who has made a thorough study of the subject in this laboratory, was able to confirm that both parties were right, that is to say that differences both in time and intensity contribute to directional hearing. The remarkable thing is that these contributions are additive: the angle from which a sound seems to come, owing to a time difference, becomes greater or smaller as the effect of a superimposed intensity difference works in the same or in the opposite direction. It is even possible to let the two effects compensate each other; thus, a sound-source apparently heard from a certain angle as a result of a time difference, can be made to "return" to the centre by means of an opposing intensity difference.

Stereophony

A second remarkable effect noted by K. de Boer is that the sound stimuli which the ears receive from two loudspeakers, placed some yards apart and each connected to a microphone on an artificial head, are mentally interpreted as a single apparent sound-source, which appears to lie in between the two real ones. This helped to clarify the mechanism involved in obtaining three-dimensional acoustic effects using two loudspeakers instead of headphones (*fig. 3*). Stereophonic reproduction giving the impression that the sounds come from various directions was achieved by Fletcher and Stokowski as long ago as 1933 [3]).

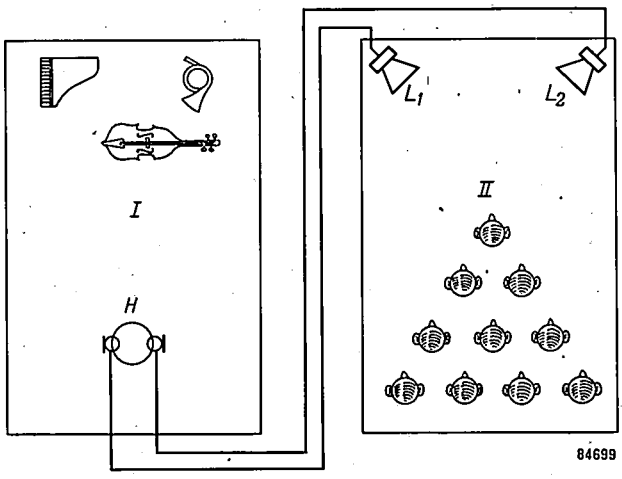


Fig. 3. Stereophony. Each microphone on the artificial head (H) in hall I is connected to its own loudspeaker (L_1, L_2) in hall II. Here too, the audience in II receives a spatial impression of the sound.

The explanation of the effect given at the time, however, is not entirely convincing. It may be briefly restated as follows. Suppose that in a concert hall a curtain is hung which is impervious to sound and which is provided, at the side facing the orchestra, with a large number of microphones. Each microphone is connected to a loudspeaker at the other side of the curtain (*fig. 4*). Sound waves picked up by the microphones are then radiated at the other side by the loudspeakers, and thus proceed as if the curtain were not present. If such a curtain with loudspeakers is set up in another hall, the same sound field will be set up as behind the curtain in the first hall. As there is a limit to the number of microphones, transmission channels and loudspeakers which can be installed, one has to make do with a rough approximation, for which three microphones and three loudspeakers were found to be sufficient.

Listening to the result of such an arrangement, and moreover on learning that the result is particularly good when using only two channels (two microphones and two loudspeakers), it is difficult to understand, if the above explanation were complete, how such a good approximation is obtained with so few channels. In our opinion we cannot leave out of consideration the psychological phenomenon that the four sound-stimuli received by both ears from the two loudspeakers are interpreted as coming from one single source [4]). It is thus not necessary to produce a

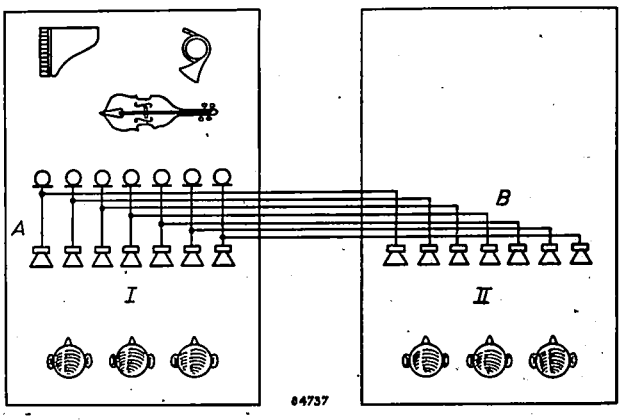


Fig. 4. Explanation of stereophony according to an American view. A is an imaginary "curtain", one side being fully taken up by microphones, the other by loudspeakers. Each microphone is linked via its own channel to the corresponding loudspeaker at the other side of the curtain, and also to a similar loudspeaker on "curtain" B in hall II. Thus, the same sound field is produced in hall II as in hall I. Practical stereophony, necessarily using only a small number of channels, would only be a rough approximation to the ideal case outlined.

[2]) K. de Boer, Stereophonic sound reproduction, Dissertation, Delft, 1940.

[3]) Symposium on wire transmission of symphonic music and its reproduction in anditory perspective: H. Fletcher, Basic requirements, Bell Syst. tech. J. **13**, 239-244, 1934.

[4]) K. de Boer, Stereophonic sound reproduction, Philips tech. Rev. **5**, 107-114, 1940.

replica of the sound in the room; it is enough if we aim at supplying the two ears with a set of signals that are perhaps different from the original ones but still create the same impression. According to the earlier explanation, three channels would give a better approximation than two, four a better approximation than three, and so on, whereas our experience is that two channels give a clearer and, above all, a "sharper" sound image than three. To avoid misunderstanding we should add that the use of more than two loudspeakers can still be advantageous, e.g. in order to produce a stereophonic effect for a *large* audience.

If the two microphones are mounted without the artificial head between them, the differences of intensity between the signals which they pick up become much smaller, these differences having been mainly the consequence of the shielding effect of the artificial head. The stereophonic effect must now rely primarily on the time differences. To suggest the same direction these must be strongly exaggerated, which can be done by making the distance between the microphones about three times as large as in the artificial head. It is surprising that the ear is able to interpret as directions such large time differences, although it can never have had the opportunity to acquire the faculty. In natural directional hearing the time difference is always less than 0.6 milliseconds and contributes only about 10% to the perception of direction. An artificial head thus supplies less abnormal signals than two separate microphones, and moreover, as a compact unit, it is easier to handle. The size of the artificial head is only in special cases the same as that of the human head. It can best be chosen in accordance with the set up and the size of the orchestra, following this empirical rule: if φ is the angle (in degrees) subtended at the head by the orchestra (*fig. 5*), then the horizontal diameter of the artificial head must be $(2000/\varphi)$ cm, and the distance between two freely-mounted microphones must be $(6000/\varphi)$ cm [5]).

In the reproduction of music the *conscious* perception of direction, as made possible by stereophony, does not play a very significant role: after all, for the audience in a concert hall, appreciation of the performance is not critically dependent on the precise positioning of all the instruments. It is therefore not so very important that the directions are reproduced accurately. The reason why stereophonic reproduction gives a significant improvement is that the instruments are heard distinct from each other in space instead of as a muddle of sounds. Our "hole in the wall" has now become as it were a large window that enables the listener to survey aurally the whole of the orchestra.

[5]) K. de Boer, The formation of stereophonic images, Philips techn. Rev. 8, 51-56, 1946.

Another remarkable fact is that with stereophonic reproduction one can concentrate effortlessly on sound from a certain direction and detach one's attention from unwanted sounds (noise, hiss, reverberation, etc.) from other directions. It is striking how loud the noise in the studio seems when heard from one loudspeaker, how much more aware one becomes of reverberation, and how soon sounds become unintelligible when two or more persons are talking at the same time, whereas these phenomena are hardly noticed if one is in the studio oneself or if one listens to a stereophonic transmission.

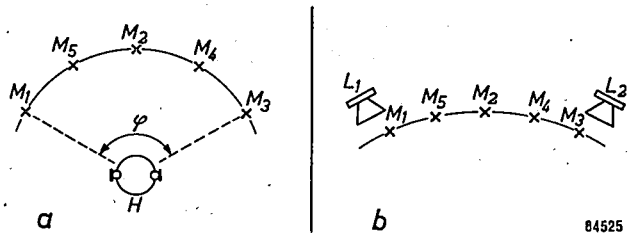


Fig. 5. *a*) The artificial head H "sees" an ensemble of musicians at a subtended angle φ. The most favourable horizontal diameter of the artificial head is $2000/\varphi$ cm (φ in degrees); the acoustic representation of the five musical instruments $M_1 \ldots M_5$ is then as shown in (*b*), between the loudspeakers L_1 and L_2.

Room acoustics

May we now expect that the stereophonic reproduction of orchestral music will be indistinguishable from the original — assuming, of course, that every attention has been paid to all details of the electro-acoustical installation?

This will certainly not be the case if the reproduction takes place in a hall or room whose acoustic properties are inadequate. It is generally recognized that the acoustics of a concert hall constitute an important element in the appreciation of music; we cannot therefore expect the imitation of an orchestra to do without this support.

It might be presumed that certain shortcomings in the acoustics of the room where the music is to be reproduced could be rectified by resorting to electro-acoustical aids. We shall leave this subject to a subsequent paper and proceed now to describe tests in which "live" and reproduced music were compared together. As both were played in the same hall, the influence of room acoustics was excluded.

Comparative tests

In order to ascertain objectively to what extent a stereophonic installation is capable of imitating actual musical instruments, we invited 300 persons to the Philips Laboratory to make a comparison

between the reproduction of stereophonically recorded pieces of music and the same pieces played by a small ensemble, seated behind a thin but opaque curtain. The greatest care was spent on the recording of the music, so that it would really be the best possible replica of the ensemble, particular attention being paid to the setting up of the artificial head. Naturally, the reproduction had to be just as loud as the actual music. The entire reproduction apparatus remained in operation all the time in order to prevent switching clicks and perhaps a change, however small, in the level or the character of the room-noise, from giving some listeners a clue. The recordings (on tape) were therefore made with long periods distributed arbitrarily between them; during these blank periods, the tape still running, the live music was played by the musicians.

The procedure was a follows. The same, short piece of music (15 to 30 seconds) was rendered twice, both stereophonically and by the musicians, but in arbitrary order, each piece of music being indicated in chronological order as "reproduction A" and "reproduction B" ("reproduction" thus includes the live performance. The listeners were not aware that "live" music would be played). Immediately afterwards, there followed, as "reproduction X", a repetition of either A or B. The listeners then had to answer, within about one minute, the following questions presented to them on a questionnaire:

I) Was X the same as A or the same as B?

II) Which of the two reproductions, A or B, was the more natural?

The first question merely requires the listener to hear a difference between the actual and the reproduced music, whereas the second question requires that he should moreover have an idea of how "live" music sounds. At each session, ten pieces of music of different kinds (chamber music, dance music) were played, the number of musicians varying slightly.

After the experiments were concluded, the right answers returned were marked with the number 1 and the wrong ones with 0 and the means taken of the totals. The result was 0.75 for the first question and 0.71 for the second question. If it is borne in mind that an average of 1 for the first question would mean that every participant in the experiment always heard a difference between "live" and reproduced music, and that a result of 0.5 means that no participant heard any difference whatsoever and everybody was therefore guessing, it appears that the result obtained (0.75) lies midway between these two extremes. Thus, the average of 75 correct answers out of a hundred cases can be comprised of 50 correct "decisions" and 25 lucky guesses. Relatively few persons (16%) were able to hear the difference infallibly, and even they found it rather difficult. To recognize the "live" music as the more "natural" appeared to be even more difficult, as was shown by the average of 0.71 for the second question.

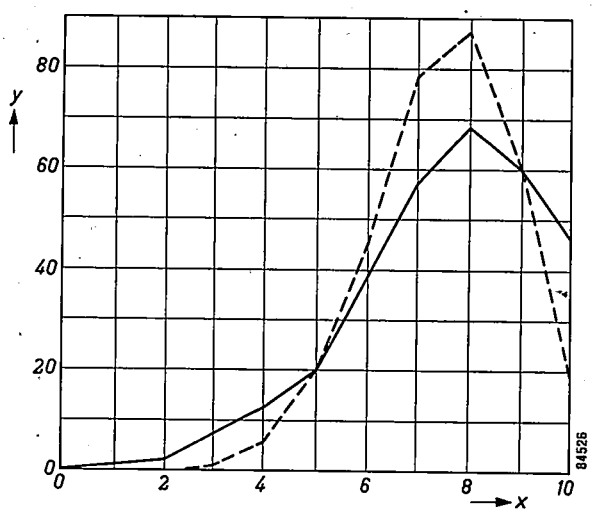


Fig. 6. Full curve: number of persons y, on an average per series of ten tests, that gave x correct answers to question I, plotted as a function of x. Dashed curve: probability distribution (binomial curve) taking 0.75 as an average of all answers to question I. Total number of participants in the tests: 310.

The results of a statistical treatment of the answers are represented graphically by the full lines in *fig. 6* (question I) and *fig. 7* (question II). The number of listeners, y, who gave x correct answers to the ten questions is plotted as a function of x; a total of more than 300 persons participated in the experiments. If there had been no mutual differences between the participants, and no differences in

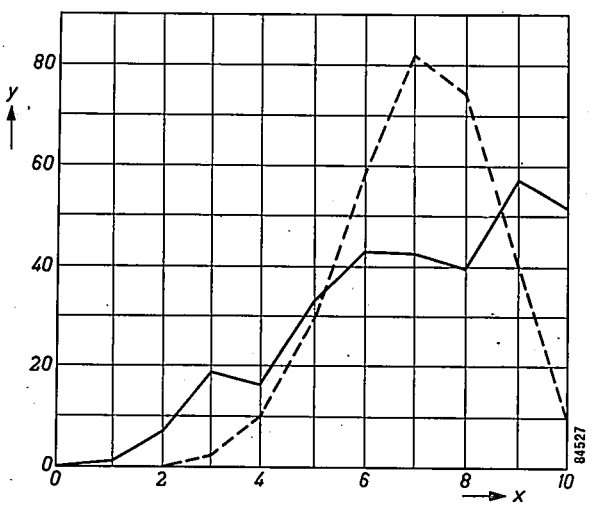


Fig. 7. As fig. 6, but applicable to question II, and with 0.71 as average. Total number of participants: 308.

difficulty between the successive tests, we should have had a curve determined exclusively by chance, that is to say a binomial curve. This curve is shown by a dashed line in both figures, for an average result of 0.75 and 0.71 respectively. It can be seen that the full curve in fig. 6 follows very roughly the binomial distribution. On the extreme right the observed curve lies somewhat above the binomial curve. This means that a group of persons had a more than average power of discrimination and gave the right answer relatively often. On the other hand there is a large group that often heard no difference at all and a small group that systematically gave the wrong answers. A more detailed analysis leads to the conclusion that for the persons individually the chance of giving a correct answer varies between 0.55 and 0.95.

The much greater disagreement between the curves for question II is doubtless attributable to the fact that, to answer this question correctly, considerable familiarity with the sounds of musical instruments is required, a familiarity which can, on the whole, only be expected of professional and amateur musicians. The curve in fig. 7 lies, at the left, appreciably above the binomial, which means that a number of persons did indeed notice a difference, but systematically took the reproduction to be the live music, and conversely.

It might be objected that the foregoing conclusions are drawn from tests made with only a small ensemble and therefore may not be extended out-of-hand to apply to a large orchestra, because its dynamic range is so much larger and consequently more difficult to deal with. In a subsequent article we shall discuss experiments in which a large orchestra was involved — although, it may be added, the purpose of the experiments was not to reproduce the music in another hall, but to improve the acoustics of the hall in which the orchestra was playing. On this occasion recording and reproduction were again stereophonic. No systematic inquiry was held on the results, so that no figures can be offered, but the opinion of the listeners gave us the impression that this reproduction too was deceptively like the real thing. We therefore feel justified in concluding that it is possible to keep the imperfections of electro-acoustical equipment at such a low level as to make them almost imperceptible, even in the case of a large orchestra.

Summary. The author poses the question: why, in spite of the technical progress made in electro-acoustics, is there still an audible difference between the music played in the concert hall and its reproduction via a loudspeaker. The answer should not be sought in the first place in minor technical imperfections, but rather in the two following facts: 1) the instruments of the orchestra are not heard separated because the sound emerges from the small opening of one loudspeaker, and 2) the room where the music is reproduced is often acoustically inadequate.

As has long been known, the first drawback can be remedied by stereophonic reproduction: the sound is picked up by two microphones — preferably placed on an artificial head — and is reproduced, via separate channels, by at least two loudspeakers.

To ascertain in how far stereophonic reproduction can be distinguished from "live" music, comparative tests were carried out, in which strict vigilance was exercised to ensure that the persons taking part in the tests (more than 300) were given no clue as to whether they were listening to "live" music (a small ensemble, concealed from view) or to a stereophonic reproduction thereof. Ten tests were made per session and, for each test, the participants had to give written answers to two questions. Question I concerned the ability to discriminate between the "live" music and the imitation, and question II the "naturalness" of the music. The answers were treated statistically. The general conclusion reached is that, of the average of 75 correct answers out of a 100, 50 are attributable to discernment of the difference and 25 to guessing. Relatively few people (16%) can identify the difference with certainty, however, and then only with difficulty.

A postscript to this article reports on similar experiments carried out in the Amsterdam Concertgebouw and in the "Academisch Genootschap" building in Eindhoven.

Postscript. After the above article had been prepared, two somewhat differently arranged demonstrations with "live" and reproduced music were made, on the instigation of G. Slot and with the cooperation of the Apparatus Division's Acoustical Development Laboratory, in the "Kleine Zaal" (Small Auditorium) of the Amsterdam Concertgebouw and in the "Academisch Genootschap" building at Eindhoven.

As gramophone music was also involved in the demonstrations, no stereophony was employed, but an attempt was made by means of a specific arrangement of the loudspeakers to create the best possible impression of auditory perspective. The installation comprised a type EL 3500 tape recorder (tape speed 76 cm/s), a 60 W amplifier of very high quality and two AD 5002 loudspeaker sets. Each of these sets consisted of an acoustical box with two 9720 loudspeakers (diameter 21 cm) for the frequency range from 30 to 400 c/s, and two high-note projectors, each equipped with one 9710M loudspeaker, for the frequency range from 400 to 20 000 c/s.

For certain pieces, a mixture was played of direct and reproduced music. There was, for example, a piano-piece for four hands, one part of which had been previously recorded and was reproduced while the other was actually being played. Then there was the "farewell" piece, which had been recorded in such a way that the musicians were able to leave the platform one by one during the performance, while the music went on without interruption. They did this, not as soon as their part had been taken over by the reproduction, but some bars later, making sham movements in the meantime. It appeared that the audience found it almost impossible to indicate with certainty the moment of take-over.

In the "Academisch Genootschap" building at Eindhoven an enquiry was held after the interval, which is briefly reported here. Four pieces were played by a small ensemble, of which one or two of the instruments had been previously recorded. The non-playing musicians again made sham movements and the hall was in partial darkness. The persons present were requested after each number to indicate on a questionnaire for each instrument whether they believed it had actually been playing or whether it had been reproduced.

When judging the results given below it should be borne in mind that, as the reproduction was not stereophonic, some clues were given by the different directions from which the sounds of the musical instruments and of the loudspeakers reached the audience. In the front half of the hall especially, this factor was by no means negligible.

Out of an audience of 130 persons, 107 completed questionnaires were returned. The total number of wrong answers amounted to 378. If the 107 persons had only guessed, the number of wrong answers would have been 720. We set out below the results compiled for the instruments individually.

Double bass. With the low notes produced by this instrument it is very difficult to discern the direction from which the sound originates, so that the results in this case give the fairest picture of the quality of reproduction. It is therefore not surprising that the largest number of wrong answers, viz. 150, were returned for the double bass. In the following table, the actual figures are set out in the column headed "In reality", and the figures based on pure chance are given in the column headed " If guesswork". It can be seen that the differences are slight, so that we may assume that the audience was mainly guessing.

	In reality	If guesswork
Bass taken for reproduction . .	61	54
Reproduction taken for bass . .	89	90
Reproduction recognized as such	18	18

Piano. For the piano, 74 wrong answers were returned. In the following table we give an additional column headed: "If half guessed"; the figures shown in this column, which are very close to the actual figures, would apply if half the audience had only guessed.

	In reality	If guesswork	If half guessed
Piano taken for reproduction	27	63	27
Reproduction taken for piano	47	84	45
Reproduction recognized as such	60	21	62

Accordeon. The number of wrong answers given for the accordeon was 62, against 74 for the piano. The distribution was approximately the same as for the piano.

Saxophone. Only 48 wrong answers were returned for the saxophone. There is reason to believe that the recording was not so good as it might have been. Later tests have in fact confirmed this, showing that in this case a great deal depended upon the positioning of the microphone [6]).

Percussion instruments. Although the directional effect as regards cymbals and brushes was very distinct, there were nevertheless 44 wrong answers.

Only three out of 130 persons returned a fully correct questionnaire. It may be assumed that the 23 persons who failed to hand in a questionnaire were unable to discern any difference, so that the actual result was probably even better than emerges from the above figures.

An interesting detail worth mentioning in conclusion is that the average number of wrong answers given by the 42 active music-lovers present amounted to 3.0, as against 3.8 given by the 65 others. No difference worth mentioning could be ascertained between the answers given by the technical (mainly radio and sound engineers) and non-technical members of the audience.

[6]) In the high register, the saxophone has a very pronounced directional effect. See H. F. Olson, Musical Engineering (Mc. Graw Hill, New York 1952), page 234.

A SIMPLE ULTRAMICROTOME

by H. B. HAANSTRA.

578.67:621.385.833

The cutting of very thin sections with a microtome, long practised in optical microscopy, is now being applied to specimens for the electron microscope. Since the resolving power of the electron microscope is about 100 times greater, ultra-microtomes are necessary which produce sections about 100 times thinner than the normal microtomes, viz. of a thickness of the order of 0.01 microns (100 Å). It is remarkable that it has been found possible to do this with comparatively simple apparatus.

Specimens which are to be studied in the microscope by transmitted light, must be sufficiently transparent to light and must not contain too many discernable details behind one another in the direction of vision. The permissible thickness is thus related to the resolving power. For the optical microscope, with a resolving power of the order of $^1/_2$ to 1 micron, specimens a few microns in thickness are required.

Cutting such thin slices (sections) of objects is an old technique, particularly for biological work. Instruments for the making of sections were in use as early as the 18th century (cylinder microtomes). With the present types of microtome, which are part of the standard equipment of biological laboratories, one can make fairly reproducible sections of a thickness of a few μ; in some cases of 1μ.

In the development of electron microscopy, an analogous problem has been met with from the outset. The image is formed in a somewhat different manner; the contrast occurs through the varying degree of electron scatter due to the varying density of the object; moreover, the apertures of the incident beams are extremely small so that there is a much greater depth of focus than in optical microscopy. The criteria for the permissible object thickness are, however, analogous to those for optical microscopes: the parts of the object in which details must be distinguishable, must transmit sufficient electrons of the energy employed (in most microscopes 50 to 100 keV), and there must not be too many details piled up behind one another. In view of the different manner of image formation the latter condition is even more important here than in optical microscopes: in parts of the object lying behind others, the electrons may suffer repeated scattering, so that they "forget" the information which they should transmit and only contribute a general veil to the image. At worst, the detail looked for is obscured; at best, the superimposed sharp images of details lying in different planes makes interpretation more difficult.

Even for early electron microscopes, which reached a resolving power of a few 100 Å, sections of 1μ (10 000 Å) were much too thick. For a long time, therefore, research with the electron microscope was limited to objects which did not need cutting, for example, very small isolated bodies or objects in the form of very thin films (organisms such as bacteria or viruses, vapour-deposited materials or collodion films, etc.), or surfaces of which very thin replicas could be made. Innumerable techniques have been developed for this purpose and a great deal of research carried out in this way [1]). In some fields, however, in particular anatomy and its subsidiary field histology (study of tissues), no advantage could be taken of the high resolving power of the electron microscope since their specimens cannot very well be prepared for investigation other than by making sections. With a resolving power of 50 Å, as quite normally attained in present electron microscopes, sections of a thickness of 200 Å or even less are needed.

Refinements of the classical microtome made possible sections of a thickness of 0.1 μ (1000 Å) [2]), and in recent years Porter et al. in America and Sjöstrand in Sweden have succeeded in constructing apparatus which will make sections of less than 200 Å in a reproducible manner [3]). It is not unreasonable to say that this has meant the beginning of a new era for the electron microscope. Already a vast amount of research has been done with the help of the new technique.

The cutting of successive slices of, for example, 100 Å thickness means, when one considers it, that one is setting out to cut large organic molecules by mechanical means. It will be immediately assumed that the "ultramicrotomes" for this pur-

---

[1]) For a survey, see: D. G. Drummond, The practice of electron microscopy, Royal Microscopical Society, London 1950.

[2]) D. C. Pease and R. F. Baker, Proc. Soc. Exper. Biol. Med. **67**, 470-474, 1948.

[3]) K. R. Porter and J. Blum, Anat. Record **117**, 685, 1953. F. S. Sjöstrand, Experientia **9**, 114, 1953.

pose must be instruments requiring the utmost precision in construction and thus rather costly. Surprisingly enough, however, it is possible to achieve this performance with a very simple instrument, the construction of which involves no exceptional precision work. A prototype of the instrument to be described here, which will be marketed in due course, is shown in *fig. 1.*

Construction of the new ultramicrotome

The ultramicrotomes made by Porter, Sjöstrand and others, distinguish themselves from the classical microtomes by the manner in which the motion of the object is obtained. In the classical method, where the displacement is produced by a screw which causes the specimen holder to move along guides, the reproducibility is limited principally by the variation in thickness of the oil film between the moving surfaces which require lubrication. This effect is entirely avoided in the ultramicrotomes of Porter and Sjöstrand by obtaining the displacement from the thermal expansion of a metal rod, at the end of which the specimen is mounted [4]). Starting at room temperature and supplying heat at a constant rate, one obtains for some time a very uniform, continuous change of length of the rod. If the extremity of the rod with the mounted specimen is then periodically allowed to carry out the cutting motion past the fixed knife, very thin sections of constant thickness are obtained (*fig. 2a*).

A complication occurs here just because the change of length is continuous. After a cut, the specimen has to be brought back to its original position with respect of the knife. Between the moment of cutting and that of return, the forward motion of the specimen has continued. The specimen may not therefore move back along the same path as it travelled in the cutting motion, since it would scrape along the knife and the surface of the next section would be destroyed before it was cut off.

In one of the existing ultramicrotomes this problem was solved by mounting the object excentrically on a vertical rotating disc (fig. 2*b*). The specimen makes contact with the knife on the downward part of its path and not on the upward part. This construction, however, once more makes use of journal bearings, which threaten the reproducibility through their lubrication and possible vibration. To minimize this source of error the most extreme precision in finish was necessary.

In the new microtome described here the displacement is also obtained by thermal expansion. The return stroke, however, follows the same path as the cutting stroke, the necessary clearance between

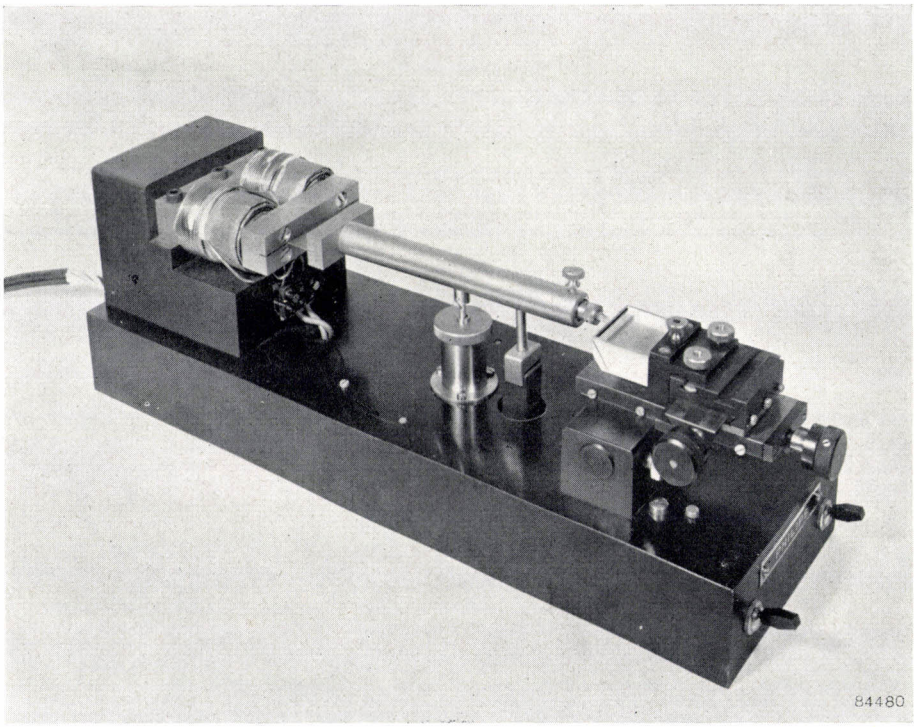


Fig. 1. The new ultramicrotome designed at Eindhoven.

[4]) This principle was first applied by G. B. Newman, E. Borysko and M. Swerdlow, J. Res. Nat. Bur. Stand. **43,** 183-199, 1949.

specimen and knife on the return journey being produced by magnetostriction. The rod on which the specimen is mounted is magnetized after each cutting stroke, before it moves upwards again; the magnetization causes shrinkage, so that the specimen is pulled back slightly; when it reaches the top position, the magnetic field is switched off and the rod re-assumes exactly its original length — except for the thermal extension, which has continued in the meantime. The diagram in *fig. 3* illustrates schematically the successive motions.

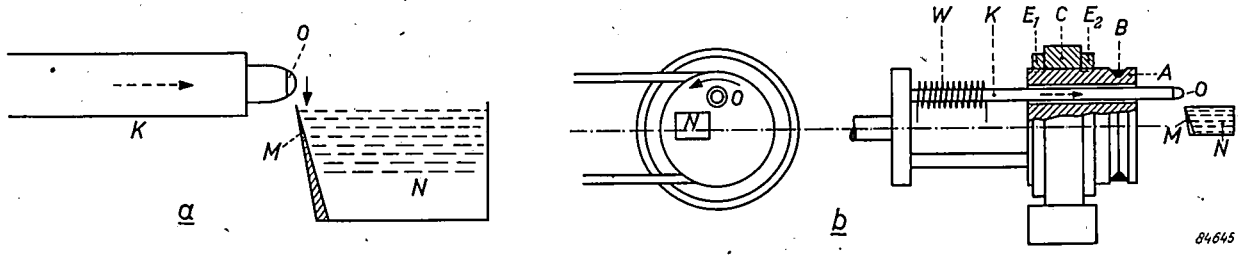


Fig. 2. *a*) Simplified diagram of an ultramicrotome with displacement by means of thermal expansion. The knife is a safety-razor blade, (*M*) sharpened and polished in a special way, or in some cases, a knife specially made for this purpose of glass or diamond. The knife is fixed. The specimen (*O*) is mounted on a metal rod (*K*) which holds it above the knife. The thermal espansion of the rod (——→) produces the forward motion of the specimen, so that when the rod moves downwards a thin section is cut. A small dish (*N*) is fixed to the knife and is filled with liquid (usually 20% alcohol) up to the cutting edge of the knife. The sections, when cut, slide off the knife on to the surface of the liquid and usually form a chain. They can easily be transferred from the liquid surface to a formvar film on a specimen carrier. *b*) In an instrument developed by Sjöstrand the rod *K* rotates with the specimen *O*. This is effected by the pulley *D* driven by the belt *B*. The pulley runs in bearings in the fixed ring *C* and is fixed in the axial direction by the rotating rings E_1 and E_2. Electric heating (winding *W*) causes the rod to expand. The knife *M* and the dish *N* are so placed that the specimen only touched the knife on its downward path. On the upward path, the specimen passes behind the knife.

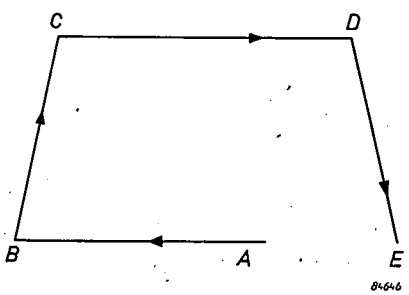


Fig. 3. Schematic diagram of the motion of the specimen in the new ultramicrotome (not to scale).
$AB \approx$ 1-2 μ: retraction by magnetostriction.
$BC \approx$ 3 mm: upward movement of specimen carrier.
$CD = AB$: magnetostriction released.
$DE = BC$: cutting stroke (specimen carrier drops).
The distance $AE \approx$ 50-200 Å corresponds to the thermal expansion during the whole cycle *ABCDE*.

The whole arrangement now becomes extremely simple, because the same current is used to produce the gradual heating and the magnetostriction. The rod, whose thermal expansion produces the displacement, is made of nickel, which exhibits a strong magnetostrictive effect. A coil is wound on the rod and through it flows an intermittent current of several amperes, which produces the periodic change of length according to the diagram of fig. 3. At the same time, the heat dissipated in the winding warms the rod to the desired degree. The intermittent supply of heat means that the distribution of temperature within the rod does not change uniformly throughhout; a slight ripple is superimposed on the heat passing into the rod. However, the thermal capacities of the system are so large that the resulting thermal expansion occurs uniformly with the time.

To conclude this general description it may be noted that in this instrument the unfavourable effects of lubricated bearings, sliding surfaces, etc., have been eliminated entirely, even in the up-and-down motion of the specimen. Instead of the more obvious hinged construction, use is made of a broad piece of spring strip; the up-and-down movements are produced mechanically. *Fig. 4* gives a picture of the arrangement of the whole instrument. The nickel rod or core *K*, surrounded by a coil, produces the required displacement of the object by thermal expansion and periodic magnetostriction. The

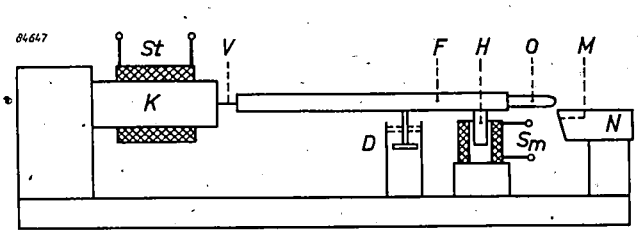


Fig. 4. Diagram of the complete instrument. *M* knife, fixed; *O* specimen carried by the arm *F*, which is fixed to the nickel core *K* by the spring strip *V*; *St* coil for heating and magnetizing the core *K*; *Sm* solenoid by means of which the permanent magnet *H* is periodically raised; *D* oil damping.

specimen (*O*) is mounted on a brass arm (*F*) of constant length connected to the nickel rod by the spring strip (*V*). The arm *F* carries near its free end a ferroxdure magnet (*H*), which is inserted into a vertically fixed solenoid. An intermittent current is passed through the solenoid, forcing the magnet and the end of the arm upwards at intervals and then releasing them. During the downward movement the specimen (*O*) moves past the fixed knife (*M*) and a section is cut. It is of importance to notice that the cutting motion is initiated by gravity and the elasticity of the spring, two forces which may be assumed to act repeatedly in a reproducible manner. To prevent the arm *F* from carrying out free oscillations, damping is introduced in the form of a piston moving in an oil cylinder (*D* in fig. 4). By adjusting the magnitude of this damping, the speed at which the specimen passes the knife in cutting can also be controlled.

The excitation current for the solenoid and that for the heating and magnetostriction coil are both derived from a D.C. source, which is periodically switched on and off by two small commutators. The required intermittent action is thus obtained, and since the two commutators are mounted on the same spindle which is driven by a small motor, the necessary synchronization between the two currents is ensured: the ferroxdure magnet is raised each time after the nickel core has been magnetized and the specimen retracted; the magnet is released, and the section cut, after the core has been demagnetized.

The thickness of the sections

The energy dissipated in the heating coil, and thus the speed at which the nickel core expands, is independent of the speed of rotation of the commutators. However, the number of sections made per second is proportional to this speed of rotation, and by altering the speed of the motor, the thickness of the sections can be varied quite simply. Another method is to alter the current in the heating coil. (This also causes a change in the magnitude of the magnetostriction, but in practice this is always so much larger than the thickness of a section, see fig. 3, that there is never any fear of the specimen touching the knife on its return journey.)

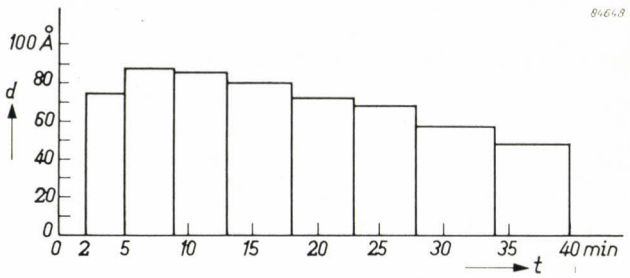


Fig. 6. Average thickness *d* of the sections obtained during thermal expansion of the nickel core over a period of 40 minutes. The thickness is calculated from the measured displacement of the specimen over the given periods and the number of cuts made (in the present case, 49 per minute.) The sections cut during a period of about 20 minutes are of fairly constant thickness, viz. 80 ± 10 Å.

It is naturally important to be able to estimate the thickness of the sections. This is very difficult to do during the actual cutting. The following method was therefore introduced. First, the knife was observed through a microscope to ensure that, at a particular power dissipation in the heating coil and a particular speed of rotation, sections were cut continuously for about 40 minutes. After the nickel core had cooled down, the knife was replaced by the pick-up of a displacement meter type PR 9300; the solenoid (*Sm*, fig. 4) was removed (but

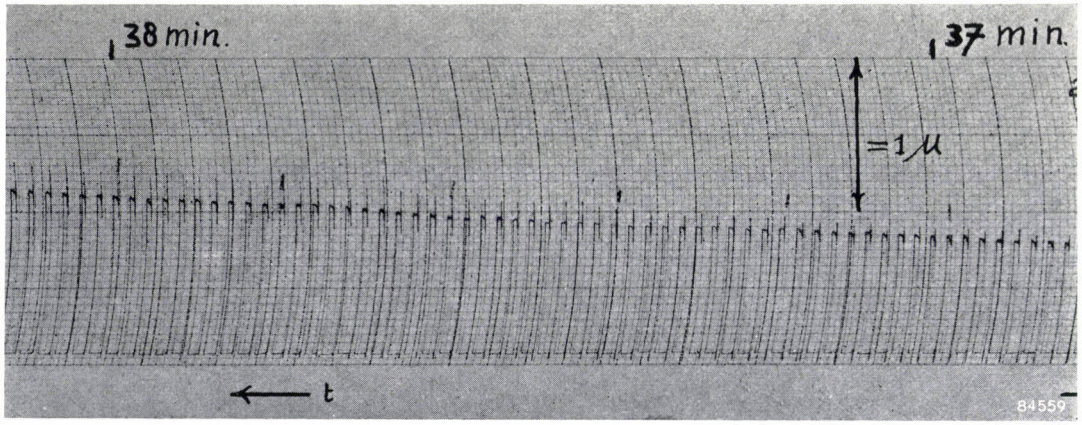


Fig. 5. Section of strip chart recording the displacement of the object during the 40 minutes in which the nickel core was heated. The displacement was measured with an inductive displacement pick-up and meter.

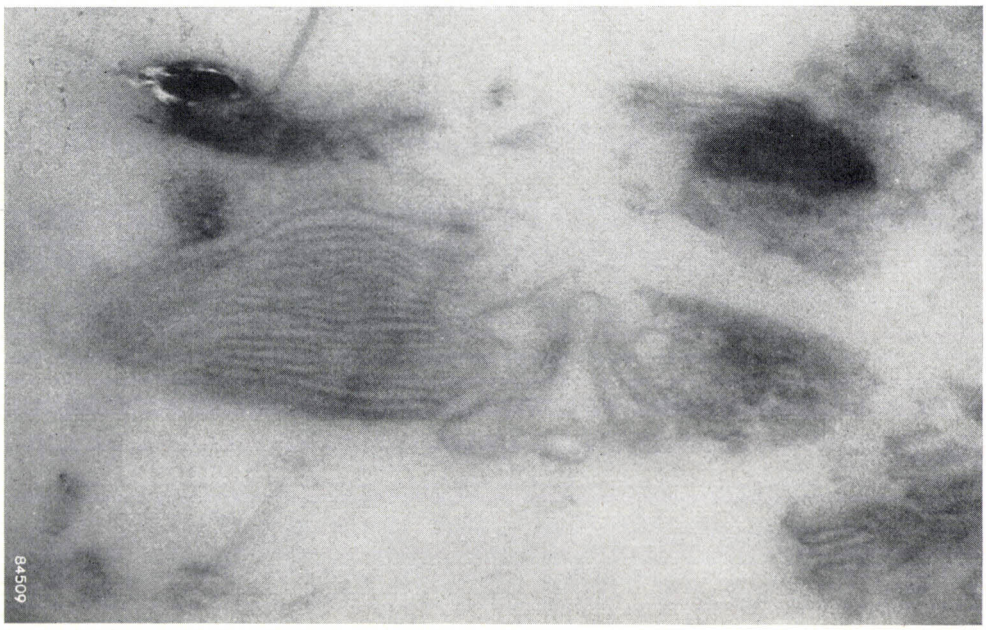


Fig. 7. Section of a chloroplast from a type of lettuce (Lactuca sativa). The chloroplasts are treated with a 1% solution of osmium tetroxide (OsO_4) of $pH = 7.4$ at 4 °C and then dehydrated with ethanol/water mixtures rising to 100% ethanol. From the ethanol they are transferred to a mixture of 95% butylmethacrylate and 5% methylmethacrylate monomers and then polymerized at 45 °C with the aid of 1% benzoyl peroxide as catalyst. Thus embedded, the chloroplasts are cut in the ultramicrotome; after that the sections were transferred to a specimen carrier.

When treated with the OsO_4, the latter is more reduced in the fat-containing parts of the tissue than in the other parts. The photograph shows a chloroplast grain, built up of alternating layers of lipoid (dark) and protein. Magnification 67 000 ×.

remained connected, so that the electrical circuit was unchanged). For about 40 minutes, the displacement measured by the pick-up was recorded on a recording instrument. A small section of the recording strip is reproduced in *fig. 5*. The very uniform thermal expansion with the magnetostrictive change of length superimposed on it can be seen. About 2 minutes are required after switch-

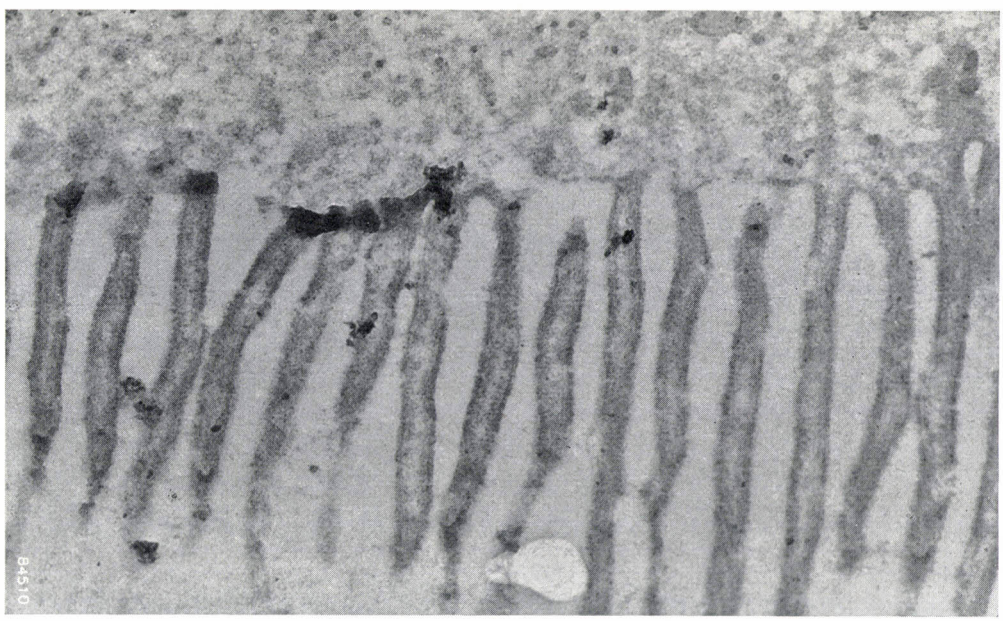


Fig. 8. Section of the fringed edge of the intestinal epithelium of a type of Ascaris, a parasite living in the intestines of pigs. Treated as in fig. 7; magnification 15 000 ×. (Specimen provided by Prof. L. H. Bretschneider, Zoological Laboratory of the University of Utrecht.)

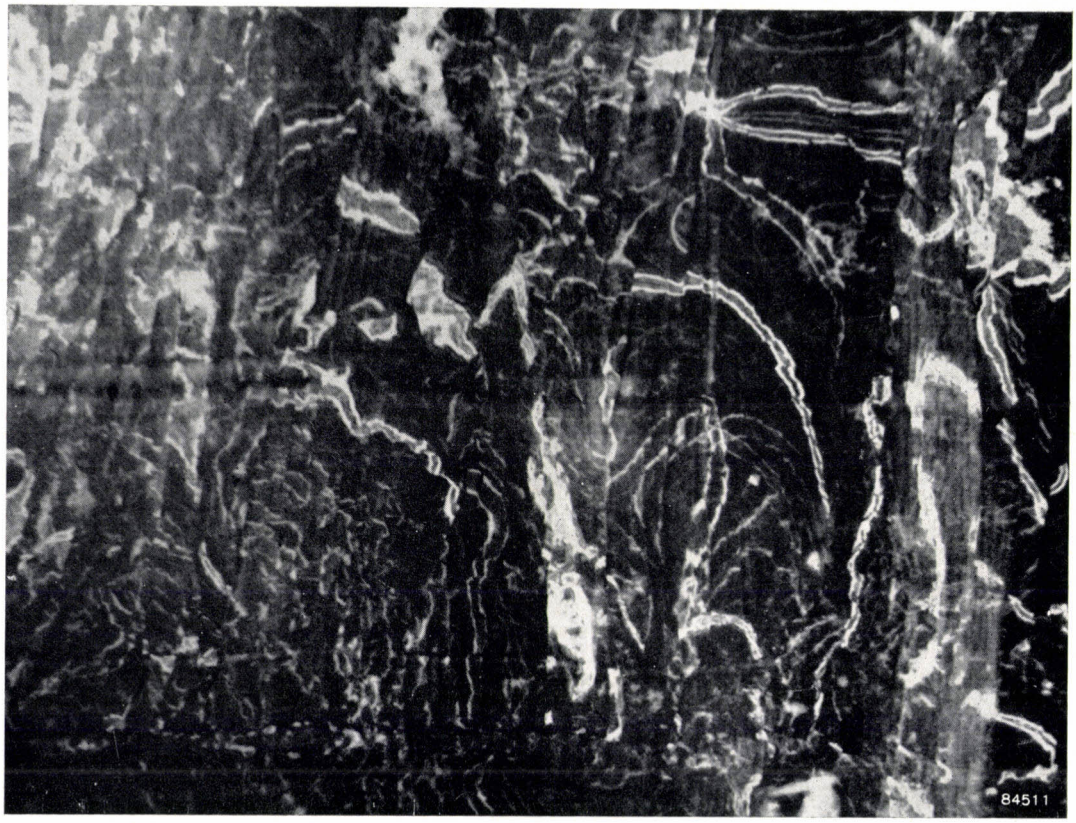


Fig. 9. Section of aluminium. An aluminium rod was shaped so that a small pyramid (about 0.3 mm high, base 0.3 × 0.3 mm) stuck out at one end. The tip of this was cut with a diamond knife. The angle between knife and specimen was very critical; the best adjustment was found empirically. No further particulars can yet be given about the structure seen in the photograph. Magnification 24 000 ×.

ing on before the heating coil transmits sufficient heat to the nickel core: the expansion was only 0.2 μ in the first 2 minutes. In the following 3 minutes, the core expanded 1.08 μ. The cutting speed being 49 sections per minute, the average thickness of the sections during these 3 minutes was 73 Å. The average thickness, calculated in an analogous way, for a number of successive intervals, is shown in *fig. 6*. It is seen that the instrument, during a period of about 20 minutes, cut sections of a reasonably reproducible thickness of 80 ± 10 Å. The figure shows, incidentally, that it is quite possible to make sections of 50 Å.

Figs 7, *8* and *9* show a few electron microscope exposures of sections made with the ultramicrotome described. Details are given in the captions.

Summary. Biological specimens, such as parts of tissues, can only be made accessible to microscopic investigation by cutting sufficiently thin sections. The development of ultramicrotomes (by Porter and by Sjöstrand) with which sections of 200 Å and less can be made, makes it possible to take advantage in the above field of the enormous resolving power of the electron microscope. In these ultramicrotomes the displacement is effected by thermal expansion of a metal rod which carries the specimen. In order to prevent the knife from scraping the specimen surface on the return stroke, a rather complex mechanical arrangement is used, which must be constructed with the utmost precision. In the new ultramicrotome described in this article, the clearance is obtained in a very simple way by magnetostriction of the nickel core which carries the object. The periodically switched magnetizing current produces at the same time the heating necessary for the displacement by thermal expansion. The up-and-down movement of the specimen for the cutting is provided for by mounting the specimen carrier arm on a spring strip fixed to the nickel core. The instrument thus has no bearings or sliding surfaces. Measurements with an inductive displacement meter have shown that in a particular case, about 1000 sections were made in 20 minutes, of thickness 80 ± 10 Å. Sections down to about 50 Å in thickness are attainable.

A HIGH-VACUUM TAP WITH SHORT OUTGASSING TIME

621.646.6:621.52

The better the vacuum achieved in the pumping-out of electronic tubes, the better, as a rule, the quality of the tubes; they will have a longer life, better uniformity and — of prime importance with transmitting valves, rectifiers and X-ray-tubes — be able to withstand higher voltages. Improvements in vacuum technique are therefore constantly being sought.

One weak point in the high-vacuum installation is the taps, which by their very nature are indispensable items. A glass tap of normal construction is shown in *fig. 1*. In order to make it easy to turn the tap plug in the housing, and at the same time to provide a vacuum-tight seal, it must be lubricated from time to time with a special tap-grease. After the greased plug is set in place and the pumping

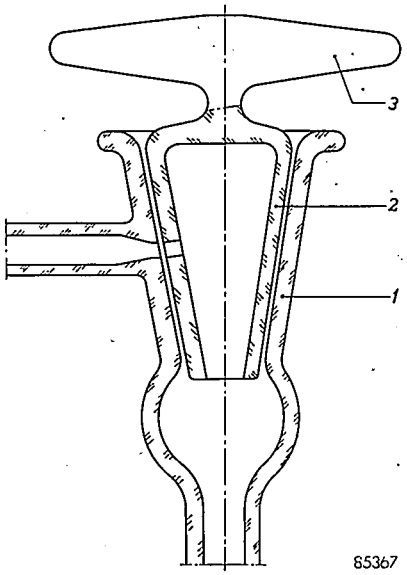


Fig. 1. Glass high-vacuum tap of normal construction. *1* tap housing. *2* plug. *3* handle.

installation has been put to work, some considerable time ensues before a vacuum of, say, 10^{-6} mm Hg is reached; this may quite well involve a period of some ten days. Depending on the type of vacuum grease used, it is necessary to renew this at least every six months, and sometimes even within one month; several days are then again necessary to re-attain the vacuum.

The cause of this long pumping-time is the fact that the air, which is partly absorbed in the grease-layer and partly within it in the form of air-bells, is liberated only very slowly. Only a very small section of the grease-layer is in contact with the evacuated space, and the air is rapidly extracted from this section. The remainder is enclosed between plug and housing, and can thus only find an outlet along a very narrow path, some centimetres long [1]).

An improved vacuum tap has now been designed in which *the whole greased surface can be exposed to the vacuum*. The air then nowhere has to find an

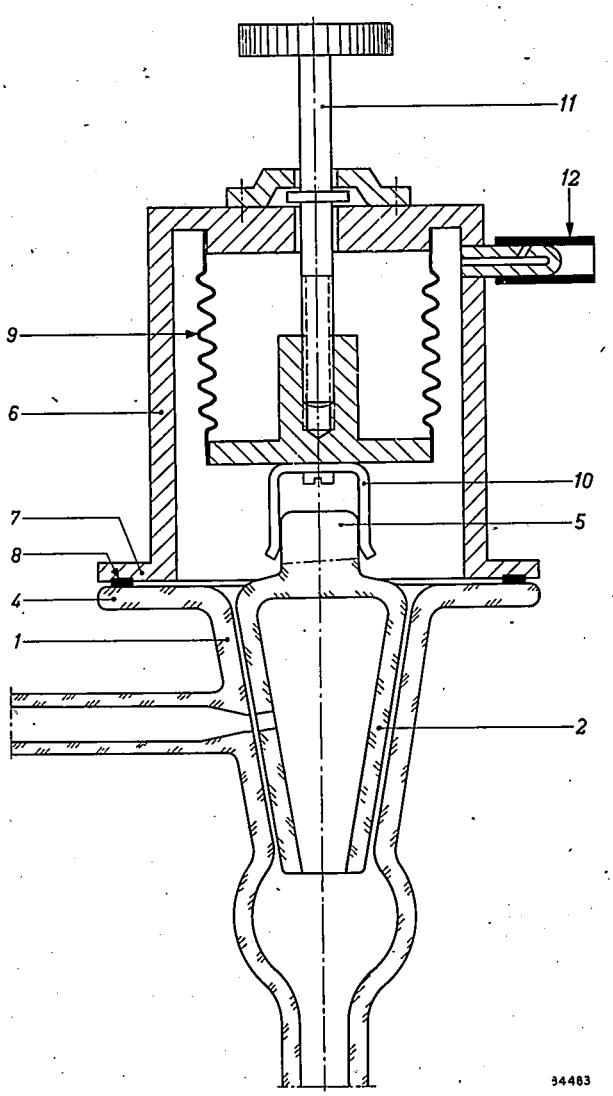


Fig. 2. New high-vacuum tap with short outgassing time. *1* tap housing with glass flange *4*. *2* plug with square, solid head *5*. *6* metal cap, with flange *7*, carrying sealing ring *8*. *9* metal bellows. *10* sprung fork. *11* adjusting screw. *12* valve. After the tap is outgassed, the cap is removed.

[1]) The whole process may be somewhat speeded up by turning the plug slightly now and again during pumping; this at least gives part of the grease in other places an opportunity to lose its air rapidly (this turning stretches out the air-bells, which may be seen as stripes through the glass), and reduces also the sudden increases in pressure which later occur when turning the tap.

escape route longer than the thickness of the grease-layer, which is only of the order of 0.1 mm.

Fig. 2 shows a cross-section of the new tap. The tap housing is provided with a flange, which is ground flat. The greased plug is inserted in the housing, turned round a few times (to distribute the grease evenly) and taken out again, so that the excess grease can be removed; it is then replaced in the housing. A metal cap is placed over the tap; the flange of the cap rests on the glass flange, to which it is temporarily fixed with spring clips. The inside of the cap carries a metal bellows, bearing a sprung fork which, when the cap is attached, grips the square head of the plug protruding from the plug-housing. By means of an adjusting screw, the bellows may be contracted, thus withdrawing the plug from its housing. The pumping installation is now set working. Since the plug is withdrawn, the metal cap will be evacuated, and pressed down firmly onto the glass flange through atmospheric pressure; a greased ring of synthetic rubber set in a groove in the cap-flange ensures a good seal. The vacuum used during this outgassing process need not be particularly high, e.g. 10^{-4} mm Hg.

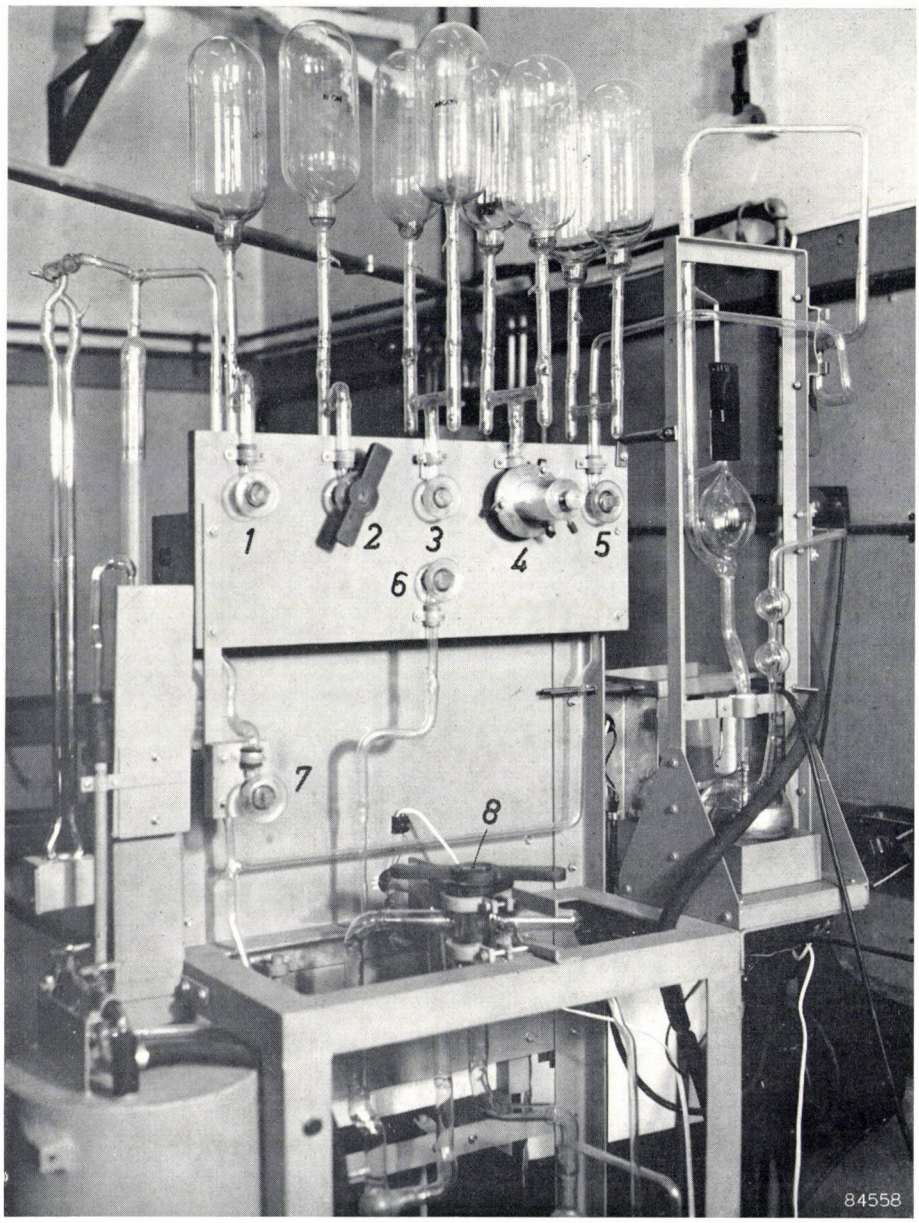


Fig. 3. A high-vacuum installation with nine taps as in fig. 2, eight of which (*1* - *8*) may be seen. A cap is attached to tap *4*, and operating keys to *2* and *8*.

The whole of the greased surface of the plug is now exposed directly to the vacuum. Due to this

fact, the time for the grease to lose nearly all its air is now a question of only about an hour.

When this time has been reached, the adjusting screw is screwed back in, so that the bellows expand and the plug is once more seated in the housing. Opening a small valve lets air into the cap, which may then be easily removed. The time now required to attain a vacuum of 10^{-6} mm Hg is very short.

On the first occasion that the tap is turned after one side of it has come into contact with the outer air, a further slight outgassing will occur; the pressure-increase due to this is appreciably less than that experienced with a normal tap after six months' use. The new tap is completely transparent after outgassing; this is a further indication that the grease in the tap contains no air.

It is also an important advantage that the new tap allows the use of types of grease having the best lubrication properties; these are less suitable for normal taps, since they are difficult to outgas.

As mentioned above, the plug is not provided with a handle for turning (fig. 1) but with a solid head (fig. 2). This is done primarily to keep the dimensions of the cap small. The plug is turned by using an operating key made of compressed cardboard, toughened fibre or a similar material. This is, indeed, also to be recommended for normal taps: glass handles are easily broken off. With a compact solid head and a key of tough material, this is virtually impossible.

Fig. 3 shows a high-vacuum installation, using a number of the new taps.

J. HORSELING.

ABSTRACTS OF RECENT SCIENTIFIC PUBLICATIONS OF 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 Philips Research Laboratory, Eindhoven Netherlands.

2199: W. Ch. van Geel: Über die electrolytische Gleichrichtung (Contribution to Halbleiterprobleme I, edited by W. Schottky, Vieweg, Braunschweig 1954). (On the electrorectifier; in German).

The properties of the rectifier system aluminium — aluminium oxide — electrolyte are considered. The relation between current and voltage during the formation of the oxide layer agrees with theoretical expectations. Analysis of the behaviour of the impedance of the system as a function of frequency shows that the oxide layer is inhomogeneous in the direction of its thickness. The layer nearest the metal is well-conducting, whilst that nearest the electrolyte is poorly conducting. This asymmetry is closely related to the operation of the rectifier; without it, there would be no rectifying action. Rectification, as well as most of the other properties of the system, remains if the electrolyte is replaced by a semiconductor. An attempt is made to explain the rectification qualitatively.

2200: A. J. W. M. van Overbeek: Enige schakelingen met transistoren (T. Ned. Radiogenootschap **19**, 231-260, 1954, No. 5). (Some circuits using transistors; in Dutch).

After a short introduction to the physical principles of transistor action, the characteristics of junction transistors at low frequencies are given. The variation of the small signal parameters with operating conditions is described. At frequencies of 1-10 Mc/sec the equivalent circuit is already as complicated as the equivalent circuit of radio valves at frequencies a hundred times higher. Some low frequency circuits are given and the circuit diagram of a broadcast receiver is drawn. The bandwidth varies automatically with the signal strength so as to provide a narrower band for smaller signals. The variation of parameters with current can be described as a non-linear phenomenon. This gives information about modulation-distortion and cross-modulation of transistors compared with valves. In trigger circuits there is a natural limit to the operating speed, given by the frequency cut-off of the current amplification factor. A circuit containing a *pnp* and an *npn* transistor has properties resembling those of a gas discharge tube with adjustable ignition voltage, a short ignition time, a very low voltage drop and low discharge noise.

2201: S. Woldring and A. G. Th. Becking: Skin impedance and chronaximetry (Acta physiol. pharm. Neerl. **3**, 458-459, 1954, No. 3).

It is shown that the skin has an effect on the form of the intensity-duration curve of muscles and nerves even if the curve is obtained with a constant current stimulator.

2202: W. A. M. van Bergeijk and S. Woldring: Localization and range of the respiration centre of the carp (Acta physiol. pharm. Neerl. **3**, 460, 1954, No. 3).

With the aid of unipolar micro-electrodes, the medulla oblongata of the carp was systematically searched for potentials in the respiratory rhythm. It is established that two respiratory centres can be distinguished; their location is described.

2203: W. J. Oosterkamp: Nieuwe aanbevelingen voor radiologische eenheden en voor stralenbescherming (Ned. T. Geneesk. **98**, 2263-2264, 1954, No. 32). (New recommandations for radiological units and radiation protection; in Dutch).

Summary of new International recommendations on radiological units and protective measures adopted by the International Commissions for Radiological Units (I.C.R.U.) and for Radiological Protection (I.C.R.P.) in Copenhagen, July 1953.

2204: L. Schultink, H. Spier and A. van der Wagt: The abrasion of diamond dies (Appl. sci. Res. **A5**, 1-11, 1954, No. 1).

See Philips tech. Rev. **16**, 91-97, 1954/1955 (No. 3).

2205: J. Haantjes and K. Teer: Multiplex television transmission (Wireless Engineer **31**, 225-233 and 266-273, 1954).

Systems for the transmission of several television signals within a single television channel are described in this article; they are based on the use of signal components which cancel out in two successive pictures. A distinction is made between the sub-carrier and the dot-interlace systems. The typical characteristics of both systems are determined and, in particular, those characteristics which affect the separation of the signals at the receiver. Their application to colour television is considered and the conclusion is drawn that for this the sub-carrier system is to be preferred.

2206: A. Claassen and L. Bastings: Notes on the extraction of nickel-dimethylglyoxime by chloroform and on the photometric determination of nickel by the glyoxime method (Rec. Trav. chim. Pays Bas **73**, 783-788, 1954).

The extraction of nickeldimethylglyoxime by chloroform has been investigated as a function of *p*H. In pure nickel solutions extraction is complete in the *p*H range 4·7-10, in tartrate solutions in the range 4·8-12 and in citrate solutions in the range 7·2-12. Interferences by copper and cobalt are discussed. For the photometric determination of the extracted nickel as the oxidized nickeldimethylglyoxime complex, an improved procedure has been developed resulting in a colour system of great stability.

2207: W. J. Oosterkamp: Image intensifier tubes (Acta radiologica, suppl. **116**, 495-502, 1954).

See Philips tech. Rev. **17**, 71-77, 1955/1956.

2208: H. B. Haanstra: Quelques applications du microscope électronique aux Philips Gloeilampenfabrieken N.V. à Eindhoven (Rep. European Congr. Appl. Electron Microscopy, Ghent, 1954).

The author gives a series of examples of the application of the electron microscope in industrial scientific investigations, together with a brief description of preparation techniques. The examples cover the investigation of the shape and size of particles in powders and the appearance of surfaces.

2209: A. M. Kruithof: Some remarks on the measurement of furnace temperatures by thermocouples (Rep. Third Int. Congr. Glass, Venice 1953, publ. Rome 1954, pp. 529-539).

It is always difficult to judge the accuracy of given temperature-measuring instruments under given conditions. The author takes as an example an instrument long used in glass technology, viz. a suction pyrometer, in which the sensitive element is a thermocouple. When properly designed and set up, the suction pyrometer gives reliable results. Mounting, calibration and use are discussed. Careful attention must be paid to screening from radiation. The use of the instrument to investigate the recuperator of a glass furnace is described. The results obtained with the suction pyrometer are compared with those obtained with thermocouples without radiation shields.

R 258: F. K. du Pré: On the microwave Cotton-Mouton effect in ferroxcube (Philips Res. Rep. **10**, 1-10, 1955, No. 1).

A microwave beam traversing a ferrite will be split up into two beams with different velocities when a transverse magnetic field H is applied. This is the Cotton-Mouton effect. The ordinary beam has its magnetic field parallel to H, the extraordinary beam perpendicular to H. The beams will emerge with a phase difference. This phase difference is determined for Ferroxcube 4B as a function of the field. The frequency used is 9350 Mc/sec. By applica-

tion of Polder's theory to the simplified case of plane-wave transmission through a ferrite having a small damping loss, fair agreement with the experimental results is obtained.

R 259: W. L. Wanmaker, A. H. Hoekstra and M. G. A. Tak: The preparation of calcium halophosphate (Philips Res. Rep. **10**, 11-38, 1955, No. 1).

An investigation has been carried out on the reactions occurring in the preparation of calcium halophosphate, a phosphor used in fluorescent lamps. We started from a mixture of the following raw materials: $CaHPO_4$, $CaCO_3$, Sb_2O_3, CaF_2 and $SrCl_2$, this mixture being fired at different temperatures. In order to trace the behaviour of the activators three other mixtures were prepared, namely a mixture in which Sb_2O_3 has been replaced by Sb_2O_3, another one without Mn and a third without Sb. The transformations of the separate components and simple mixtures were studied by means of differential thermal analysis. During firing, the first reaction that occurs is the decomposition of the $MnCO_3$. At a higher temperature the $CaHPO_4$ is converted into $Ca_2P_2O_7$. This reaction is followed by the dissociation of the $CaCO_3$. The formation of the phosphor begins at a temperature of about 800 °C. The conversions of the activators are important for the quantum efficiency. A substantial part of the antimony volatilizes in the firing process, either as Sb_2O_3 or as $SbCl_3$. At 700 °C the majority of the antimony is converted into calcium antimonate. This compound can be transformed at higher temperatures, with production of the trivalent antimony required. During the firing in air part of the manganese oxidizes to Mn^{3+}, in consequence of which the fluorescent colour becomes relatively more blue. The calcium antimonates isolated by us appeared to have the composition $CaO.Sb_2O_5$. This substance fluoresces weakly and has a long afterglow.

R 260: F. A. Kröger, H. J. Vink and J. Volger: Temperature dependence of the Hall effect and the resistivity of CdS single crystals (Philips Res. Rep. **10**, 39-76, 1955, No. 1).

The D.C. dark-conductivity and the Hall effect have been measured with single crystals of CdS, pure and doped with Cl and Ga, from 25 to 700 °K. Peaks in the curves of the Hall constant as a function of the temperature indicate that two energy bands for conduction electrons are present. Analysis of the results shows that the data can best be interpreted by assuming that, apart from their mobility in the normal conduction band, the electrons can also move in a band situated at the donor energy level. In all cases electrons are found to be the current carriers. At room temperature the electron mobility in the normal conduction band, limited by thermal scattering, is found to be $\mu_c \approx 210$ cm²/V sec. The mobility in the donor band is found at low temperatures. It increases both with the concentration of donors and with the number of electrons in the donor band. Thermo-electric-power data at room temperature and the variations of the Hall constant with temperature can be explained by an effective electron-mass ratio $m^*/m_0 \approx 0.2$—0.3. The depth of donor levels below the conduction band in CdS-Cd and CdS-Cl is 0·01-0·02 eV at low donor concentrations, and decreases to zero at concentrations exceeding 10^{18} cm^{-3}. For insulating crystals, the conductivity and the Hall effect were measured upon illumination with the green mercury line (0.5 W/cm²). Irradiation with infrared in addition is found to cause a decrease in the concentration of free electrons. Capacity measurements with insulating crystals lead to a value of the static dielectric constant $\varepsilon_s = 11.6 \pm 1.5$. The temperature variation of the electron mobility as limited by lattice scattering above 200 °K can be quantitatively accounted for by optical and accoustical modes. With $m^*/m_0 = 0.2$-0.3, longitudinal optical modes give a characteristic temperature of $\Theta_0 = 250$-300 °K; for scattering at acoustical modes the electron mobility is given by the well-known formula $\mu_{ac} = aT^{-3/2}$, in which a is found to lie between 3×10^6 and 8×10^6 cm² degree$^{3/2}$/volt sec.

VOL. 17 No. 7-8, pp. 189-232 JANUARY-FEBRUARY 1956

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 "SCENIOSCOPE", A NEW TELEVISION CAMERA TUBE

by P. SCHAGEN, J. R. BOERMAN, J. H. J. MAARTENS and T. W. van RIJSSEL.

621.383.2:621.385.832:621.397.611

The image iconoscope is a television camera tube with one or two notable advantages: the picture quality is very good, the tube is not very much affected by stray electric or magnetic fields, and the field of vision can be varied electrically. On the other hand, the sensitivity of this tube is not as high as is desirable for many practical applications. Accordingly, an investigation was made with the aim of designing another camera tube which would combine the above-mentioned advantages with higher sensitivity. These efforts have resulted in a new type of camera tube, the "Scenioscope", which in many respects, is similar to the image iconoscope in design. Its operation, however, is altogether different and it therefore merits a new name. This new tube is so sensitive as to give an acceptable, though not entirely noise-free, picture with an object illumination of only 100 lux, and an excellent picture almost entirely free from noise with 300 lux.

Introduction

A few years ago, two articles on the image iconoscope 5854, a television camera tube designed in the Philips laboratories, were published in this Review [1][2]). This tube has a number of very valuable features, particularly its excellent picture quality with adequate scene illumination, and its facility for continuous variation of the field of vision by purely electrical means.

The disadvantage of the image iconoscope, on the other hand, is that the picture quality deteriorates when the illumination level drops below roughly 1000 lux. This is mainly owing to increasing interference from spurious signals (see later). Compensation is possible to some extent, by line-by-line clamping of the black level [3]), which largely suppresses the spurious signals superimposed on the picture signal and thus leaves the received picture — at least that side of it at which scanning begins — almost free from this type of interference.

Although the picture quality at low illumination levels is very much improved by this periodic clamping of the black level, it is nevertheless inadvisable to take the illumination very far below 400-500 lux; apart from a renewal of interference from residual spurious signals in the line direction, the picture is also affected by *noise* (mainly due to the first amplifier valve in the camera pre-amplifier).

Although an illumination level between 1000 and 1500 lux is readily procured in a television studio, it is associated with a heat output high enough to cause discomfort to the performers. Also, in the televising of stage productions, flood-lit sports, and so on, the installation of special extra lighting for the occasion involves major problems. To transmit such programmes, then, European television stations often employ the *image orthicon*, which is the camera tube employed throughout the United States. However, the picture quality of this more sensitive tube is in some respects inferior to that of the image iconoscope when the latter is operated at an adequate illumination level. Attempts have therefore been made everywhere to develop a more sensitive camera tube, based on the image iconoscope and retaining the high quality of its picture.

[1]) P. Schagen, H. Bruining and J. C. Francken, Philips tech. Rev. **13**, 119-133, 1951/52.
[2]) J. C. Franken and H. Bruining, Philips tech. Rev. **14**, 327-335, 1952/53.
[3]) See e.g.: P. Schagen, T. Ned. Radiog. **16**, 236 etc., 1951, and "On the image iconoscope, a television camera tube", Thesis, Vrije Universiteit, Amsterdam 1951, page 58 etc.; H. Bruining, Le Vide **7**, 1255, 1952; J. Haantjes and Th. G. Schut, Philips tech. Rev. **15**, 300, 1953/54.

In England, these experiments produced the "P.E.S. Photicon" (or "photo-electrically stabilized Photicon") [4]) known in Germany as the "Riesel-Ikonoskop" [5]), and in the Philips laboratories at Eindhoven a tube known as the "Scenioscope" (*fig. 1*). Despite a superficial similarity in principle

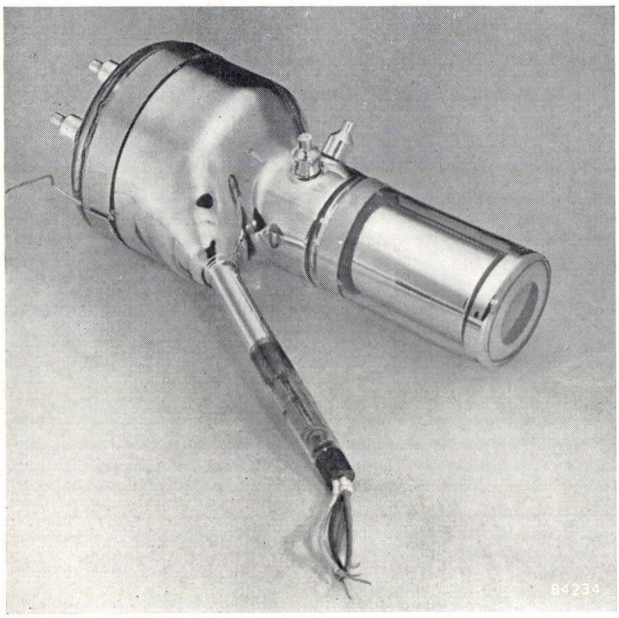


Fig. 1. The "Scenioscope" (identical in appearance with the image iconoscope type 5854). The photo-cathode is on the right, and the arm of the electron gun producing the scanning beam projects into the foreground.

between the P.E.S. photicon and the "Scenioscope" the two are entirely different in design. The "Scenioscope" is more sensitive and, at the same time, much simpler.

The operation of the "Scenioscope", which is identical in appearance to the image iconoscope 5854, will now be explained and its characteristics compared with those of the image iconoscope.

The image iconoscope

Preparatory to explaining the operation of the "Scenioscope" let us consider one or two of the principal features of the image iconoscope as described in article [1]). An image of the scene to be televised is formed, by a special lens, on a transparent photo-cathode (*fig. 2*). The photo-electrons emitted by this cathode are deflected by two fields, one electric and one magnetic, so as to strike points on a mica target corresponding to the image points from which they were emitted, thus producing on the target a pattern or "image" of positive charges, which is then scanned by an electron beam. As they are scanned one after another, the individual picture elements assume a potential which corresponds to a secondary-emission factor of unity. This equilibrium potential is slightly higher than the potential of the collector (that is, the earthed metal coating inside the bulb; the potential of the photo-cathode is —1000 V). Now, between consecutive scannings, the potential of the individual target elements drops, owing to the fact that they intercept slow secondary electrons released from other points on the target by the fast photo-electrons and the equally fast electrons of the scanning beam; this interception effect is known as *redistribution*. The probability that the secondary electrons released from a particular target element by the photo-electrons will escape to the collector or to other parts of the target increases as the above-mentioned potential decreases. Hence the effective secondary-emission factor δ_{eff}, or average number of secondary electrons that can escape from the target per incident primary photo-electron, increases; thus the photo-electrons contribute more and more to the charge pattern on the target. To illustrate this point, *fig. 3a* shows δ_{eff} plotted against the potential of a target element. Similarly, fig. 3*b* shows the variation of the potential of an element as a function of the time constituting one complete frame period τ (i.e. 1/25 second in European television

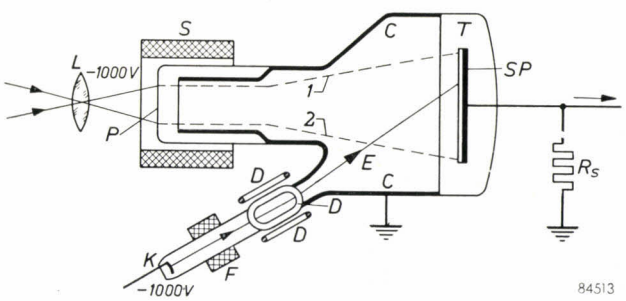


Fig. 2. Schematic cross-section of the image iconoscope. *L* lens; *P* photo-cathode; *S* coil of the magnetic electron lens; *T* target (mica); *C* collector; *E* scanning beam supplied by an electron gun of which only the cathode *K* is shown in the diagram; *F* focusing coil; *D* deflecting coils; *SP* signal plate; R_s signal resistor.

systems), with and without illumination, V_0'' and V'_0 representing respectively the potentials of the particular element at the end of period τ; the difference $V_0''-V_0'$ governs the contribution of the element to the video signal.

For an image iconoscope operating ideally, each photo-electron should produce on the target a

[4]) J. E. Cope, L. W. Germany and R. Theile, Improvements in design and operation of image iconoscope type camera tubes, J. Brit. Inst. Rad. Engrs. **12**, 139-149, 1952.

[5]) R. Theile, Die Signalerzeugung in Fernseh-Bildabtaströhren, Arch. elektr. Übertragung **7**, 15-27, 281-290 and 328-337, 1953.

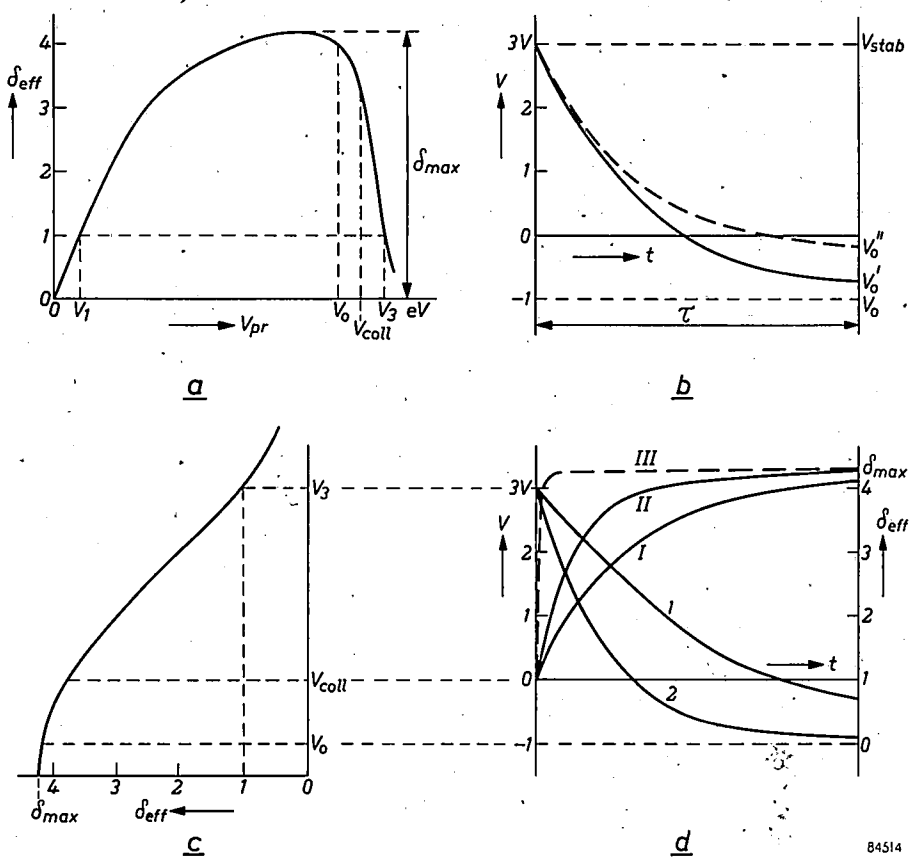


Fig. 3. *a*) Effective secondary emission factor (δ_{eff}) of an insulator, plotted against the energy (V_{pr}) of the primary electrons. $\delta_{eff} = 1$ when $V_{pr} = V_1$ or $V_{pr} = V_3$. Potential V_3 is one or two volts higher than the collector potential V_{coll}. V_0 is the potential at which the flow of secondary electrons to the particular target element is completely cut off.
b) Potential V of a target element in the image iconoscope, plotted against the time t, in the frame period τ between two scannings. Full curve: photo-cathode not illuminated; dotted curve: photo-cathode illuminated. Both curves start at the equilibrium potential V_{stab} ($\approx V_3$, see *a*) and approach V_0 asymptotically. By the end of interval τ, V drops to V_0' or V_0'' respectively. The contribution to the output signal is $V_0''-V_0'$.
c) Right-hand part of the curve shown in *a*, rotated 90°, showing relation to *d*), where *1* and *2* are $V = f(t)$ curves like the full curve in (*b*). From *1* we derive, with the aid of (*c*), curve *I* for $\delta_{eff} = f(t)$, and from *2* curve *II*. Curve *2*, sloping downward more steeply than *1*, produces higher values of δ_{eff} (*II* is above *I*, that is, closer to the ideal curve *III*), and is therefore more favourable than *1*.

positive picture charge of magnitude $(\delta_{max}-1)e$, where δ_{max} is the maximum value attainable by δ_{eff} (fig. 3*a*), and e is the charge on the electron. The signal current output of the tube would then be roughly 1.3 ($\delta_{max}-1$) times the photo-current corresponding to the brightest parts of the picture; the factor of 1.3 is included to allow for the fact that although the photo-current is continuous, the video current flows only 75% of the time, owing to the line and frame suppressions.

In reality, however, the signal current remains very much smaller, for two reasons. Firstly, immediately after a scanning, the secondary-emission factor δ_{eff} differs only slightly from unity, and afterwards rises only gradually (fig. 3*d*, curves *I* and *II*). Secondly, it is precisely the potential of the most brightly "illuminated" target elements that drops the least (fig. 3*b*): therefore the δ_{eff}

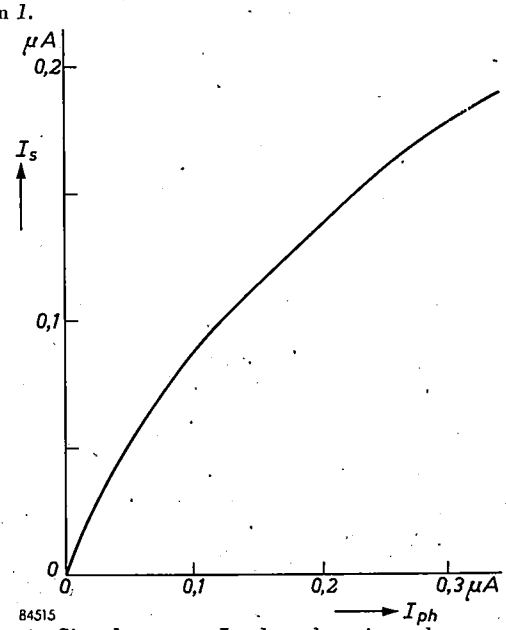


Fig. 4. Signal current I_s plotted against photo-current I_{ph} for image iconoscope 5854.

of these elements invariably falls short of the maximum. Although δ_{max} in image iconoscope 5854 is roughly 4, which gives rise to a maximum signal current 1.3 $(4-1)\approx 4\times$ the photo-current, it is found in practice that, owing to the above-mentioned effects, the maximum video current is at most roughly equal to the photo-current at low illumination levels. Moreover, with an increasing light level the effect of the consequent increase in photo-current becomes smaller and smaller, so that the characteristic of the image iconoscope (signal current against photo-current) bends in the manner shown in *fig. 4*, with a "gamma" smaller than unity.

Possible methods of increasing the sensitivity

That the actual sensitivity of the image iconoscope is appreciably lower than the ideal sensitivity apparently arises from the fact that the potential of the target elements *drops too slowly* to enable the photo-electrons to contribute their utmost to the construction of the potential image immediately after a scanning (fig. 3*d*). If this drop in potential could be accelerated, a stronger signal could be obtained. There are three possible methods of accelerating the drop, i.e.:

1) intensifying the redistribution effect;
2) flooding the target with low-velocity electrons;
3) dissipating the charge by conduction.

They will now be discussed one by one.

Intensifying the redistribution effect

The drop in potential already referred to is caused by redistribution (that is, the interception of low-velocity electrons released by secondary emission from other parts of the target). It is therefore reasonable to suppose that the sensitivity can be increased by increasing the number of redistribution electrons; this can be done, for example, by increasing the scanning current (*fig. 5*). In fact this method does produce a stronger signal [6]), provided that the scanning current is not taken above a certain limit. Beyond this limit, however, the signal strength diminishes again owing to the fact that the erasing of positive video charges, accumulated at the "illuminated" points on the target, becomes more and more intensive as the number of redistribution electrons increases. Another disadvantage is that the spurious signals also become stronger as the number of redistribution electrons increases (a spurious signal, or "shadow signal", may be defined as a signal produced by a camera tube without illumination, and visible on the picture tube of the receiver — for example, the bottom right-hand corner of the screen appearing brighter than the top left-hand corner. It arises mainly from a lack of uniformity of redistribution over the target elements; see article [1])).

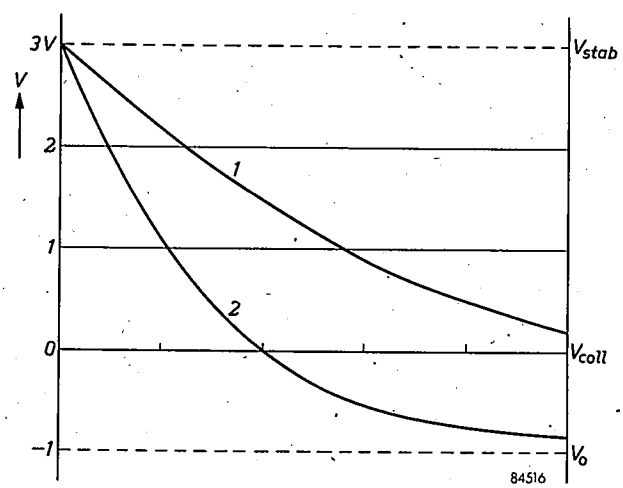


Fig. 5. $V = f(t)$ curves (*1* and *2*) for a non-illuminated-image iconoscope, *2* referring to a heavier scanning current than *1*.

Intensifying the redistribution effect does not therefore greatly improve matters.

Flooding with low-velocity electrons

In England, the "P.E.S. Photicon" is based on the flooding of the target with low-velocity electrons, a method already suggested [7]) in 1934 as a means of improving the ordinary iconoscope. The effect of such flooding is very much the same as that of increasing the number of redistribution electrons, that is, it accelerates the drop in potential of the target elements. However, the advantage of flooding is that it does not increase, and may even reduce, the strength of the spurious signals. Variable-bias auxiliary electrodes arranged all round the target enable the low-velocity electrons to be directed to the parts of the target where redistribution electron concentration is lowest.

In the "P.E.S. Photicon", flooding with low-velocity electrons is achieved with the help of a transparent auxiliary photo-cathode applied to the inside of the bulb and maintained at collector potential (see articles referred to in notes [4]) and [5])). Incandescent lamps arranged around the tube enable the photo-cathode illumination and therefore the "rain" of low-velocity electrons (whence the German name: "Riesel-Ikonoskop") to be varied

[6]) 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.

[7]) A. W. Vance and H. Branson, U.S. Patent 2 147 760, of 1934.

as required. A schematic cross-section of the "P.E.S. Photicon" is shown in *fig. 6*.

The gain in sensitivity procured in this tube is mainly attributable to the almost complete elimination of spurious signals, enabling a good-quality picture to be obtained at a lower illumination level

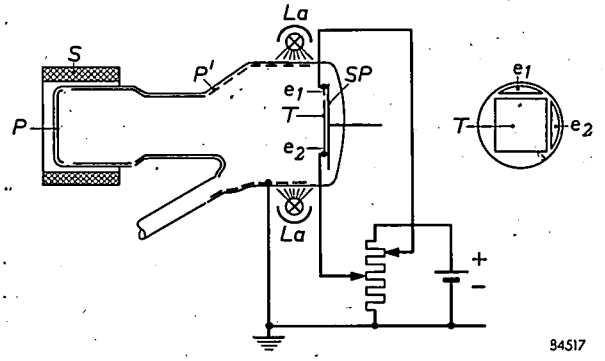


Fig. 6. Schematic cross-section of a "P.E.S. Photicon" (reproduced from fig. 7, page 145, of the article referred to in note [4]). P' auxiliary photo-cathode, illuminated by a ring of incandescent lamps *La*. Segment-shaped control electrodes e_1 and e_2 with variable voltage, direct the low-velocity photo-electrons from P' to parts of the target plate T where the concentration of redistribution electrons is low. This suppresses the spurious signals. Meanings of the other letters as in fig. 2.

than with the ordinary image iconoscope, subject to a limit imposed by noise. It is stated in the literature [8]) that the practical gain in sensitivity as compared with the ordinary image iconoscope is roughly a factor of 2.

In principle, the sensitivity could probably be increased a little more by augmenting the current of low-velocity electrons and reducing the potential of the auxiliary photo-cathode slightly in relation to that of the collector; this would enable the potential of the target elements to drop still lower (i.e. to below the value V_0 at which the flow of secondary electrons to the target element is suppressed, see fig. 3*b*). However, practical experience has shown that under these conditions it is impossible to control the electrons (with the auxiliary electrodes referred to) so as to avoid spurious signals. Moreover, it is evident that the low-velocity electrons proceed preferentially to the "brightest" parts of the target, where the potential is highest; this likewise imposes a limit on the current of low-velocity electrons, since too strong a current would erase the charge image too completely.

Owing to these effects it is most unlikely that the flooding method will produce a sensitivity equal to that of an image iconoscope operating ideally to give a signal current roughly 1.3 (δ_{max}—1) times the photo-current.

[8]) R. Theile, Arch. elektr. Übertragung 7, 333, 1953.

Dissipation of the charge by conduction

In essence, the above two methods of accelerating the drop in potential of the target elements both consist in supplying negative charge to the surface of the target from an outside source, that is, either from electrons of the scanning beam, or from others emitted by an auxiliary photo-cathode. Let us now consider another method, i.e. supplying negative charge *through the target*, or, in other words, *dissipating* positive charge energy through the target to the signal plate.

This idea was suggested in Germany as long ago as 1938, as a means of increasing the sensitivity of the ordinary iconoscope. It was employed in what is known as the "Halbleiter-Ikonoskop" [9]). Here, maximum sensitivity is obtained at a certain value of the scanning current, but the target elements are stabilized at a potential remaining appreciably lower than that of the collector. Owing to this stabilization, for reasons which cannot be discussed here, the images of moving objects in the scene become vague. Hence this method was not then adopted for ordinary television.

In the Philips Laboratories, the principle of charge-dissipation through the target has now been applied to the image iconoscope. These experiments led to the development of the "Scenioscope," which will now be described more fully.

The "Scenioscope"

The essential difference between image iconoscope 5854 and the "Scenioscope" lies in the target. A glass plate (*fig. 7*) having a certain conductivity is substituted for the insulating mica plate employed in the image iconoscope. The back of it is covered with a layer of metal, acting as a signal plate and assuming a negative potential with respect to the collector. As in the image iconoscope, the scanning beam periodically stabilizes the surface elements of the target at a potential slightly higher than that of the collector. The potential of a particular element drops in the interval between successive scannings, partly owing ito the interception of redistrihuton electrons, but mainly because of the leakage current through the target. The dominance of the leakage current as compared with the redistribution is conducive to the elimination of spurious signals.

The leakage current can be varied by varying the D.C. voltage on the signal plate. The size of this current governs the maximum white level of the tube output signal, since the photo-current at the

[9]) G. Krawinkel, W. Kronjäger and H. Salow, Über einen speichernden Bildfänger mit halbleitendem Dielektrikum, Z. techn. Phys. 19, 63-73, 1938.

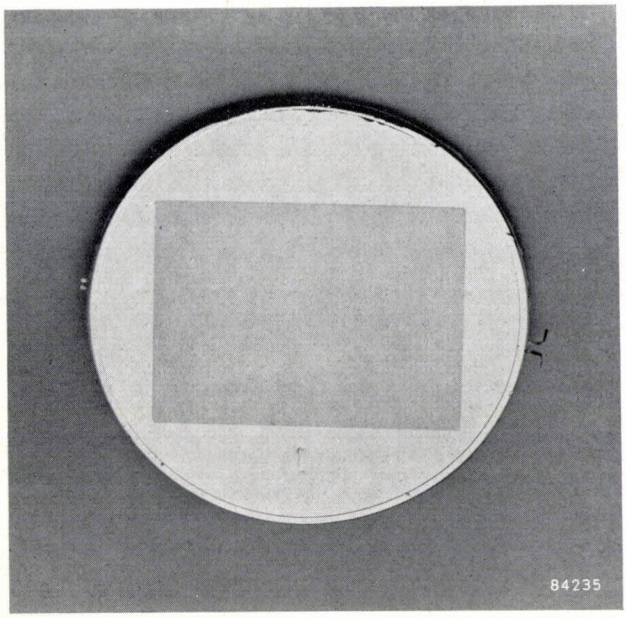


Fig. 7. The target of the "Scenioscope" is a thin plate of slightly conductive glass, from 50 to 70 μ thick, on a metal ring. The size of the rectangular area scanned is 45 × 60 mm. The signal plate is a layer of metal on the back of the target.

"brightest" points can at most only compensate for the charge which is dissipated. In theory the maximum quantity of charge that a target element can collect is equal to the charge increment supplied during scanning to non-illuminated elements; increasing the leakage current, by giving the signal plate a more negative bias, therefore enables a heavier picture-charge to be stored.

Considered superficially, it may well seem odd to make the target conductive, since this enables a certain amount of the charge pattern to leak away. As will be explained below, however, careful design reduces the loss of sensitivity owing to this leakage to a mere fraction of the gain resulting from the fact that the accelerated potential drop enables the photo-electrons to contribute sooner to the build-up of the potential image.

Before discussing the sensitivity of the "Scenioscope", we shall consider the following points: the elimination of spurious signals, the measures taken to overcome certain difficulties (amongst other things, the charge pattern leakage already referred to) and the requirements imposed on the glass target.

The avoidance of spurious signals

As the potential V of a target element drops in the interval between successive scannings, the number of redistribution electrons trapped by the element decreases, until finally the potential is affected only by the leakage current (*fig. 8*). With a negative signal plate, this current forces V below the cut-off limit V_0 for secondary electrons proceeding to the target element. Accordingly, the leakage current does not depend upon the position of the element on the target plate and, unlike the redistribution effect, does not produce spurious signals. To avoid such signals, then, the effect of redistribution should be kept as small as possible compared with that of leakage current (here is a fundamental difference in operating conditions between the image iconoscope and the "Scenioscope": in the one the redistribution effect is essential, whereas in the other it is merely incidental). Since the redistribution current to a particular target element decreases sharply with the potential of the element, the need is to force the potential below the value at which redistribution still occurs (that is, roughly between V_{coll} and V_0, fig. 8) as quickly as possible. Apart from increasing the sensitivity, then, a sharp drop in potential also tends to suppress spurious signals.

The potential drop can be accelerated in two ways, viz. by increasing the leakage current, or by reducing the capacitance of the particular element with respect to the signal plate. The permissible

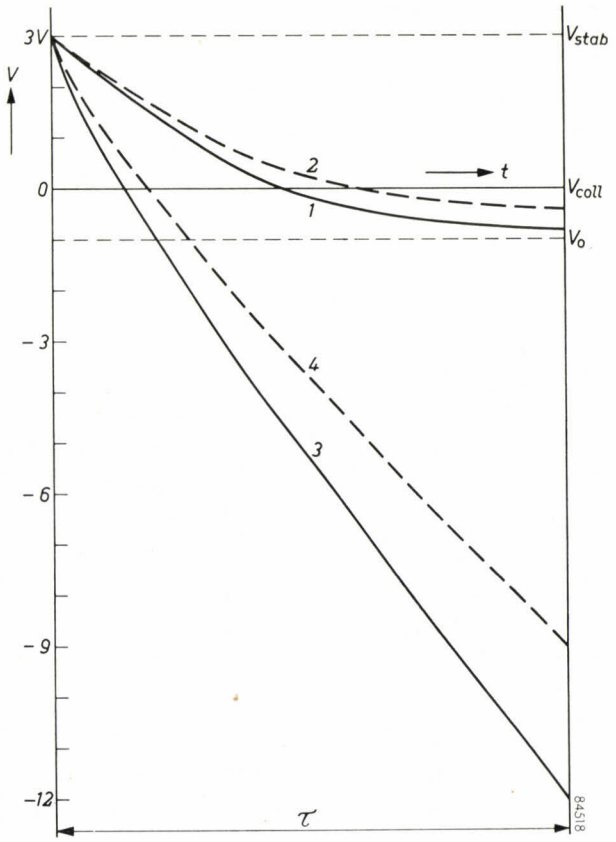


Fig. 8. *1* and *2* are $V = f(t)$ curves, as shown in fig. 3*b*, for the image iconoscope. *3* and *4* refer to the "Scenioscope" (slightly conductive target, signal plate potential negative), *3* for a non-illuminated, and *4* for a faintly illuminated photo-cathode.

increase in leakage current is limited owing to the fact that it produces undue electrolysis in the glass, thus shortening the life of the tube. With a heavy leakage current, moreover, any local variation in the thickness of the target plate, however slight, would cause a local difference in this current, proportional, of course, to the overall leakage, thus producing spurious signals.

Similarly the reduction of capacitance by increasing the thickness of the target plate is limited by the high resolving power required (see below) and, as will be explained later, the requirement as to shape imposed on the tube characteristic.

Leakage of the charge pattern

Theoretical considerations have shown[10]) the extent to which the charge pattern is affected by the conductivity of the glass and the thickness of the target plate. The conductivity of the glass affects the output signal in two ways: by conduction perpendicular to the surface of the target (i.e. in the axial direction of the camera tube) and by transverse conduction, parallel to the surface. Despite the fact that it increases the sensitivity, then, this conductivity must not be too high.

Consider first the axial conduction. This dissipates part of the impressed picture-charge before the next scanning enables it to contribute to the output signal. The amount of charge thus dissipated depends upon the relaxation time (R_0C_0) of the target plate (R_0 being the resistance and C_0 the capacitance, both per unit area, between the front of the target and the signal plate). *Fig. 9* shows the

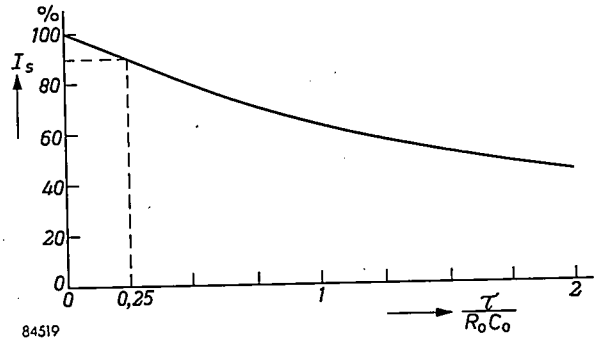


Fig. 9. Signal current (I_s; on relative scale), plotted as a function of τ/R_0C_0. The values of R_0 and C_0 are chosen so as to give an I_s equal to 90% of the maximum value ($\tau/R_0C_0 = 0.25$)

signal on a relative scale plotted against τ/R_0C_0, τ being the frame period ($^1/_{25}$ second in European television systems).

Owing to this effect, a certain loss of sensitivity must be accepted and, as will be seen from fig. 9, to limit this loss to 10% we must ensure that τ/R_0C_0 does not exceed 0.25. Now, for τ/R_0C_0 we may write: $\tau/\varepsilon_0\varepsilon_r\varrho$, where ε_0 is the dielectric constant of free space, ε_r the relative dielectric constant of the glass and ϱ the volume resistivity of the glass. Substituting $\tau = 1/25$ second, $\varepsilon_0 = 8.86 \times 10^{-12}$ F/m and $\varepsilon_r = 8$, we find that for $\tau/R_0C_0 = 0.25$, the volume resistivity ϱ must be $> 0.23 \times 10^{12}$ ohm cm to ensure that the loss of sensitivity owing to leakage will not exceed 10%.

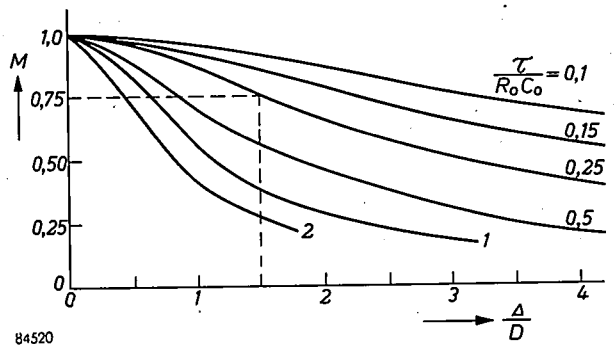


Fig. 10. Relative modulation depth M of the potential pattern on the target, plotted against Δ/D (Δ = thickness of target, D = width of bands in a picture comprising black and white bands), for various values of τ/R_0C_0. With $\tau/R_0C_0 = 0.25$, Δ/D must not exceed roughly 1.5 if the modulation depth is to be maintained at a reasonable value.

Secondly, consider the transverse conduction between a given picture element and adjacent elements with a relatively lower illumination level. Such leakage causes a loss of definition, all the more serious, of course, for the smallest details of the picture, where it also gives rise to loss of contrast. For a picture comprising light and dark bands, for example, the result of transverse conduction will be that the narrower the bands, the smaller the percentage modulation of the charge pattern (that is, the weaker the output signal).

To explain this effect more fully, we shall define a certain relative modulation depth M as the ratio of the signal associated with a band of given width D to that with a picture containing only one black-to-white transition. We then calculate M as a function of Δ/D (Δ being the thickness of the target) for different values of τ/R_0C_0. The results of this calculation[10]) are shown in *fig. 10*, from which it is seen that the thinner the glass (Δ) and the higher its volume resistivity (ϱ), the higher the resolving power.

What, then, are the most suitable values of ϱ and Δ? Fig. 10 requires that a high value of ϱ is chosen with a view to the resolving power, and fig. 9 requires a minimum value 0.23×10^{12}

[10]) P. Schagen, Limiting resolution due to charge leakage in the "Scenioscope", a new television camera tube, Philips Res. Rep. **10**, 231-238, 1955 (No. 3).

ohm cm in order to limit the loss of sensitivity. Accordingly, 0.3×10^{12} ohm cm is taken as the minimum, and 10^{12} ohm cm as the maximum limit; a higher value of ϱ would give rise to undue interference from edge flare, an effect which will be described later.

The glass used has a negative temperature coefficient of resistance. The above-mentioned limits of ϱ hold good for temperatures between 35 and 45 °C; hence the working temperature of the camera must be maintained within this range. At room temperature, ϱ should be between roughly 10^{12} and 3×10^{12} ohm cm.

As regards the thickness (Δ) of the glass, it is seen from fig. 10 that when $\tau/R_0C_0 = 0.25$, the ratio Δ/D must not be very much higher than 1.5, in order to maintain a reasonably large modulation depth (say, $M = 0.75$). The diameter of the smallest resolvable detail in the image equals that of one picture element; with the "Scenioscope" in a 625-line system, it is roughly 75 μ. Hence we see that, even for the smallest practical value of D, that is d (= 75 μ), a glass thickness between 50 and 70μ, as here employed, is consistent with a Δ/D ratio well below the prescribed limit (1.5). Accordingly, we may expect the resolving power to be high; in fact, experiments have shown that with ϱ between 0.3×10^{12} and 10^{12} ohm cm, it is very nearly equal to that of image iconoscope 5854. *Fig. 11* shows the measured modulation depth at the centre of the image plotted against the number of lines per frame.

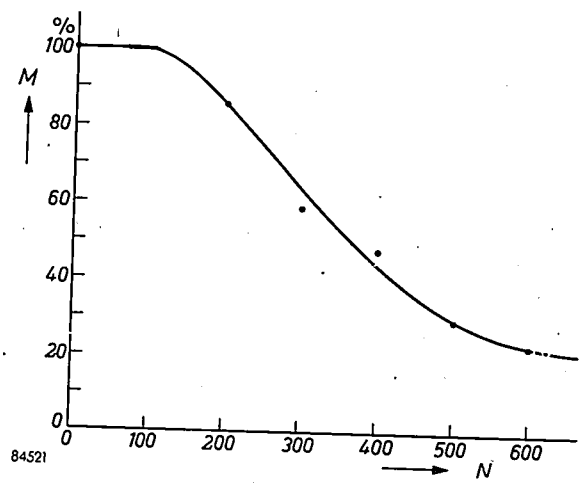


Fig. 11. Relative modulation depth M in the output signal of the "Scenioscope" for the centre of the picture, plotted against the number of lines per frame N.

Edge flare

Although the overall diameter of the circular target is roughly 80 mm, the area actually scanned is a rectangle only 45 by 60 mm (fig. 7). Experiments have shown that it is best to bound this area with a conductive frame whose potential can be taken slightly lower than that of the collector. Without this frame, the part of the target not scanned assumes the potential of the signal plate, owing to the conductivity of the glass. This potential being very much lower than that of the area scanned the elements at the edge of the image are prevented from producing their full secondary emission and therefore assume a relatively low equilibrium potential. The elements at the edge thus acquire a lower potential relative to the signal plate than the centre ones, and hence a smaller leakage current. The charge increment supplied during scanning is therefore smaller at the edge than in the middle of the image, that is, white edges appear in the received picture (edge flare). It is evident that the conductive frame will suppress this effect provided that it is given a suitable potential.

If the limit of scanning does not coincide exactly with this frame, the outer picture elements will continue to be affected to some extent by the adjoining, un-scanned target elements. Hence the potential of the signal plate must not be taken too low; as already explained, this also imposes a maximum limit on the volume resistivity of the glass.

If the potential of the signal plate with respect to the collector is, say, —50 V, and the scanning area is properly covered by the beam, the screening frame will keep the picture almost entirely free from perceptible edge flare. In practice it is found that a leakage current of 0.5 μA through the target gives very good results. With a volume resistivity of, say, 0.5×10^{12} ohm cm, the signal plate bias required to produce this current is roughly —50 V. For a target of the prescribed thickness, viz. 50-70 μ, and a leakage current of 0.5 μA, the potential-variations of the non-"illuminated" target elements are in the region of 7 V. Under the above-mentioned conditions, the redistribution effect, noticeable mainly during the traversing of the first two or three volts below the equilibrium potential, is so small that the received picture contains virtually no perceptible spurious signals. Accordingly, the sensitivity limitation when the light-level is reduced is imposed solely by noise.

The characteristics of the "Scenioscope"

With adequate leakage current, the potential of target elements having a low illumination-level drops quickly enough to enable the coefficient δ_{eff} of the photo-current to reach its maximum very soon after a scanning. Initially, then, the signal current (I_s) versus photo-current (I_{ph}) characteris-

tic of the tube is a straight line (gamma = 1). With increasing photo-current, however, the potential drop of the target elements in the intervals between scanning becomes smaller and smaller; hence the δ_{eff} of the photo-electrons takes longer and longer to reach its maximum, and finally falls short of it. The characteristic therefore exhibits a gradual saturation. The shape of the part of the characteristic where this saturation begins is governed mainly by the capacitance of the target. The lower this capacitance (that is, the thicker the target), the higher the photo-current at which saturation begins (the leakage current being constant); the characteristic may then show quite a sharp bend. With a higher capacitance, the drop in the potential of the target elements will be smaller, and a more gradual saturation will begin at a relatively lower photo-current; it reduces the sensitivity of the tube slightly (smaller signal current for the same photo-current) but at the same time offers two advantages.

Firstly, the gamma is then smaller than unity over a relatively larger part of the characteristic, as in the image iconoscope. This largely compensates for the opposite curvature of the picture tube characteristic (luminance versus control voltage), thus enabling gamma correction to be dispensed with.

Secondly, the more gradual saturation means that the maximum signal is obtained only at a far higher photo-current, enabling the tube to reproduce a higher ratio of scene contrast.

The leakage current through the target governs the photo-current level at which the characteristic begins to bend. It is found that with the prescribed leakage current of 0.5 μA, a glass-thickness of 50-70 μ produces a very suitably shaped characteristic. *Fig. 12* shows characteristics of the "Scenioscope" for three values of the leakage current, and, for comparison, the characteristic of image iconoscope 5854.

The sensitivity of the "Scenioscope"

The higher sensitivity of the "Scenioscope" as compared with the image iconoscope is attributable mainly to the conductivity of the glass target, which accelerates the potential drop of the target elements in the intervals between scannings and thus brings the ratio I_s/I_{ph} closer to the theoretical ideal $1.3(\delta_{max}-1)$.

However, a not inconsiderable part of the improvement is due to an increase in the secondary emission factor δ_{max}. In the image iconoscope 5854, as explained in article [1]), $\delta_{max} \approx 4$. By virtue of certain special features of design, the δ_{max} of the "Scenioscope" is raised to between 8 and 10. It is seen from the characteristics shown in fig. 12, that the sensitivity of the "Scenioscope" is not very far below the maximum value theoretically attainable with such a system at the above value of δ_{max}: at the start of the characteristic, I_s/I_{ph} is roughly 7, whereas a value in the region of 10 is theoretically possible.

The lowest light-level at which good picture quality is obtained is limited by the noise level, not, as in the image iconoscope, by the spurious signals. The compensating signals which have to be applied to the amplifier of an image iconoscope to suppres spurious signals, are therefore not necessary for the "Scenioscope".

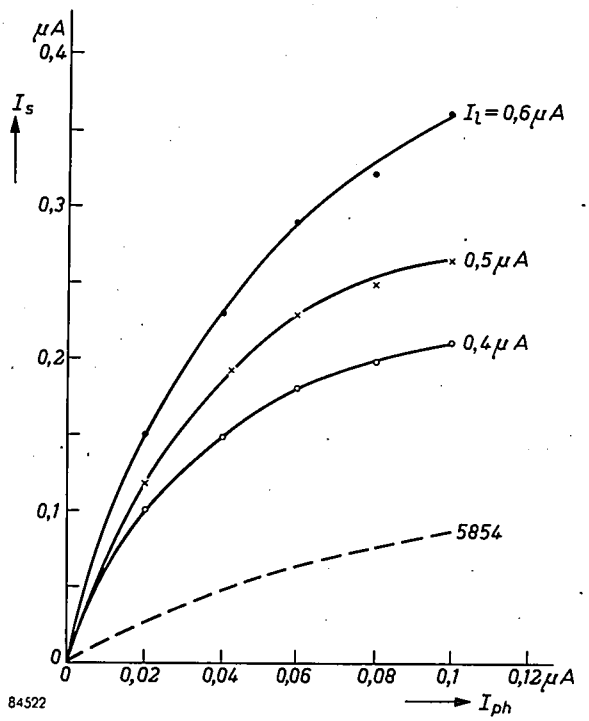


Fig. 12. Full curves: I_s as a function of I_{ph} for the "Scenioscope" for various values of the leakage current I_l through the glass target (I_l measured in darkness). Broken curve: $I_s = f(I_{ph})$ for image iconoscope 5854 (from fig. 4).

With a properly matched amplifier producing very little noise, acceptable pictures are obtained with only 100 lux of illumination on the scene, and under practical operating conditions roughly 300 lux is enough to ensure a virtually noise-free picture (lens aperture $f:2$).

It will be evident that, apart from the difference in sensitivity, the characteristics of the "Scenioscope" and the image iconoscope are very much the same; the photo-cathodes, electron-optical image-forming systems and electron guns of the two tubes are identical. The "Scenioscope" therefore also

admits the facility of continuous variation of the field of view by electrical means (see the article referred to in note [2]).

Summary. Experiments with the image iconoscope have shown that the sensitivity of such a tube can be increased if the potential drop of the target elements in the intervals between scannings can be accelerated. In certain camera tubes now on the market, this is accomplished by "flooding" the target with low-velocity electrons from an auxiliary photo-cathode provided with variable illumination. Another method, simpler and more effective, is employed in a new tube developed by Philips. In this tube, the "Scenioscope", a target of glass having a certain conductivity is employed instead of one made of non-conductive mica. The volume resistivity of the glass and the thickness of the target are considered. The effect of redistribution, a source of spurious signals, is minimized by increasing the leakage current through the glass. Hence the minimum limit of illumination is imposed not by interference from spurious signals, but merely by noise. Whereas redistribution is essential in the image iconoscope, it is merely incidental in the tube considered here. The signal current versus photocurrent characteristic of the "Scenioscope" has a curvature opposite to that of the luminance versus control voltage characteristic of a picture tube, thus enabling gamma correction to be dispensed with.

The gain in sensitivity as compared with the image iconoscope is attributable partly to the conductivity, and partly to the relatively higher secondary emission factor of the target. A picture of acceptable quality is obtained with only 100 lux of illumination on the scene, and in practice 300 lux is enough to ensure a virtually noise-free picture (lens aperture $f:2$).

THE "TL"F LAMP, A TUBULAR FLUORESCENT LAMP WITH A DIRECTIONAL LIGHT DISTRIBUTION

by J. J. BALDER and M. H. A. van de WEIJER.

621.327.534.15

This article introduces a newly developed type of "TL" lamp, which, especially in dusty locations, provides a considerable improvement in light output.

Normal tubular fluorescent lamps ("TL" lamps) display a uniform brightness over the whole of their cylindrical surface (if we discount the short lengths at the tube extremities, where end-effects appear). The light distribution curve of such a lamp in a plane perpendicular to the tube axis is thus a circle (*fig. 1a*). For many applications, however, it is desirable that the light radiated should mainly be concentrated within a given solid angle. It is, for example, very often desired that the light be directed for the most part onto the working plane. This can be achieved by use of a suitable lighting fitting. With such a reflecting fitting, for example, a light distribution curve like that shown in fig. 1*b* can be obtained.

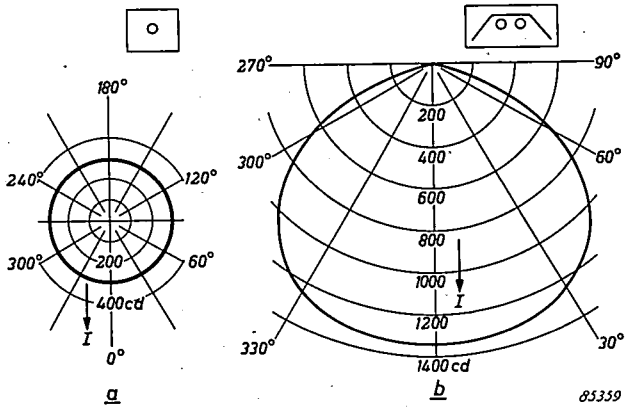


Fig. 1. *a*) Light distribution curve (polar diagram of the luminous intensity in candelas as a function of the direction) in a plane perpendicular to the axis of a normal 40 W "TL" lamp. The curve is a circle.
b) The same for a trough fitting (see inset) with two normal 40 W "TL" lamps. The reflector directs the light downwards.

Influence of dust and dirt

The diagrams of fig. 1 hold for a lamp, or lamps and fittings, which are clean. These diagrams are unfavourably influenced, however, by the inevitable fouling of lamp and fitting occurring in practice due to dust etc. On the lamp itself, in the course of time, a layer of dust settles on the upper surface. This occurs, though to a lesser extent, even when a reflector is placed above the lamp. The influence of an even layer of dust with a transmission factor $\tau_v = 0.5$ coated on the upper half of a normal "TL" lamp can be seen in *fig. 2a* and *b*. Even in places which are not excessively dusty, τ_v reaches the value of 0.5 within a few months of the insertion

or cleaning of the lamps. In very dirty places it is even found that the dust layer becomes practically opaque. The light distribution curve of such a dirty lamp ($\tau_v = 0$) is also shown in fig. 2*a*.

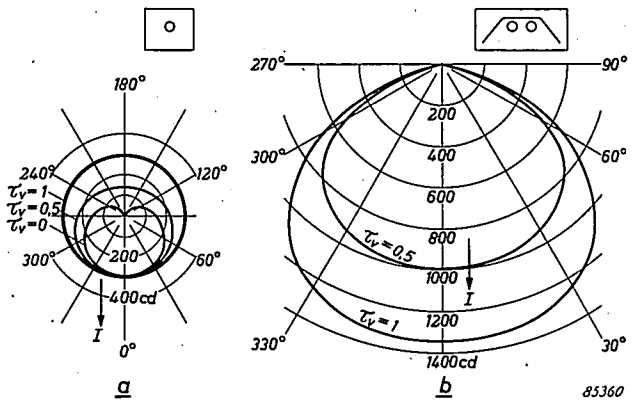


Fig. 2. *a*) Light distribution curves for normal "TL" lamps coated on the upper surface with a uniform layer of dirt, having a transmission factor τ_v. The curve for $\tau_v = 1$ (clean lamp) is identical to that in fig. 1*a*.
b) Light distribution curves for a two-lamp trough fitting holding normal "TL" lamps evenly coated with dirt on the upper surface. The curve for $\tau_v = 1$ is identical to that in fig. 1*b*.

"TL" lamps with directional light distribution ("TL"F lamps)

In the case of the newly developed "TL"F lamp, which forms the subject of this article (F is the type-indication), fouling has much less influence than with the usual lamp types. In the new lamp the light distribution curve is no longer circular, the light being mainly directed to one side (*fig. 3a*). Its is thus possible to mount the lamp in such a way that only a relatively small proportion of the light is radiated upwards. A layer of dust on the top of the lamp then has only a very slight influence on the total luminous flux radiated.

The directional effect can be obtained by making the fluorescent powder layer, with which the inside wall of the tube is coated, thicker than normal over a portion (about 230°) of the tube circumference. The reflection factor of a powder layer increases with the thickness of the layer, so that we obtain in this way a (diffuse) reflector inside the tube. In the lamp described in the present article the increase in reflection factor is obtained by incorporating, between the fluorescent powder layer and the glass wall, a separate layer of powder which does not fluoresce but which possesses very good reflection characteristics. That part of the lamp where fluorescent powder alone is applied forms as it were a slot through which the greater portion of the luminous flux emerges (*figs. 4* and *5*). This part of the radiating surface will be referred to below as the "slot".

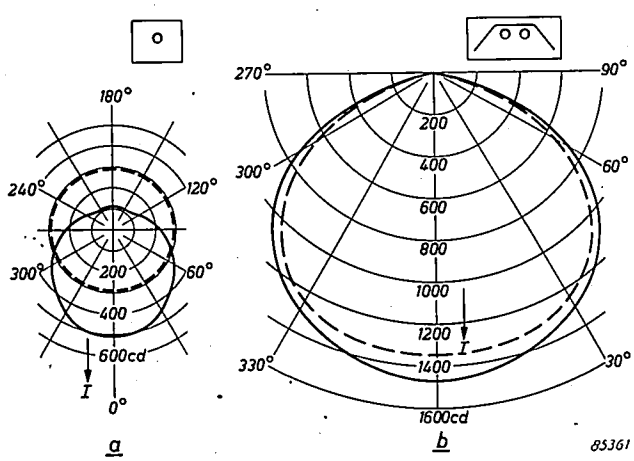


Fig. 3. Light distribution curve of the new 40 W "TL"F lamp with directional radiating characteristics (*a*) and for two of these lamps in a trough fitting (*b*). For comparison purposes the corresponding curves for the normal lamp, taken from fig. 1*a* and *b*, are indicated with broken lines.

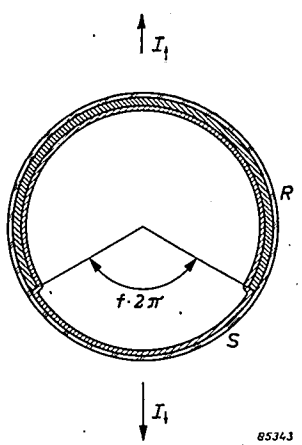
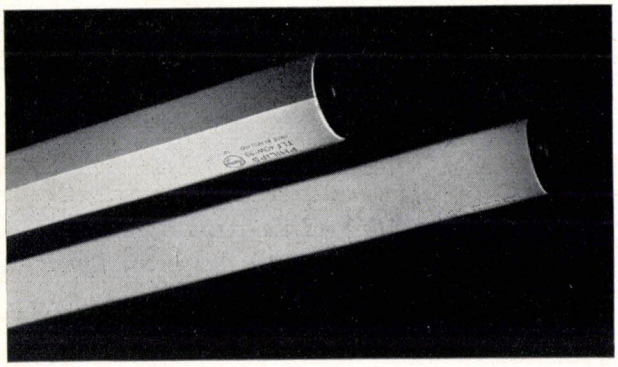


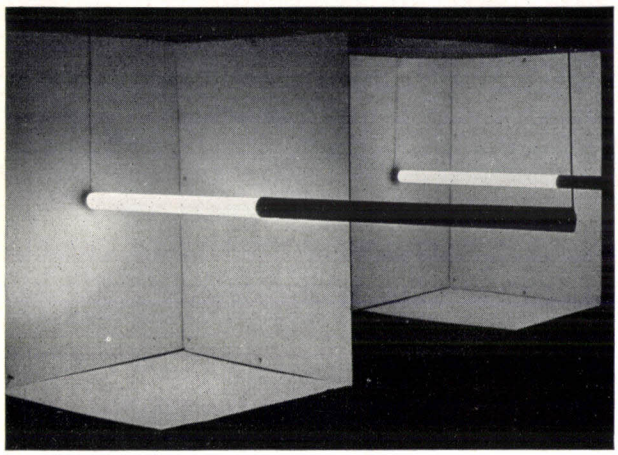
Fig. 4. Schematic cross-section of a "TL"F lamp. The portion S, subtending an angle $2\pi f$, is coated only with a normal fluorescent powder layer. This portion forms the so-called slot. Over the rest of the circumference, i.e. the portion R, a layer of powder of good reflection properties is applied between the fluorescent powder and the glass wall. The portion R has thus a higher reflection factor, and also a greater absorption factor than the portion S. $I\downarrow$ represents the luminous intensity vertically downwards, and $I\uparrow$ the luminous intensity vertically upwards.

Fig. 6 gives light distribution curves for the new lamp, compared with those of a normal lamp, in both clean and dirty conditions.

We shall now firstly consider how the radiated luminous flux in a normal "TL" lamp is produced. It will then be possible to give the results of calculations for the "TL"F lamp without discussing them in great detail. Secondly, we shall give the results of the comparative measurements on normal "TL" lamps and "TL"F lamps, in the form of graphs and in a table.



a



b

Fig. 5. *a*) "TL" lamps in operation; a normal lamp and a lamp with directional distribution ("TL"F lamp, in the foreground). In the "TL"F lamp the brightness is low where the reflecting layer is present between the fluorescent powder layer and the glass wall. The light is mainly emitted through the bright slot (angular width about 130°).
b) Illumination with a "TL"F lamp (foreground), compared with illumination with a normal "TL" lamp (background), in both cases without a fitting. With the "TL"F lamp the lower portion of the cabinet is more strongly illuminated, and the upper portion less so, than is the case with the normal lamp.

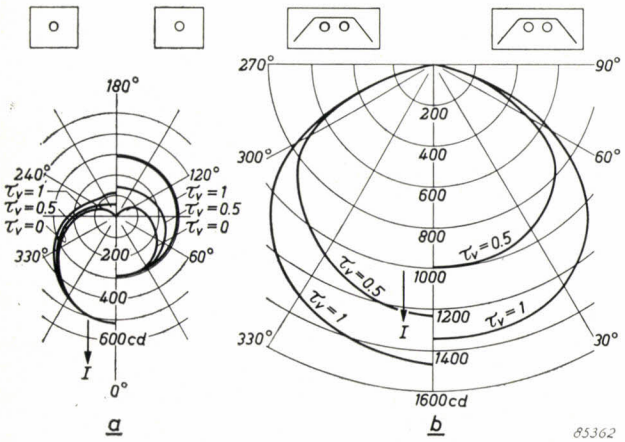


Fig. 6. *a*) The light distribution curves for "TL"F lamps with dirt layers on the upper half having the same values of τ_v as in fig. 2*a* are shown on the left. The slot is directed downwards. For comparison purposes, the curves of fig. 2*a* for the normal lamp are reproduced on the right.
b) The same for a two-lamp trough fitting with dirty "TL"F lamps. The curves on the right of the figure are for normal "TL" lamps as already given in fig. 2*b*.

Light production of a normal "TL" lamp

When light falls on a glass surface coated with a uniform layer of powder, a fraction τ is transmitted, a fraction a absorbed, and a fraction ϱ reflected. τ, a and ϱ are termed respectively the transmission, absorption and reflection factors. Clearly, $\tau + a + \varrho = 1$. We will consider only a powder layer for which both the transmitted and the reflected light are diffuse.

The powder coating on the inside wall of a "TL" lamp fluoresces under the influence of the ultraviolet radiation produced by a gas discharge maintained inside the tube. Due to the high absorption factor of the powder for the ultraviolet radiation, the visible fluorescent light is excited in a thin layer on the surface of the powder. If we assume that this surface layer is infinitely thin — for our purpose a sufficiently close approximation to the truth — then of the total primarily excited luminous flux Φ_p, a half is sent out in the direction through the powder layer. This half of the flux can be regarded as if it fell on the powder layer from inside the tube. The other half is radiated from the surface of the powder towards the inside of the tube, but must, since the surface is in itself closed (a cylinder), nevertheless meet the powder layer again somewhere else. (The tube is assumed long compared with its diameter so that the loss of light at the extremities of the tube may be neglected.) For the purposes of calculation, therefore it may be taken that all the primary luminous flux Φ_p falls on the inside wall of the tube.

Of this incident luminous flux, $\varrho\Phi_p$ will be reflected and thrown back onto the wall of the tube. This time $\varrho^2\,\Phi_p$ will be reflected, and so on. The total luminous flux Φ_i falling on the tube wall is thus:

$$\Phi_i = \Phi_p\,(1 + \varrho + \varrho^2 + \ldots) = \frac{\Phi_p}{1-\varrho}. \quad . \quad . \quad . \quad (1)$$

The fraction τ of Φ_i is transmitted and the fraction a is absorbed. The useful luminous flux Φ emerging from the tube is thus:

$$\Phi = \tau\Phi_i = \frac{\tau}{1-\varrho}\,\Phi_p. \quad . \quad . \quad . \quad . \quad (2)$$

The absorbed luminous flux Φ_a is

$$\Phi_a = a\Phi_i = \frac{a}{1-\varrho}\,\Phi_p. \quad . \quad . \quad . \quad . \quad (3)$$

Clearly, $\Phi + \Phi_a = \Phi_p$, the primary luminous flux generated inside the tube.

A thick powder layer causes, in two ways, an increase in the absorption and thus a decrease in

the useful output of the lamp. Firstly, a itself becomes greater, since the distance to be traversed through the layer by the light increases, and, secondly, ϱ also becomes greater, resulting in an increase in Φ_i and Φ_a (One may not, of course, deduce from (2) that an increase in ϱ produces an increase in Φ, the emergent luminous flux, since an increase in ϱ is accompanied by a decrease in τ). It follows from this argument that an infinitely thin layer of powder would be the most favourable. In actual fact there is an optimum thickness of layer, since a certain thickness is needed to absorb the ultraviolet radiation and also, with very thin layers, gaps occur in the fluorescent powder as a result of the finite grain size of the powder.

Light production of the "TL"F lamp

As mentioned earlier, the "TL"F lamp is coated over a fraction f of its circumference (viz. the so-called "slot") with normal fluorescent powder only, whilst on the remainder of the circumference, a layer with good reflection characteristics is applied under the fluorescent powder layer. Quantities associated with the slot will be distinguished by the suffix S and those associated with the reflecting layer by R; values associated with a normal lamp (i.e. a lamp for which $f = 1$) will be given the suffix n.

The ratio of various characteristics of the "TL"F lamp to the corresponding values for the normal "TL" lamp, are set out in *fig. 7* as a function of the slot width f. We shall not deal here in detail with the calculations upon which these curves are based, but confine ourselves to an explanation of the general form of the curves.

We shall begin with the values of the *luminance* B observed at the outside surface of the tube. The luminance at a given place is proportional to the emergent luminous flux per sq. cm., and thus to the internal illumination multiplied by the transmission factor τ_S or τ_R ($\tau_S > \tau_R$) at that place. If the internal illumination is everywhere uniform in the lamp, then the luminance values are proportional to the transmission factors.

If we imagine the thicker powder layer as being so narrow (f practically equal to 1) that it does not appreciably influence the internal reflections, then the internal illumination remains equal to that in a normal lamp, so that $B_S = B_n$, whilst $B_R = (\tau_R/\tau_S)B_n$.

In the other extreme case, where the thicker powder layer covers practically the whole of the circumference (f practically equal to 0), the internal illumination is almost equal to that which would occur if the slot width was actually zero. The total luminous flux falling on the internal surface is given in this case (see equation (1)) by $\Phi_p/(1-\varrho_R)$. In the normal case ($f = 1$) this was $\Phi_p/(1-\varrho_S)$, so that it has been increased by a factor $(1-\varrho_S)/(1-\varrho_R)$. ($\varrho_R$ is $> \varrho_S$). The internal illumination

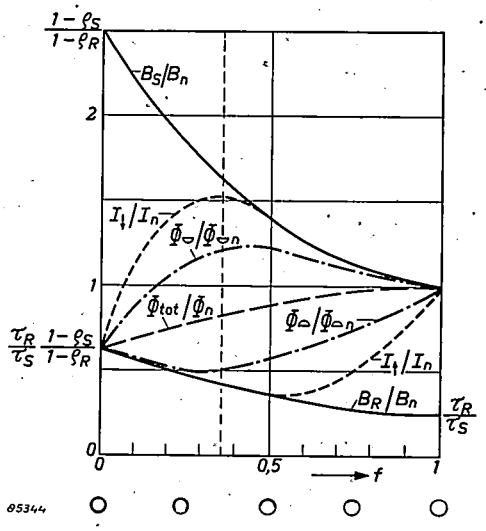


Fig. 7. Theoretical curves of various illumination quantities, as functions of the relative slot width f, all expressed as ratios with respect to the corresponding values for a normal lamp ($f = 1$, suffix n). The following values are plotted: B_S and B_R the luminance of the slot and of the remainder of the lamp surface, respectively; the luminous intensity $I_\downarrow$ vertically downwards and the luminous intensity $I\uparrow$ vertically upwards; the luminous flux radiated into the lower hemisphere and that radiated into the upper hemisphere, $\Phi_\triangledown$ and $\Phi_\triangle$ respectively, and the total luminous flux Φ_{tot}. For the reflection, absorption and transmission factors of the slot and the remainder of the lamp surface, we have assumed, for the purpose of this example, the values $\varrho_S = 0.69$, $\alpha_S = 0.01$, $\tau_S = 0.30$ and $\varrho_R = 0.875$, $\alpha_R = 0.05$, $\tau_R = 0.075$, these being the actual values for some of the powder layers used in the experiments.

is, therefore, now also a factor $(1-\varrho_S)/(1-\varrho_R)$ greater than in the normal case. The luminance of the (infinitely narrow) slot is also increased by the same factor:

$$(B_S)_{f=0} = \frac{1-\varrho_S}{1-\varrho_R} B_n.$$

B_R is again smaller than B_S: for $f = 0$, $B_R = (\tau_R/\tau_S)B_S$, so that

$$(B_R)_{f=0} = \frac{\tau_R}{\tau_S} \frac{1-\varrho_S}{1-\varrho_R} B_n.$$

As f increases from 0 to 1, i.e. as the slot width increases, the average reflection factor gradually decreases from ϱ_R to ϱ_S and thus the total luminous flux falling on the inside of ths lamp wall correspondingly decreases from $\Phi_p/(1-\varrho_R)$ to $\Phi_p/(1-\varrho_S)$. The luminances B_S and B_R will thus decrease. Since the rotational symmetry is lost when f differs from

0 or 1, we may not expect that the internal illumination will remain independent of the position on the circumference of the lamp. The internal illumination will not even remain constant within the arc covered by S or by R, so that this will neither be the case for the luminances B_S and B_R. It proves, however, in practice, from both calculations and measurements, that the variations in luminance within the portion S and within the portion R are only small, so that by approximation one can nevertheless speak of "the" luminance B_S and "the" luminance B_R. These luminances, which are actually averages over the surfaces to which they refer, both decrease gradually as the slot width is increased, according to the curves when shown in fig. 7.

We will now consider the *luminous intensities* $I_\downarrow$ and $I_\uparrow$, i.e., the luminous intensity in the vertically downward direction and that in the vertically upward direction (with the lamp positioned as in fig. 4). As long as $1 \geqq f \geqq 0.5$, the whole of that half of the lamp visible from below has the luminance B_S and thus $I_\downarrow$ is proportional to B_S, i.e. the curve for $I_\downarrow/I_n$ coincides with that for B_S/B_n. As soon as f becomes smaller than 0.5, the outer edges of the surface visible from below acquire the much lower luminance B_R. This effect has the tendency to cause $I_\downarrow$ to decrease. Initially, the increase of the luminances B_S and B_R retains the upper hand and the curve for $I_\downarrow/I_n$ continues to rise, although less steeply than that for B_S/B_n. Gradually, however, the decrease in the area of the bright slot gains in effect; the luminous intensity vertically downwards reaches a maximum and then decreases until, for $f = 0$, the value

$$(I_\downarrow)_{f=0} = \frac{\tau_R}{\tau_S} \cdot \frac{1-\varrho_S}{1-\varrho_R} I_n$$

is reached. As will be seen from fig. 7, $I_\downarrow$ reaches its maximum value at a slot width of about 130° ($f \approx 0.36$), which is used in practice.

For $I_\uparrow$, the luminous intensity in the direction vertically upwards, an analogous reasoning applies.

The total *luminous flux* Φ_{tot} decreases as the slot is made narrower. That this must be the case can most easily be seen by remembering that the average reflection factor and the average absorption factor of the powder layer both become greater, two effects which favour an increase in absorption (see formula (3)). The ratio of the total luminous flux between the extreme cases of zero slot width and a normal lamp ($f = 0$ and $f = 1$ respectively), is of course the same as that of the luminances. The curve for Φ_{tot}/Φ_n thus commences for $f = 0$ at the same point as the curve for B_R/B_n.

Two further quantities of importance are, finally, the luminous fluxes emitted in the lower and in the upper hemispheres, indicated by $\Phi_\cup$ and $\Phi_\cap$, which are also shown in fig. 7.

For uniform lamp luminance, as obtained with $f = 0$ and $f = 1$, $\Phi_\cup$ and $\Phi_\cap$ are of course mutually equal and proportional to the luminance. For this reason the starting points ($f = 0$) of the curves for $\Phi_\cup/\Phi_{\cup n}$ and $\Phi_\cap/\Phi_{\cap n}$ coincide with the starting-point of the curve for B_R/B_n, and their terminal points ($f = 1$) with the end-point of the curve for B_S/B_n. The curve for $\Phi_\cup/\Phi_{\cup n}$ shows, in the neighbourhood of $f = 1$, a tendency to follow the curve for B_S/B_n. Since also the upper part of the lamp radiates some light into the lower hemisphere, the darker portion R of the lamp begins to exert an influence on $\Phi_\cup$ as soon as f differs from 1. Hence $\Phi_\cup/\Phi_{\cup n}$ is immediately less than B_S/B_n. The maximum can be explained in the same way as that in the curve of $I_\downarrow/I_n$. The curve of $\Phi_\cap/\Phi_{\cap n}$ can be explained qualitatively in a similar manner.

Fig. 8 gives the light distribution curves of a series of experimental lamps in which the influence of the slot width was investigated. It can be seen, for example, how the luminous intensity in a vertically downwards direction, reaches a maximum for a

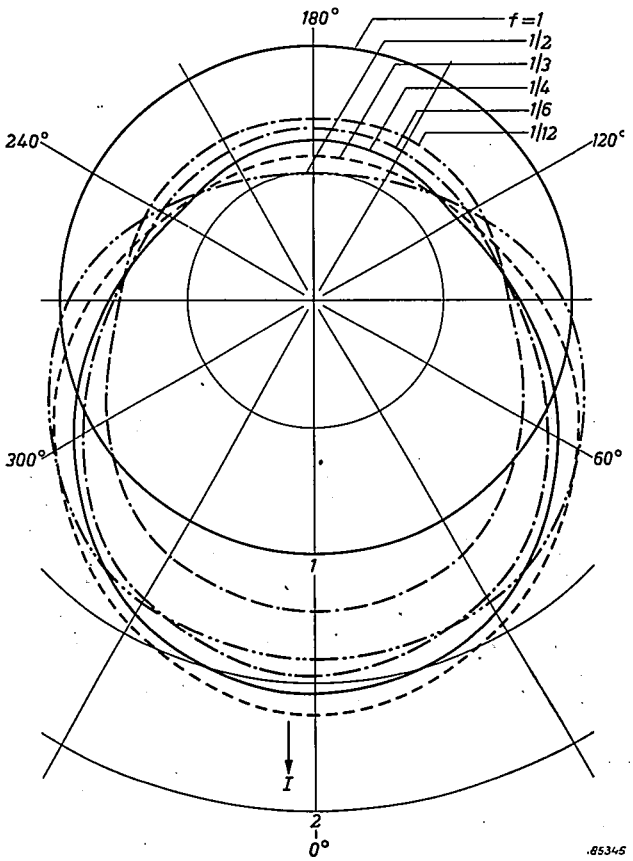


Fig. 8. Light distribution curves for a series of experimental lamps with increasing slot width, f. In these preliminary experiments the optimum properties of the powders had not yet been chosen. The luminous intensity of a normal lamp ($f = 1$) is taken as unity.

value $f < 0.5$, in agreement with the theory. No special attention was paid, in these preliminary experiments, to achieving an optimum combination of powder properties.

Comparative measurements on normal "TL" lamps and "TL" F-lamps

Lamps without fittings

A comparison of the light distribution curves of normal lamps and "TL"F lamps, in clean condition and in two degrees of dirtiness has already been given in fig. 6*a*. *Fig. 9*, which is derived from a combination of measurements and calculations, provides some further details regarding the influence

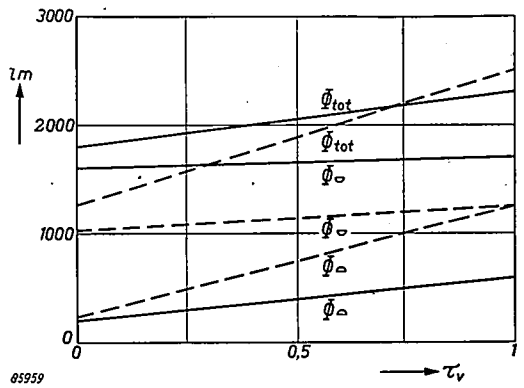


Fig. 9. Effect of a dirt layer on the upper surface on the performance of "TL"F lamps (full lines) and of normal "TL" lamps (dotted lines). The transmission factor τ_v of a uniform layer of dirt on the upper surface is plotted horizontally, whilst the total luminous flux Φ_{tot}, and the luminous flux radiated into the lower and upper hemispheres, $\Phi_\cup$ and $\Phi_\cap$ respectively, are plotted vertically.

of dirt on the two types of lamps. In this figure the total luminous flux and the luminous flux sent out into the lower and into the upper hemispheres are plotted as a function of the transmission factor τ_v of the dust layer on the upper surface of the lamps. In agreement with theory, the normal "TL" lamp delivers, in a clean condition, a somewhat greater total luminous flux than the "TL"F lamp: at zero burning-hours, normal "TL" and "TL"F lamps of 40 W rating deliver 2500 and 2300 lumens respectively. Dirt has, however, much more influence on the total luminous flux of the normal lamp than on that of the "TL"F lamp, so that even with a moderate degree of dirtiness the latter shows an advantage. All graphs and data refer to lamps at zero burning hours.

Lamps in fittings

The light distribution curves of a trough fitting with two normal "TL" lamps and with two "TL"F lamps were compared in fig. 6*b*, for both clean and dirty lamps. These curves apply for a clean *fitting*. It will be apparent that, if the reflecting surface which has to cast the light downwards becomes dirty, the use of "TL"F lamps can be advantageous. *Fig. 10a* shows for normal and for "TL"F lamps in the same fitting as in fig. 6*b*, how the effective luminous flux delivered depends on the reflection

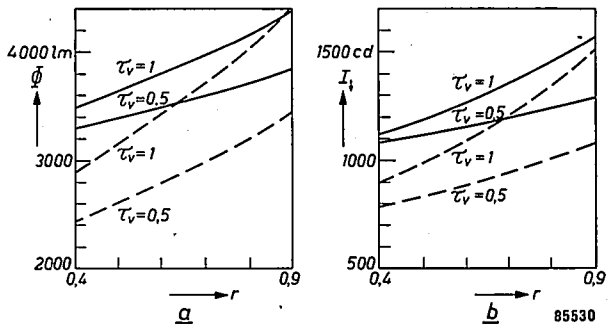


Fig. 10. *a*) Effect of the reflection factor r of the reflector of a two-lamp trough fitting on its luminous flux. The full lines refer to "TL"F lamps, the dotted lines to normal lamps. Both types of lamp were measured in a clean condition ($\tau_v = 1$) and with a layer of dirt on the upper surface ($\tau_v = 0.5$).
b) As for (*a*), but showing the effect on the luminous intensity vertically downwards, $I\downarrow$.

factor r of the external reflector. Here, again, the results are shown for both clean lamps and lamps coated on the upper side with a dust layer for which $\tau_v = 0.5$. For a good, clean reflector the value of r is about 0.8. Even with clean lamps and a clean reflector the use of "TL"F lamps thus gives a higher luminous flux from the fitting than is obtained with

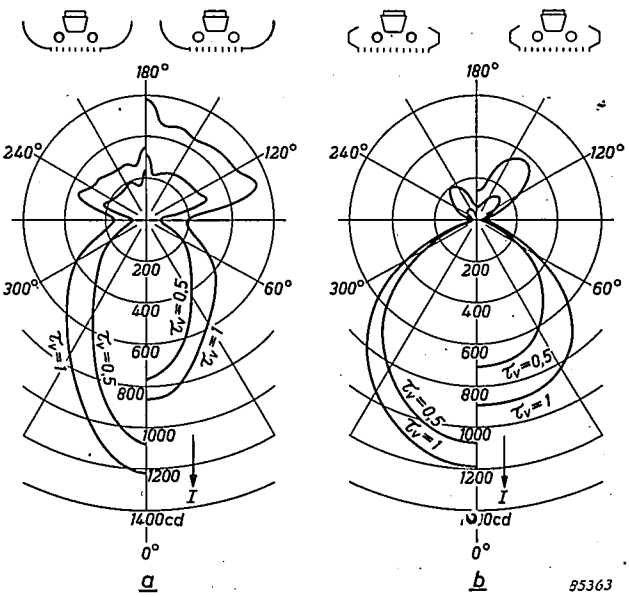


Fig. 11. *a*) Light distribution curves for a louvred fitting with translucent side-panels (see insets). The left-hand curves are for clean "TL"F lamps ($\tau_v = 1$) in a clean fitting, and for dirty "TL"F lamps ($\tau_v = 0.5$) in a correspondingly dirty fitting; the right hand curves are for normal "TL" lamps under the same conditions.
b) As for (*a*) but using a two-lamp louvred fitting with metal side-panels (see insets).

normal "TL" lamps, as can also be seen from fig. 6*b*; the effect of a layer of dirt on the lamps enhances still more the superiority of "TL"F lamps in this type of fitting. The same applies to the luminous intensity vertically downwards, as can be seen in fig. 10*b*.

In *fig. 11a* and *b* the light distribution curves for normal and for "TL"F lamps are compared for two other important types of fittings, viz. a louvred fitting with translucent sides and a louvred fitting with metal sides, both two-lamp fittings. The influence of a dirt layer is again demonstrated by plotting the curves for clean lamps in a clean fitting and for lamps covered on the upper surface with a uniform dirt layer ($\tau_v = 0.5$) in a correspondingly dirty fitting.

Illumination on the working-plane

A very important quantity in practice is the illumination on the working-plane. To compare this for "TL"F lamps and for normal "TL" lamps, 12 normal 40 W "TL" lamps in a certain type of fitting were mounted in a room 2.85 metres high with a floor area of 5 × 5.5 square metres. This corresponds to a lamp consumption of 17.5 watts per square metre. The illumination in a working-plane 0.75 metres above floor-level was then measured for clean lamps and for dirty lamps. Here, again, the latter had been coated on the upper surface with a uniform layer of dirt with a transmission factor $\tau_v = 0.5$. The measurements were then repeated using the same number of "TL"F lamps of the same wattage. The tests were carried out successively with various types of fittings and with lamps mounted directly on the ceiling. When mounted in fittings the lamps hung 0.35 metre below the ceiling; the distance from lamps to working-plane was thus 1.75 metres. Using ceiling-mounted lamps this distance was 2 metres. The reflection factor of the ceiling was 0.73 and that of the side walls 0.43. For ceiling-mounted lamps, the measurements were repeated with a ceiling reflection-factor of 0.55. In this latter case, as was to be expected, "TL"F lamps, in the clean condition, proved to be somewhat more advantageous than was the case with the higher ceiling-reflection factor.

A number of the measurements with dirty lamps were made with correspondingly dirty fittings. The results are presented in *Table I*.

General observations

The insensitivity of "TL"F lamps to dirt layers results, in most practical cases, in smaller costs per lux-hour on the working-plane.

The "TL"F lamp is, as already mentioned, in effect a tubular lamp with built-in reflector. A reflecting metal-surface is not used, since a higher reflection factor can be obtained with a suitable layer of powder. The powder, moreover, in contrast to a metallic mirror, transmits a proportion of the light not reflected, so that it is more efficient than a mirror.

Table I. Illumination values in the working-plane, 0.75 m above the floor, in a room of floor-area 5 × 5.5 m² and height 2.85 m. Illumination by twelve 40W "TL" lamps (a consumption of 17.5 watts for each square metre of floor-area), evenly spaced over the area and in cases A, B, C and D mounted on 6 two-lamp fittings. Lamps 0.35 m below the ceiling, distance from lamps to working-plane 1.75 m. In cases E and E′ the 12 lamps were mounted directly on the ceiling. The distance from lamps to working-plane was then 2 m. In all cases the ceiling reflection-factor was 0.73 (very good white), except in case E′ where it was 0.55. The reflection factor of the walls was 0.43, except for the top 0.4 m where the walls where white with a reflection factor of 0.7. The numbers give the illumination (lux) averaged over the whole measuring surface; the numbers in brackets are the maximum values measured directly under the fittings. The indication "dirty lamp" implies that the upper half of the lamp was evenly coated with a dirt-layer having a transmission factor $\tau_v = 0.5$.

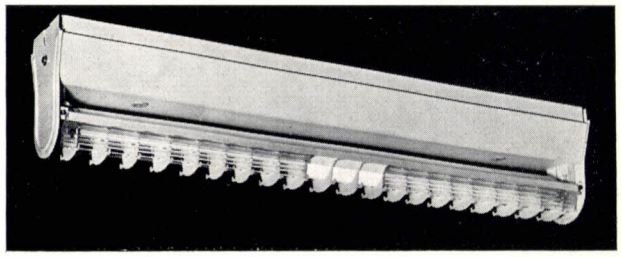
	Type of fitting	Condition of fitting	Reflection factor ceiling	With normal "TL" lamps		With "TL"F lamps	
				clean lamps	dirty lamps	clean lamps	dirty lamps
A.	Trough reflector (closed) (fig. 6b)	clean refl.	0.73	780 (850)	525 (575)	785 (860)	675 (745)
B.	Trough reflector open top	clean refl.	0.73	770 (850)	530 (580)	785 (865)	685 (765)
C.	Open louvred fitting (fig. 11a) with translucent side-panels	clean	0.73	685 (765)		700 (800)	
		dirty	0.73		415 (480)		495 (580)
D.	Louvred fitting (fig. 11b)	clean	0.73	640 (700)		675 (755)	
		dirty	0.73		380 (430)		505 (575)
E.	Mounted direct on ceiling	clean	0.73	715	540	735	685
E′.	Mounted direct on ceiling	clean	0.55	630		675	

The high luminous intensity of the "TL"F lamp in the preferred direction can also be of direct benefit in some cases. It is sometimes impossible to fit a good external reflector because of limitations of space. The "TL"F lamp then provides the obvious solution.

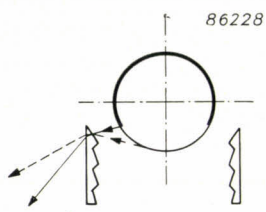
By the use of "TL"F lamps, furthermore, saving can often be made in the cost of fittings, without a serious loss of light. In general, however, it will not prove desirable to dispense entirely with the use of fittings when employing "TL"F lamps. This is because the luminance of the slot is so high (about 13×10^3 cd/m^2), that, although the apparent area of the slot seen from many directions is quite small, screening remains advisable to avoid objectionably high brightnesses in the field of view. Suitable screening can be provided in the usual way by the use of louvres or flaps to cut off the direct view of the lamp from end or side-on. In the case of the side-on direction, protection can also be achieved by employing refracting side-panels adjacent to the lamp which deflect the light radiated sidewards into a downwards direction and out of the direction of viewing (*fig. 12*).

When "TL"F lamps are mounted "bare" against fairly dark ceilings — a practice not often followed with normal "TL" lamps because of the resultant loss of light — account must be taken of the fact that long objects parallel to the lamps will cast multiple sharp shadows, which can be objectionable. Reflections of the high brightness of the slot in polished surfaces in the field of view must be avoided, especially if these are located in fairly dark surroundings, just as care is also advisable in this respect with normal lamps.

The luminance of the darker part of the tube is sufficiently low (about 2.5×10^3 cd/m^2) to require no screening. It is possible, for example, to illuminate vertical panels with the bright portion of vertically mounted lamps directed at the panels, and to leave the visible darker portions unscreened. This technique can, of course, also be applied in horizontal reflectors with horizontal "TL"F lamps mounted "upside down", i.e. with the gap on top. Owing to the enhanced effect of dust, however, this is only advisable with dust-free fittings, in dust-free rooms, or where frequent cleaning is feasible.



a



b

Fig. 12. Method of reducing the high brightness of the gap of a "TL"F lamp by means of refracting side-panels. The light radiated sidewards is refracted downwards. In the photograph (*a*) a portion of a side-panel in the centre has been removed. It can be seen that the brightness is much greater there than elsewhere. The principle is illustrated in the diagram (*b*).

Summary. By introducing a powder layer of good reflecting properties over a portion of the circumference of a "TL" lamp, between the fluorescent powder and the glass wall, the lamp is provided, as it were, with an internal reflector. In this way the light is radiated mainly in a preferred direction. This offers advantages when the lamps are mounted with the internal reflector at the top: deposition of dust on the upper surface then has a much smaller influence on the luminous output than is the case with normal lamps. The results of a theoretical investigation into the effect of an internal reflector of this type on various quantities of interest in illuminating engineering are discussed. On varying the angle taken up by the reflector, a maximum luminous intensity in the preferred direction occurs when this angle is approximately 230°. It is shown from the results of measurements, that even with a slight degree of dirtiness of the top part of the lamp, and in some cases even with clean lamps, the new type provides considerably better lighting than the normal type.

X-RAY INTENSITY MEASUREMENTS WITH COUNTER TUBES

by W. PARRISH *)

548.734:621.387.4:621.374.32

The use of counter tubes, which nowadays play such an important part in research, forms the basis of the method of X-ray diffraction analysis known as diffractometry. The essential advantage of this method is the relative ease with which radiation intensity measurements, for example in quantitative chemical analysis, are performed. Nevertheless, quantitative radiation intensity measurements with counter tubes have some peculiar problems, even if the radiation is nearly monochromatic as in the case of X-ray powder diffraction analysis. These problems are discussed in the present article with special reference to the various procedures of intensity measurement possible with the "Norelco" diffractometer.

Introduction

In the well known Debye-Scherrer method of X-ray diffraction analysis, the relative intensity of diffraction lines is measured by their density in the photographically-recorded diffraction pattern. In X-ray diffractometry a counter tube is substituted for the photographic film as a radiation detector. The number of quanta counted by the tube in unit time (counting rate) serves to indicate the relative diffracted intensity for a narrow Bragg angle region. The counter tube is mounted on a goniometer to permit scanning of the pattern.

A description of the "Norelco" X-ray diffractometer based on this principle was given in this Review some time ago [1]). In that article — hereinafter referred to as I — some general characteristics of the intensity measurements were discussed, e.g: the need for a very stable X-ray source; the reduction of the "background" of short wavelength continuous X-radiation by the use of an argon-filled Geiger counter tube which has a low sensitivity to these wavelengths; the accuracy limitation due to counting statistics; the inaccuracy of diffracted intensities due to the relatively small number of crystallites present in the specimen used; etc. Moreover, a number of methods available for performing the actual counting of the quanta and for recording the results were mentioned. The methods which can be selected in order to conform best to specific purposes will now be considered in greater detail. The means provided with the "Norelco" diffractometer for applying these methods will also be described. *Fig. 1* shows the cabinet containing all the circuitry and accessories used in X-ray diffractometry and also in X-ray spectrochemical analysis (to be dealt with in this Review shortly).

A short recapitulation of counting statistics will be included after a description of the first method; a general discussion of what may be called the "counting strategy" will be given following the description of the second method.

Point-by-point measurements with fixed time

A straightforward method consists of operating the counting device at a fixed Bragg angle position 2Θ of the goniometer for a predetermined length of time t, after which the number of counts N is read on a mechanical register. The procedure is repeated for the same time interval after moving the counter tube to another Bragg angle position, and so on. The counting rate

$$n = N/t \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

is computed and manually plotted against the angle read on the goniometer.

In view of the limited speed of the mechanical register, it is customary to "scale down" electronically the pulses delivered by the radiation detector before feeding them into the register. The scaling circuits used here are of a well-known binary type. Each stage therefore gives a reduction factor of 2. Eight such stages are provided with the "Norelco" diffractometer, enabling the selection of scaling factors 1, 2, 4, ... 128, 256 by press-button operation. The number of counts delivered by the detector is equal to the register reading multiplied by the selected scale factor. If required, interpolation between register readings is made readily by means of neon lamps included in each scaling stage which

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

1) W. Parrish, E.A. Hamacher and K. Lowitzsch, The "Norelco" X-ray diffractometer, Philips tech. Rev. **16**, 123-133, 1954/55 (No. 4). For general information on diffraction analysis and diffractometry the reader may refer to: H. P. Klug and L. E. Alexander, X-ray diffraction procedures for polycrystalline and amorphous materials, Wiley & Sons, New York 1954; X-ray diffraction by polycrystalline materials, edited by H.C. Peiser, H. P. Rooksby and A. J. C. Wilson, The Institute of Physics, London 1955.

are lit by the even pulses and extinguished by the odd pulses arriving at each stage.

For good reproducibility, the counting time interval is controlled by a clock which is preset to the desired time period. A useful method consists of setting the fixed time interval (seconds) equal numerically to the selected scaling factor. In this case the number recorded on the mechanical register is equal to the counting rate in counts/sec. For time intervals exceeding 64 sec, counting must be controlled manually using a stop-watch.

Long counting time intervals are desirable in order to minimize the influence of the terminating errors of the clock ($^1/_5$ sec). At the same time, long counting periods decrease the error caused by the statistical fluctuations in the rate of arrival of the quanta at the detector. Since these fluctuations are of fundamental importance we shall expand on this point in the next section.

Straightforward though the point-by-point measurement and manual plotting may be, this method has the drawback that it makes an excessive demand on the operator's time. It can evidently be useful for scanning only a small portion of the diffraction pattern — and in fact the possibility of selecting small portions of a pattern for measurement constitutes one of the important features of diffractometry.

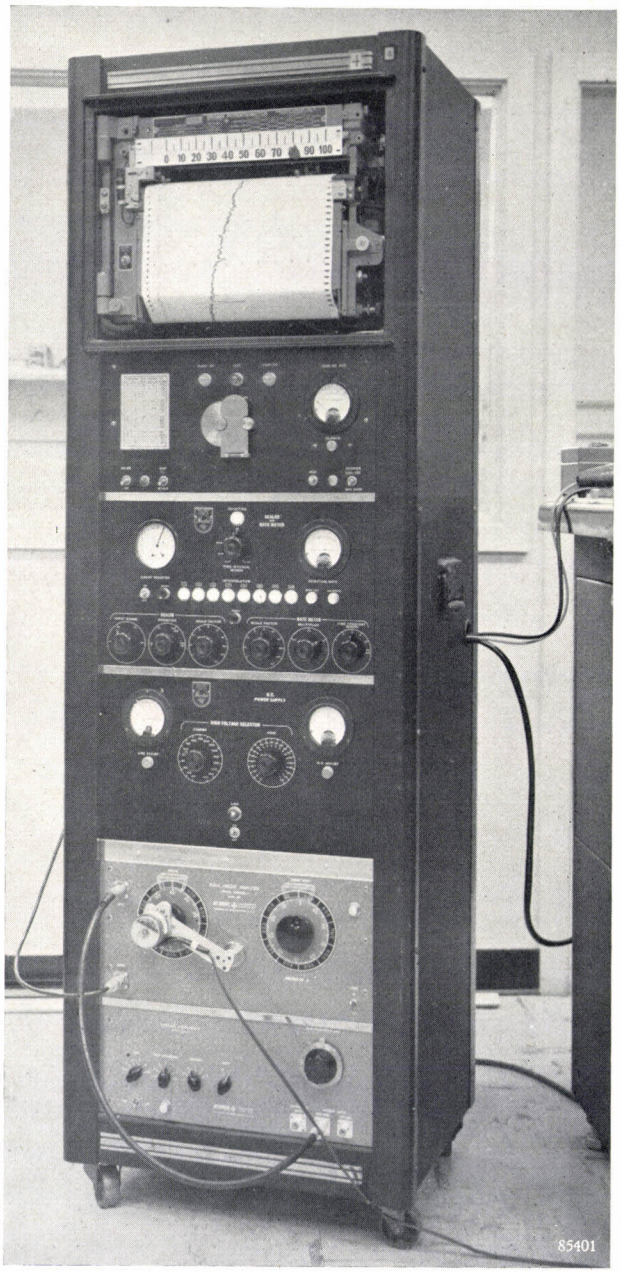


Fig. 1. Rack containing circuitry and accessories for the measuring and recording of diffracted X-ray intensities with the "Norelco" X-ray diffractometer.

Counting statistics [2])

Two measurements of a constant intensity, in which counting is performed during equal times t, will not in general yield the same number N of counts, owing to the random distribution of the quanta in time. A very large number of experiments would be required in order to obtain an average value $\bar{N}$ and the corresponding average value of the counting rate $\bar{n} = \bar{N}/t$ which is to be considered as a true measure of the X-ray intensity. The spread of the results N of individual experiments is approximately given (cf. Appendix) by a Gaussian distribution (*fig. 2*) of mean value $\bar{N}$ and standard deviation

$$\sigma = \sqrt{\bar{N}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

An individual measurement N has a certain probability p of deviating from the true value by an amount $(N - \bar{N})$ smaller than a prescribed value ΔN_p. For the Gaussian distribution it can be shown that the "probable" error, i.e. the error ΔN_p for which $p = 50\%$, is

$$\Delta N_{50} = 0.67\sigma = 0.67\sqrt{\bar{N}} \approx 0.67\sqrt{N}.$$

The probable *fractional* error ε_{50} of the calculated counting rate n is therefore

$$\varepsilon_{50} = \frac{\Delta n_{50}}{n} = \frac{\Delta N_{50}/t}{N/t} = \frac{0.67\sigma}{N} = \frac{0.67}{\sqrt{N}} \quad . \quad . \quad . \quad (3)$$

Thus, for example, 4500 counts are necessary to make $\varepsilon_{50} = 0.01$, i.e. to have a 50% probability that the result deviates less than 1% from the true value. In some cases a result giving a higher confi-

[2]) For a more extensive account of the statistics of counting and a bibliography, see : L. J. Rainwater and C. S. Wu, Nucleonics **1**, No. 2, 60-69, Oct. 1947; **2**, No. 1, 42-49, Jan. 1948.

dence in the prescribed accuracy is desired, with say $p = 90$ or 99%. This will necessitate longer counting intervals. The theory shows that

$$\varepsilon_{90} = \frac{1.64}{\sqrt{N}} \quad \ldots\ldots \quad (3a)$$

and

$$\varepsilon_{99} = \frac{2.58}{\sqrt{N}}, \quad \ldots\ldots \quad (3b)$$

so that about 27 000 counts are necessary to make $\varepsilon_{90} = 1\%$, and about 67 000 counts for $\varepsilon_{99} = 1\%$.

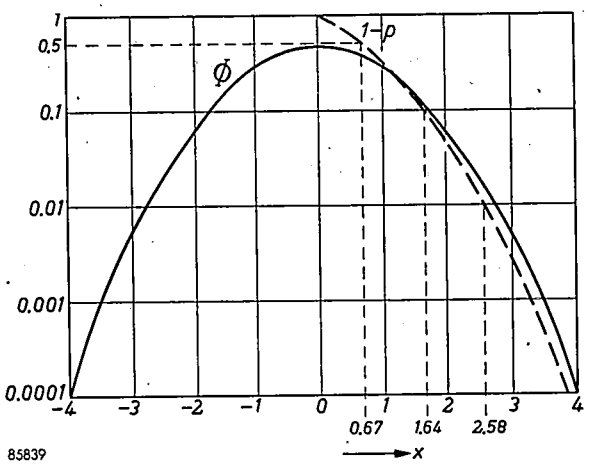


Fig. 2. Gaussian distribution function $\Phi = (1/\sqrt{2\pi}) \exp(-x^2/2)$. For a large count, $\Phi(x)dx$ gives the probability that the result will be between $\overline{N} + x\sigma$ and $\overline{N} + (x + dx)\sigma$, where $\overline{N}$ is the "true" value and σ is the standard deviation, $\sigma = \sqrt{\overline{N}}$. The dotted curve, obtained by integration from the distribution function, shows $1 - p$, where p is the probability that the deviation of a count from the true value $\overline{N}$ is less than x times the standard deviation. For the probabilities $p = 50\%$, 90% and 99% the values of x are 0.67, 1.64 and 2.58 (see eq. 3, 3a, 3b). (From: L. J. Rainwater and C.S. Wu, loc. cit. [2]).

For these numbers of counts, ε_{50} is 0.4% and 0.26% respectively. *Fig. 3* is a graph of formulae (2), (3), (3a) and (3b).

The accuracy of an individual intensity measurement is considerably affected by the "background" of the diffraction pattern. If, at a certain Bragg angle, a total number of N pulses have been counted, while the background in the same angle region contributed N_B pulses in the same time, the diffracted intensity is calculated from $N - N_B$. The statistical error of this difference is given by

$$\sigma = \sqrt{\sqrt{N}^2 + \sqrt{N_B}^2} = \sqrt{N + N_B}$$

or

$$\varepsilon_{50} = \frac{0.67\sqrt{N + N_B}}{N - N_B}.$$

If $N = 500$ and $N_B = 0$, the error would be $\varepsilon_{50} = 3\%$. For $N = 600$ and $N_B = 100$, ε_{50} of the same resulting intensity is increased to 3.5% and for $N = 1000$, $N_B = 500$ it is increased to 5.2%.

It should be noted that the statistical error involved in point-by-point measurements will be markedly diminished in measuring the complete profile of a diffraction line, since a smoothed curve can generally be drawn through the individual measured points. The resulting error of an individual value, e.g. the peak intensity of the line, is difficult to assess but is certainly smaller than that given by (3). On the other hand, the errors discussed in this section refer only to the random arrival of the quanta and do not include possible errors due to instability of the X-ray source, finite resolving time of the counter tube and its circuits, and other causes, which will be briefly discussed at the end of this article.

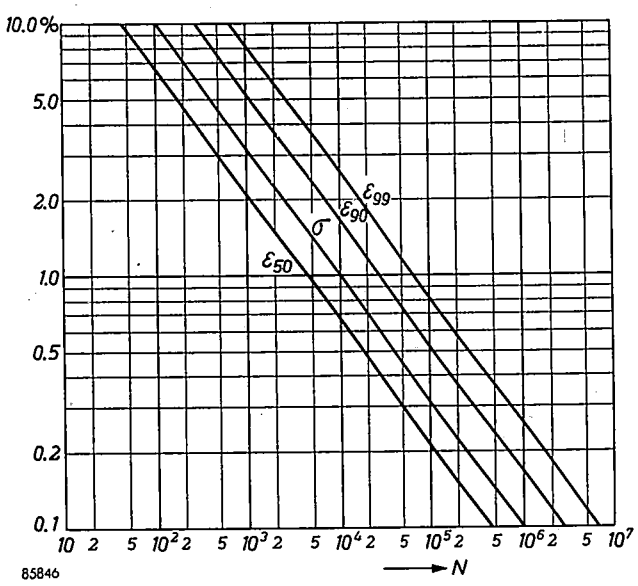


Fig. 3. Probable fractional error ε_{50} of counting rate, and 90% and 99% confidence limits ε_{90} and ε_{99}, as functions of the accumulated number of counts N, for a Gaussian distribution.

The recording counting-rate meter

Counting rates can be determined directly by measuring the mean value of the current composed of the separate counting pulses occurring in the detector. The basic circuit is illustrated in *fig. 4*. The voltage on the resistor R, which is smoothed by the condenser C, depends on the sum of the pulse current contributions received during the RC-time of the circuit. Since all pulses delivered to this circuit are of equal height and duration [3]), the voltage on R is a direct measure of the counting rate, averaged over about the RC-time.

[3]) In order to comply with this condition, averaging in the rate meter is preceded by equalizing the pulses in a pulse shaper.

Owing to the constant averaging time, the method is similar to the preceding one ("fixed time" measurements). Since the counting-rate meter is a direct and continuous indicating device, however, point-by-point measurement and plotting are no longer necessary. The detector is moved by the goniometer to scan the diffraction pattern *continuously*, while the rate meter reading is recorded on a strip chart moved synchronously with the goniometer. This combination of strip chart recorder (uppermost panel in fig. 1) and counting-rate meter is the most convenient and commonly used method in diffractometry since the operations are automatic and the accuracy is sufficient for the large majority of applications, such as qualitative chemical analysys, identification of surface layers, etc.

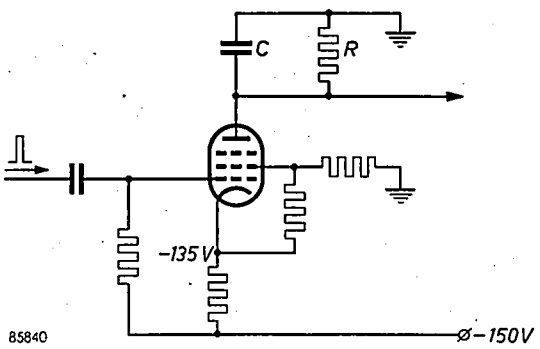


Fig. 4. Basic circuit of the counting-rate meter. The pulses arising in the counter tube are fed to the control grid of a pentode. Since the anode current of the pentode is nearly independent of the anode voltage, a constant charge is passed through the pentode for every pulse, independent of the charge already present on the capacitor C. The voltage produced across the resistor R is smoothed by C.

The accuracy of the relative intensity measurements obtained with the counting-rate meter is also subject to the fundamental limitation due to the statistical fluctuations. Measuring the counting rate n with the counting-rate meter which has a time constant RC can be shown to be equivalent to totalling the number of quanta arriving during a time interval $2RC$ (the operating time of the rate meter of course must be several times in excess of $2RC$ in order that this statement be valid). Thus, an individual counting-rate meter reading contains the probable fractional error

$$\varepsilon_{50} = \frac{0.67}{\sqrt{2n\,RC}} \quad \ldots\ldots\ldots \quad (4)$$

As mentioned earlier, the statistical error will be greatly diminished by drawing a smoothed curve through the more or less ragged strip chart recording [4]). The resultant accuracy of a recorded pattern will depend hardly at all on the time constant but only on the total number of quanta detected during the recording, i.e. on the scanning speed and the width of the receiving slit of the counter tube.

Apart from this statistical error, the counting-rate meter will introduce a certain *distortion* of the line shape when the diffraction pattern is continuously scanned. This distortion is due partly to the width of the receiving slit just mentioned — an effect similar to that encountered in scanning photographically-recorded diffraction patterns with a microphotometer slit — and partly to the rate meter output lagging behind the input by a time roughly equal to the time constant RC [5]). The time lag effect evidently will be more noticeable the higher the rate of change of the counting rate, i.e., the steeper the profile of the scanned line and the higher the scanning speed. The distortion of the line shape results in a shift of the peak in the direction of scanning, a reduction of the peak height and an increase in the width of the line. These effects are demonstrated in *fig. 5a-c* which shows a diffraction line (a doublet) recorded with different scanning speeds. In each of the three cases illustrated, the line is first scanned in the direction of increasing Θ and then in the direction of decreasing Θ (this is easily effected by means of the limit stops on the goniometer which reverse the direction of scanning, see I). It will be noted that the recorded profile of the line differs considerably for the two scanning directions. This difference is due to the fundamental asymmetry of the diffraction line, caused by geometrical aberrations [6]) and by the overlapping of $K\alpha_1$ and α_2 lines. (With regard to the latter point, it should be remembered that the $K\alpha$ line of copper or of whatever target material is used in the X-ray tube, consists of a doublet, the two components α_1, α_2 differing slightly in wavelength and producing two superimposed diffraction patterns; the α_1 and α_2 lines will overlap at small Bragg angles and be separated at large angles.)

In fig. 5*a* to *c* progressive deterioration of the line shape is caused by increasing the scanning speed. The same deterioration occurs when using a fixed scanning speed but increasing the time constant of the rate meter: see *fig. 6a-c*. In fact both effects are equivalent and the distortions in the chart recording

[4]) See: H. S. Peiser, H. P. Rooksby and A. J. C. Wilson, l.c. [1]), p. 223.

[5]) W. Parrish, E. A. Hamacher and M. Gabin, Amer. Cryst. Assoc. Meeting, Penn. State College, 11 April 1950. M. Tournarie, J. Phys. Radium (Supp. No. 1) **15**, 16A-22A, 1954.

[6]) These include the effect of the "line" being ring-shaped, the so-called flat-specimen error and others, which cannot be discussed further here (see I, and A. J. C. Wilson, J. sci. Instr. **27**, 321-325, 1950).

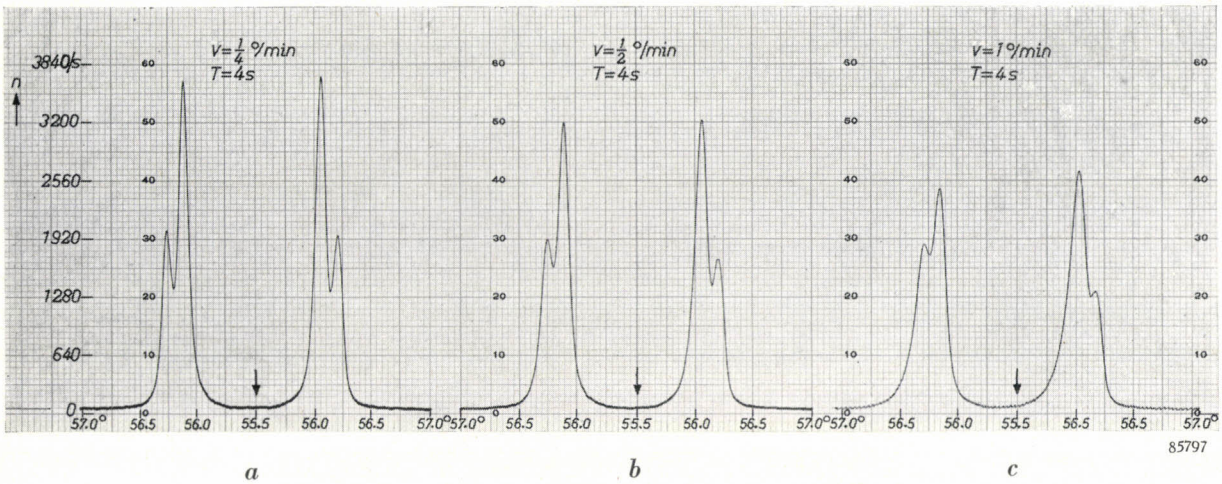


Fig. 5. *a-c*) Diffraction line (silicon, reflecting plane 311) recorded with the counting-rate meter at three different scanning speeds v; time constant $T = 4$ seconds. Unfiltered CuK radiation, angular aperture 1°, receiving slit width 0.15 mm.
Each recording shows the line scanned in both directions (reversal of direction at the arrow). The line profile is seen to deteriorate as the scanning speed is increased. The speed of the recorder chart, which can be varied independently of the goniometer speed v, was selected to make the width of the recordings *a*) - *c*) equal. The number of counts accumulated in any given angular interval naturally decreases with increasing v.

depend only on the *product* of scanning speed and time constant. This can be seen by comparing the line profiles of fig. 5*a* with 6*a*, 5*b*-6*b* and 5*c*-6*c*. Each pair of recordings was obtained with different time constants and scanning speeds, their product, however, being identical.

The resulting distortion of the profile of a particular line (Si 111, CuKα radiation) recorded with a receiving slit of 0.15 mm width, is shown quantitatively in *fig.* 7, where the peak value I_p, the peak shift $\Delta 2\Theta$ and the line width W at one-half peak height are plotted against the product (time constant) $\times$ (scanning speed).

The time constant may be varied in the "Norelco" diffractometer by selecting one of a set of capacitors C in the rate meter circuit (fig. 4). A number of scanning speeds, viz. $^1/_8$, $^1/_4$, $^1/_2$, 1 and 2° (2Θ) per min, can be selected by the use of changeable spur gears in the goniometer drive (cf. I). For general powder diffraction problems a scanning speed of $^1/_4$° per minute and time constant 4 sec have proved most satisfactory. The peak shifts are then less than 0.01° (2Θ) so that peak position corrections (from a table giving correction as a function of Bragg angle) are usually unnecessary. A smaller time constant will decrease the peak shift and other distortion but

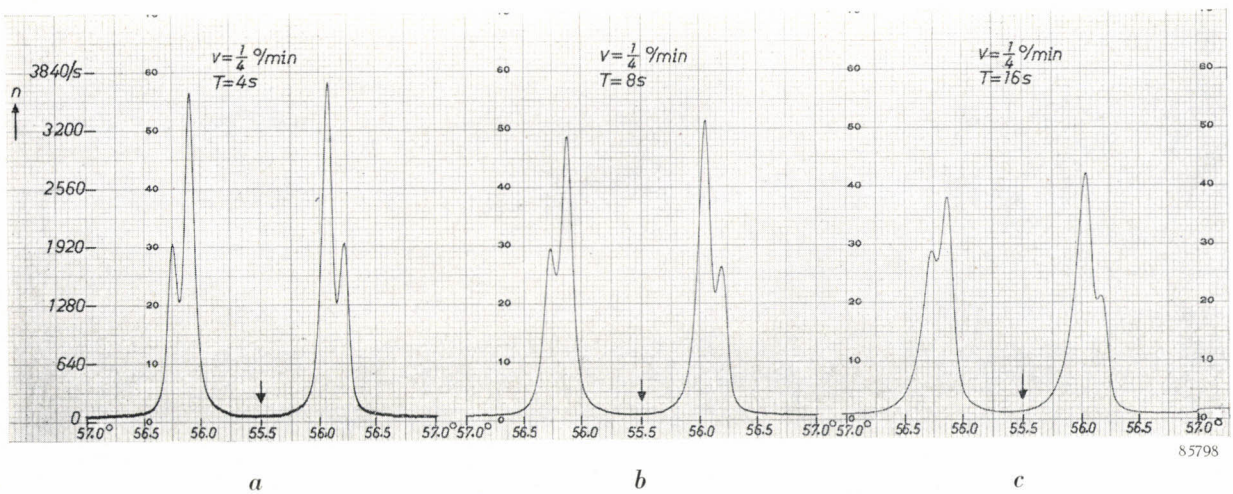


Fig. 6. *a-c*) The same diffraction line as in fig. 5, recorded with the counting-rate meter at a scanning speed $v = \frac{1}{4}$° per min and three values of the time constant T. All other conditions same as in fig. 5. Note that the deterioration of line shape with increasing T is similar to that in fig. 5. The number of counts in corresponding angular intervals, however, is the same for all three recordings.

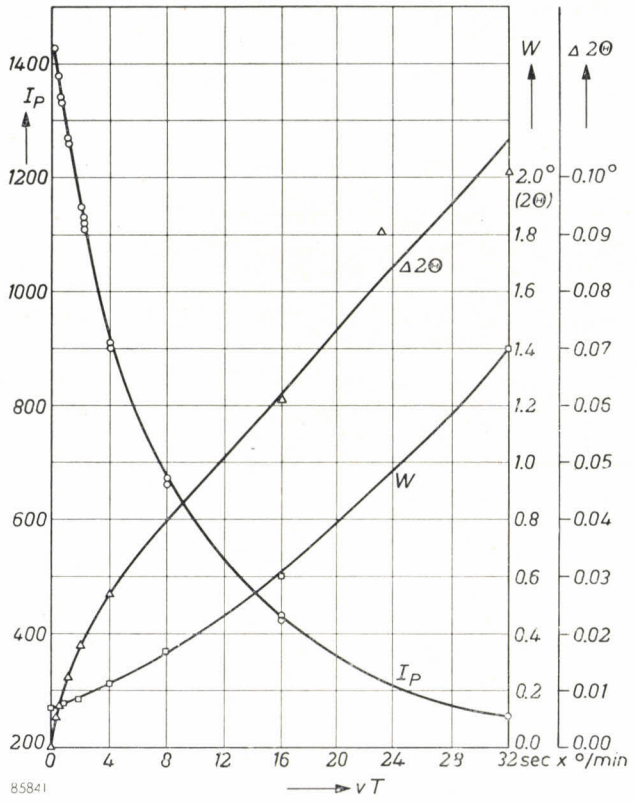


Fig. 7. Effect of the line distortion caused by scanning speed v and time constant T, on the peak height I_P (after subtraction of the background intensity I_B), the angular peak location (shift $\Delta 2\Theta$) and the with W at one-half peak height. The effects depend only on the product vT. These measurements were made on the Si 111 line with CuKα radiation (Kα doublet not resolved) and receiving slit width 0.15 mm. The curves are valid only for the specified experimental conditions and will vary with line shape and receiving slit width. Lines with similar profiles in the same pattern, will be affected by the same percentage peak reduction and therefore need no further correction.

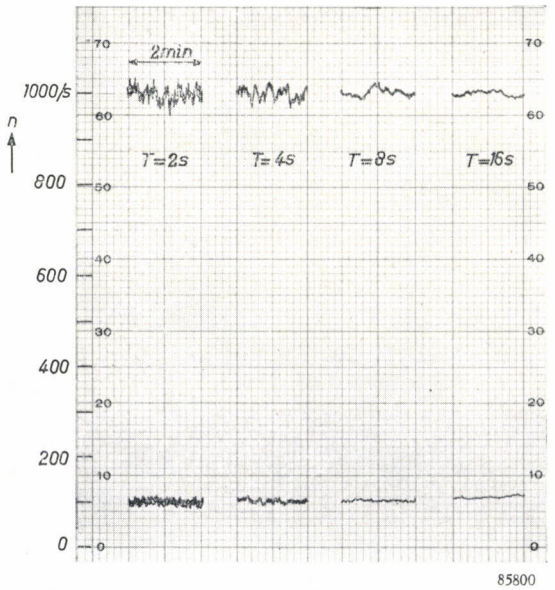


Fig. 8. The "jitter" of the recorded trace, which is due to the statistical fluctuations of the counting rate, diminishes with increasing time constant T. The recordings shown were made at a fixed position of the detector, with counting rates $n = 1000$/sec and $n = 100$/sec.

increase the "jitter" of the recorded trace. This is seen in fig. 6 and is demonstrated more clearly by *fig. 8*. A higher scanning speed, although desirable for more efficient use of the equipment and of the operator's time, will affect the accuracy of the intensity values as was shown above. The effect may be compensated by using a wider receiving slit, thus restoring the total number of counts to the original level. This procedure is frequently employed for rapid surveys of diffraction patterns which are not too complicated, i.e. where the loss in angular resolving power (and the reduced peak to background ratio, see fig. 18) due to the wider slit can be tolerated. The difference in appearance between a recording made for good accuracy and a recording made for a rapid survey is shown in *fig. 9*.

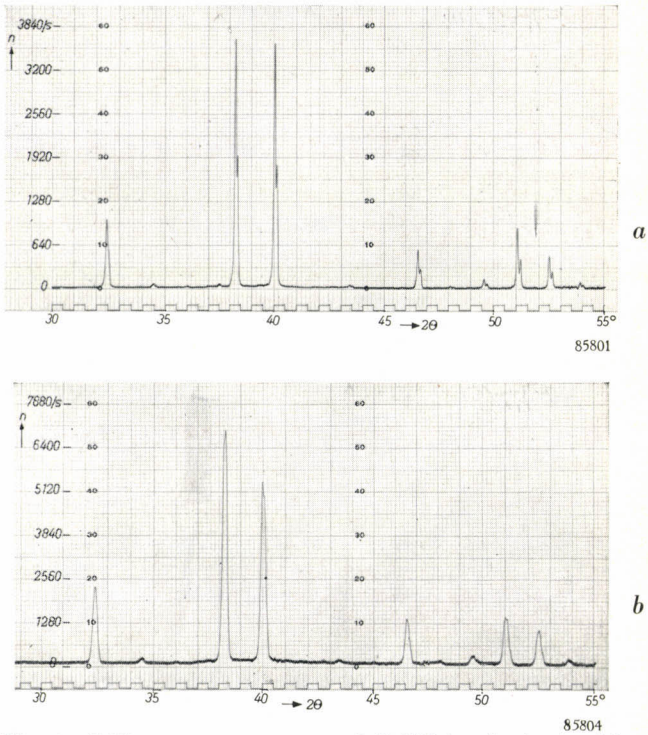


Fig. 9. Diffractometer pattern of $Sr(NO_3)_2$ obtained with CuKα radiation, 30 kV peak, 20 mA.
a) Recorded with 0.15 mm receiving slit at scanning speed $v = \frac{1}{4}°$/min and with time constant $T = 4$ sec.
b) Recorded with 0.60 mm receiving slit, $v = 1°$/min and $T = 2$ sec.
The total number of counts is approximately the same in both cases (92 290 and 88 260 respectively, after subtraction of background), but the angular resolution of record (*b*), which took only $^1/_4$ of the time necessary for (*a*), is much poorer. Peak-to-background ratio for (*a*) 96.5, for (*b*) 42.7 (cf. fig. 18).

It should be pointed out that the errors inherent in continuous scanning with a device which is subject to a time-lag can be greatly reduced by scanning step-wise and stopping at every step for a period several times longer than the *RC*-time. The "Norelco" goniometer is provided with a step-scanning device (see I). The resulting line profile, corresponding in accuracy to a profile measured at very low continuous scanning speed, is illustrated by *fig. 10*.

A novel feature of the "Norelco" counting-rate meter is its relatively small basic range, the full scale deflection corresponding to 50 counts/sec. In order to cover a wide range of intensities, the impulses delivered by the detector are scaled down by the binary scaling circuits mentioned above

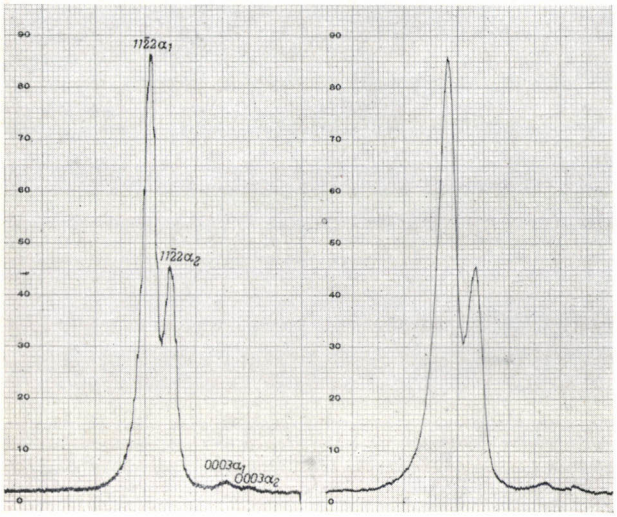


Fig. 10. Diffraction lines (quartz $11\bar{2}2$ and 0003) recorded with the counting-rate meter: *a*) step scanning with 0.02° steps, counting time interval for each step 30 sec, time constant T = 4 sec; *b*) continuous scanning at $v = {}^1/_8{}^\circ$ per min.

before being applied to the rate meter. Thus the full scale range of the ratemeter can be adjusted from 50 to 12800 counts/sec by exact factors of 2. An additional switch changes the basic range to 0.8 × 50 and to 0.6 × 50 counts/sec, so that there are 27 full scale ranges available. The advantage over the usual wide-range type of rate meter is that greatly differing relative intensities (recorded on different ranges) may be accurately compared. Moreover, calibration is simplified by using the power line frequency (mains frequency) as a standard for any of the three basic ranges; all other ranges are automatically calibrated by multiplication by exact factors of 2.

Another advantage of scaling the pulses down before averaging is that higher counting rates can be precisely measured, since counting losses caused by the finite resolving time of the actual counting circuit will be reduced (see the section on Linearity at the end of this article). This favorable influence of the scaling-down procedure is enhanced by its so-called regularizing effect: the scaled down pulses arriving at the counting circuit are more regularly spaced than the original ones, so that their number per second without undue losses may approach more closely the critical frequency existing for a strictly periodic sequence [7]).

[7]) Cf. e.g. E. J. van Barneveld, Fast counter circuits with decade scaler tubes, Philips tech. Rev. **16**, 360-370, 1954/55 (No. 12).

The rate meter circuit is designed so that its output can be combined with that of an additional rate meter to give a voltage proportional to either the sum, the difference or the ratio of two intensities measured. Only one recorder is needed to record this composite result. Such a facility is useful for various monitoring techniques.

The "counting strategy"

The dependence of the accuracy on the accumulated number of counts, see eq. (2) and (3), is the factor dominating all intensity measurements made with Geiger tubes or other counter tubes [8]). One consequence is that when using *fixed* time intervals for counting — as in the methods described above — the relative error is inversely proportional to the square root of the counting rate, i.e. of the intensity.

The question arises whether or not this "natural" variation of accuracy with intensity should be accepted or not. Different variations of accuracy with intensity can obviously be obtained by using *variable* time intervals for counting. In order to answer the above question, the final use that has to be made of the recorded intensities must be taken into account. Two main cases will be considered.

a) For some problems, *ratios* of intensities must be calculated to yield the solution. This applies to the study of strain in cold-worked metals, of particle size distribution, e.g., in clay minerals, of imperfect structure in crystals, etc., where broad diffraction lines are found; information in these cases is obtained from a Fourier expansion of the line profile, whose exact form ("local" relative intensity at each angle) should then be precisely measured [9]). It also applies to quantitative chemical analysis, where the ratio of the percentages of components in a mixture is proportional to the ratio of the intensities (peak values) of their respective diffraction lines. In these cases the local (or peak) values which constitute the numerator and denominator of the ratio to be computed, should be known with approximately the same *relative* accuracy: decreasing the relative error in only one of them will not significantly improve the accuracy of the ratio.

[8]) Strictly speaking, it is also present in photographic intensity measurements but the enormous number of quanta involved renders the statistical error so much smaller than the errors caused by the techniques of density measurements, that it was overlooked for a long time. Only recently, with the advent of X-ray image intensifying devices in which quanta effects are preponderant, have investigators grown more aware of the role of these effects in photographic recording (see e.g. Philips tech. Rev. **17**, 71-77, 1955/56, No. 3).

[9]) B. E. Warren and B. L. Averbach, X-ray diffraction studies of cold work in metals, Chap. V in Imperfections in nearly perfect crystals, Wiley & Sons, New York 1952; A. J. C. Wilson. X-ray optics, Methuen & Co., London, 1949.

For this type of problem, therefore, the counting strategy should be such that a constant relative error is obtained at all counting rates. From eq. (3) it is seen that to attain this result a *fixed number N of counts* has to be accumulated at every Bragg angle position. This is equivalent to making the counting time inversely proportional to the counting rate; see *fig. 11*.

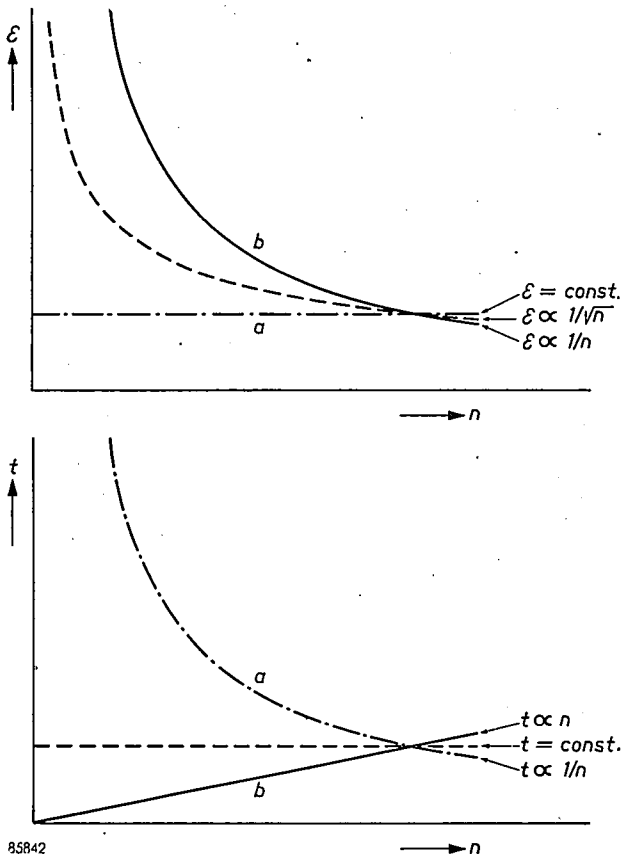


Fig. 11. The relative statistical error ε and required counting time t, as functions of the counting rate n, for three different types of counting strategy.

b) in every measurement of a single peak intensity, the *difference* of two counting rates has to be calculated, viz. that between peak and background. In this case, these two counting rates should be known with approximately the same *absolute* accuracy: decreasing the absolute error in either of them will only slightly improve the accuracy of the difference.

A constant absolute accuracy will also be desirable in ascertaining the profile of a line, when the relative error has been decreased to a degree where the absolute error is equal to that due to the width of the line written by the recorder. The width, of course, will affect all local intensity values by a constant absolute error (about 0.2% of the recorder range).

In order to obtain a constant absolute error at all counting rates, i.e., a relative error inversely proportiona to the counting rate, $\varepsilon \propto 1/n$, the counting time t must be made *proportional to the counting rate*, as is seen from eq. (3) and (1). This is also illustrated by fig. 11.

The "natural" counting strategy, using fixed time intervals for counting, regardless of the counting rate, stands in between cases (*a*) and (*b*). The situation is summarized in *Table I*.

Table I. Different types of "counting strategy".

	Dependence of relative error ε on counting rate n	Dependence of counting time t on counting rate n
Constant relative error ("fixed count", case *a*)	$\varepsilon = \text{const.}$	$t \propto \frac{1}{n}$
Fixed time	$\varepsilon \propto \frac{1}{\sqrt{n}}$	$t = \text{const.}$
Constant absolute error (case *b*) . . .	$\varepsilon \propto \frac{1}{n}$	$t \propto n$

The realization of case (*b*), to our knowledge, has never been attempted in an automatic recording instrument. Case (*a*), on the other hand, represents the well known "fixed count" method. Ways of applying this method will be described in the next sections.

It should be pointed out that still other varieties of counting strategy are conceivable and have been used. The method described by Cooke-Yarborough is such a variety [10]), compromising between the fixed count and the fixed time methods.

The above statements concerning the effect of statistical errors on the ratio and on the difference of two measured values can be understood as follows.

If two counting rates n_1 and n_2 ($n_1 > n_2$) have the relative errors ε_1 and ε_2, the absolute errors are $n_1\varepsilon_1$ and $n_2\varepsilon_2$ (it does not matter whether the "probable" errors ε_{50} or the 90% or 99% confidence limits ε_{90} or ε_{99} are considered).

The relative errors ε_r of the *ratio* n_1/n_2 and ε_d of the *difference* $n_1 - n_2$ will be:

$$\varepsilon_r = \sqrt{\varepsilon_1^2 + \varepsilon_2^2},$$

$$\varepsilon_d = \frac{\sqrt{\varepsilon_1^2 + \left(\frac{n_2}{n_1}\varepsilon_2\right)^2}}{\left(1 - \frac{n_2}{n_1}\right)}.$$

If ε_1 is fixed, it is seen that in both cases the second term under the root should not be appreciably larger than ε_1^2, since this would greatly increase the resulting error (the latter can be visualized as the hypotenusa of a right-angle triangle). On the other hand, little profit is gained by making the second term much smaller than ε_1^2, since in that case the resulting

[10]) F. H. Cooke-Yarborough, The counting of random pulses, J. Brit. Inst. Radio Eng. **11**, 367-380, 1951.

error would remain approximately equal to ε_1. A suitable procedure, therefore, is to make the second term approximately equal to the first. Thus for the ratio, it is desirable to have $\varepsilon_2 \approx \varepsilon_1$ (equal relative errors of both counting rates) and for the difference it is desirable to have $n_2\varepsilon_2 \approx n_1\varepsilon_1$ (equal absolute errors).

In a more rigorous treatment of the problem[11]), one would have to take into account that a decrease of ε_2, the error of the smaller of the two counting rates, will require a greater expenditure in time than an equally important decrease of ε_1. The most "economic" proposition would be to ensure that a small improvement in ε_r (or ε_d) would cost the same additional counting time whether it be achieved by decreasing ε_1 or by decreasing ε_2. A simple calculation shows that the counting time would have to be chosen proportional to the square root of the counting rate in order to meet the said requirement for ε_r, and inversely proportional to the square root of the counting rate in the case of ε_d.

Point-by-point measurements with fixed count

Point-by-point measurement of a diffraction pattern according to the "fixed count" method is performed in the following manner: in each goniometer position, the clock-type mechanical register actuates a count stop circuit after it has made 100 steps, i.e. after one complete revolution of the main pointer. Thus the scaling circuits are in operation for the time necessary to accumulate a predetermined number of counts N, viz., 100, 200, . . . 12 800 or 25 600, and the time interval t is measured by an electric stop clock (*figs. 12* and *13*). The value t is used to calculate the counting rate $n = N/t$, which can then be plotted against the Bragg angle.

Because the number of counts is preselected at a fixed value, the relative accuracy of all measured intensities will be equal and have a preselected value (eq. 3). Of course the method has the same disadvantage of being time-consuming as the point-by-point measurement with "fixed time" counting. In fact the disadvantage is even more pronounced because the time required for each measurement varies with the intensity and becomes very long at low counting rates (fig. 11). We must not complain of this peculiarity of the method, however, since it really represents *le défaut de ses merites*, and its

[11]) The following considerations were pointed out to us by Dr. H. C. Hamaker of the Philips Laboratory at Eindhoven. We are also indebted to Dr. P. M. de Wolff of the T.N.O., Delft, for general suggestions concerning the comparison of counting stategies.

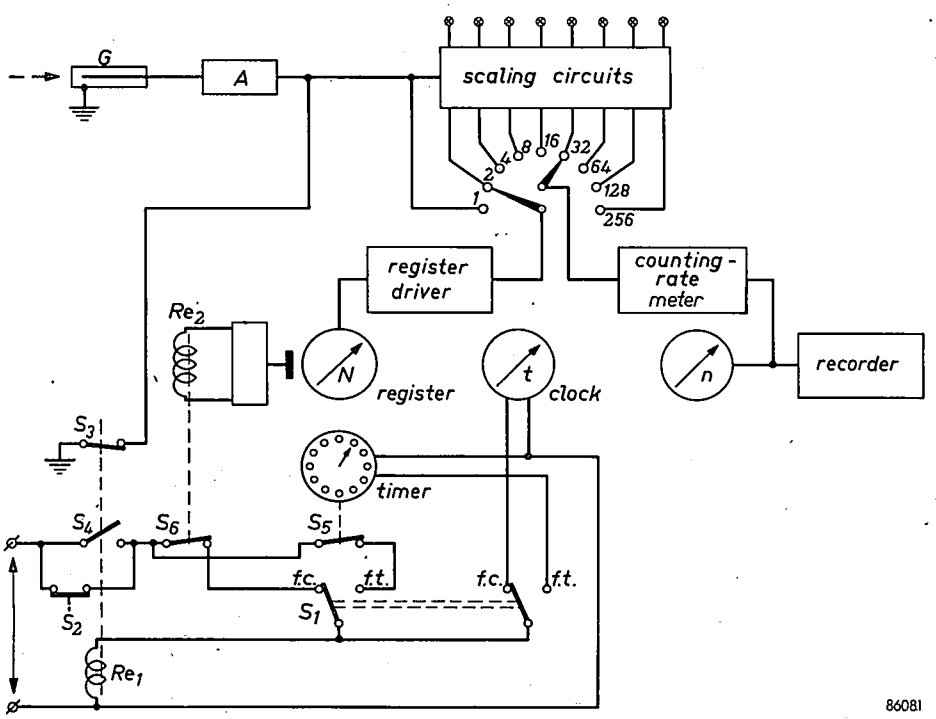


Fig. 12. Schematic block diagram of circuits for measuring intensities by different methods. The pulses produced in the detector G (e.g., a Geiger counter tube) pass through a pulse shaper and amplifier A to the scaling circuits consisting of eight binary stages. *Fixed time methods:* Pulses from any selected stage of the scaling circuits are counted in the register N for a time selected on the timer by setting the operation selection switch S_1 to position f.t. ("fixed time"). The measurement begins by pressing push button S_2; this will start the timer and energize relay Re_1 which closes the holding contact S_4 and opens switch S_3, so that counts are passed from the amplifier to the scaling circuits and the register. After the preset time, the timer opens switch S_5; this de-energizes Re_1, closes S_3 and opens S_4, thus stopping the operation. For making a rate meter recording, S_2 is kept closed by means of another switch (not shown) parallel to the push button so that counting continues indefinitely. If the limit switches on the goniometer are used, they actuate microswitches (not shown) which operate S_2. *Fixed count methods:* Selector switch S_1 is switched to position f.c. ("fixed count"). This measurement again begins by pressing S_2. When the main pointer of register N has completed one revolution, a contact (not shown) is closed, relay Re_2 is energized via a relay tube, so that switch S_6 opens and relay Re_1 is de-energized, thus completing the cycle. The counting rate is calculated from the time on the clock which starts and stops simultaneously with the register.

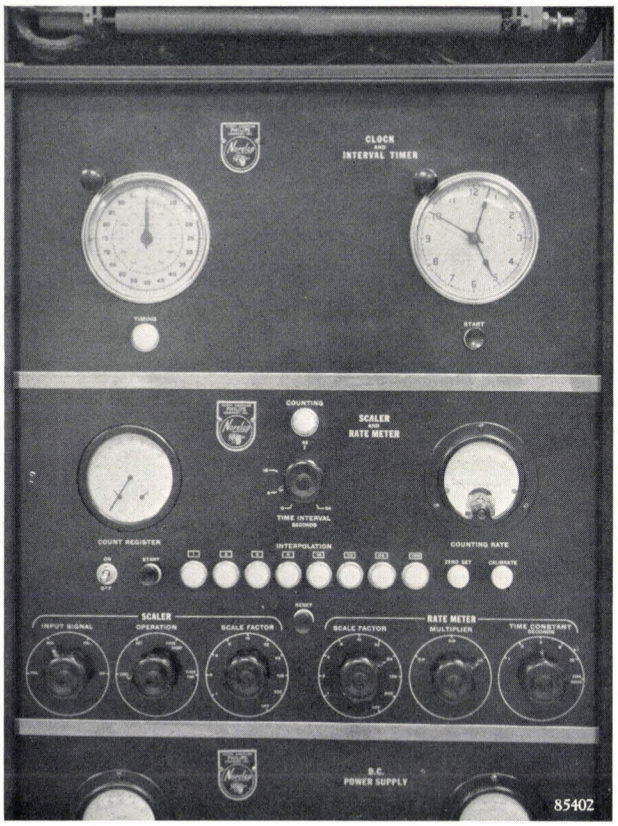


Fig. 13. Front view of the panels which contain the scaling and rate meter circuits (below, cf. also fig. 1), the interval timer and clock (above).

sting can be removed by applying the device described in the following section.

The counting-rate computer

To eliminate the labor involved in making accurate intensity measurements with the fixed count method, the "Norelco" counting-rate computer has been developed [12]). This device automatically performs the sequence of operations required and records the data. It will be described at some length in an article in the next issue of this Review, and only its basic characteristics and performance will be mentioned here.

The computer chassis is about the size of the scaling circuit chassis and fits in the electronic circuit rack in place of the clock and interval timer (fig. 1, second panel from top). The computer is of the analogue type; it is based on Ohm's law and delivers an electric current inversely proportional to the time required to accumulate a predetermined number of counts, at a given Bragg angle position of the detector. This current is fed to the strip chart recorder so that a plot of the counting rate (on a linear scale) is obtained on the recorder. After one value has been recorded, the instrument advances the chart a fixed increment, advances the counter tube a fixed angular increment and then starts the counting circuits automatically to repeat the cycle. A step-wise recording is obtained in this way. The angular increment may be selected as 0.01°, 0.02° ... 0.05° (2Θ). The statistical error (total number of counts for one position) and full scale intensity are also selectable.

A typical recording obtained with the counting rate computer is shown in *fig. 14*. About $2\frac{3}{4}$ hours was required to record the pattern in the limited Bragg angle region shown, but the operation was entirely automatic.

The Cooke-Yarborough method

In fixed count measurements, time consumption is largest for the lowest intensities and especially for the background. The desirability of not spending too much time on one recording (though it may be only instrument time) has led to the development of several modified "fixed count" methods, which sacrifice some of the accuracy at the lower intensities. One method consists of increasing the angular increment when scanning regions containing only the background. For example, steps of 0.05° could be used in such regions and 0.01° for the lines. Another method is to reduce the scaling factor for the background measurements, thereby increasing their relative statistical error. Both methods can be combined, but the gain in speed is limited by the "recycling time" of the computer, i.e., the time not required for the actual counting but for the recording and the chart and goniometer movement. This shows that the first-mentioned expedient (increase of step width) is more effective since it diminishes the number of

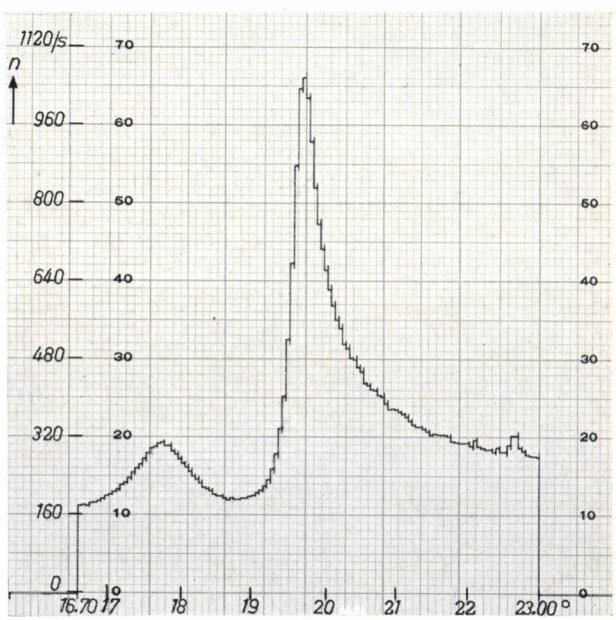


Fig. 14. Part of recording of a diffraction pattern ($16.7° < 2\Theta < 23.0°$) made with the counting rate computer. Montmorillonite (dry specimen), CuKα radiation with nickel filter, steps 0.05° with $N = 6400$; recording time $2\frac{3}{4}$ hours.

[12]) W. Parrish and E. A. Hamacher, Geiger counter X-ray diffraction spectrometer: instrumentation and techniques, Trans. Instr. Meas. Conf. Stockholm 1952, p. 95-105.

recycling periods; the second method, moreover, has the drawback of changing the amplitude recorded on the chart paper.

Both modifications have the disadvantage of requiring the operator's attention to change scale factors or angular increments at the appropriate time. To avoid this and to return to completely automatic operation, the method devised by Cooke-Yarborough [10]) may be employed. He recommended adding periodic pulses at a constant, controlled, low rate to the random pulses that are to be measured. At high counting rates the added pulses, say 100 per second, hardly play a role so that the high and uniform relative accuracy characteristic of the fixed count method will remain unaffected in measuring the peaks. At low counting rates the time required for a measurement (i.e., for accumulation of the preselected number of counts) obviously approaches a fixed value, dependent upon the added constant pulse rate. The Cooke-Yarborough method therefore stands between the fixed count and the fixed time counting methods.

When using the Cooke-Yarborough method in conjunction with the counting-rate computer, it is an important consideration that the intensity scale should remain linear. It will be shown in the article on the computer that this is obtained by selecting a pulse rate, f, which is a fixed fraction, x, of the full scale counting rate: $f = x(n+f)_{max}$. A pulse generator has been designed which is synchronized with the power line and delivers pulses on a basis of $x = 0.075$, so that for $(n+f)_{max}$ varying from 50 to 3200 counts/sec (the upper limit for the computer) pulse frequencies of 3.75, 7.5, ..., 240/sec are provided.

The gain in time can be seen by referring to *fig. 15*. The dotted lines show the time required in normal operation for accumulating the predetermined number of counts $N(= 100 \times$ scale factor), as a function of the counting rate (X-ray intensity), for different values of N. The minimum time, being equal to 8 sec for all values of N, corresponds to the maximum output current of the computer. The solid curves show the time

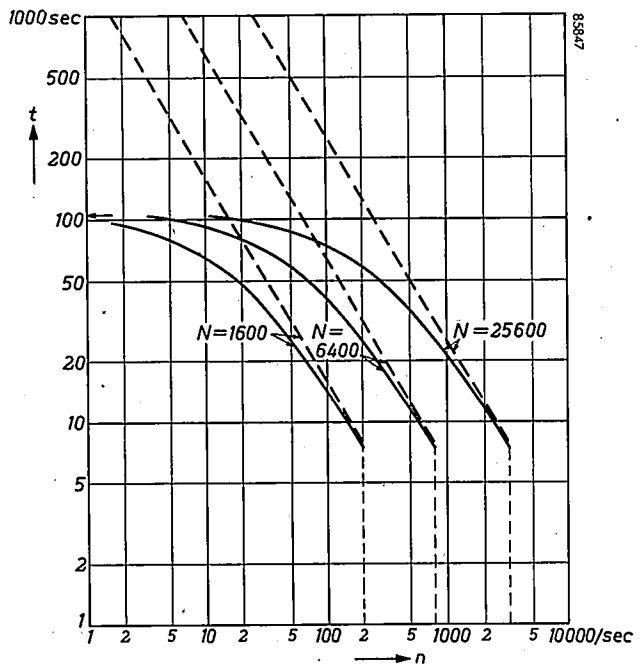


Fig. 15. Time t required to accumulate a number of counts N, in normal operation (dotted lines) and in the Cooke-Yarborough method (solid lines). The added pulse frequency f is 0.075 times the full scale counting rate.

required by the Cooke-Yarborough method with $x = 0.075$. The time at low counting rates is considerably reduced and approaches a maximum of 106.6 sec in all cases. The time for the highest counting rates is not greatly affected: it is decreased from 8 to 7.45 sec.

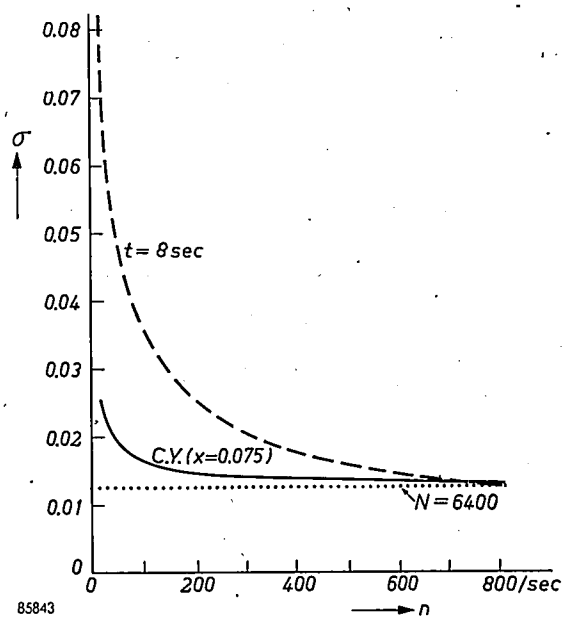


Fig. 16. The standard deviation σ as a function of the counting rate n, for the Cooke-Yarborough method (C.Y.), for fixed time measurements with $t = 8$ sec (— — —) and for fixed count measurements with $N = 6400$ (. . . . .).

The actual time required to scan a given portion of a diffraction pattern will depend upon the background level and the number of lines and their intensities. In one specific case, the counting-rate computer required 8 hours to record a pattern which included a sharp line profile and a large background region; the use of the Cooke-Yarborough method reduced the time for the same pattern by a factor of about 2. The expected saving in time is not fully realized because of the recycling time of the computer (see above) which amounts to 60 sec per cycle.

The loss in statistical accuracy involved by the Cooke-Yarborough method must also be considered. The standard deviation with this method is [10]):

$$\sigma_{C.Y.} = \frac{1}{\sqrt{N}} \sqrt{\frac{n+f}{n}} \quad \ldots\ldots \quad (5)$$

$\sigma_{C.Y.}$ as a function of n, for $f = 60$ and $N = 6400$ (range $(n+f)_{max} = 800$) is shown in graphical form in *fig. 16*; the σ-lines for normal fixed count operation ($N = 6400$) and for fixed time operation (8 seconds) are drawn in for comparison. The Cooke-Yarborough modification resembles fixed time operation in that as the counting rate decreases the relative statistical error increases. At high counting rates, all three methods give essentially the same statistical accuracy.

Integrated intensity measurements

The peak height of a diffraction line is often tacitly assumed to represent the line intensity. Although quite satisfactory and convenient for the routine identification of substances, peak heights are

less suitable for precise quantitative analysis. For this purpose, the "line intensity" should be a quantity strictly proportional to the percentage of the substance present. We have seen that this is not the case for the peak height, which in rate meter recordings is markedly dependent upon the scanning

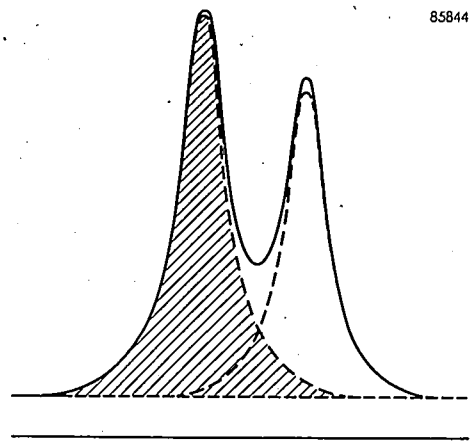


Fig. 17. The area of slightly overlapping diffraction lines reflected from different lattice planes can be integrated by sketching in the "tails" of the lines in the overlapping region and using a planimeter.

speed and even upon the scanning direction. A quantity which physically represents the line intensity more correctly is the total diffracted energy, i.e. the intensity integrated across the entire width of the diffraction line. This integrated intensity (area of a diffraction line) is not affected by the distortion of line profile and change of peak height inherent in the rate meter recording [12]). It is also of importance for structural analysis, since the line intensity thus defined is directly related to the structure of the crystal.

Practical considerations which favor the use of integrated intensities are the following: 1) the statistical error inherent in the measurement of the area of a line is smaller than that in the measurement of the peak, owing to the much larger number of counts summed; 2) the problem of the overlapping of $K\alpha_1$ and $K\alpha_2$ lines is avoided. In the overlapping region the measured peak height of a line will be larger than the α_1-peak height (which is about double the α_2-peak height) by a factor which depends upon the extent of overlap. On integrating the line intensity, the total area including the α_1 and the α_2 contributions is always taken, so that it does not matter whether α_1 and α_2 are separated or not.

Table II, which lists the relative values of integrated intensity and of peak intensity of a few lines of one specimen, illustrates the effect of the α_1 - α_2 separation.

Of course, integration is only possible where there is no severe overlapping of lines reflected from different lattice planes. In cases of slight overlapping, the "tails" of the lines may be sketched in as shown in *fig. 17*. The integration is then performed by means of a planimeter. If no overlapping occurs, a very easy method is to use a receiving slit wide enough to include the entire line and count at the peak position (*fig. 18*). For a total line width of $0.01°(2\Theta)$, a 1.2 mm slit would be required. Another easy possibility for performing integrated intensity

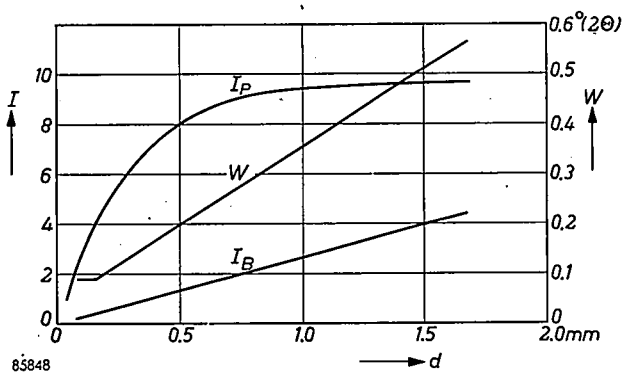


Fig. 18. When a diffraction line is scanned with receiving slits of increasing widths d, the recorded peak intensity I_P (with background subtracted) also will increase. When the receiving slit becomes wider than the line, the intensity no longer increases and the peak value indicates the integrated line intensity. The background I_B and the width W (degrees 2Θ at half peak height) of the recorded line profile continue to increase with receiving slit width. These curves were obtained using the 111 line of silicon with Cu $K\beta$ radiation using an angular aperture $2\alpha = 1°$. Wider lines and $K\alpha$-doublet separation will change the given values.

Table II. Comparison of area and peak height of a few silicon lines to show influence of $K\alpha_1$-α_2 separation. Unfiltered CuK radiation was used with an angular aperture of 1°; receiving slit width 0.15 mm; scanning was done with speed $\frac{1}{4}°(2\Theta)$ per min, time constant 4 sec, scaling factor 128. All values have been corrected for the background.

Line	2Θ - range	Total number of counts (area of line)	Counting rate at peak of line counts/sec	Relative intensity derived from	
				area	peak
111	29.5 - 27.5°	48 040 ± 0.5%	$\alpha = 1336 \pm 0.5\%$	100	100
220	48.5 - 46.5°	36 590 ± 0.5%	$\alpha_1 = 903 \pm 1.1\%$	74.0	67.6
331	57.0 - 55.0°	21 376 ± 0.8%	$\alpha_1 = 511 \pm 0.9\%$	44.4	38.3

measurements is provided with the "Norelco" equipment. The goniometer scans the pattern continuously, and the scaler is started manually at one side of a line and shut off after the line has been scanned. During scanning, the mechanical register sums the counts, the total number being proportional to the area of the scanned portion of the pattern.

The integrated background intensity should of course be subtracted. This is measured separately, for example by counting the total number of background pulses in a region adjacent to the line and of the same width as the scanned line. It is also possible to correct in a simple way for a background which varies slowly with the angle.

In cases were the line is weak it may be scanned back and forth several times to accumulate sufficient counts for the desired statistical accuracy.

When the relative intensities of two or more lines are to be compared, the same receiving slit width and scanning speed must be used or allowances made for any differences in these factors.

It was stated above that the area of a line is more suitable than the peak height for representing the line intensity. Similarly for the exact angular position of a line, the *center of gravity* defined by

$$2\Theta_1 = \frac{\int n \times 2\Theta \, \mathrm{d}(2\Theta)}{\int n \, \mathrm{d}(2\Theta)} \quad \ldots\ldots \quad (6)$$

is a better indication than the angular peak position. Since the measured line profiles are unsymmetrical and the peaks have a finite breadth, there is always some doubt about how to define the precise position of the peak; these doubts do not interfere with the determination of the center of gravity. Moreover, the effect of the most important systematic errors (which depend on the Bragg angle 2Θ and which have been calculated) can be easily allowed for with the center of gravity, whereas the corrections for the peak positions are so complicated they have been worked out for only a small angular region [13]). The center of gravity is readily calculated from a counting rate computer record, and although this procedure takes considerable time, it is useful in the determination of lattice parameters and other cases where the highest precision is required.

Incidentally, eq. (6) demonstrates that the angular accuracy depends on the accuracy of the intensity measurements. In a typical case, an accuracy of a few thousandths of a degree was obtained when measuring the intensities to an accuracy of $\varepsilon_{50} = \pm 0.8\%$, in 0.01° to 0.05° steps, across lines in the back reflection region. (The goniometer is also capable of setting and reading separate angles to this accuracy.)

Linearity of intensity measurements

All the above methods of mapping the intensity vs. angle relation are based on a linear relationship between diffracted X-ray intensity and number of counts. As pointed out in this Review on previous occasions [7]) [14]), such a linear relationship, even when background counts due to undesired radiation have been subtracted, exists only to a limited extent. Non-linearity is caused by the "dead time" of the detector and of the scaling circuit, which prevents quanta arriving within too short a time interval from being counted separately ("counting loss"). The resolving time of the Geiger counter tube (type No. 62019) usually used in the "Norelco" diffractometer is 170 microseconds. The effective resolving time when using the counter for detecting Cu Kα radiation produced by a 35 kV peak full-wave rectified X-ray source [15]) was found to be $\tau = 270$ μsec. This was established by measuring the response of the Geiger counter (observed counting rate n) to a known X-ray intensity (true pulse rate n_0); the response should obey the well-known relationship

$$n = \frac{n_0}{1 + \tau n_0} \quad \ldots\ldots \quad (7)$$

The measured response is plotted in *fig. 19*. This

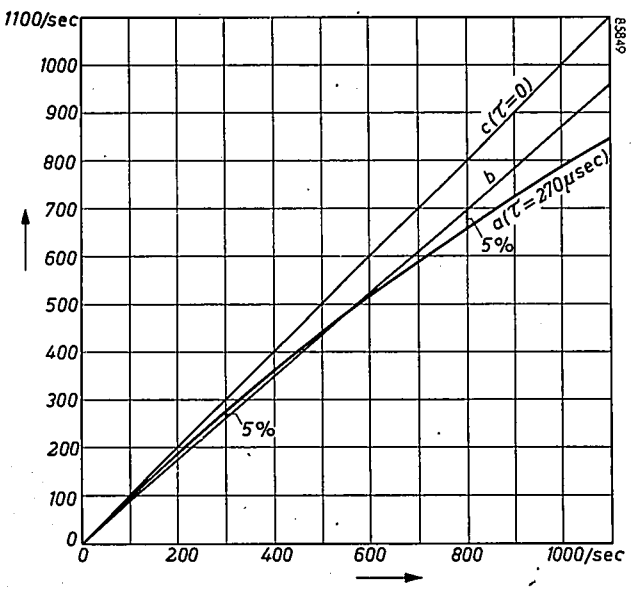


Fig. 19. Response (a) of the Geiger counter tube to random quanta as a function of the true average count rate. The deviation from linearity is due to the finite resolving time for successive quanta. The measured effective "dead time" of the "Norelco" No. 62019 Geiger counter is $\tau = 270$ μ sec for Cu Kα radiation from an X-ray tube operated with full wave rectification at 35 kV peak. With the X-ray tube operated at constant potential a smaller effective dead time is found; cf. note [15]). For $\tau = 0$, the straight line c would be obtained.

[13]) W. Parrish and A. J. C. Wilson, Precision measurement of lattice parameters of polycrystalline specimens, International Tables for X-ray Crystallography, Vol. II, in press.

[14]) Philips tech. Rev. **10**, 1-12, 1948/49.

[15]) The effective resolving time of the Geiger counter is longer than its true resolving time owing to the unfavorable bunching of the quanta: CuK radiation is produced only during a portion of the time, viz., whenever the undulating voltage exceeds the 8.8 kV excitation potential of Cu.

curve (*a*), or eq. (7), which are valid up to about $n = 1500$ counts/sec, can be used for correcting individual counting rates. When no corrections are made and a linear response is assumed, the straight line (*b*) drawn through the measured response curve in fig. 19 shows that the error is less than $\pm$ 5% for intensities whose observed counting rate is below 800 counts/sec.

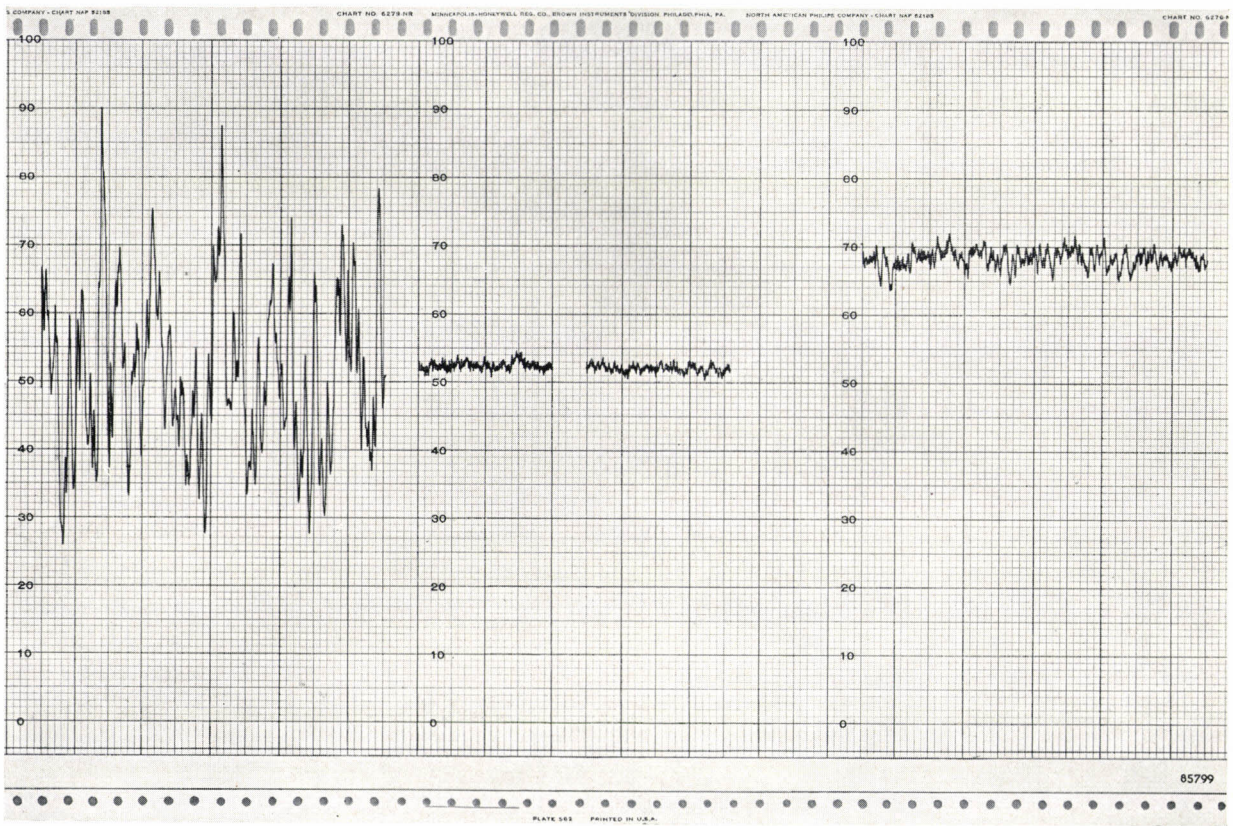
New radiation detectors have been developed recently for use with the "Norelco" diffractometer, viz., the proportional counter and the scintillation counter [16]). Both have been designed in such a way that they can be substituted for the Geiger counter tube by the mere plugging-in of detectors and amplifier stages. (The lowest two panels of the rack shown in fig. 1 are used in conjunction with these detectors.) Both the new detectors have resolving times so short that the linearity of the counting system is limited principally by the dead time of the scaling circuit, which in the "Norelco" equipment is about 7 μsec. In this case, the deviation from linearity is 0.7% at an observed counting rate of 1000 counts/sec.

In applying the Cooke-Yarborough method, non-linearity is enhanced by the additional counting losses caused by the added pulses. Calculation of the largest loss, occurring at the highest added pulse frequency (240/sec), showed that with the 7 μsec resolving time, the deviation from linearity at $n = 1000$ counts/sec is increased only to 0.9%.

Specimen and X-ray source errors

To conclude this article, we should further qualify the statement that either the measured peak height of a diffraction line or its area will

[16]) J. Taylor and W. Parrish, Absorption and counting efficiency data for X-ray detectors, Rev. sci. Instr. **26**, 367-373, 1955 (No. 4).



a *b* *c* *d*

Fig. 20. *a*) Changes in diffracted X-ray intensity recorded at a fixed Bragg angle position of the detector when a specimen consisting of relatively large crystallites (30 - 50 μ) is slowly ($^1/_7$ rev/min) and continuously rotated in its own plane; silicon 111 line with Cu Kα radiation. *b*) By *rapid* rotation of the same specimen in its own plane (77 rev/min) the large intensity changes are averaged out. The remaining fluctuations are similar to those obtained with a stationary specimen (*c*). (The latter recording was made with a position of the specimen selected to give the same average intensity as obtained in *b*.) *d*) Slow rotation of a specimen with a much larger number of crystallites (size 0 — 5 μ) shows only small variations of intensity. (The 0-5 μ particles give a higher average intensity than the 30 - 50 μ particles probably owing to closer packing of the surface of the former specimen.)

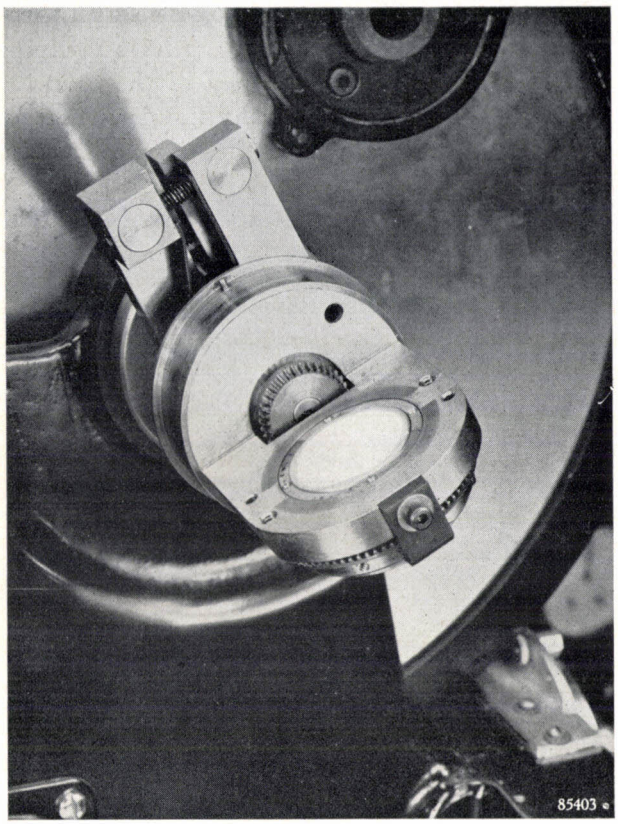


Fig. 21. Device developed by K. Lowitzsch of the Irvington Laboratory for rotating the specimen in its own plane.

offer a reliable indication of the concentration or of structural characteristics of a substance. Even if errors in the alignment of the goniometer or positioning of the specimen are disregarded (they do not greatly affect the integrated intensity of a line) the above statement will be true only for an ideal specimen. Preferred orientation, for example, should be entirely avoided, and the number of crystallites should be large enough for every possible reflecting position to be adequately represented. If there are too few crystallites, large fluctuations in diffracted intensity may be found in a fixed goniometer position when the flat specimen is slowly rotated in its own plane (*fig. 20a*). The remedy for this effect is to average out these fluctuations by *rapidly* rotating the specimen in its own plane during the measurements (fig. 20*b*). An accessory apparatus designed for the purpose (flat specimen spinner) is shown in *fig. 21*. Spinning the specimen, of course, will not generally eliminate errors due to a preferred orientation of crystallites. Because of the difficulty of preparing specimens entirely free of preferred orientation, an accuracy of 1% of the intensity measurements is therefore usually sufficient.

This 1% is actually also the long-time variation of the intensity of the X-ray source provided for the "Norelco" diffractometer. In cases where specimen conditions warrant a higher accuracy, it will be necessary to re-adjust the slowly varying X-ray intensity during one recording; this readjustment can be made in a simple manner by inserting a standard specimen and adjusting the X-ray tube current to give a prescribed counting rate for a given diffraction line. An X-ray source having a long-time stability better than 1% (PW 1010) is provided with the version of the diffractometer made by Philips in Europe.

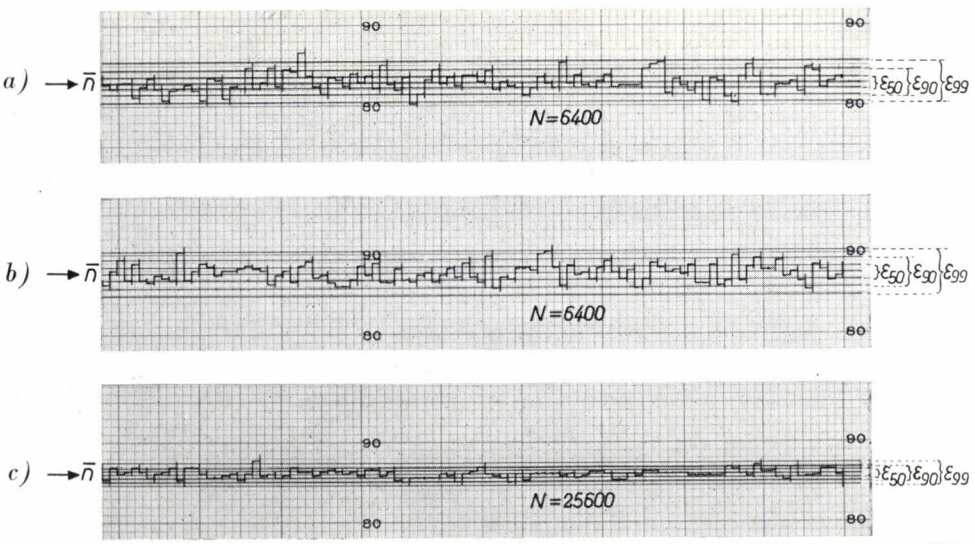


Fig. 22. Measurements of a constant X-ray intensity with the counting-rate computer, in order to check the statistical fluctuations of the counting rate. In each of the three cases, 100 measurements of a fixed number N of counts are performed.
a) Diffracted X-ray beam from a Cr-target, 13 kV peak, full wave rectification. $N = 6400$.
b) X-radiation from a radioactive source. $N = 6400$.
c) Diffracted X-ray beam from a Cr-target, 37 kV peak, full wave rectification. $N = 25\,600$.

Appendix: Statistics of X-ray quanta

The standard deviation $\sigma = \sqrt{N}$ (eq. 2) is valid for counting experiments performed on ideally random sequences of quanta (the results show a practically Gaussian distribution). In many cases the X-ray quanta are not produced at a strictly random rate, since the tube is not operated at constant potential but on a half or full-wave rectified voltage. Moreover the random sequence is marred to some extent by the finite resolving time of the counter tube and circuits. For a correct assessment of the accuracy obtained in the intensity measurements it is therefore important to check whether formula (2) is applicable to these measurements.

For a rough check, recordings involving 100 scanning steps with the counting rate computer have been made *a*) of a diffracted X-ray beam (Cr-target, 13 kV peak), with $N = 6400$; *b*) of the X-radiation of a radioactive source (where the quanta are emitted at a strictly random rate), with $N = 6400$; *c*) of a diffracted X-ray beam (Cr-target, 37 kV peak), with $N = 25\,600$. The recordings are shown in *fig. 22*. In fig. 22*a* and *b* according to eqs. (3), (3a), (3b), 50 steps should deviate less than $\varepsilon_{50} = 0.84\%$ from the average height, 90 steps less than $\varepsilon_{90} = 2.06\%$ and 99 steps less than $\varepsilon_{99} = 3.22\%$. The real numbers of steps found within these limits were 51, 78 and 97 for the X-ray tube and 42, 90 and 99 for the radioactive source. In fig. 22*c* the corresponding numbers of steps found within ε_{50}, ε_{90} and ε_{99} are 51, 91 and 99. These numbers agree reasonably well with the theory, although one number (78) is found which deviates from the theoretical number rather more than might be expected — from statistical considerations — in 3 groups of 100 measurements. However, a much larger number of measurements would be necessary to make sure whether significant deviations really occur.

Summary. In X-ray diffractometry the measurement of X-ray intensities is achieved by counting quanta with a radiation detector such as the Geiger counter, proportional counter or scintillation counter. The counting rate is computed from the number of counts (which is subject to a statistical error) and the counting time interval. Different systems of "counting strategy" are discussed. Counting during a fixed time at every Bragg angle position of the detector will yield a relative statistical error inversely proportional to the square root of the counting rate, whereas measuring the time necessary for a fixed number of counts will make the relative statistical error independent of the counting rate. A third possibility, yielding a constant absolute error in the counting rates, is also mentioned. The recording counting-rate meter and the counting-rate computer, being part of the "Norelco" X-ray diffractometer equipment, are described as representatives of the fixed time and fixed count methods. Facilities can also be provided for the Cooke-Yarborough method, which is a compromise between both methods. For precise quantitative chemical and structural analysis, the integrated intensities of diffraction lines have to be determined. Corrections for non-linearity of the intensity measurements caused by the dead time of the detectors (counting losses) are discussed. The influence of inadequate crystallite orientation in the specimen can be eliminated by spinning the specimen in its own plane.

FUNGICIDE RESEARCH

by M. J. KOOPMANS *).

632.952

Many agents are known for the control of fungus infections in cultivated plants, but apart from their specific useful properties, all have disadvantages of one sort or another. An active search for new agents is therefore in progress in many parts of the world — including Boekesteyn). This, the third article in the series on the work of this laboratory, gives a description of the research in the field of fungous control.*

General properties of fungi

Research relating to agents for controlling animal and vegetable pests in agriculture and horticulture is among the most important work carried out in the Boekesteyn Agrobiological Laboratory of N.V. Philips-Roxane. A review of this work has already been given in an introductory article [1]), and a second article was specially devoted to the combating of animal pests [2]). We now wish to discuss measures against another danger to which cultivated plants are exposed, namely, attack by the lower-organized plant organisms, known as fungi.

The losses ensuing from attack by fungi may be very appreciable. Thus in the United States annual harvest losses amounting to 2000 million dollars still occur as a consequence of inadequate control of fungous diseases. In tropical regions, a constant crop loss of 30% for such plants as banana, coffee and cocoa is by no means exceptional. In countries for which less dramatic figures can be reported it often happens that only the unceasing application of preventive measures keeps fungous infections within limits.

It is beyond the scope of this article to give a survey even of the most important fungal diseases, but it may nevertheless be useful to try to describe the general properties of these injurious organisms.

Many pathogenic fungi are propagated by their wind-borne spores. Examples are the rust fungus on wheat and the mildew mould on grapes, fruit trees, cereals, etc. Other fungi are propagated by splattering raindrops which carry away their spores, as is the case in scab on fruit and late blight in potato. In general both these groups cause leaf and fruit diseases.

Further, there are species whose spores adhere to the seeds of the host, so that the young plant carries the disease germs with it from the outset. We may cite in this category: bunt in wheat, black leg in beets and various diseases in peas.

Finally there is yet another category of plant-pathogenic fungi, whose members live in the soil and principally affect seedlings, although the symptoms of the disease may only become manifest at a later stage in the life of the plant. Some important crops which are prone to attack as seedlings, are beets, cucumbers and peas.

The vast majority of plant-pathogenic fungi thrive at high temperatures (24 °C) and in humid atmospheres. That is why fungous diseases constitute such a serious problem in the tropics. On the other hand, infection is a matter not only of the virulence of the fungus, but also of the power of resistance of the plant, and it may thus so happen that it is precisely low temperatures which promote infection of the plant. This is particularly true of soil fungi which cause diseases in seedlings.

Review of existing fungicides

Agents which protect plants against fungous diseases, are called fungicides. In the majority of cases their action is prophylactic, that is to say their presence precludes infection: the germination of the fungal spores is prevented. Actual cure of the plant, in the sense that an infection which has already developed, is driven out, or even merely checked, is practically never found in the control of fungal diseases. This problem will be broached from another angle in a subsequent article.

Besides their unmistakable advantages, various known fungicides have serious disadvantages. Thus lime sulphur (calcium polysulphide [23] [3]), which has been in use since 1888 for controlling, inter alia, scab in fruit trees, is a good fungicide, but it unfavourably influences the development of the trees and the appearance of the fruit to such an

*) Boekesteyn Agrobiological Laboratory, N. V. Philips-Roxane, 's-Gravenland, Holland.

1) R. van der Veen, Philips tech. Rev. **16**, 353-359, 1954/55 (No. 12).

2) J. Meltzer, Philips tech. Rev. **17**, 146-152, 1955/56 (No. 5).

3) The numbers in square brackets refer to the structural formulae in the appendix. The numbering follows on from that of the previous articles [1]) and [2]).

extent that it is in fact no longer used at the present time for the new varieties. More recent preparations, such as colloidal sulphur and the organic sulphur compounds thiram [1], zineb [2] and captan [5] [4]) are welcome successors. The latter even has a positive effect on the appearance of the fruit, enhancing its colour and sheen.

Copper in the form of Bordeaux mixture (basic copper sulphate [24], and the related copper oxychloride [25], is still a very popular fungicide, although it has been in use since 1882. For combating numerous diseases it still does excellent duty, but nevertheless, in many cases, a substitute is readily accepted. In many varieties of fruit, spraying with copper gives rise to undesirable russeting; in ornamental plants the bluish-green colour of the deposit is disadvantageous, and added to this the leaves are frequently damaged. As a result, Bordeaux mixture has found competitors in zineb and captan. Zineb has certain advantages over copper, since plants sprayed with it retain their greeness longer, this leading in many cases to a greater crop. Zineb adheres less tenaciously to the plant (see below) than copper however, as a result of which the protection may be of shorter duration.

The excellent fungicidal properties of the organic mercury compounds, which have also been used for many years, go hand in hand with a powerful leaf-damaging action and a high toxicity to warm-blooded animals. Thus an intensive search is in progress for substitutes for this group.

For various fungous diseases — mildew in apple trees for example — no satisfactory combating agent has yet been found. Here then is a virgin field for the development of new preparations. This and the above-mentioned defects of existing fungicides are the reasons for the current large-scale search for new fungicides.

Properties to be considered in fungicide research

According to the form in which they will be used, the control agents sought may be divided into the following four groups:

a) Wettable powders, i.e. preparations which may be readily suspended in water and, with the aid of this vehicle, may be sprayed over the plant.

b) Dustable powders, i.e. mixtures of an active component with inert carrier powders, which can be dusted over the plant. Agents (*a*) and (*b*) find application in combating leaf and fruit diseases.

c) Dry powder preparations to disinfect sowing-seed, which are mixed in a small percentage with the seed. They kill the fungal spores on or in the seeds and protect the seedling against soil fungi.

d) Soil treatment agents, which, mixed with the top-soil, serve to combat the sources of root and seedling diseases.

Irrespective of the final form of the fungicide, a large number of factors must always be taken into account in deciding its composition. A number of these factors are cited in random order below:

1) Fungicidal activity of the active component; by this is meant the intrinsic fungus-killing power of the compound. This power must of course conform to a certain minimum requirement.
2) Resistance of the preparation to external influences (moisture, temperature, light). This can never be too great.
3) Solubility of the active component. There is an optimum value: if the solubility be too low the toxic concentration cannot be reached and there will thus be no fungicidal action; if it be too high the fungicide will be washed away too rapidly.
4) Adhesion of the preparation once it has been applied to the leaf: the better the adhesion, the more slowly the fungicide will be washed away by rain or dew.
5) Distribution of the preparation over the plant. One factor determining this is the wetting agent content. Here too there is an optimum value: too little wetting agent gives poor distribution, too much leads to loss by draining.

The experiments relating to fungicide research conducted in the Mycological department at Boekesteyn can be divided into two categories. In the first only one living organism is involved, for example the spores or the mycelium [5]) of a fungus, and the action of a preparation is judged by the extent to which it is able to prevent the germination of the spores or the growth of the mycelium. The second category involves two living organisms, namely, both the disease-producing fungus and a host that has been artificially infected. We shall now discuss these two categories more closely.

[4]) Thiram, captan, etc., are names of two or three syllables, usually derived from the initial letters of some of the components and used by mutual agreement to replace complex chemical descriptions. Such names are coined only when an agent has shown that it is going to assume some practical significance.

[5]) Mycelium is the thread-like or fluffy vegetative part of a fungus.

Experiments involving one living organism

The spore germination test

For determining fungicidal activities under glass, use is often made of the spore germination test. In principle the test is performed as follows. A series of diluted solutions in acetone are prepared from the compound under investigation, and a known number of 0.01 millilitre drops of each solution are placed on a glass slide (*fig. 1*) with the aid

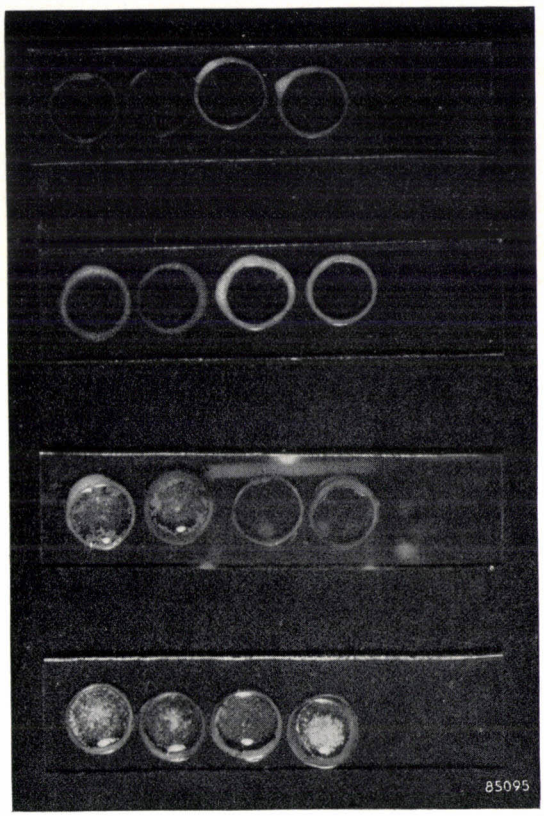


Fig. 1. Germination slides, with small paraffin wax rings, into which a known amount of the compound to be investigated and a drop of an aqueous suspension of the fungal spores are introduced.

of a micro syringe which permits of very accurate dispensing. To ensure that the solution is distributed over a known area, paraffin wax rings are placed on the glass slide. When the solvent has evaporated off, a known amount of the compound under investigation is left behind within each ring in the form of a more or less uniform layer. An 0.05 ml drop of an aqueous suspension of fungal spores (usually those of Fusarium culmorum, *fig. 2a*) is then placed on this layer. A little cherry extract is added to the suspension to render the environment favourable for the subsequent germination process. The glass slides are then placed in tubular containers, which are closed with a plate covered with moist filter paper (*fig. 3*). After a period of about 20 hours at 20 °C, it is ascertained whether or not the spores have germinated (fig. 2).

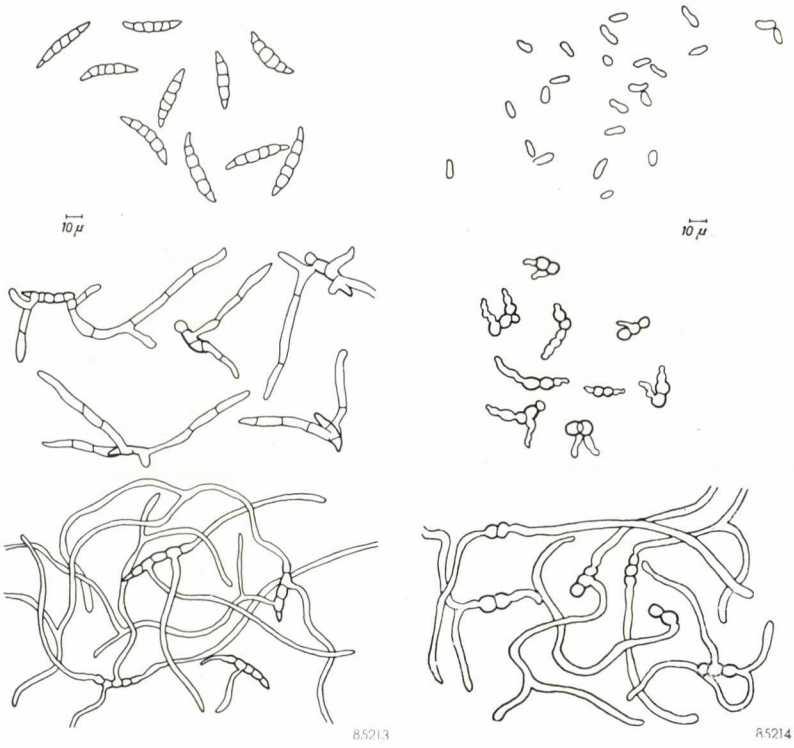
It would take us too far from our subject to discuss the spore germination test in detail. It will suffice to say that the concentration that is just able to suppress the germination of the spores completely or almost completely, is usually taken as a measure of the fungicidal activity of the compound. Use is made of a concentration series in which each solution of fungicide is of half the strength of the next highest. If greater accuracy is required, as for example in standardizing a preparation, then the dilution steps must lie closer together and the percentage germination must be determined for each of the dilutions, so that an "LD50" may be arrived at; LD50 is the designation given to the concentration which is lethal to 50% of the individual spores (LD = lethal dose). This value is obtained by determining the percentage germination for each solution of the concentration series used, and analysing the results graphically. Investigation in this manner is more accurate than the former method but requires a great deal of time, since it necessitates the individual examination of a large number of spores.

The spore germination test can equally well be employed for compounds which are soluble in water or can be suspended in it. The concentration series are prepared, using the spore suspension as diluent, and 0.05 ml drops of the various concentrations (which thus already contain the spores) are measured out onto the germination slides.

The spore germination test is essentially an analytical method. It may be used not only to determine the fungicidal value of unknown compounds, but can be adopted as means of establishing the content of a known fungicide. So employed, the spore germination test forms the cornerstone of research relating to resistance and adhesion (see below), and preparations of unknown composition can be analysed with it, provided the nature of their active component is known. Generally speaking the spore germination test can be said to be an aid to the search for new combating agents and better "formulations" (i.e. the form in which the compound is used: dusting powder, spray, etc).

Investigation of the effect of soil organisms

The extent to which fungicides are decomposed by soil organisms is investigated in the following manner. The compounds are mixed with soil rich in humus and the mixture allowed to stand for several weeks under conditions of humidity, oxygen



a *b*

Fig. 2. *a*) Ungerminated spores (above), germinated spores (centre) and further mycelium growth, of the fungus Fusarium culmorum. (*b*) The same for the fungus Mycosphaerella pinodes.

supply and at the temperature (24 °C) that is optimum for the development of the microflora. The compounds are then extracted with acetone or another solvent, and the activity of the extract is determined by means of the spore germination test. For the purpose of eliminating secondary influences of the soil (adsorption, activity of soil components) a determination is always carried out on a mixture of the soil with the compound under investigation, immediately after this mixture has been prepared.

Determination of the sticking power

A method for the determination of the tenacity or sticking power of the preparations should also be mentioned. Samples introduced onto spore germination slides are exposed to an artificial rainfall of 1 cm per hour, in the apparatus shown in *fig. 4*. *Fig. 5* shows the working

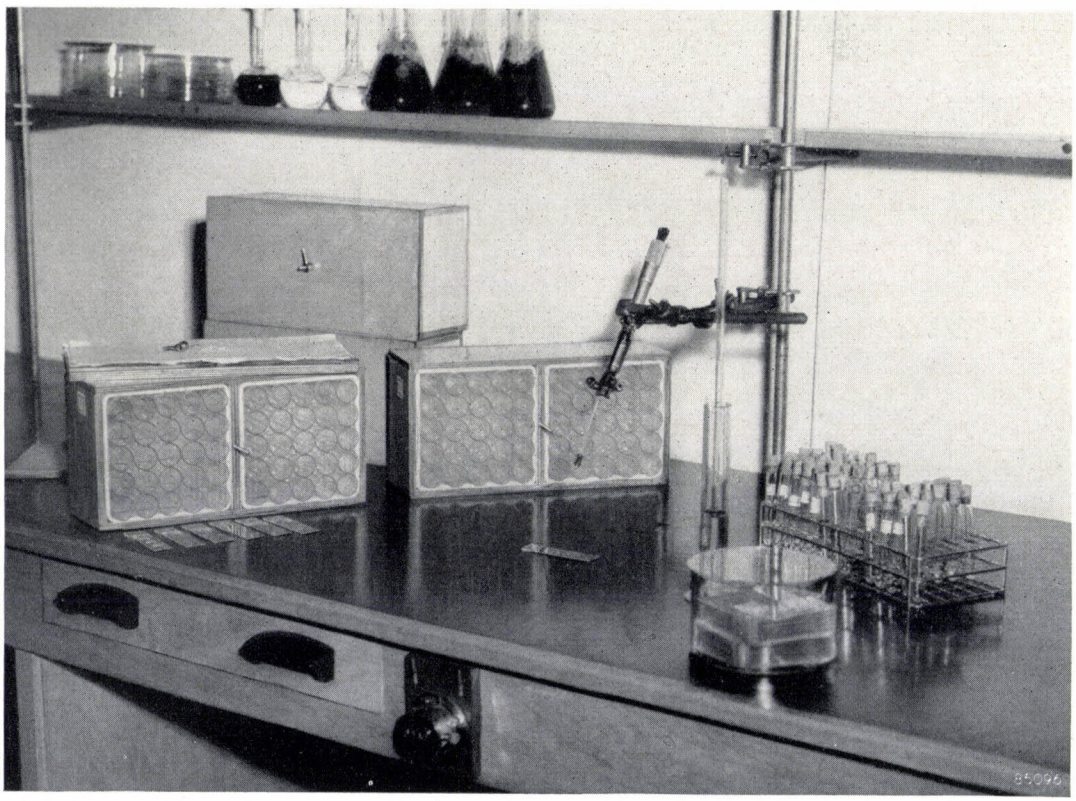


Fig. 3. The germination slides are inserted into glass cylinders (left), whose ends are covered with moist filter paper, Right: test tubes containing fungal suspensions; in the centre: micro syringe for accurate dispensing.

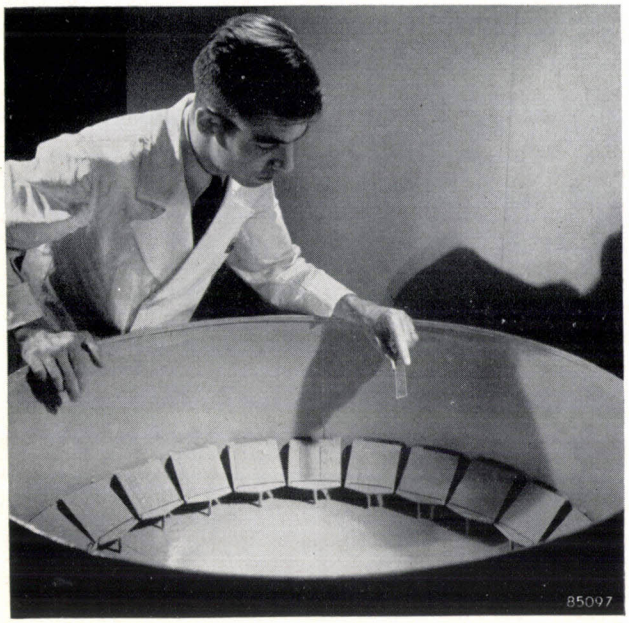


Fig. 4. An apparatus for determining the extent to which fungicidal preparations remain fixed to glass during rainfall.

principle. After 1 to 2 hours of rain, the fungicidal activity of the layers on the slides is compared with that of layers on slides which have not been exposed to rain. The values of the sticking power found in this way are compared with those of standard preparations.

The sticking power is also determined with leaves (e.g. potato leaves, see below) as the sub-surface; here the protective action of the fungicide against a subsequently applied artificial leaf infection serves as the yardstick. Experiments of this kind belong to the second category (two living organisms).

Investigation of the effect of toxicants on the growth of mycelium

In some cases it is desired to examine the influence of a toxicant on the growth of mycelium. A method is available for this purpose, which makes use of so-called notched tubes. A notched tube is a culture tube with an indentation just behind the mouth, which enables an agar culture medium to set in a layer parallel to the longitudinal axis (*fig. 6*). The culture medium is mixed beforehand with the compound to be investigated. A piece of mycelium from the fungus is placed in the centre of the agar layer. After a time (5 to 20 days depending upon the species), the spread of the mycelium is determined by comparison with a control experiment.

Investigation of bactericidal activity

In some cases an investigation is made of the bactericidal activity of new compounds. Use is made here of a score of plant-pathogenic bacteria. The experiment is the same as the spore germination test, save that a culture medium favourable for bacteria is used and the criteria by which the test is judged are "growth" (solution becomes turbid) and "no growth" (solution remains clear).

Experiments involving two living organisms

The second category of experiments for determining the fungicidal activity of preparations, namely that in which a host is artificially infected with a disease-producing fungus, approximates more closely to actual practice than the first. Nevertheless the second category is still a laboratory method, in which many factors obtaining in practice (rain, irradiation by the sun, wind, nature of the soil, manuring, irrigation, etc) are deliberately eliminated or standardized.

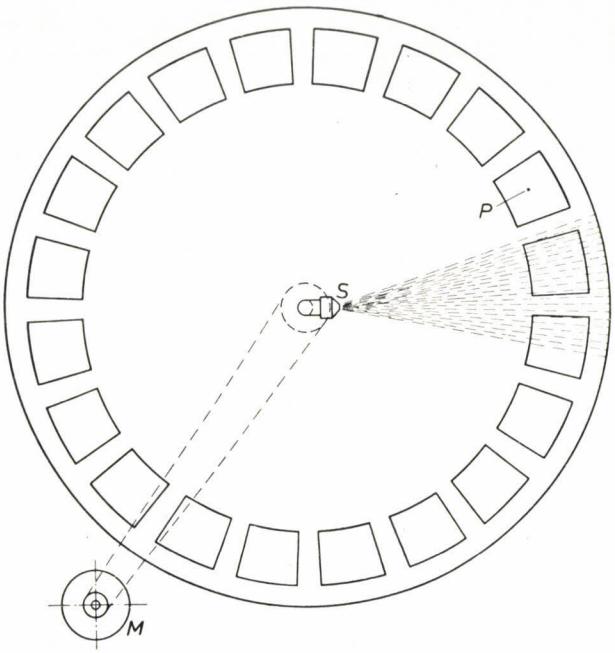


Fig. 5. The apparatus in fig. 4 is seen from above and in section. The sprayer S is rotated about a vertical axis by an electric motor M, and plays successively on the glass slides P which have been treated with the preparations whose sticking power is to be investigated. A rainfall of 1 cm per hour is employed for a period of 1 to 2 hours.

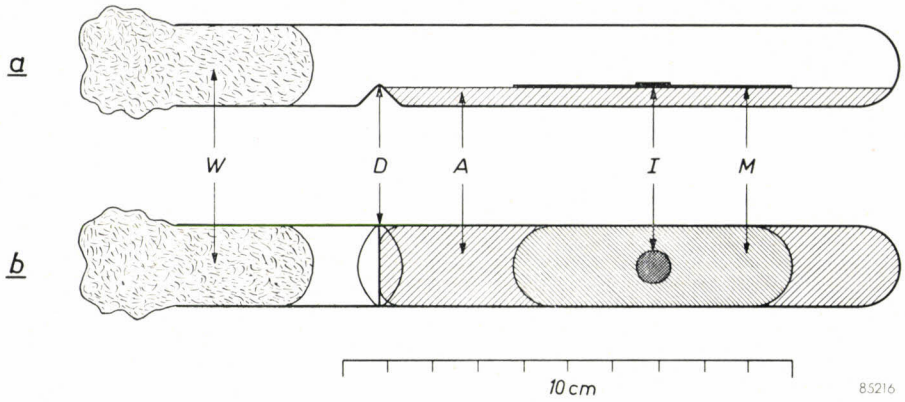


Fig. 6. Notched tube (*a*, seen from the side, *b* from above), with indentation *D*. *A* agar culture medium, mixed with the compound under investigation. *I* inoculum (piece of mycelium from the fungus). *W* cotton-wool plug. After a certain time the spread *M* of the mycelium is compared with that of a control test.

Experiments involving artificial infection usually take the following form. The plants (or parts of the plants) are sprayed with a spore suspension (with a density of the order of 100 000 spores per millilitre), or dusted with air-dried spores (in the case of powdery mildew). The plants are then placed for 24 hours in a room in which the relative humidity is always near 100% and at the optimum temperature for germination and the infection process (usually around 20 °C for the species used). This gives the disease opportunity to establish itself. The plants are then transferred to an environment favourable for the disease to develop. The temperature varies according to the species between 18 and 25 °C, a relative humidity of 75% will generally suffice, and in addition care must be taken that there is adequate lighting, so that the conditions are favourable for the development of the plant.

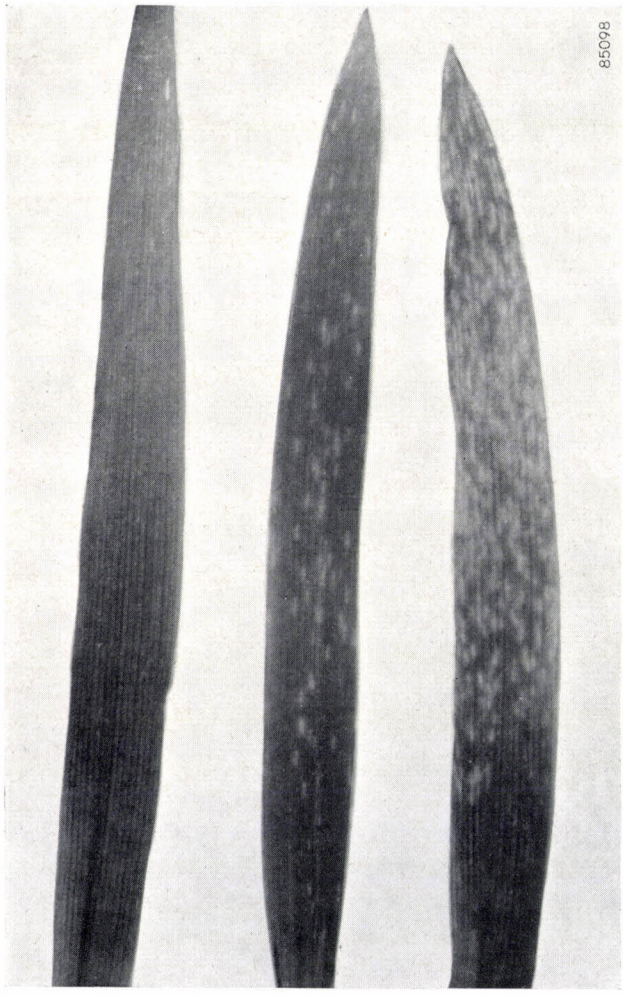


Fig. 7. Leaves from a young barley plant. Left to right: healthy leaf, leaf lightly attacked and leaf severely attacked by a species of powdery mildew (in this case Erysiphe graminis).

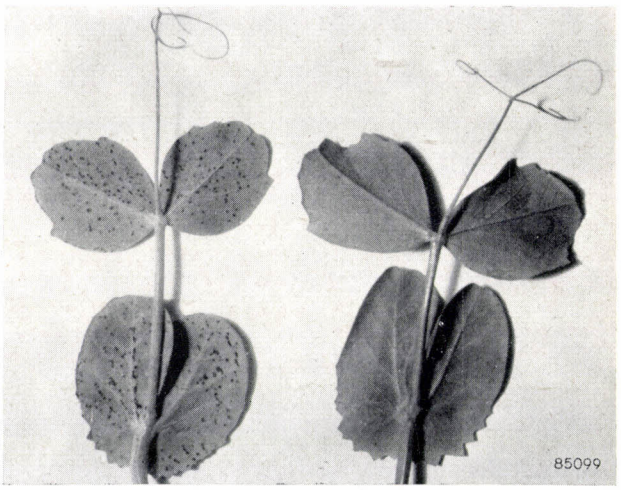


Fig. 8. Left: leaves of a young pea plant attacked by Mycosphaerella pinodes (fig. 2*b*). Right: healthy leaves.

The protective action of fungicide preparations is now investigated by starting from plants that prior to infection have been sprayed with the compound to be examined. It is then observed that the degree in which the symptoms of the disease occur, depends upon the concentration of the fungicide that had been applied to the plant.

Fig. 9. Tomato leaves infected with Phytophthora infestans. From left to right: untreated plant, plants treated with 0.2%, 1% and 5% copper oxychloride [25].

A number of plant diseases are currently employed for this purpose. We cite the following:

a) Young barley plants (about 8 cm tall) begin to show small white powdery specks about five days after infection, as a result of the development of the spores of powdery mildew (Erysiphe graminis), see *fig. 7*.

b) Young pea plants (also about 8 cm tall), sprayed with the spores of Mycosphaerella pinodes (fig. 2*b*), are covered after a few days with brown flecks (*fig. 8*).

c) Leaves cut from the potato plant, with their stems placed in water, die within five days of having been artifically infected with the spores of Phytophthora infestans, the causative organism of the familiar late blight. The leaves shrivel, and if the humidity of the air be sufficiently high, become covered with a white fluff of mycelium.

d) Tomato plants (cut leaves or young plants) infected with Phytophthora infestans show a similar picture (*fig. 9*).

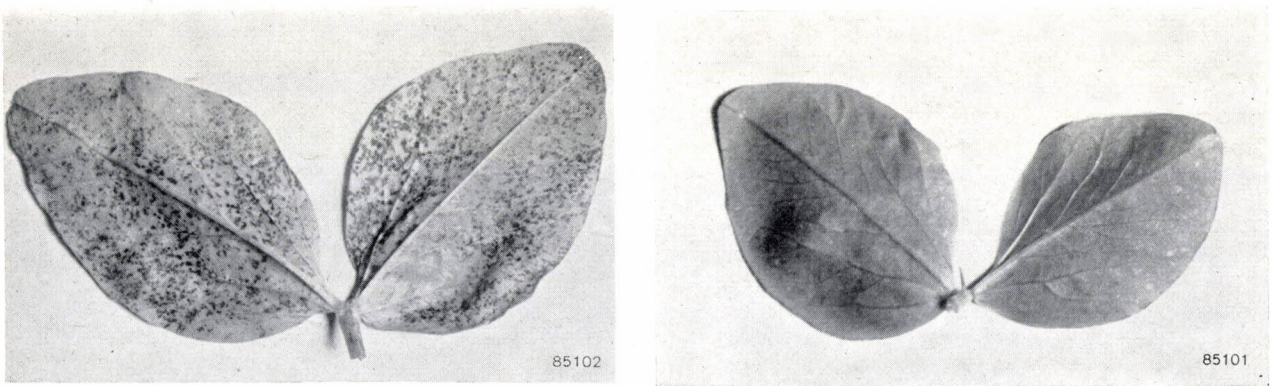


Fig. 10. Broad-bean leaves, left: infected with Botrytis fabae, right: healthy.

e) Another favourite infection is chocolate leaf spot in broad-beans. Within about two days, the causative fungus (Botrytis fabae) gives rise to round dark-brown spots on the broad-bean leaf, that stand out sharply against the somewhat greyish-green leaf (*fig. 10*).

f) The cucumber also shows disease symptoms when the young plants are treated with Colleto-

trichum lagenarium. After a few days, irregular, light-brown spots appear and finally the leaf completely withers.

In all infection tests, the activity of the preparation under investigation is expressed in terms of the activity of one or more standard preparations. Copper oxychloride, zineb and organic mercury preparations are some of the standards used.

If the provisional infection tests prove favourable, they are repeated, but this time with a practical preparation worked up from the compound. The results obtained serve as a pointer for the ultimate field tests, in which the efficacy is assayed in orchard, field, kitchen-garden or greenhouse.

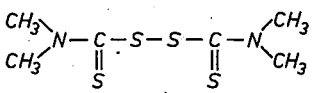
Concluding remark

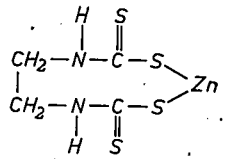
In the foregoing we have outlined the task of the Boekesteyn mycological department, as the link between chemical research and field tests. In general the work described has been the application of methods to be regarded as belonging to the routine work of the laboratory. Apart from this, however, investigations of a more fundamental nature are undertaken. Just recently, for example, a study of the penetration of substances (primarily fungicides) into the tissues of both fungi and the higher plants has been initiated. In this investigation, use is made of compounds labelled with radio-active isotopes. We hope that the results will give a better understanding of the mechanism of the leaf-damaging action of some fungicides.

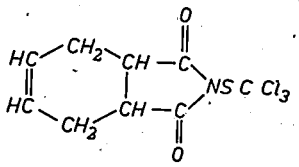
APPENDIX: STRUCTURAL FORMULAE

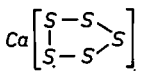
Below are given the structural formulae of the compounds mentioned in the text.

Sulpher preparations

[1] Tetramethylthiuram disulphide (thiram, TMTD) . . . . . 

[2] Zinc ethylene bisdithiocarbaminate (zineb, dithane) . . . . . 

[5] N-trichloromethyl tetrahydrophthalimide (captan, orthocide) [6] 

[23] Calcium polysulphide, the principal constituent of lime sulphur (The precise structure of this compound is still uncertain; one of the more recent ideas is given here.) . . . . . 

Copper preparations

[24] Basic copper sulphate (principal constituent of Bordeaux mixture) . . . . . $[Cu(OH)_2]_3CuSO_4$

[25] Copper oxychloride . . . . . $[Cu(OH)_2]_3CuCl_2$

Mercury preparations

[26] R-Hg-X, in which R is an aliphatic or aromatic group (preferably C_2H_5 or phenyl) and X an acid radical (usually Cl or phosphate, or an organic acid radical, such as p-toluene sulphonamide or dichlorocresol). Examples: ethyl mercury sulphide ("New Improved Ceresan"), phenyl mercury dichlorocresolate.

Summary. The principal task of the Boekesteyn mycological department of N.V. Philips-Roxane consists in the testing of compounds for their fungicidal action. This work forms a link between chemical research work and field tests. It may be divided into two categories. In the first, the only living organism involved is the fungus, whose reaction under glass to various dilutions on the compound under investigation is examined. The spore germination test (followed by descriptions of investigations of the influence of soil organisms, and on adhesion to the plant) and the investigation of the effect on the growth of the mycelium are discussed. The second category employs two living organisms: a fungus and a plant artificially infected with it. Besides this routine work the mycological department also undertakes research of a more fundamental nature, in which radio-active isotopes are employed to investigate the penetration of control agents.

[6]) In article [1]), page 358, an incorrect formula was given for this compound.

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

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

2210: M. L. Huggins and J. M. Stevels: Comparison of two equations for calculation of densities of glasses from their compositions (J. Amer. Ceramic Soc. **37**, 474-479, 1954).

Two different relationships have been proposed by the authors, individually, for the calculation of the densities of glasses from their compositions. These relationships are here compared with regard to accuracy, composition limitations, relationships between their constants, etc. The Stevels equation is useful in supplying values (accurate to within about 2%) of the density of glasses having a high proportion of network formers (e.g., Si, B, and Al). Whether or not it is applicable, with different constants, when the content of network modifiers (e.g., Na, Ca, and Pb) is high, has not been determined. The Huggins relationship involves more empirical constants, including two which have different values, depending on which of four composition ranges is pertinent. It is considerably more accurate for well-annealed glasses of accurately known composition. Difficulties arise in applying it to glasses having a high content of network formers other than silicon or to glasses having components for which accurate values of the necessary constant have not previously been deduced. It is interesting to note that the limiting conditions of the Stevels equation lead, at least for the sodium silicate system, to the same breaks of the volume-concentration curves as found by Huggins. Comparison of the two relationships yields a better understanding of the reasons for the long-realized fact that the presence of certain elements, such as Li, Be, and Ti, invalidates the Stevels equation.

2211: W. P. van den Blink, E. H. Ettema and P. C. van der Willigen: A new proces of stud welding (British Welding J. **1**, 447-454, 1954).

A new process of stud welding is described, in which the timing is regulated by a cartridge on the end of the stud. This cartridge is a semi-conductor; it starts the arc between stud and plate, and also determines the distance between them. The composition of the cartridge resembles that of a normal electrode coating in so far as metallurgical and arc-stabilizing functions are concerned, and rimmed-steel studs and A.C. can be used. A simple stud-welding gun containing only a spring, pressing the studholder on to the work piece, can be used.

2212: P. J. L. Scholte, C. C. Kok van Alphen and B. Combée: Treatment of the cornea with a new lilliput roentgen tube (Acta Radiologica **42**, 316-328, 1954).

A description is given of a new lilliput X-ray tube energized at a potential of 25 kV and the basic principles of its design are discussed. Certain experiments carried out on the cornea of rabbits are reported and the results of treatment of the human cornea with the Philips 50 kV contact therapy apparatus and a 25 kV unit with the new tube are presented and commented upon. Both in "keratitis" and vascularisation of transplanted cornea, the results appeared to be most promising.

2213: M. E. Wise: The ratio of two factorials and some fundamental probabilities (Kon. Ned. Akad. Wet. Amsterdam **A 57**, 513-521, 1954).

Expansions are obtained for the ratio $N!/(N-n)!$ of two factorials, and for the logarithm of this ratio. It is shown how several well-known results in probability theory and statistical mechanics are thereby obtained much more easily than by the standard approach using Stirling's approximation for the factorials. In fact the latter seems to be even incorrect in one important case, viz. in finding how a binomial distribution approaches its limiting Gaussian form. Arising out of this, a new continuous probability function is found that is close to the binomial over its whole range, and may be useful in empirically fitting some observed distributions when it is important to estimate their behaviour accurately at the tails (i.e. extreme values).

2214: Th. A. J. Payens: Influence of salt on the spreading pressure of films of long-chain weak acids (Kon. Ned. Akad. Wet. Amsterdam **B 57**, 529-533, 1954).

The ionization of incompletely ionized monolayers at an air-water or oil-water interface depends not only on the ionization constant of the monolayer electrolyte but also on the electrical potential at the interface. As a consequence, the surface pressure of long-chain weak electrolytes may be increased by other ionizing substances, such as salt, dissolved in the water. Experimental evidence of this is presented with monolayers of stearyl phosphoric acid

at an oil-water interface and with caprylic acid at an air-water interface.

2215: J. W. L. Köhler and C. O. Jonkers: De koudgaskoelmachine (De Ingenieur **66**, 0 103-0 110, 1954). (The gas refrigerating machine; in Dutch).

See Philips tech. Rev. **16**, 69-78 and 105-115, 1954/1955.

2216: J. L. H. Jonker: Secondary emission (T. Ned. Radiogenootschap **19**, 267-281, 1954).

The following principal properties of secondary emission are dealt with: the secondary emission of metals, the secondary emission of insulators, secondary emission as a function of the angle of incidence of the primary electrons, the energy distribution of the secondary electrons, and the angular distribution of the secondary electrons. Some properties can be qualitatively understood by means of a simple classical theory. At the end some conditions for practical application of the secondary emission of metals, insulators and semi-conductors are discussed.

R 261: J. Bruijsten: Graphical determination of reflex-klystron characteristics (Philips Res. Rep. **10**, 81-96, 1955, No. 2).

The small-signal theory of reflex klystrons has been treated very thoroughly in literature. The article shows that some of the results obtained can be derived by a graphical method. Starting from an expression of normalized efficiency as a function of the bunching parameter, a triangular diagram is derived of normalized efficiency as a function of resonator loss conductance, load conductance and small-signal electronic conductance. This diagram can be applied for the construction of an electronic tuning diagram and a theoretical Rieke diagram. A method is described of evaluating the parameters from experimental data with the aid of the diagrams.

R 262: M. E. Wise: Formulae relating to single-sample inspection by attributes (Philips Res. Rep. **10**, 97-112, 1955, No. 2).

Operating characteristics (O.C.) for batch inspection by single attribute samples are considered; mathematical formulae for them are restated and used to derive accurate expressions for the 50-per-cent point of an O.C. and its slope at this point, and also some simple approximations which are compared with empirical ones. They are given both for infinite batches (with binomial distributions of the number of defectives found in a sample of n) and for finite ones (hypergeometric distributions). The general treatment of O.C.'s in terms of their behaviour at the 50-per-cent point is discussed. In a numerical example an attribute sampling system for a finite batch is given and an equivalent variable system is calculated; the two O.C.'s are compared.

R 263: J. Smit and H. G. Beljers: Ferromagnetic resonance absorption in $BaFe_{12}O_{10}$, a highly anisotropic crystal (Philips Res. Rep. **10**, 113-130, 1955, No. 2).

The ferromagnetic resonance absorption has been measured at 24 000 Mc/s for a single crystal of hexagonal $BaFe_{12}O_{19}$. The resonance conditions are severely influenced by the crystalline anisotropy and, for fields too small for saturation, also by the Weiss-domain structure. The theory predicts, for a varying magnetic field perpendicular to the hexagonal axis, at most three absorption peaks, which have been observed at elevated temperatures. The spectroscopic splitting factor g and the anisotropy field are evaluated. The g-factor has the spin-only value. The crystalline anisotropy is suggested to be caused by dipole-dipole interaction.

R 264: K. F. Niessen: Magnetic anisotropy and Van Vleck's relation for antiferromagnetics (Philips Res. Rep. **10**, 131-140, 1955, No. 2).

For the influence of magnetic anisotropy on the susceptibility of antiferromagnetics a formula (12) is derived which is more general than that given by Nagamiya, since anisotropy coefficients are supposed to be different for the two kinds of magnetic spins and also a small difference in the spin moments is taken into account. A corresponding extension (15) of Van Vleck's relation between the powder susceptibilities at 0 °K and at the Curie temperature is given.

R 265: T. Tol, W. J. Oosterkamp and J. Proper: Limits of detail perceptibility in radiology particularly when using the image intensifier (Philips Res. Rep. **10**, 141-157, 1955, No. 2).

Report in more detail of the investigation described in Philips tech. Rev. **17**, 71-77, 1955, No. 3.

R 266: H. Koelmans: Suspensions in non-aqueous media (Philips Res. Rep. **10**, 161-193, 1955, No. 3).

The stability of suspensions in solvents of very low dielectric constant ($\varepsilon<5$) is dealt with in the first three sections. Theoretical considerations lead

to the conclusion that quite modest electric charges and ζ-potentials are sufficient to stabilize suspensions of coarse particles (>1 μ), whereas hardly any stabilization can be expected from adsorbed layers of non-ionized surface-active molecules. Experiments on the setting-times of suspensions in xylene confirm that only ionized surfactants give rise to stability: detergents that do not increase the conductivity of the xylene do not give rise to a sufficient ζ-potential of the particles and do not improve the stability very much. The behaviour of suspensions in polar organic media is dealt with in the fourth and fifth sections, in relation to the phenomenon of electrophoretic deposition. It is shown that the particles are accumulated near the electrode by the applied field, but that the formation of an adhering deposit is caused by flocculation, introduced by the electrolyte formed as a result of the electrode reaction.

R 267: G. Diemer: Light patterns in electroluminescent ZnS single crystals activated by diffusion of Cu (Philips Res. Rep. **10**, 194-204, 1955, No. 3).

Non-activated ZnS single crystals can be made electroluminescent by providing them with copper electrodes through evaporation in vacuum. It is shown that at temperatures little above room temperature the copper diffuses rapidly along the surface of the crystal and along certain imperfections into the bulk of the crystal. During electroluminescence in an a.c. field, light is emitted only from narrow lines, which with high optical magnification prove to consist of a series of nearly equidistant dots. These "dotted lines of light" always have the same direction as the main direction of growth of the hexagonal crystals.

R 268: P. Zalm, G. Diemer and H. A. Klasens: Some aspects of the voltage and frequency dependence of electroluminescent zinc sulphide (Philips Res. Rep. **10**, 205-215, 1955, No. 3).

Experiment shows that the relation between the luminous emittance H of an electroluminescent cell and the applied r.m.s. voltage V is given by $H = H_0 \exp(-c/V^{\frac{1}{2}})$. A mechanism is proposed that may explain both the well-known linear frequency dependence of the emittance at constant r.m.s. voltage and the observed voltage dependence.

R 269: G. Thirup: Design of low-pass amplifiers for fast transients (Philips Res. Rep. **10**, 216-230, 1955, No. 3).

By means of network synthesis a broad-band amplifier is designed, special attention being paid to a good transient response. The design leads to a novel type of interstage network. Some details of a 50 Mc/s amplifier are given: measured phase and amplitude curves as well as transient response are shown. In an appendix details of the design of the phase-correction network are given.

R 270: P. Schagen: Limiting resolution due to charge leakage in the scenioscope, a new television-camera tube (Philips Res. Rep. **10**, 231-238, 1955, No. 3).

The conductivity of the target material in the scenioscope, a new television pick-up tube, results in leakage of picture charge through the target. The potential distribution at the surface of the target between successive scans is calculated for the case where picture charge is supplied to the target on alternate illuminated and dark bars with a width d. The potential difference between the centre of illuminated and dark bars immediately before the next stabilization by the scanning beam appears to be a function of τ/R_0C_0 (where τ represents the frame period and R_0C_0 the RC-time of the target material) and of the relative picture detail d/D, where D is the target thickness. Formulae are derived for two effects of the target characteristics on the picture signal: (1) Leakage of picture charge to the signal plate results in a loss of sensitivity, determined by τ/R_0C_0. (2) Leakage of picture charge parallel to the surface of the target results in a decreased depth of modulation for smaller picture details d/D. Limiting values for τ/R_0C_0 and d/D follow from the requirement of a negligible influence of these two effects.

VOL. 17 No. 9, pp. 233-268 MARCH 1956

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

SIMPLE THEORY OF THE JUNCTION TRANSISTOR

by F. H. STIELTJES and L. J. TUMMERS.

537.311.33:621.375.4

In the future development of the electronic industry, the transistor is undoubtedly destined to play a more and more prominent role. It is quite conceivable that the use of this crystal amplifier will come to be as frequently discussed in this Review as the electronic tube. It therefore seems desirable to devote a series of articles to the basic theory of the transistor and its behaviour in circuits. This first article deals with the simple theory of the transistor, applicable for small current densities and low frequencies (audio frequencies).

Introduction

A transistor is a crystal amplifier which can be used to amplify electrical signals. It is manufactured from semi-conducting material, for which, so far, germanium has chiefly been used. The "junction transistor" derives its name and its properties from the so-called *P-N* junctions between the different layers of which it is composed [1]). The properties of these junctions have been discussed in an earlier article in this Review [2]) (further referred to as I), in which the rectifying action of a *P-N* junction was explained. So far as they are necessary for an understanding of the transistor mechanism, we may recapitulate the ideas given in I.

P and N regions in germanium; P-N junction

In germanium the electrical conductivity is due to two kinds of charge carriers: the "conduction electrons" and the "conduction holes", subsequently referred to simply as "electrons" and "holes". The holes manifest themselves as positive, mobile charge carriers. In the absence of outside influences, at every point in the crystal, the product of the concentration of holes and the concentration of electrons is equal to a constant (which increases rapidly with rising temperature), characteristic for germanium. The presence of traces of certain elements in the crystal lattice makes the concentration of holes very high and, in view of the constancy of the product, that of the electrons low; this gives us *P*-germanium. Other elements raise the concentration of the electrons and thus lower that of the holes; one then speaks of *N*-germanium. The charge carriers in the higher and lower concentrations are called the majority and minority charge carriers respectively; the minority carriers are thus electrons in the *P*-region and holes in the *N*-region.

In the simple theory of the *P-N* junction and also of the transistor, we may assume that if the foreign chemical elements present are homogeneously distributed in their respective regions, every volume element is electrically neutral [3]). Together with the condition that the product of the concentrations of holes and electrons be constant, the neutrality condition determines the values of these two concentrations in homogeneous regions of the crystal in the absence of external influences. The values of the concentrations thus fixed are known as the "equilibrium concentrations".

P and *N* regions may be present in the same

[1]) The idea of the junction transistor was originated by Shockley, who, after setting out the theory of the *P-N* junction, showed how the junction transistor could be constructed and predicted its properties. See Bell Syst. tech. J. **28**, 435-489, 1949.

[2]) J. C. van Vessem, Theory and construction of germanium diodes, Philips tech. Rev. **16**, 213-224, 1954/55 (No. 8).

[3]) Strictly speaking, space charges can occur at the extreme outer edges of a homogeneous crystal region. The distances over which these space charge regions extend, however, are always so small that by comparison with the other distances involved in the theory of the transistor, they may be neglected.

crystal. At the junction between P and N regions large space charge densities may occur. The space charge region, which is very thin (see also note [3]), and which one can visualize as being bounded by two more or less sharply defined planes *1* and *2*, is called the "*P-N* junction" or "barrier". *Fig. 1a*

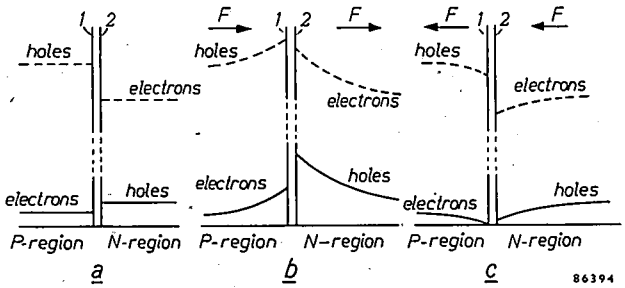


Fig. 1. Concentrations of minority charge carriers (full lines) and of majority charge carriers (broken lines), as functions of position on either side of a *P-N* junction in a germanium crystal. The concentration of majority charge carriers in practice is between 10^4 and 10^6 times that of the minority charge carriers. *1* and *2* denote the boundary planes of the junction.
a) No voltage is applied to the crystal; everywhere outside the junction the concentrations assume uniform equilibrium values. By the addition of suitable impurity elements in the proper proportions, both in the *P*-region and the *N*-region, the equilibrium concentrations can be varied within wide limits (on condition that the product of minorities and majorities in any one region remains constant).
b) A voltage is applied to the crystal which raises the potential on the *P*-region with respect to the *N*-region. The electric field F, whose direction is shown by arrows, drives majority charge carriers (holes in the *P*-region and electrons in the *N*-region) towards the *P-N* junction.
c) The polarity of the voltage is opposite to that in (*b*). The field now drives majority charge carriers away from the junction.

represents the equilibrium concentrations of majority and minority charge carriers in the P and N regions, supposed homogeneous, on opposite sides of such a barrier. How the concentrations vary within the barrier is not of importance for the simple theory of the transistor. A typical value of the concentration of the minority charge carriers is, for example, 10^{11} per cm^3; that of the majority charge carriers is then 5×10^{15} per cm^3 (since the constant product is about 5×10^{26} at room temperature).

In the following discussion, we shall consider principally the minority charge carriers. We shall see that the properties of the *P-N* junction and of the transistor are largely determined by the behaviour of the minority charge carriers, although the majority carriers are indispensable to the physical processes to be described. Furthermore it will be seen that the currents of minority carriers (or, briefly, minority currents) can be directly calculated and that conclusions can then easily be drawn about the majority currents.

The concentration of the minority charge carriers (or, briefly, minority concentration) will be denoted in general by g and its equilibrium value by G. Where necessary a suffix will indicate to which region the concentration applies.

Concentration changes under an applied voltage

If a voltage is applied between a P and an N region, then deviations from the equilibrium concentrations occur in the neighbourhood of the junction (fig. 1*b* and 1*c*). If the portion of the potential difference which is developed between the two boundary planes *1* and *2* of the *P-N* junction be V (V is considered positive if the potential of the *P*-region is raised with respect to that of the *N*-region), then the minority concentrations in these boundary planes are given by:

$$g = G\, e^{Vq/kT} \quad \ldots\ldots\ldots \quad (1)$$

where $-q$ is the charge of the electron, k is Boltzmann's constant and T the absolute temperature [4]). For each of the two boundary planes the appropriate equilibrium concentration must be substituted for G.

For a given *P-N* junction the minority concentrations in the boundary planes are thus fixed by the external voltage across the junction. It can be shown that when a voltage is applied, the neutrality condition remains valid for homogeneous regions of the crystal outside the barrier layer [5]), but that the condition that the product of the minority and majority concentrations be constant then ceases to apply. The neutrality condition means that outside the barrier layer, at every point in a homogeneous crystal region where the minority concentration is altered, the majority concentration must also be altered by an equal amount.

Diffusion currents of minority charge carriers; forward and reverse direction

As explained at length in I, pairs of electrons and holes continually appear and disappear: so-called generations and recombinations. Generations will be dominant where the concentrations lie below the equilibrium values, and recombinations will dominate in the reverse case. Concentrations differing from the equilibrium values will be set up at the barrier planes when a voltage is applied across the junction. Some distance from the junction, the equilibrium concentration will be maintained by

---

[4]) In accordance with the increasing practice in transistor literature, the symbol q is used for the electronic charge in place of the conventional e. In article I, q/kT was introduced as an unspecified constant a; V corresponds to $-\Delta\varphi$ in article I, while (1) corresponds to equations (16*a*) and (16*b*) in that article.

[5]) This neutrality condition was used in I to deduce equation (1).

generation and recombination: thus a concentration gradient is set up which is steepest near the junction and trails off into the crystal. Whenever a gradient exists in the concentration of particles that take part in the thermal motion, the thermal agitation sends, on an average, more particles in the direction from high to low concentration than vice versa. This is the well-known phenomenon of diffusion. The magnitude of the diffusion current density at any point is (apart from the sign) proportional to the concentration gradient at that point; the proportionality factor D is called the "diffusion constant".

The charge carriers are subjected to the action of the electric field as well as to that of diffusion: the resulting current is a superposition of the field current and the diffusion current. Where the majority current is concerned, the field and diffusion work in opposite directions; it can be seen in fig. 1*b* that the field drives holes to the right, while the concentration gradient in the P-region drives them to the left. The field, however, which set the whole process in motion, retains the upper hand. For the minority carriers, the field and diffusion reinforce each other. *The minority current resulting from the electric field is, however, everywhere negligible with respect to the total current (majority plus minority currents).* The reason for this lies in the great difference in the concentrations: since the field current is proportional to the concentration, a field which produces a reasonable current of majority carriers will result in no appreciable minority current. When the minority current makes an appreciable contribution to the total current, this contribution must be derived almost entirely from diffusion. *Where the minority current is concerned, therefore, we need to consider only the diffusion current.* Thus the flow of holes which, in fig. 1*b*, crosses the P-N junction into the N-region is almost purely diffusion current. This current is maintained by a continuous supply of holes from the P-region. In fig. 1*c*, the diffusion current of holes in the N-region is directed towards the P-N junction. Those holes which reach the junction are carried off into the P-region by the electric field against the diffusion action. The *electron* currents in fig. 1*b* and 1*c* are brought about in an analogous manner.

That the minority charge carriers are almost exclusively moved by diffusion, can be seen in a more quantitative way as follows.

The current of holes I^+ consists of a portion I_F^+, supplied by the field, and a portion I_d^+, supplied by diffusion. The same applies to the electron current I^-, which consists of I_F^- and I_d^-. Suppose that in the N-region, the field supplies a fraction A of the minority current (i.e. hole current). Then:

$$I_F^+ = A\, I^+,$$

and thus

$$I_d^+ = (1-A)\, I^+.$$

The field will give rise to a majority current of electrons, which is greater than the field hole current by a certain factor B, thus:

$$I_F^- = B\, I_F^+ = BA I^+.$$

We are here considering only cases where the majority concentration is everywhere very much greater than the minority concentration. The factor B is then a very large number: for example of the order of magnitude of 10^5.

Since the concentration of electrons and holes follow parallel curves, the concentration gradients for the two sorts of charge carriers are the same in any cross-section. The diffusion constant for electrons is about twice as great as that for holes, and the diffusion current of electrons is thus a factor $M \approx 2$ greater than that of holes. As a result of the difference in sign of the charge, the currents, regarded electrically, are in opposite directions:

$$I_d^- = -M\, I_d^+ = -M\,(1-A)\, I^+.$$

The total current is

$$I = I^+ + I^- = I^+ + I_F^- + I_d^- = \{1 + BA - M(1-A)\}\, I^+,$$

so that

$$\frac{I^+}{I} = \frac{1}{1 - M + (B+M)\,A}. \qquad \ldots\ldots \quad (2)$$

At some distance from the P-N junction, where the concentration gradient is rather small, the current of minority charge carriers may be as much as 1% field current, i.e. $A = 10^{-2}$. Then (taking $B = 10^5$ and $M = 2$) I^+ is about one thousandth of the total current and the whole minority current is thus negligible. This case is of no interest. Everywhere where I^+ is not negligible compared to the total current, however, A is very small. It can easily be calculated that $A \approx 10^{-3}$ when $I^+/I = 1\%$, and A decreases as I^+/I increases. If the minority current is appreciable, then it is almost entirely a diffusion current.

At a great distance from the P-N junction, where the equilibrium concentrations are substantially the equilibrium values, the concentration gradient is very small. Diffusion is then quite negligible and the minority current is therefore purely a field current. However, the whole minority current is then negligible, in accordance with the conclusions drawn from (2).

The fact that the behaviour of the minority carriers is determined entirely by diffusion is the reason why their behaviour is much simpler to analyse than that of the majority carriers. In article I (p. 221) it was deduced for the one-dimensional case, with which we are also concerned here, that the minority concentration approaches the equilibrium value as an exponential function of distance x. The distance L in which the deviation from the equilibrium value changes by a factor e is the

so-called diffusion length [6]). The fact that we are here dealing with an exponential function gives us the following relationship (see *fig. 2*):

$$\frac{dg(x)}{dx} = \frac{g(x) - G}{L} \quad \ldots\ldots (3)$$

For the minority current [7]), which is actually a diffusion current (I_d),

$$I_d = q\,D\,\frac{dg(x)}{dx},$$

or, in combination with (3):

$$I_d = q\,D\,\frac{g(x) - G}{L} \quad \ldots\ldots (4)$$

In these formulae no account is taken of the sign of $dg(x)/dx$ and of I_d; the sign of I_d can be decided by a glance at the figure. The diffusion constant for holes D_p, or that for electrons D_n, must be substituted for D according as the minority current consists of holes or electrons. D_n is about twice as large as D_p. The value of L is strongly dependent on lattice defects in the crystal and on the impurities present, and its value can vary widely as between the various homogeneous regions of the crystal. In the following discussions we shall, where necessary, indicate by a suffix to L, the region to which this quantity relates [8]).

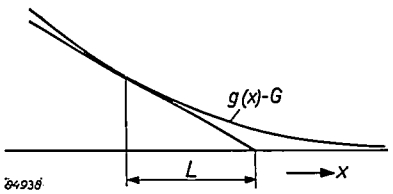


Fig. 2. The difference $g(x) - G$ between the concentration $g(x)$ of minority charge carriers and their equilibrium concentration G follows, in the one-dimensional case, an exponential curve. The sub-tangent of such a function is constant and represents the distance in which the function changes its value by a factor e (here, this is the diffusion length L) Thus for every value of x, ignoring the sign, $dg(x)dx = \{g(x) - G\}/L$.

Since $g(x)$, for the boundary planes, is given by (1), we can express the minority currents in these planes in terms of the voltage across the *P-N* junction:

$$I_d = q\,\frac{DG}{L}\,(e^{Vq/kT} - 1) \quad \ldots\ldots (5)$$

The total current is, of course, the sum of hole and electron currents calculated at the same place. Consider the cross-section *2* at the boundary plane of the junction (fig. 1). Here the holes form the minority current, the magnitude of which is given by (5). The electron current at *2*, i.e. the majority current, is unknown. However, we now make use of the fact that, because the *P-N* junction is very thin, the effect of generation and recombination within the junction is negligible: hence the electron current across *2*, where it is the majority current, must be equal to that across plane *1*, where it is the minority current, and is given by calculating the latter from (5).

Fig. 3a and *b* show the concentrations of the minority charge carriers at a *P-N* junction and their relation to the diffusion lengths, for both voltage directions. From these diagrams it is simple to deduce the minority currents at the junction, and thus also the value of the total current. Fig. 3*c* and *d* show how the hole current and the electron current change continuously on going from the *P*-region to the *N*-region and vice versa, so that their roles as majority and minority currents are interchanged.

As V becomes increasingly negative, it is seen from (5) that the diffusion current approaches a saturation level. Also from fig. 3*b* it is immediately clear that saturation must occur: the minority concentrations cannot fall below zero. Since at room temperature $kT/q \approx {}^1/_{40}$ volt, the saturation value of the current has been reached long before the potential drop across the *P-N* junction is −1 volt. The negative sign indicates that the potential is in the reverse direction. If V is positive, the current increases rapidly with the voltage; this is the forward direction.

The following remarks will serve to illustrate that the electrical conductivities of the adjoining *P* and *N* regions have a very unexpected effect on the current across the junction. As we have seen, the total current can be evaluated by adding the minority diffusion currents at the junction as given (5). by At the junction, each type of carrier therefore contributes an amount proportional to its minority equilibrium concentration G to the total current (the latter, of course, is the same throughout the crystal). Now, the minority equilibrium concentrations G will be smaller in a crystal of high conductivity since then the concentration of majority carriers is high (product of equilibrium concentrations is constant). Hence we find the apparent paradox that for a given voltage across the junction, the current is higher the worse the conductivities of the *P* and *N* regions on either side of it. This again illustrates that it is not the electrical conductivity but the diffusion which determines the current.

[6]) Diffusion-recombination length would perhaps be a more appropriate name since L is the average distance which minority charge carriers move in the x-direction by diffusion, before they disappear by recombination.

[7]) Strictly speaking, not the current, but the current density is of importance here. One can imagine simply that we are here dealing with a *P-N* junction of unit cross-sectional area.

[8]) Once more we deviate from the notation in I. In I the suffices to the diffusion lengths L_n and L_p mentioned on p. 221 refer to the type of charge carrier.

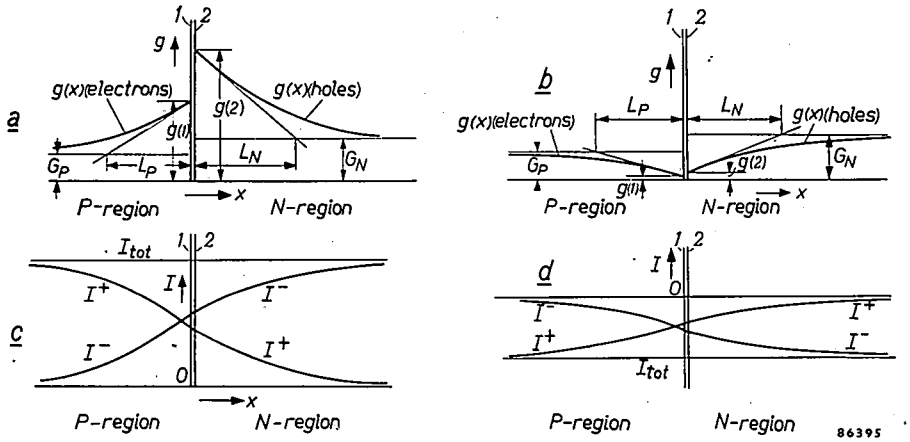


Fig. 3. Variation of the concentration g of the minority charge carriers (electrons in the P-region and holes in the N-region) at the P-N junction. The concentration gradients occurring at the boundary planes *1* and *2* of the junction can be read off the figure, after which the minority currents themselves can be written down directly. L_P and L_N are the diffusion lengths in the P and N regions respectively.
a) The voltage across the junction is in positive direction of x; the field drives holes to the right (same case as in fig. 1*b*).
b) The voltage is of the same magnitude, but negative; the field drives holes to the left (same case as in fig. 1*c*).
c) Hole current I^+ and electron current I^- for the P-N junction in situation (*a*). In the direction of increasing x, the hole current gradually gives way to the electron current, so that the total current remains constant.
d) Hole current and electron current in situation (*b*). Both currents are now negative and, furthermore, much smaller than in situation (*a*): the voltage is in the reverse direction, whereas in (*a*) it is in the forward direction.

Action of the junction transistor

Basis of amplification by a transistor

A transistor consists of a single crystal of germanium in which a P-N junction facing in the x-direction is followed by an N-P junction. There are thus two P-regions, separated by an N-region, known as the base. Each region is fitted with an electrode (*fig. 4*).

Before investigating the details of the action of the transistor, we shall first summarize its essentials.

At one boundary plane of the base (cross-section *2*, see *fig. 5*) the minority concentration (holes) is fixed at a value $g(2)$ lying above the equilibrium value by applying a voltage across the P-N junction in the *forward* direction. In the other end-plane (section 3) the concentration of holes is fixed at zero by a sufficiently large voltage across the N-P junction in the *reverse* direction. A large concentration gradient in the x-direction now exists in the base.

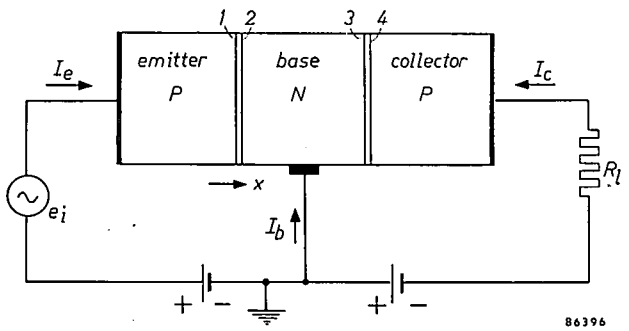


Fig. 4. Schematic representation of a P-N-P-transistor. Between the P-regions to the left and right, respectively designated "emitter" and "collector", is the "base" of N-germanium. I_e, I_c and I_b are respectively the emitter, collector and base currents. Currents in the direction of the arrows are considered positive. *1*, *2*, *3* and *4* mark the boundary planes of the P-N and N-P junctions. e_i A.C. input source; R_l resistance in the collector circuit, across which the amplified voltage appears.

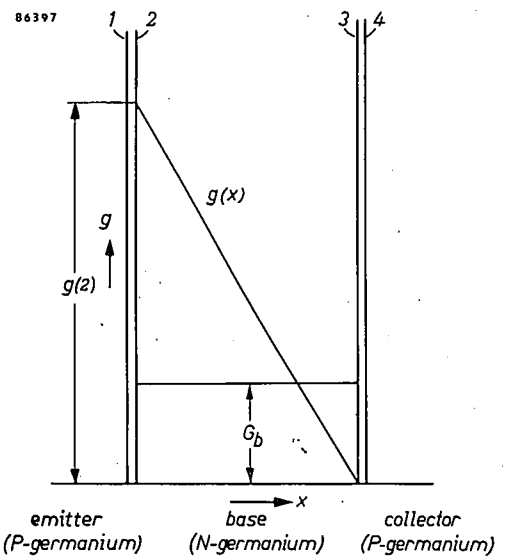


Fig. 5. To explain the essential principle of transistor action, the concentration of only the minority charge carriers $g(x)$ in the base of a junction transistor is given as function of position x. G_b represents the equilibrium concentration of the minority charge carriers in the base.

Since the current of minority charge carriers is a diffusion current, the concentration gradient results in a current of holes in the x-direction. This flow of holes is maintained by a continuous supply of holes from the P-region on the left; this P-region is therefore called the "emitter". It is often said that the emitter "injects" holes into the base. The holes disappear over the N-P junction into the P-region to the right; this P-region is therefore known as the "collector". The current of holes through the base thus makes a contribution to the current through the collector circuit. Since the emitter voltage is in the forward direction with respect to the base, a small variation of this voltage gives a large change in the concentration of holes at section *2* (see eq. 1). This also causes a change in the concentration gradient and thus in the flow of holes in the base and across the collector junction. The result is a change of current in the collector circuit and thus in the voltage drop over a load resistance R_l which is included in the circuit (see fig. 4). This is accompanied by a change in the potential of the collector with respect to the base, but since this potential is in the reverse direction, the collector voltage can vary quite considerably before the concentration of holes at section *3* deviates appreciably from zero. It is in fact possible to make R_l so large that a small change in the voltage between emitter and base reappears, amplified, across R_l. In the circuit shown (fig. 4), the transistor thus gives voltage amplification.

Closer examination of the situation in a transistor

In the above explanation of the amplification produced by a transistor, no mention was made of the electrons, nor of generation and recombination. We shall now fill in these gaps. In *fig. 6a* it is assumed that the base thickness w is considerably greater than the diffusion length of holes in the base ($w \gg L_b$). The variation of the minority concentrations in emitter, base and collector is shown: in the emitter and collector this is the concentration of electrons; in the base, that of holes. Again, the minority concentration at the emitter junction is raised to a level above the equilibrium value by a voltage in the forward direction, while at the collector junction it is held almost at zero by a voltage in the reverse direction. The currents across the two junctions are independent of each other; they are simply the currents across single P-N junctions as discussed earlier. The difference between the currents across the emitter and collector junctions is supplied via the base contact. Because a change of voltage between emitter and base has no effect on the concentration gradient at the collector junction, there is, in this case, no transistor action.

In fig. 6*b* it is assumed that the base thickness is small with respect to the diffusion length ($w \ll L_b$), as should be the case in a good transistor. In contrast to the case shown in fig. 6*a*, the concentration gradient for holes in the base near the collector (section 3) is now no longer equal to G_b/L_b as would follow from (eq. 3), but greater, because the concentration of holes $g(2)$ at section *2* exerts an influence (of magnitude dependent on the thickness w of the base). The current of holes crossing the collector junction will not, therefore, have the normal saturation value corresponding to the given G_b and L_b, but will be larger. A change in $g(2)$, occasioned by a change of voltage between emitter and base, will now certainly influence the flow of holes across

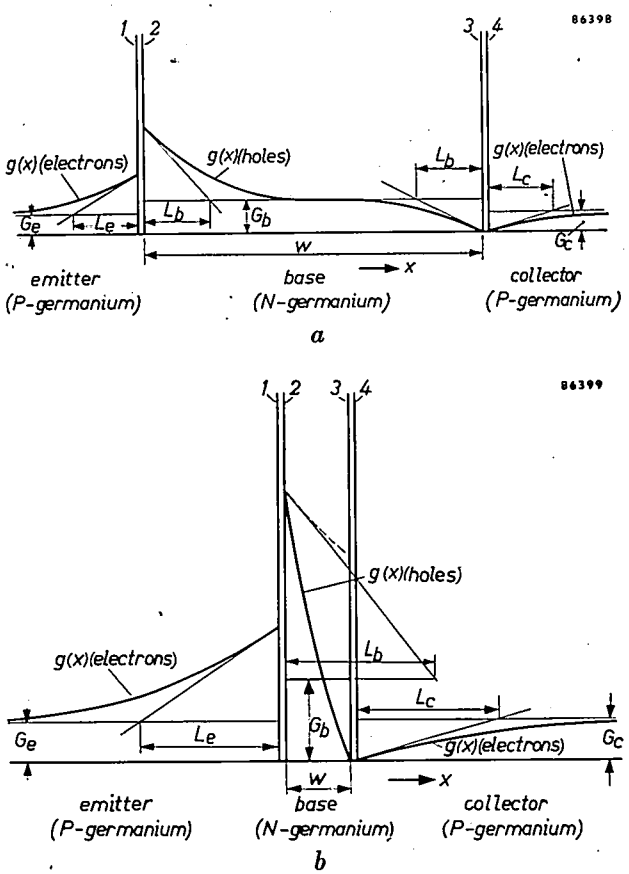


Fig. 6. Minority concentrations $g(x)$ in the emitter and in the collector (electrons) and in the base (holes) of a P-N-P transistor, as functions of position x.
a) The base thickness w is large with respect to the diffusion length L_b in the base. The concentration curve is determined by the generation and recombination processes, just as in separate P-N junctions. The concentration gradient of holes at section *3* is G_b/L_b and is not influenced by the hole concentration in section *2*. There is no transistor action.
b) The base thickness w is small compared to L_b. The hole concentration in section *3* is now affected by the concentration at section *2*, and transistor action occurs. In emitter and collector, the concentration curves remain the same as in (*a*). The broken line shows the hole concentration curve that would occur in the base if the N-P junction was much further away from the P-N junction.

the collector junction. The *electron* current in the collector is not altered because the concentration gradient for electrons in the collector is not altered. A variation of the hole current across the collector junction thus means that the same variation occurs in the *total* current through the collector circuit, for there is no flow of holes to or from the base contact (the concentration gradient is directed substantially along the x-axis and the hole current follows this gradient).

If the thickness w of the base is very small compared with the diffusion length, generation and recombination in the base can be neglected. The hole current, given by $qDdg/dx$, is then constant throughout the base, and since q and D are constants, the same applies to the concentration gradient dg/dx, which becomes $g(2)/w$. This gives the situation shown in fig. 5. The bend in the concentration curve, as shown in fig. 6*b*, is caused by the fact that generation and recombination do have some effect.

At the emitter, the current change which accompanies the voltage change is supplied partly by electrons and partly by holes. Since only holes contribute to the current change in the collector circuit, the transistor, in the circuit shown in fig. 4, gives a current amplification of less than 1. However, the product of current amplification and voltage amplification can, by careful design of the transistor and suitable choice of R_l and of the D.C. bias voltages of the emitter and collector circuits, be made much larger than unity, so that power amplification occurs.

Emitter efficiency γ, base efficiency β and current amplification factor α

For a small increase in the emitter voltage V_e (considered positive when in the forward direction), the current across the emitter junction increases by an amount ΔI_e, consisting of a hole portion ΔI_e^+ and an electron portion ΔI_e^-. Since only ΔI_e^+ is effective in producing a change of collector current, the transistor works better the larger the fraction ΔI_e^+ makes up of the total current change ΔI_e. For constant collector voltage V_c, this fraction is called the emitter efficiency γ, thus

$$\gamma = \left(\frac{\Delta I_e^+}{\Delta I_e}\right)_{V_c}. \quad \ldots\ldots (6)$$

If generation and recombination in the base cannot be neglected, it should be noted that recombination will be dominant because the average hole concentration in the base lies above the equilibrium value (fig. 6*b*). Only a fraction β of ΔI_e^+ reaches the collector to bring about an increase ΔI_c^+ in the collector current. This fraction, again measured at constant collector voltage, is called the base or transport efficiency:

$$\left(\frac{\Delta I_c^+}{\Delta I_e^+}\right)_{V_c} = -\beta. \quad \ldots\ldots (7)$$

(The negative sign comes from the convention given in fig. 4 of calling currents positive if they are directed towards the crystal.) The electron current across the collector, as we have said, is constant because the concentration gradient for the electrons in the collector remains unchanged. This applies in the case considered here (V_c = constant) exactly. Thus:

$$\Delta I_c = \Delta I_c^+ = -\beta \Delta I_e^+ = -\beta\gamma \Delta I_e.$$

This relation fixes the current amplification which occurs at constant collector voltage. The positive product $\beta\gamma$ is called the current amplification factor α; if we now change over to differentials, this may be written:

$$\alpha = -\left(\frac{\partial I_c}{\partial I_e}\right)_{V_c}. \quad \ldots\ldots (8)$$

Also,

$$\alpha = \beta\gamma. \quad \ldots\ldots (9)$$

From the definition of β and γ it follows that the current amplification factor $\alpha < 1$. In practice V_c will not be constant and the real current amplification will be smaller than α. Efforts are made to make β and γ independently as close to unity as possible in order to make α as large as possible.

Approximate calculation of γ

Neglecting generation and recombination in the base, it is easy to calculate γ. It follows from (1) that the ratio of the minority concentrations on opposite sides of a P-N junction is equal to that of their equilibrium concentrations, and thus independent of the applied voltage. When the voltage is changed, therefore, the accompanying concentrations changes will also be in the same proportion. For the emitter junction we thus have the relationship:

$$\frac{\Delta g(1)}{\Delta g(2)} = \frac{G_e}{G_b}. \quad \ldots\ldots (10)$$

From fig. 6*b*, assuming that the hole concentration curve is linear in the base, it can be deduced that:

$$\gamma = \frac{\Delta I_e^+}{\Delta I_e^+ + \Delta I_e^-} = \frac{\frac{\Delta g(2)}{w} D_p}{\frac{\Delta g(2)}{w} D_p + \frac{\Delta g(1)}{L_e} D_n},$$

from which, with the help of (10), we find for the emitter efficiency:

$$\gamma = \frac{1}{1 + \frac{G_e D_n w}{G_b D_p L_e}}. \quad \ldots . \quad (11)$$

Some requirements for emitter, base and collector

Equation (11) shows various possibilities for bringing γ close to unity. D_p and D_n are constants of the pure material of the transistor, in the present instance, germanium. L_e is made as large as possible by allowing the crystal to grow as regularly as possible; irregularities promote recombination and generation and thus reduce the diffusion length. The most important possibilities for increasing the value of γ lie in making both G_e/G_b and w small. Now a small minority equilibrium concentration necessarily implies a large majority equilibrium concentration and vice versa, and since the conductivity is determined by the majority concentration, the conductivity of the emitter should be large and that of the base small, to get a low value of G_e/G_b. The requirement that w be small (thin base) is not only to obtain a large emitter efficiency, but is, of course, also favourable for the base efficiency, since in a thin base only few holes are lost. For the same reason it is advantageous if the diffusion length in the base is long. In the base, too, the crystal must therefore be as regular as possible. The loss of the transistor action if the thickness of the base is several times the diffusion length has already been pointed out (fig. 6*a*). In such a case, the base efficiency $\beta = 0$.

The electron current across the collector is a leakage current, which is desirable to keep small; this can be done by making G_c small. Therefore the collector should also be of high conductivity.

Approximate calculation of β and α

If the loss through recombination in the base is small, and β therefore little less than 1, it is easy to deduce an expression for β. The concentration curve then deviates only slightly from the straight line in fig. 5. The number of holes in the base, summed over the length w (per cm² cross-section), is thus given fairly well by the area of the triangle under this concentration line and is equal to $\frac{1}{2}wg(2)$. Now minority charge carriers have an average lifetime of τ (in the base τ_b), which is proportional to their probability of recombining with a majority charge carrier. The lifetime does depend on the majority concentration though not in a simple way, owing to the intermediate steps involved in the recombination process. Apart from this, however, since we are considering concentration changes which are small compared to the equilibrium concentration of the majority carriers, and the other factors which influence the life do not change, we can consider the majority concentration as being constant and therefore the same applies for τ. The concept "average lifetime" implies that on the average, a number of holes, equal to the number present at any given moment disappear by recombination within a time τ. Thus, in the base, $\frac{1}{2}wg(2)/\tau_b$ holes disappear per second. For a variation $\Delta g(2)$, brought about by a change of voltage across the emitter junction, the loss of holes caused by recombination changes by $\frac{1}{2}w\Delta g(2)/\tau_b$. Since the number of generations per second is constant (see I), $\frac{1}{2}w\Delta g(2)/\tau_b$ is also the change in the difference between the number of holes brought in from the emitter per second and that led off to the collector. This means that:

$$\Delta I_c^+ = -\left\{\Delta I_e^+ - \frac{q}{2\tau_b} w\Delta g(2)\right\}.$$

Assuming again that the hole concentration in the base falls away linearly, we have again:

$$\Delta I_e^+ = qD_p \frac{\Delta g(2)}{w}.$$

Hence we obtain that:

$$\beta = -\frac{\Delta I_c^+}{\Delta I_e^+} = 1 - \frac{w^2}{2\tau_b D_p}. \quad \ldots . \quad (12)$$

This formula shows the quantitative effect of the base thickness w and the lifetime τ_b on the base efficiency.

Since equation (11) is still approximately valid when there are not too many recombinations in the base, one can find from (9) an approximate formula for the current amplification factor α, viz.:

$$\alpha = \frac{1 - \frac{w^2}{2\tau_b D_p}}{1 + \frac{G_e D_n w}{G_b D_p L_e}}. \quad \ldots . \quad (13)$$

In this equation the applied voltages V_e and V_c do not appear. If we neglect the complication which arises because w is somewhat dependent on the applied voltage (Early effect), we may then consider α as a constant of the transistor. We shall see that with this assumption we can easily predict the characteristics of the transistor.

It is clear that the diffusion constant and the lifetime of the minority charge carriers are related to the diffusion length. In fact,

$$L = \sqrt{\tau D} \quad \ldots\ldots\ldots (14)$$

This equation is frequently used in the literature on transistors and semi-conductors in general. The constant $\sqrt{\tau D}$, which plays an important part in the theory, is called the diffusion *length* because of its significance in the one-dimensional case [9]).

Substituting from (14), we can write (12) as

$$\beta = 1 - \frac{w^2}{2L_b^2}, \quad \ldots\ldots\ldots (15)$$

from which it is seen that the ratio of base thickness to diffusion length determines the efficiency of the base.

Exact calculation of α, β and γ for the one-dimensional case

To calculate α, β and γ exactly, one must start from the two equations which govern the behaviour of holes in the base. The first equation expresses the fact that the increase per second of the hole concentration at an arbitrary point is equal to the difference in the number of holes which flow in and out of unit volume per second, less the difference between the number of recombinations and generations per second, i.e.

$$\frac{\partial g}{\partial t} = -\frac{1}{q}\frac{\partial I^+}{\partial x} - \frac{g-G}{\tau}. \quad \ldots\ldots\ldots (16)$$

The second equation expresses the fact that the hole current is caused by diffusion:

$$I^+ = -qD\frac{\partial g}{\partial x}. \quad \ldots\ldots\ldots (17)$$

We assume here that all concentration changes take place so slowly that at any moment we are dealing with a stationary state (i.e. the analysis does not apply to high frequency changes). Then $\partial g/\partial t = 0$. Combining this with (16) and (17) gives

$$D\frac{d^2g}{dx^2} - \frac{g-G}{\tau} = 0. \quad \ldots\ldots\ldots (18)$$

Apart from this differential equation, g must also satisfy the boundary conditions determined by the emitter voltage and collector voltage (see fig. 6*b* and eq. 1) in the limiting planes of the base [10]).

Solution of (18) gives the hole concentration as a function of x and by differentiating with respect to x, also the concentration gradient; substituting in (17) gives the hole current I^+ as a function of x. Next, by substituting the values of x for the boundary planes *2* and *3* (fig. 6*b*), we find I_e^+ and $-I_c^+$ respectively, both as a function of V_e and V_c. Differentiating with respect to V_e at constant V_c gives ΔI_e^+ and $-\Delta I_c^+$ after which β can be calculated from (7). The result is:

$$\beta = \operatorname{sech}\frac{w}{L_b}. \quad \ldots\ldots\ldots (19)$$

To determine γ, we must calculate $\Delta I_e = \Delta I_e^+ + \Delta I_e^-$ (see (6)). I_e^- is found directly by application of (5). Differentiation of the result with respect to V_e produces ΔI_e^-. We then find for γ:

$$\gamma = \frac{1}{1 + \frac{G_e D_n L_b}{G_b D_p L_e}\tanh\frac{w}{L_b}}. \quad \ldots\ldots\ldots (20)$$

The product of (19) and (20) gives α. For small values of w/L_b, the equations (19) and (20) reduce to the approximate equations (12) and (11) found earlier.

In practice, the situation is less favourable than would be indicated by (20). As we have already said, the diffusion length is strongly dependent on the lattice defects and certain impurities in the crystal, since the lifetime of the minority charge carriers is generally shortened by such disturbances. At the outside surface, the periodicity of the lattice is completely disrupted where the crystal ends. The recombination which occurs here sucks holes to the crystal surface. This suction is further influenced by the presence of foreign atoms adsorbed on the surface, and reinforced because mechanical processes, (grinding, scouring and polishing) have often deformed the surface. Usually, the recombination at the surface is more important than that inside the base. A special etching process is used to reduce these deleterious surface effects as much as possible.

The base current

We have seen how some of the holes injected into the base by the emitter disappear as a result of the recombination excess, while the remainder flows into the collector. There is no flow of holes to or from the base contact since, as pointed out carlier, the concentration gradient is directed substantially along the x-axis and the hole current follows this gradient. The current of holes could therefore be dealt with uni-dimensionally. This is not the case with the electron current. An electron current flows into the base from the collector and electrons are led off into the emitter. These currents are determined by the concentration gradients at the junctions in the collector and emitter respectively (where the electrons are minority carriers), and hence by the voltages across the collector and emitter junctions. In general, more electrons are led off across the emitter junction than flow into the base from the collector. Furthermore, electrons are needed to compensate for the recombination surplus. A continuous supply of electrons via the base contact is therefore needed. The current which results is called the "base current", I_b. Since the collector electron current is a (constant) saturation current, the change in the required supply of electrons which accompanies a change in emitter current must come entirely from the base current. The role which the base current plays is discussed in the next section.

---

[9]) The fact that, in the one-dimensional case, $\sqrt{\tau D}$ is the distance in which the deviation from the equilibrium concentration changes by a factor e, is derived in I, p. 221. A constant $1/b$ is used there in place of τ. The connection between b and the lifetime was not mentioned.

[10]) Equation (18) was also used for the calculation in I (p. 238) of the minority current across a *P-N* junction. The second boundary condition there was different: the concentration had to approach equilibrium values at $x = \infty$.

Here too, we have drawn conclusions about the majority charge carriers from their known behaviour in adjoining regions where they are minority carriers.

Circuit with common base and circuit with common emitter; the current amplification factor α'.

In *fig. 7a* the circuit shown in fig. 4 is reproduced, this time with the usual symbol for a *P-N-P* transistor. The emitter current I_e flows through the A.C. input source; I_e is thus the input current. The A.C. input source can also be included in the circuit in the manner shown in fig. 7*b*, the base current I_b is then the input current.

The D.C. sources used to bias the transistor, form a short circuit to the alternating currents which occur in the network. Where the A.C. characteristics of the network are concerned, then, we can leave these D.C. sources out of consideration. Thus, in fig. 7*a*, the A.C. input source e_i and the load resistance R_l are both connected directly to the base lead

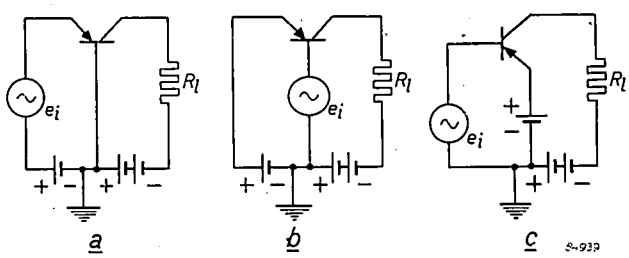


Fig. 7. *a*) Transistor circuit with common base. If the internal resistances of the D.C. sources are neglected, then A.C. input source e_i and load resistance R_l are both connected directly to the base lead of the transistor.
The *P-N-P* transistor is here indicated by the usual symbol. The emitter is distinguished from the collector by an arrow; the arrow is directed towards the crystal because the D.C. flowing to the emitter is in this direction.
b) Transistor circuit with common emitter. Disregarding the D.C. sources, the A.C. input source and the load resistance are both directly connected to the emitter lead.
c) Usual way of drawing a common emitter circuit. The circuit is identical with that in (*b*).

of the transistor. This method of connection is therefore known as "circuit with common base", or, since the common base is usually grounded, "grounded base circuit". In fig. 7*b*, neglecting again the D.C. sources, the A.C. input source and the load are both connected directly to the emitter lead of the transistor; this circuit is therefore known as "circuit with common emitter" or "grounded emitter". In this case the circuit is usually drawn as shown in fig. 7*c*.

For a circuit with common emitter, the current amplification is the ratio of the change in collector current to that of the base current, and for this circuit the current amplification factor α' is defined analogously to that for the circuit with common base, viz. [11]):

$$\alpha' = \left(\frac{\partial I_c}{\partial I_b}\right)_{V_c} \quad \ldots\ldots\ldots \quad (21)$$

Taking into account the sign convention for current (see fig. 4), we have:

$$\Delta I_e + \Delta I_c + \Delta I_b = 0. \quad \ldots\ldots \quad (22)$$

At constant collector voltage (see eq. 8), $\Delta I_e = -\Delta I_c/\alpha$; hence from (22),

$$\left(\frac{\Delta I_c}{\Delta I_b}\right)_{V_c} = \frac{\alpha}{1-\alpha},$$

so that

$$\alpha' = \frac{\alpha}{1-\alpha}. \quad \ldots\ldots\ldots \quad (23)$$

For a good transistor α is not much less than 1, and from (23) it is seen that α' is therefore much greater than 1. A reasonable value of α is for example $\alpha = 0.98$; then $\alpha' = 49$.

Input and output impedances of transistor circuits with common base and common emitter

Practical importance of input and output impedance of an amplifier

For a given amplifier and a given signal source, the aim in principle, is to obtain the maximum transference of power from the source to the amplifier and from the amplifier to the load. To achieve this, source, amplifier and load must be matched. Matching is attained when the ratio of the A.C. input voltage to the input current of the amplifier (input impedance) is equal to the internal resistance of the source of the signal, and, at the same time, the ratio of the A.C. output voltage to the output current (output impedance) is equal to the load resistance. (For simplicity we restrict ourselves to the case where all the impedances involved can be taken to be pure resistances).

Whether such matching can be achieved in practice depends largely on the order of magnitude of the input and output impedances. In view of their importance in practice, we shall qualitatively compare the impedances of transistor circuits with common base and with common emitter.

[11]) The quantity α' thus defined is always positive: the addition of a minus sign in the defining equation is therefore unnecessary, in contrast with the case of the definition of α (see eq. 8).

Input and output impedances of transistors

For a transistor, in fact for amplifiers in general, it is not possible to quote an input and output impedance without further qualification. The input impedance depends on the load and the output impedance on the internal resistance of the signal source connected to the input side.

The output impedance in a circuit with common emitter is lower than with common base. To appreciate this, we must take into account that the concentrations of the minority charge carriers at the collector are not exactly zero, but that these concentrations depend somewhat on the voltage across the collector junction. In order to determine the output impedance, it is necessary to be able to vary the voltage between collector and base; an A.C. source is therefore connected between them (*fig. 8*). It is

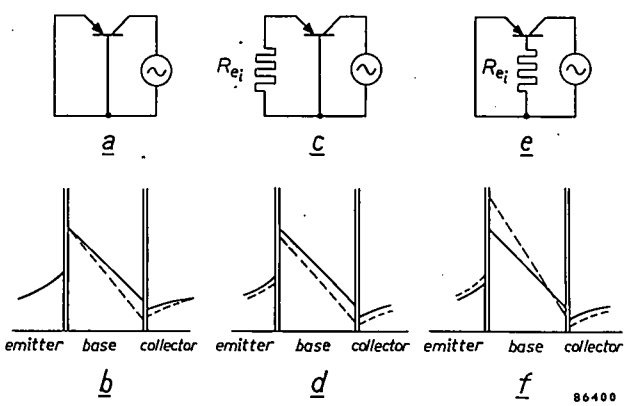


Fig. 8. As a result of the internal impedance of the signal source (here, for simplicity, represented as a pure resistance R_{ei}) the output impedance of a transistor with grounded base (*c* and *d* is higher, with grounded emitter (*e* and *f*) lower than when base and emitter are short-circuited for A.C. (*a* and *b*). The concentration distributions of the minority charge carriers corresponding to each circuit are shown at a moment when the voltage across the collector junction is slightly more in the reverse direction (broken lines) than that due to the simple D.C. bias (full lines). To simplify the diagrams the D.C. sources providing the bias are not drawn.

useful to begin the argument from the hypothetical situation represented in fig. 8*a*: here the signal source is assumed to be of zero internal resistance so that no voltage variation due to the generator in the collector circuit can appear between emitter and base. Hence the minority concentrations at the emitter remain fixed. Let us suppose that the voltage across the collector changes in such a way that the minority concentrations at the collector drop a little: an increase then occurs in the current across the collector (see fig. 8*b*, in which, for simplicity, recombination and generation in the base have been neglected). The hole contribution to this increase (the hole concentration gradient has increased slightly) necessitates an extra current via the emitter circuit; the electron contribution requires an extra current via the base circuit. Both extra currents are so directed that positive charge flows towards the crystal. Since the concentration changes at the collector are small, the changes in current are small, i.e. the output impedance is high.

In practice, in the circuit with *common base*, the internal resistance R_{ei} of the signal source appears in the emitter circuit (fig. 8*c*). The extra current in the emitter circuit causes a voltage drop in this circuit which reduces the potential difference between emitter and base. The concentrations at the emitter thus drop and the increase in the hole concentration gradient in the base is partly compensated (fig. 8*d*). The hole current thus alters less than was the case in fig. 8*b*. The change in the electron current over the collector junction has remained the same, so that the total collector current change has decreased, i.e. the output impedance is even higher than with the hypothetical short-circuited emitter circuit.

For a circuit with *common emitter*, the internal resistance of the signal source is in the base circuit (fig. 8*e*). The extra current in the base circuit (due to the voltage change across the collector junction) results in a voltage drop which serves to increase the potential difference between emitter and base, and this causes the concentrations at the emitter to rise. The original increase in the hole current as a result of the fall in concentration at the collector, is thus reinforced (fig. 8*f*). The collector current therefore also increases and the output impedance is lower than with short-circuited base circuit.

The emitter electron current also increases. Hence the whole increase in the collector electron current does not have to be carried away via the base circuit, for some of it disappears across the emitter junction. The voltage across the emitter junction changes until the balance between the electron transport to and from the base is restored.

The greater the value of R_{ei}, i.e. the more difficult it is to lead electrons away via R_{ei}, the greater is the portion of the electron current which finds its way across the emitter. The voltage difference across the emitter junction must therefore adjust itself to a correspondingly higher value; hence the rise in the hole concentration at the emitter, and thus the hole current through the base, increase still further. This means that as R_{ei} is increased, the output impedance decreases.

We have here a very remarkable situation. The current generated in the collector circuit by a source, divides itself between two parallel branches, viz. the emitter and the base. If R_{ei} is placed in the

emitter branch, the result of an increase in R_{ei} is an increase in the effective resistance of the parallel combination, as would be expected. However, if R_{ei} is in the base branch, the effective resistance becomes *less* as R_{ei} becomes greater, which is quite contrary to the normal behaviour of parallel resistances.

The output impedance with common base can easily be several megohms in practice; with common emitter some ten thousands of ohms is a normal value.

The *input impedance* of a transistor in a common base circuit is considerably lower than that in a common emitter circuit (e.g. 25 to 30 Ω for the former and 1200 Ω in the latter case). For a given alternating input voltage we have to deal in the first case with fluctuations of the emitter current, which are much larger than those of the base current; the latter, however, determine the input impedance in the second case.

Comparison of transistors with thermionic valves

Transistors and amplifying valves exhibit considerable differences in their properties even at audio frequencies. This is largely due to the fact that the input and output impedances of transistors differ so much from the values usually occurring for valves. In a pentode, for example the input and output impedances are generally very large with respect to the internal resistances that are usual for signal sources and load resistances used in practice. In most cases, therefore, the A.C. output i_a of a pentode is given by $i_a = Sv_g$, where S represents the mutual conductance and v_g the e.m.f. of the signal source. The mutual conductance thus determines the behaviour of the pentode; it makes little difference how large the input and output impedances are, provided they be large.

For a transistor too, the mutual conductance can be found. With common base it is

$$S = \left(\frac{\partial I_c}{\partial V_e}\right)_{V_c};$$

With common emitter, the same expression applies, but with a minus sign, since an increase of the input voltage in the latter case means a decrease of the voltage V_e across the emitter junction.

Now:

$$\left(\frac{\partial I_c}{\partial V_e}\right)_{V_c} = \left(\frac{\partial I_c}{\partial I_e}\right)_{V_c}\left(\frac{\partial I_e}{\partial V_e}\right)_{V_c} = -\alpha\left(\frac{\partial I_e}{\partial V_e}\right)_{V_c}.$$

If one assumes that the hole concentration in the base is linear, it is possible with the help of fig. 6*b* and equation (1) to deduce an expression for I_e as function of V_e; by differentiating this one finds that for normal biassing of transistors ($g(1) \gg G_e$):

$$\left(\frac{\partial I_e}{\partial V_e}\right)_{V_c} = \frac{q}{kT} I_e.$$

Ignoring the sign, the mutual conductance of a transistor is thus:

$$S = \alpha \frac{q}{kT} I_e.$$

Apart from the direct current (I_e) through the emitter junction, S depends only on α. For a good transistor α is almost equal to 1, and all transistors have thus the same mutual conductance for the same D.C. biassing. For a normal value of I_e, for example 1 mA, the mutual conductance is 40 mA/V, which is much greater than can be achieved with pentodes.

However, a comparison of a transistor with a tube simply on the basis of the mutual conductances is of little value, because for a transistor the behaviour is largely governed by the input and output impedances. In a circuit with common base, the output impedance is very high (as for a pentode), but the input impedance is then only a few tens of ohms. The latter, for constant V_c, is

$$\left(\frac{\partial V_e}{\partial I_e}\right)_{V_c},$$

and this, as we have seen above, is equal to kT/qI_e; at room temperature and for $I_e = 1$ mA, its value is 25 Ω. In the circuit with common emitter the input impedance is higher, but the output impedance on the other hand is much lower than that of the circuit with common base. However, a transistor in the circuit with common emitter behaves much more like a valve than it does in the circuit with common base. For a very large value of α' (very small base current) the input inpedance would be comparable with that of a valve, but the output impedance in this case would be very low.

The behaviour of transistors is clearly more complicated than that of valves. Account must be taken of the input and output impedances, and this is complicated by the fact that these impedances are not dependent solely on the transistor: as pointed out above, the input impedance depends on the load and the output impedance on the internal resistance of the signal source.

An obvious advantage of transistors as compared with valves is that for amplification of small signals, a small bias is sufficient. For small signals, the bias with valves (i.e. anode voltage) is always ten or more times greater than the collector-base voltage of the transistor, so that the energy dissipation is much greater; transistor action is thus more efficient for small signals. Furthermore, a transistor needs no filament or heater current, an advantage which speaks for itself.

Transistor characteristics according to the simple theory

A transistor is generally characterized by a set of graphs which give the collector current I_c as function of the collector voltage V_c, for various values of the input current. The latter is the emitter current I_e with a common base and the base current I_b with a common emitter circuit. We are of course interested only in that section of the graphs which corresponds to a collector voltage in the reverse direction and an emitter voltage in the forward direction.

Common base

If the emitter current is zero (emitter open-circuited) a reverse current I_{cl}) (l for "leakage") flows through the collector, which for increasing reverse potential approaches exponentially to a saturation value I_{co}) (*fig. 9a*). From the definition of α (see eq. 8) it follows, by integration:

$$I_c = I_{cl} - \alpha I_e. \quad \ldots\ldots \quad (24)$$

For a *P-N-P* transistor, which we have been

considering throughout, I_e is positive, but I_c and I_{cl} are negative, like V_c (reverse potential). In the graph we therefore plot $-I_c$ as a function of $-V_c$ for the various values of I_e.

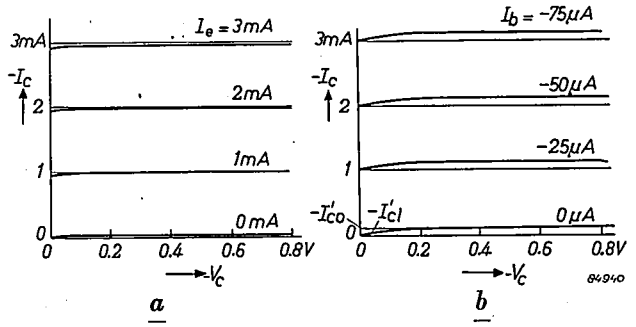


Fig. 9. Theoretical characteristics of a transistor: collector current I_c as function of the collector voltage V_c for various values of the input current. According to the sign convention (fig. 4) V_c and I_c are negative: $-V_c$ and $-I_c$ have therefore been plotted.
a) For a transistor with common base circuit. The lowest characteristic gives I_{cl}, the collector current with open-circuited emitter.
b) For a transistor in circuit with common emitter. The parameter here is the base current I_b. It should be noted that all the values of I_b shown are negative. I'_{cl} is the saturation value of the collector current I'_{cl} with open-circuited base.

The characteristic for any value of I_e is now obtained by moving the characteristic for $I_e = 0$ upwards by an amount αI_e.

Common emitter

If the base current is zero (base open-circuited), a current I'_{cl} flows which, for the same collector voltage is a factor $(1 + \alpha')$ larger than the leakage current I_{cl} in the case of the common base. This may be seen as follows. We have

$$I_e + I_b + I_c = 0.$$

Adding this to (24), we can write

$$(1-\alpha)\, I_e = -I_b - I_{cl}\,.$$

For open base circuit, $I_b = 0$ and $I_e = -I'_{cl}$ so that

$$(1-\alpha)\, I'_{cl} = I_{cl}\,.$$

From (23) it follows that $1 - \alpha = 1/(1 + \alpha')$, so that

$$I'_{cl} = (1 + \alpha')\, I_{cl} \quad \ldots\ldots \quad (25)$$

I'_{cl} also approaches exponentially to a saturation value I'_{co} at increasing reverse voltage and, following directly from (25) we have

$$I'_{co} = (1 + \alpha')\, I_{co}\,. \quad \ldots\ldots \quad (26)$$

From (21) we obtain by integration:

$$I_c = I'_{cl} + \alpha' I_b. \quad \ldots\ldots \quad (27)$$

For a *P-N-P* transistor I_c, I'_{cl} and I_b are all negative. We therefore plot $-I_c$ as a function of $-V_c$ for the various values of I_b.

The characteristic for any value of $-I_b$ is obtained by shifting the characteristic for $I_b = 0$ upwards by an amount $\alpha'(-I_b)$ (fig. 9*b*).

If the characteristics for common base and common emitter are plotted on the same scale, then, as follows from (25), at the same value of $-V_c$ the characteristics will have a gradient for common emitter which is $(1 + \alpha')$ times that for common base. The gradient of the characteristics represents the internal reciprocal resistance of the transistor at constant input current, i.e. using a signal source of infinite internal resistance. In this case the output impedance for common base is thus $1 + \alpha'$ times as great as that for common emitter.

Early effect

The actual shape of the transistor characteristics differs somewhat from the shape deduced above, as a result of the *Early effect* which was mentioned on p. 240. Up till now we have assumed that the boundary planes of a *P-N* junction, e.g. the sections *1* and *2*, or *3* and *4* in fig. 6*b*, had a fixed position in the crystal. In reality, the space charge region which exists between the two planes (see p. 234) becomes thicker as the voltage in the reverse direction increases. This results principally from a shifting of the plane which lies on that side of the *P-N* junction where the specific resistance is greatest. In a transistor it is plane *3* which undergoes the largest displacement because this plane is situated in the base which, as we have seen, usually has a greater specific resistance than the collector. With increasing reverse voltage across the collector, section *3* shifts towards the emitter and the effective base thickness decreases. This is known as the Early effect [12]). This makes α, and hence also α', larger. The transistor characteristics will therefore be less flat, and the output impedances at constant input current will be smaller, than would be expected from the simple theory. At section 2 on the emitter side of the base, the same effect occurs. There it is much less important, since the voltage variations across the emitter junction are so much smaller than those across the collector junction.

A further important consequence of the Early effect is that with increasing collector voltage, i.e. with increasing α, the emitter current increases. There is thus a feedback effect of the output voltage on the input current, which is undesirable. To limit this effect, it is necessary that the specific resistance of the material of the base is not too large.

P-N-P and *N-P-N* transistors

In unfolding the theory, we have always referred to *P-N-P* transistors. For *N-P-N* transistors — where the base is of *P*-germanium and emitter and

[12]) J. M. Early, Effects of space-charge layer widening in junction transistors, Proc. I. R. E. **40**, 1401-1406, 1953.

collector of *N*-germanium — the theory is completely analogous. In principle, the theory can be developed without mention of holes and electrons, by dealing simply with minority and majority charge carriers: the theory then becomes identical for both types of transistor. Practical differences in the behaviour of the two types are caused by the fact that the diffusion constant for electrons is about twice that for holes.

Summary. The action of a junction transistor is based on the properties of *P-N* junctions in a germanium crystal. Such a transistor is built up of a *P-N* junction from "emitter" to "base" followed at a very small distance by an *N-P* junction from "base" to "collector". The action is explained, starting from the fact that the current at a junction is determined by the diffusion currents of the minority charge carriers which can easily be calculated at the boundary planes of such a junction. It is pointed out that the voltage applied across a junction determines the concentrations of the minority charge carriers in the boundary planes. At one of the junctions this concentration is fixed at zero; a small change of voltage across the other junction then results in a strong variation of the concentration gradient, and thus of the diffusion current of the minority charge carriers, between the two junctions.

Emitter and base efficiency are discussed and calculated for the one-dimensional case; this is also done for the current amplification factors for circuits with common base and common emitter. The difference in the input and output impedances for these two circuits is explained and the characteristics which follow from the theory, are deduced.

M.F. Argo

A MACHINE FOR BEND TESTS

620.177.3

A requirement frequently imposed upon a metal is that a specimen of certain form should not crack or break when bent through a prescribed angle around a mandrel of prescribed diameter. The result of a bend test, however, depends upon the manner in which the test is carried out, and this is not always specified. The consequence may be a difference of opinion between supplier and customer as to whether the metal satisfies or falls short of the specification.

With a view to overcoming such difficulties the Materials Research Group of the Research Laboratory in Eindhoven has developed a special bending machine for testing sheet and strip metal (*fig. 1*). By means of a single hand-operation the test specimen can be bent in this machine quite smoothly and with the correct radius through any angle between 0 and 180°. The design is shown in *fig. 2*. The test specimen (*4*) is held between the clamping piece (*1*) and the mandrel in the form of an interchangeable radius plate (*3*), which is accurately rounded on the top edge. The upper face of the clamping piece is shaped in such a way as to support the specimen as close as possible to the bending point, thus preventing unwanted deformation. The bend is effected by the thrust block (*7*), which is moved by the lever (*9*). Three rollers (*6*) mounted in the thrust block ensure that, when the lever is pushed over, the block slides along the specimen almost without friction so as to exert practically no tensile load upon it. This is important, for if there were any appreciable tensile load on the specimen it would not be subjected to a pure bend and premature cracking might occur. Allowance can be made for the thickness of the specimen by adjusting with wing-nut *8* the position of the thrust block in its groove on the lever. The lever *9* turns about an axis which coincides with the centre of curvature of the radius plate; this is essential to the correct functioning of the machine. The interchangeable radius

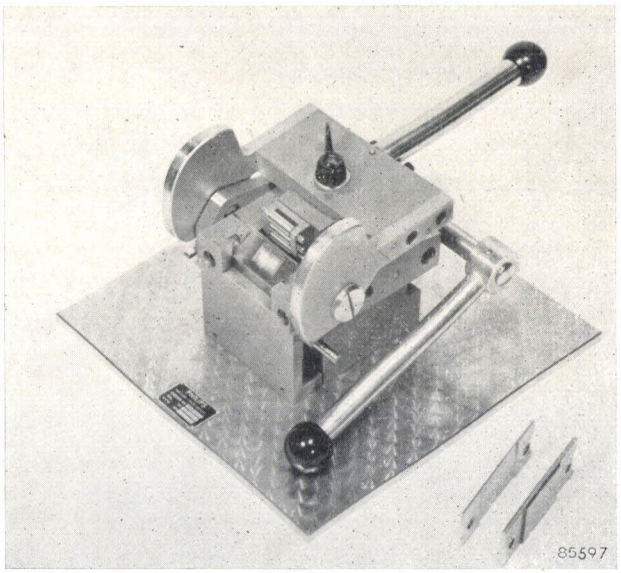


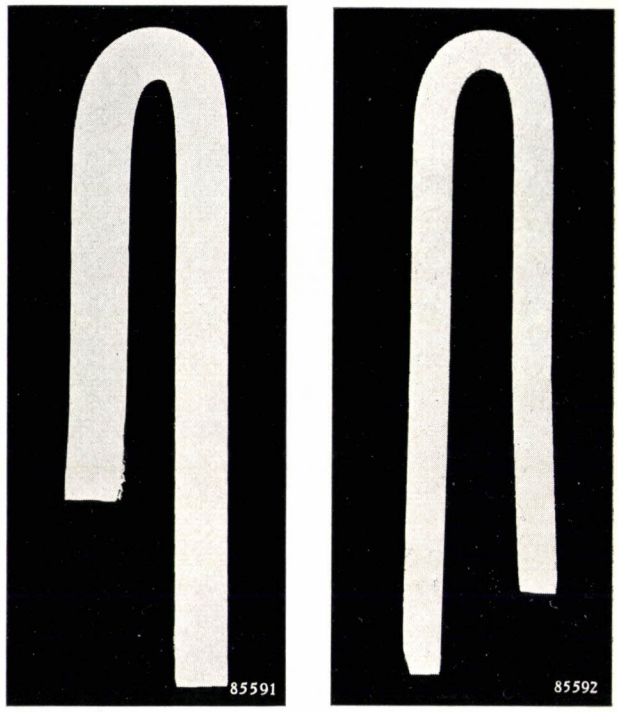
Fig. 1. The newly developed machine for testing sheet or strip metal for its bending properties. The test specimen is inserted between rollers and a radius plate. At the right can be seen two of the interchangable radius plates over which the specimen is bent.

plates are ground in such a way that the centre of curvature always lies at the same position in the machine. A support bar (*2*) takes up the horizontal component of the force exerted upon the radius plate during operation and thus prevents it from bending under the strain. When the angle of bend exceeds 90° this component reverses direction and the clamping piece (*1*) then takes up the horizontal

force. The support bar is then no longer necessary and, indeed, it forms an obstruction to further bending. For this reason it is pressed down against two springs (*12*) by the cams (*5*) on the end of the lever (*9*), which bear on both ends of the pin (*11*) when the angle of bend becomes greater than 90° (fig. 2*b*). In this way the bend test can be continued unhindered up to 180°. The angle of bend can be read from one of the cams (*5*) which is calibrated in degrees. Even at an angle as small as 60° to 70°, the outside of the bend becomes visible with the specimen still in position in the machine. It can thus readily be perceived at what angle of bend crack formation begins.

Fig. 3a shows the cross-section of a specimen of chrome steel with a tensile strength of approximately 65 kg/mm² and a thickness of 1.9 mm, which has been bent through 180° over a diameter of 0.95 mm. The diameter of the radius plate in this case was therefore half the thickness of the specimen. It can be seen that, even under such very adverse conditions for a bend test as these, the specimen nevertheless follows the radius plate quite faithfully. Fig. 3*b* shows that under more favourable conditions (mild steel with a tensile strength of 40 kg/mm² and a thickness of 1.0 mm, bent over a diameter of 1.20 mm) the inner profile follows almost exactly the curvature of the radius plate.

As stated, the machine described serves for testing sheet or strip metal, but it is also possible to design a machine on the same principles for testing metal in wire or bar form.



a *b*

Fig. 3. *a*) Section of a test specimen of chrome steel, bent through 180° by the machine shown in fig. 1 (tensile strength of specimen approx. 65 kg/mm²; specimen thickness $d = 1.9$ mm; diameter of radius plate $D = 0.95$ mm; thus $D/d = 0.5$). *b*) Section of a test specimen of mild steel (tensile strength approx. 40 kg/mm²; $d = 1.0$ mm; $D = 1.2$ mm, and $D/d = 1.2$).

In order to make clear why this rather elaborate machine has been developed for performing such a simple test, some comment sare called for on the rather primitive methods often employed for the bend test.

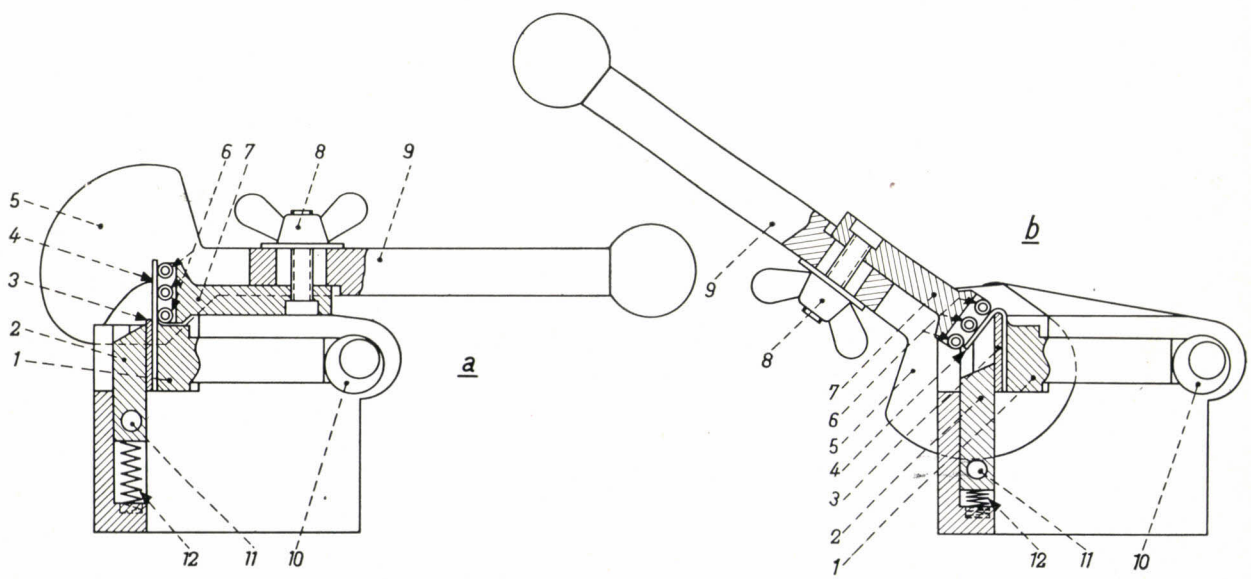


Fig. 2. Structural design: *1* clamping piece; *2* support bar; *3* radius plate; *4* test specimen; *5* cam, which depresses support bar; *6* rollers, mounted in thrust block *7*, which is fixed by wing-nut *8* to lever *9*; the excentric *10* serves to hold clamping piece against specimen; *11* pin by which the support bar is depressed against spring *12*. *a*) Machine ready for operation after insertion of test specimen; *b*) after bending through 145°.

A method in common use is to secure the test specimen together with a mandrel in a vice (*fig.4a*) and to bend the specimen over the mandrel by hand or with a hammer. If the specimen is to follow the mandrel closely, the bending force must be applied close to the mandrel (fig. 4*b* and *c*). The use of a hammer is to be deprecated, as it subjects the specimen to shock-loading of unknown magnitude, which may result in premature cracking or breaking. What is more, hammering close to the mandrel may produce undesirable work hardening of the regions subjected to impact.

If an angle of bend greater than 90° is specified, it is necessary, when using this simple method, to remove the test specimen at a 90° bend from the

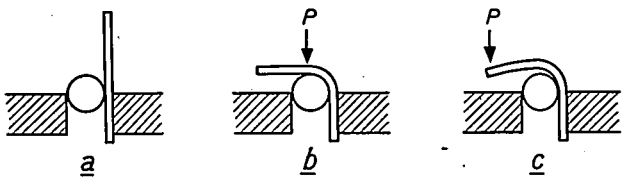


Fig. 4. Specimen and mandrel clamped together in a vice for a primitive bend test (*a*). If the bending force *P* is applied near to the mandrel, the bend radius follows the curvature of the mandrel (*b*); if *P* is applied at some distance from the mandrel, the bent radius will be too large (*c*).

vice, to secure it again in a different way (*fig. 5a*) and then to continue the bend test by tightening the vice itself. It is now almost impossible to bend the specimen closely around the mandrel, because the bending forces are necessarily applied at some distance from the mandrel, and the mandrel always slides back a little as the angle of bend increases. The consequence is a sharper bend than specified (fig. 5*b*). The second objection might be overcome by supporting the specimen directly beneath the mandrel (fig. 5*c*), but then the metal is no longer subjected to a pure bending load. Moreover, the force exerted by the support may be so considerable as to cause unwanted deformation of the specimen at the bend radius. Nor is it possible with this method to determine the point at which cracks begin to form.

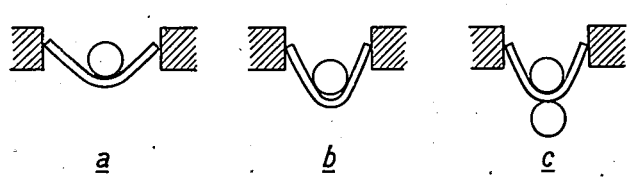


Fig. 5. Specimen from fig. 4 arranged in vice for bending through angles greater than 90°. The jaws of the vice are moved toward each other. *a*) As in fig. 4*c*, unwanted bending occurs; the mandrel also becomes displaced (*b*) causing too sharp a bend. Supporting the mandrel (*c*) avoids displacement but the specimen is then no longer subjected to a pure bending.

If an angle of bend of 180° is specified, a method frequently used is to press the test specimen into a die by means of a punch with a curvature of the prescribed radius. The aperture of the die is then equal to the thickness of the punch plus twice the thickness of the specimen (*fig. 6*). In this case an unwanted load is exerted upon the specimen by the friction occurring between it and the die. This

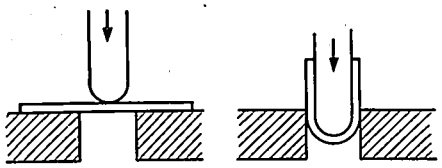


Fig. 6. Bend test employing a die and a radius plate in the form of a punch.

friction can be limited by rounding-off the edges of the die and possibly by providing the aperture with a taper, but these measures cannot be taken far without entailing the drawbacks illustrated by fig, 5*a* and *b*. The method has no practical value if the punch is thin in relation to the test specimen or if punch and specimen are both thin. In the first case the strains set up in the punch are so great as to cause it to collapse. In the second case the specimen suffers deformation as a result of the shearing forces arising.

Use is sometimes made of special tools incorporating several of the elements of the machine described here. The test specimen is secured in a clamp, one side of which is given the curvature over which the specimen is to be bent, that is to say one clamping piece acts as the mandrel (*fig. 7*). The metal is bent

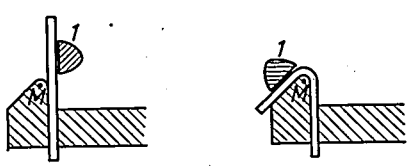


Fig. 7. Bend test employing special tools. The axis of rotation of thrust block *1* coincides with the centre of curvature *M* of the curved member over which the specimen is bent.

over this surface by means of a thrust block attached to an arm, whose axis of rotation coincides with the centre of curvature of the mandrel. With this arrangement, however, the specimen cannot be bent through more than about 110° (depending upon the requisite strength of the mandrel). Moreover a tensile load is exerted by the thrust block as it moves over the face of the specimen. These drawbacks have been substantially eliminated in the apparatus described here.

M. J. W. GEENEN.

THE "NORELCO" COUNTING-RATE COMPUTER

by E. A. HAMACHER † *) and K. LOWITZSCH *).

621.317.79:621.374.32:548.734

Analogue computers have acquired considerable importance in many fields. The computer described in this article is very simple with regard to the underlying mathematics. It may be regarded as a useful example to illustrate some of the mechanical and electrical intricacies involved, when a high accuracy is asked of this type of instrument.

Introduction

In X-ray diffractometry with the "Norelco" equipment [1]) the counting-rate computer is used for the measurement and the automatic recording of the diffracted X-ray intensity vs. Bragg angle according to the *fixed count* method. As was explained in a former article [2]) this method consists in measuring the time t necessary to accumulate a predetermined number of counts N, whereupon the counting rate n (which is proportional to the intensity of the radiation entering the detector) is computed from the formula

$$n = N/t. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

The fixed count method has the advantage that each intensity value in a pattern has the same statistical relative error (expressed, for example, by the "probable" relative error ε_{50}, cf. II).

Several counting-rate recorders for automatic fixed count measurements have been developed [3]). In these instruments, however, counting rates n are recorded on an inverse scale in one case and on an approximately logarithmic scale in another. The basic idea of the "Norelco" counting-rate computer was to obtain a recording of n on a *linear* scale by automatic evaluation of eq. (1).

The computer, which is mounted in the electronic circuits and recorder rack of the "Norelco" diffractometer, is shown in *fig. 1*.

86161

Fig. 1. Chassis and front panel of the "Norelco" counting-rate computer. This unit is placed in the electronic circuits rack of the diffractometer. The computer output can either be fed to the recorder or be read on the meter at the upper right. At the upper left a register is mounted which is not used for the computer but only in manual fixed count measurements (cf. II).

Principle of the computer

The analogue to equation (1) which is used as the basis of the computer, is Ohm's law:

$$I = E/R \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

The sliding contact of a slide-wire resistor is moved by a synchronous motor, so that the resistance R of the used portion of wire will increase with time at a constant selectable rate k:

$$R = kt \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

The current I flowing in the resistor when a fixed voltage E is applied to it will gradually decrease. When the time t in eq. (3) is made equal to that in (1), the current I evidently will be proportional to the counting rate n, and therefore may be directly fed to the strip-chart recorder to give the desired linear recording.

To make use of the analogue for the diffractometer, thus amounts to the following procedure. At a given Bragg angle position of the detector, the movement of the contact on the slide-wire is automatically

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

[1]) W. Parrish, E. A. Hamacher and K. Lowitzsch, The "Norelco" X-ray diffractometer, Philips tech. Rev. **16**, 123-133, 1954/1955 (No. 4). This article is hereafter referred to as I.

[2]) W. Parrish, X-ray intensity measurements with counter tubes, Philips tech. Rev. **17**, 206-221 1955/56 (No.7-8). This article is hereafter referred to as II.

[3]) S. W. Lichtman, E. T. Byram and H. Friedman, Strip chart recording with an autoscaler, Electronics **23**, April 1950, page 122. Another instrument was designed by the Berkeley Scientific Corp., Richmond, Calif.

started when counting begins, i.e. when the scaling circuits of the diffractometer are put into operation (cf. II, fig. 12). The movement is stopped and a fixed voltage is applied to the slide-wire when the predetermined number of counts has been accumulated, i.e. after one revolution of the mechanical register of the diffractometer has been completed (the accumulated number of counts N is then $100\times$ the selected scaling factor, thus 100, 200, . . ., 12800 or 25600). The current now flowing in the slide-wire is recorded on the strip chart. After a measurement has been made and recorded, the sliding contact is reset to its initial position, the detector is moved by the goniometer to the next Bragg angle position and the operation is repeated. A continuous trace consisting of steps of equal width is thus obtained; see *fig. 2*.

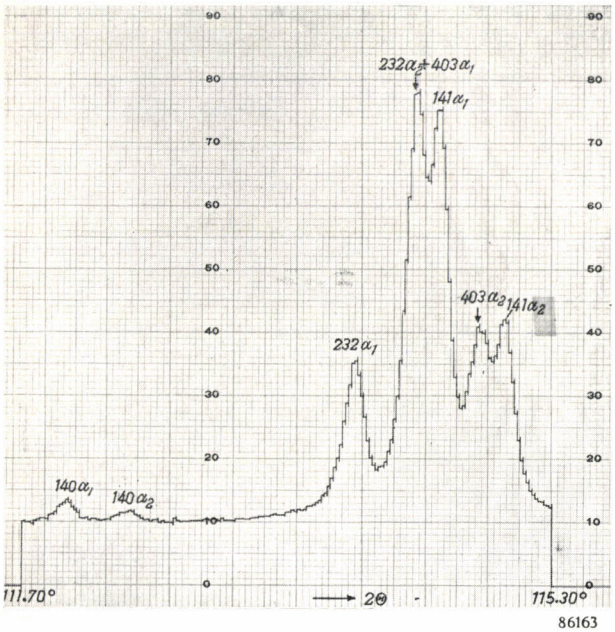


Fig. 2. Chart of diffraction pattern recorded by the "Norelco" counting-rate computer. The Bragg angle steps in this instance are $0.02°(2\Theta)$; the statistical error ε_{50} of each point is 0.8%. This recording took about 3 hours of completely automatic operation.

In order to understand the conditions involved in using the analogue, it is important first to consider the limiting case of the measurements.

Let the speed of the sliding contact, i.e. the rate k of increase of R, be given. The distance the sliding contact has travelled when N counts have been accumulated, will be larger the smaller the intensity (counting rate n) which is being measured. In fact, from eq. (1) and (3) it follows that

$$R = k\frac{N}{n}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

The *minimum* measurable counting rate, $n_{\min}$, will be that for which the whole length of wire, $R = R_{\max}$, is used; the corresponding time will be largest:

$$n_{\min} = \frac{N}{t_{\max}} = \frac{Nk}{R_{\max}} \quad . \quad . \quad . \quad . \quad . \quad (5)$$

On the other hand, in order to protect the recorder, the current fed to it must not exceed a certain permissible value, $I_{\max}$. The *maximum* recordable counting rate, $n_{\max}$, will be the one for which the fixed number of counts N is accumulated in such a short time ($t_{\min}$) that the used portion of the slide-wire will just have attained the minimum resistance value, $R_{\min}$, required to keep the current to $I_{\max}$. Thus:

$$n_{\max} = \frac{Nk}{R_{\min}} = \frac{Nk\,I_{\max}}{E}. \quad . \quad . \quad . \quad (6)$$

From (5) and (6) it is seen that the ratio

$$\frac{n_{\max}}{n_{\min}} = \frac{I_{\max}\,R_{\max}}{E}$$

is a design constant and does not depend on the values of k and N. In the actual instrument $n_{\max}/n_{\min} = 118$.

It should, of course, be possible to adjust the range ($n_{\max}$) of the instrument, and this is readily done by varying k, the speed of the sliding contact: with increased k, the total resistance of the slide wire, $R_{\max}$, according to eq. (5) will correspond to a higher counting rate $n_{\min}$. Owing to the constant ratio $n_{\max}/n_{\min}$, the range $n_{\max}$ will then be higher, too.

If N is increased in order to improve the accuracy of the measurements (i.e. to reduce the statistical error in counting, cf. II), $n_{\max}$ and $n_{\min}$ are increased (see eqs. 5 and 6). It will then be necessary to decrease k in the same proportion if the range ($n_{\max}$) is to be kept constant.

It was stated above that the resistance of the slide-wire must have attained a certain minimum value, $R_{\min}$, before the voltage E can be applied to it and the current flowing through it can be fed to the recorder. A way to ensure this would be to enforce a minimum time, $t_{\min}$, during which counting must be effected and the sliding contact moved on the wire. Since however no intermediate positions of the contact in this part of its travel would ever be used, it is simpler to substitute a resistor with fixed resistance $R_{\min}$ for this part of the slide-wire. The sliding contact (or rather its driving mechanism) whose movement is started together with the counting, should then travel a distance $\Delta = R_{\min}/R_1$ before it reaches the slide-wire and starts to move

on it; R_1 denotes the resistance per unit length of the wire. It will be noted that the "pre-travel" Δ is not affected by a change of k, whereas the time t_{min} necessary for traversing it clearly is: $t_{min} = R_{min}/k$ (cf. eq. 3).

Mechanical design of the computer

In the actual instrument, a 15 turn helical slide-wire ("Helipot") is used, along which the sliding contact is moved at a constant angular speed. *Fig. 3* shows a schematic drawing of the complete

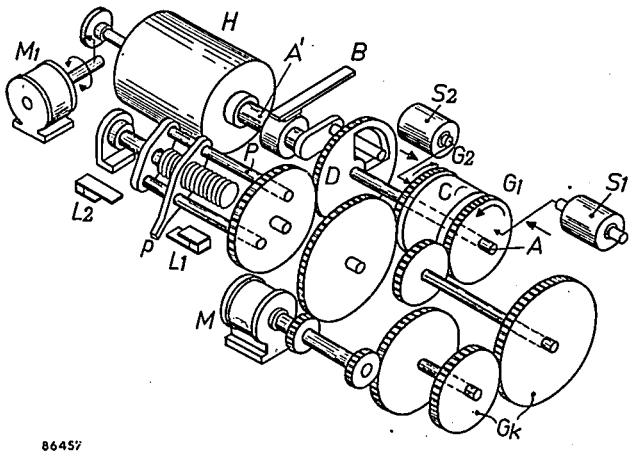


Fig. 3. Simplified drawing of the mechanical assembly. The electric motor M runs continuously, driving the gears G_1 and G_2, idling on shaft A in opposite directions. Shaft A carrying the double-face clutch drum C can be coupled either to G_1 or G_2 by the action of solenoids S_1 and S_2. H is the helical slide-wire resistor ("Helipot"), whose shaft A' is driven by A via an angular pre-travel device D; B friction brake, M_1 motor for the retractable contact mechanism (see fig. 5). P positioning device with limit switches L_1, L_2 actuated by arm p. G_k exchangeable range gears controlling the speed of movement of the sliding contact of H during the counting interval.

mechanical assembly. The 60 r.p.m. synchronous motor M turns continuously and drives the gears G_1 and G_2 idling on shaft A in opposite directions. The "Helipot" H can be coupled to either of these gears by means of the double friction clutch assembly C actuated by solenoids S_1 and S_2 respectively. Gear G_2 is used to reset the slide-wire contact to its starting position after each measurement; it turns at a constant speed of $\frac{1}{2}$ rev/sec so that the resetting will take 30 seconds if the complete length of the 15-turn "Helipot" was used (n_{min}). Gear G_1 is used to drive the slide-wire contact during the counting period; it can turn at different speeds, dependent on the selected range gears G_k, providing for different rates of change (k) of the resistance. Three pairs of range-gears are available for the computer, permitting the selection of five gear ratios and, hence, five ranges; for a scaling factor of 256 ($N = 25\,600$) these correspond to $n_{max} = 200$, 400, 800, 1600 and 3200 counts/sec.

Gear G_1 is coupled to the shaft A by the clutch at the moment when counting is started. Shaft A, however, will entrain the "Helipot" shaft A' only after it has rotated through a certain constant angle α, owing to the interposed angular pre-travel device D. The pre-travel angle α is the equivalent of the distance Δ introduced at the end of the preceding section.

The interposition of the pre-travel device has a curious consequence. In resetting the slide-wire contact, the resetting movement must be stopped by de-energizing the solenoid S_2 at the exact moment the contact has reached its initial position. It would seem natural to ensure this by arranging for a limit switch to be operated by the contact itself. Provision must be made, however, for the possibility of the counting rate being so high that in the counting operation the solenoid S_1 is de-energized and the shaft A uncoupled before the pre-travel angle has been traversed. The sliding contact has then not moved at all (the recorded intensity remaining at the top of the range) the and the limits witch could not be operated. Thus, no resetting of the shaft A would occur and in the *following* counting operation the pre-travel angle would be smaller than the required value.

For this reason — and for other reasons which cannot be discussed here — a separate positioning device P is provided which is rigidly and permanently coupled to shaft A by a pair of gears, see fig. 3. The arm p of this device describes a helical path exactly similar to that of the sliding contact of the "Helipot", except for the fact that it is longer by an angle α (distance Δ). On resetting the "Helipot", the arm p finally comes to rest on a stop assuring the accurate starting position. When touching the stop, the arm actuates a switch which disengages the reset clutch. In order to prevent the arm from recoiling from the stop, the switch turns on a time delay relay arranged to keep the reset clutch closed for about another $^1/_2$ sec: the clutch slips momentarily and the recoil energy is absorbed by the clutch and the motor M. The arm is then kept at the stop by a spring.

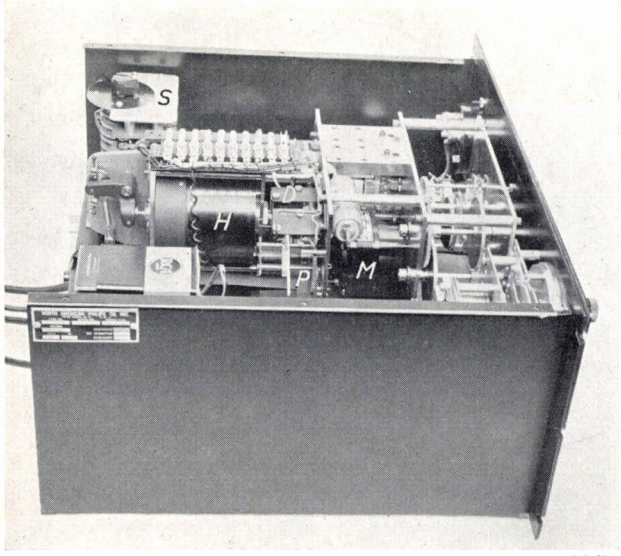
Immediately after resetting, the "Helipot" is ready for another measurement (see below). However, the synchronous motor M will have dropped back in phase a little bit (about $^1/_{15}$ of a cycle) owing to the extra load caused by the slipping clutch, and this would endanger the exact synchronism between counting time and rotating time of the "Helipot" (eq. 1 and 3). The limit switch of the positioning device therefore operates another time delay relay which prevents the starting of another counting

period for about 1 sec, by which time the motor will have caught up.

Throughout the resetting operation, the pin in the angular pre-travel device remains in contact with the driving member, and it is prevented by a friction brake B from leaving it when the resetting is completed. Thus the shaft A always has to travel through the complete pre-travel angle before starting the forward movement of the "Helipot".

A limit switch at the far end of the positioning device prevents the sliding contact of the "Helipot" from overshooting its limiting position at counting rates lower than n_{min}.

Besides controlling the counting operation and resetting, starting and stopping the "Helipot" at the correct moments with respect to the beginning and the end of a counting interval, the computer performs several other operations after the counting is completed: it feeds the final "Helipot" current to the recorder, at the same time it energizes the chart drive of the recorder to advance the chart a fixed increment and it advances the detector tube arm of the goniometer a preset 2Θ-increment (cf. I). It then switches off the "Helipot" current and the recorder chart motor and energizes the reset solenoid. All these switching operations are controlled by means of a multiple-deck motor-driven cam switch (cycling switch), which can be seen in *fig. 4* and which completes one cycle in 60 sec (recycling time, which is the lower limit for the duration of a single measurement).



86162

Fig. 4. Computer chassis. A number of elements shown in fig. 3 are visible; the lettering is identical to that of fig. 3. To the extreme left is a multiple-deck motor-driven cam switch S which controls each cycle of the operation of the computer.

Reproducibility of measurements

A number of precautions to ensure reproducible measurements have already been mentioned in the preceding section. They are aimed at maintaining the exact value of the pre-travel angle a and at making the time intervals t in eq. (1) and (3) strictly identical.

To conform with the latter condition, reliable operation of the friction clutch assembly, with a minimum and constant time lag between the energizing of the solenoids and the movement of shaft A, is of crucial importance. This was obtained by a careful construction of the linkage between the solenoids and the clutch faces. These faces are simply formed by the sides of the gears facing the drum. During axial movement of the gears, the teeth of the gears act as splines. Solenoid armatures and other moving parts are made as light as possible. The drum is faced on each end with a "Neoprene"-cork composition to assure positive driving action and a very rapid release. For the same reason the clearance is reduced to about $^1/_{64}''$ (0.4 mm); the motion is little more than enough to release the pressure. The shaft A on which the drum is fastened must, of course, have no noticeable end play.

Backlash in the gears should be very small, lest the slight shock from the solenoids should cause erratic operation. On the other hand, the gears must not be meshed too tightly causing them to move non-uniformly. A total backlash of about 0.12° is permitted, since the helical slide-wire is made up of a coiled resistance wire, so that the resistance changes in small but finite steps; each turn of the wire (having a resistance of 3 ohms) corresponds to about 0.003", which, on a 3" helix diameter, represents 0.12°.

Another important design detail affecting the reproducibility of the measurements is the sliding contact of the "Helipot". Although the contact and brush assembly normally supplied in the "Helipot" is satisfactory for many purposes, it was found to be too erratic for our application, especially in recording counting rates near the top of the selected range. In fact, our value of R_{min}, the fixed resistance in series with the slide-wire, is 869 ohms and the 15 turns of the slide-wire have a resistance of 6780 ohms each, so that the first turn of the helix will cover the large range from n_{max} ($= Nk/869$, eq. 4) to 12.8% of n_{max} (viz. $Nk/(869 + 6780)$). In this large range one simple coil step (3 ohms) of the slide-wire near the top corresponds to a change of 0.34%. It is therefore important to record the current at the exact position of the sliding contact, which should not be off by more than one coil step, i.e.

0.003″ (0.08 mm). Now, the contact will pass the first part of the slide-wire twice in every single measurement so that excessive wear of this part, which should have the most precisely constant resistance value per unit of length, is a real danger.

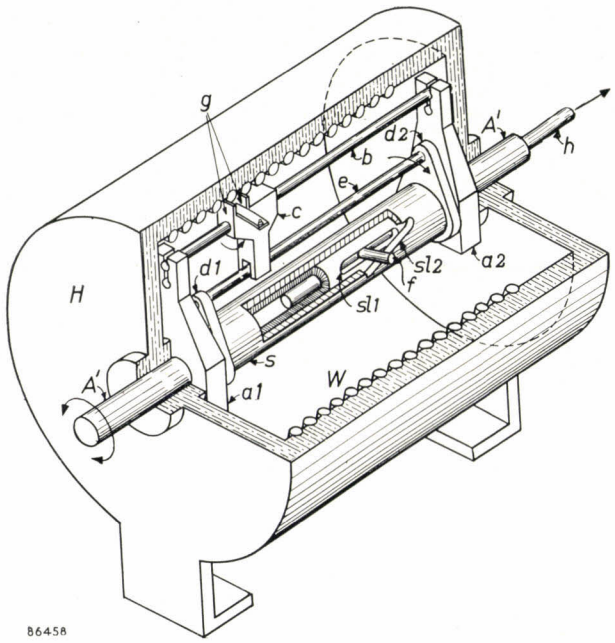


Fig. 5. Retractable contact assembly developed for the "Helipot". The helical slide-wire W is secured to the inner wall of stationary cylinder H. The hollow shaft A' rotating within this cylinder, carrying the arms a_1 and a_2 and tie-rod b, is entrained by the main shaft A of the computer as indicated in fig. 3. The "Bakelite" guides g, gliding between adjacent windings of the slide-wire, cause the contact brush holder c to traverse rod b and thus to describe a helical path parallel to the winding. The contact will touch the wire, however, only when the contact holder is slightly rotated on the tie-rod in the direction of the arrow. Such a rotation can be brought about by a slight rotation of sleeve s carrying tie-rod e secured by arms d_1 and d_2. Shaft A' and sleeve s each have a slot, sl_1 and sl_2 respectively. A pin f engaging both slots locks shaft and sleeve so that they rotate together with a certain "phase angle" between the rods b and e which depends on the position of the pin, since slot sl_2 is inclined with respect to sl_1. The pin f is carried by the auxiliary shaft h, rotating with shaft A' and sleeve s. The relative angular position of tie-rods b and e is thus controlled by the axial position of the shaft h; a movement in the direction of the arrow causes the necessary relative rotation of sleeve s to lower the contact onto the slide-wire. During such a movement, shaft A' remains stationary owing to the friction brake B of fig. 3. The axial movement of shaft h is effected by a small electric motor (M_1 in fig. 3) controlled by the cycling switch mentioned in the text (S in fig. 4).

For these reasons the contact shoe and brush assembly of the "Helipot were modified in such a manner that the contact is lifted away from the wire whenever the shoe is in motion [4]). When it is required to record the result of a measurement the contact is brought down on the wire automatically but only just long enough to complete the recording.

A cut-away view of the "Helipot" showing the inside construction with this retractable contact mechanism is given in *fig. 5*. The contact is lifted by means of a small motor (M, in fig. 3) which is controlled by the cycling switch and provides smooth reproducible action. The friction brake mentioned earlier (B in fig. 3) keeps the "Helipot" shaft from rotating while the contact is being actuated.

Performance

The over-all degree of reproducibility obtained is illustrated by *fig. 6*. The right hand part shows a recording made with the power line frequency of 60 c/s with rate k and scale factor chosen to make the full scale 100 counts/sec. There are a few steps in the line but it should be remembered that the reproducibility of the scaling circuit and associated

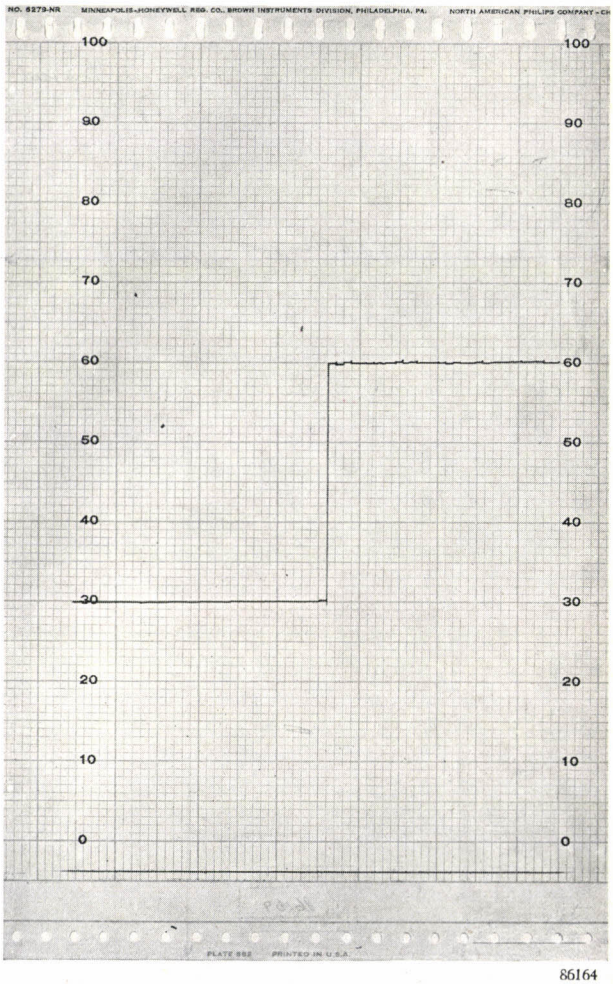


Fig. 6. In order to demonstrate the degree of reproducibility obtained, a recording of a constant counting rate (power line frequency of 60 c/s) is made. At a full scale range of 100 counts/sec (right-hand line), the recording shows a few steps. At a larger full scale range, say of 200 counts/sec (left-hand line), no steps appear, owing to the fact that for a counting rate of 60 c/s a larger part of the slide-wire is now used, so that the discrete turns of the coiled slide-wire, which determine the smallest detectable difference in counting rate, have less effect.

[4]) E. A. Hamacher and K. Lowitzsch, U.S. Patent No. 2 658 131, Nov. 3, 1953.

circuits and relays are also included in the test. Because of the smaller relative influence of the discrete turns of the coiled slide-wire at the lower portions of the scale (see above), the reproducibility is even better there: no steps appear in a recording of the 60 c/s frequency at 200 counts/sec full scale (fig. 6, left).

It will be evident from the section on the principle of the computer that deviations from the desired linear relationship between the recorded current and the counting rate will occur when the fixed resistance R_{min} and the pre-travel Δ (or pre-travel angle α) are not adjusted exactly to their (interrelated) theoretical values. Another cause of non-linearity is due to R_l, the resistance of the slide-wire per unit length, not being strictly uniform throughout. The specified tolerance for the commercially available 15 turn "Helipot" is 0.1%, i.e. 100 ohms out of a total of about 100 000 ohms. If this total deviation were to occur in the first turn, covering the range from 12% to full scale, a maximum departure of 1.5% from linearity would be found in the response of the computer in this range. Actual experience shows that the maximum departure is always less than 1%. Because of the importance of the first turn of the "Helipot", the average value of R_l in this part should be used for the calculation of the pre-travel $\Delta = R_{min}/R_l$.

Application of the Cooke-Yarborough method to the computer

The fixed count method will require very long counting times for low counting rates. Although this is a logical consequence of the condition of constant statistical relative error for all intensities, a compromise between accuracy and time consumption will often be desirable. As was explained in article II, a good compromise can be obtained by adding pulses of a *constant* controlled rate, produced by a separate pulse-generator, to the random pulses coming from the detector tube (Cooke-Yarborough method). Provisions have been made for using this method with the counting-rate computer. The changes required are quite simple and may be outlined as follows:

Summarizing the basic idea of the computer we may say that a counting rate n will give rise to a proportional current I; n_{max} and I_{max} constitute the limiting case. Owing to this proportionality, a constant counting rate f added to n will give rise to a constant additional current $I_f = If/n = I_{max} f/n_{max}$. If, therefore, we subtract this constant current I_f from I_{max}, the remaining current will again be proportional to n and can be recorded as such. The only difference from the normal procedure will be that the uppermost part of the scale on the chart, viz. from n_{max} down to n_{max} $(1 - f/n_{max})$, is not used.

In order to make the used part of the scale the same for all selected ranges n_{max} (giving an identical calibration of 0-100 for this part for all ranges), it is necessary to make f/n_{max} a constant fraction. In the present instrument this fraction is made equal to 0.075: this permits the generator of the added pulses to be synchronized with the power line for n_{max} = 800 counts/sec, since f will then be equal to 60 c/s. For other possible values of n_{max}, f will be equal to harmonics or sub-harmonics of 60 c/s. The gain in time for low counting rates is quite consid erable with these added pulse rates, while the statistical relative error is not prohibitively increased. This was illustrated by fig. 15 and 16 in article II.

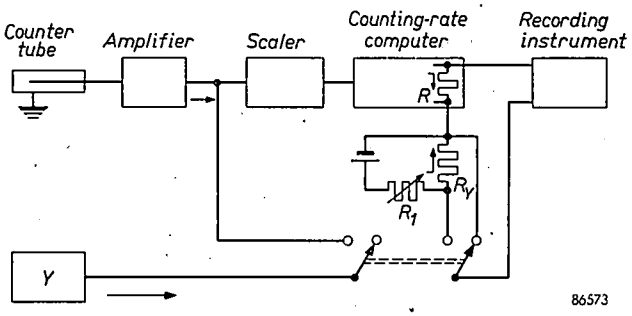


Fig. 7. Block diagram showing the provisions made for the application of the Cooke-Yarborough method to the computer. The pulse generator Y is connected to the input of the scaling circuits in parallel to the detector. At the same time a resistor R carrying a constant current (adjusted once and for all by means of a rheostat R_1) is connected in series with the resistor R through which the output current of the computer is fed. The voltage across $R + R_Y$ is fed to the recorder.

The conversion from normal operation of the computer to Cooke-Yarborough operation involves simply throwing a switch. The output of the pulse generator is thereby connected to the negative input terminal of the scaling circuits, in parallel with the pulses from the detector tube, and at the same time a circuit effecting the subtraction of current $I_f = 0.075 \times I_{max}$ from the computer output current is switched on. This is shown in the block diagram *fig. 7*.

Summary. In measuring radiation intensities by counting quanta, the statistical relative error is reduced to a predetermined value by accumulating a fixed number of counts, N. The necessary time t is measured and the counting rate (which is proportional to the intensity) computed from $n = N/t$. Application of this method with the "Norelco" X-ray diffractometer for measuring diffracted intensities at narrow-spaced diffraction angle positions (which may involve many hundreds of measurements) is bound to make excessive demands in the operator's time, unless the process can be made automatic. This has been achieved by the development of a computer, evaluating n from the above equation by means of an analogue and recording it on a strip chart with a linear scale. The analogue used is Ohm's law: the contact of a slide-wire is moved at a constant rate for the time t necessary to accumulate a pre-set number of counts N; at the end of this time, a fixed voltage is applied across the slide-wire and the current which flows is fed to the recorder. The range can be adjusted by changing the speed of movement of the contact. In the mechanical design, many precautions were necessary to ensure reproducible operation: the double clutch drive for operating and resetting the sliding contact is designed to start and stop the movement at the exact moments required; a device is incorporated for resetting the contact to its precise starting position; in addition, a special retracting mechanism for the contact is incorporated to prevent excessive wear of the slide-wire. Reproducibility to within one turn of the coiled slide-wire is obtained, which corresponds to an accuracy of 0.34% at the upper end and much better accuracy at the middle and lower parts of the scale. Provisions are also made for applying the Cooke-Yarborough modification of the fixed count method with the "Norelco" computer.

AN ELECTRICAL CLINICAL THERMOMETER

621.317.39:536.531:616-073.65

Mercury clinical thermometers require a period of the order of 3 to 10 minutes before a reliable reading can be taken. There are types available of reduced size, which permit the reading time to be brought down to 1 minute. This reduction in size, however, makes reading as well as "shaking down" more difficult.

The long reading time is caused not only by the high thermal resistance between the body tissue and the mercury, but also by the considerable heat capacity of the thermometer. This heat capacity causes the tissue in contact with the thermometer to drop in temperature. It takes a considerable time for the temperature to be restored by the blood circulation.

The above disadvantages are almost entirely overcome in the electrical resistance thermometer [1]) shown in *fig. 1*, which has a reading time of only 13 seconds. The thermometer consists of a probe connected by a flex to a small meter easily held in the hand. The temperature-sensitive element in the probe is a ceramic semi-conducting material with a high negative temperature coefficient of resistance [2]) (NTC resistor).

The body of the resistor is oval in shape; the longitudinal axis is approximately 1 mm long and the diameter is approximately 0.4 mm. Two connecting wires are baked into the body of the resistance parallel to the longitudinal axis. The resistance measured between these two wires is a measure of the temperature. The outside of the NTC resistor is glazed to protect it from the influence of the atmosphere; this makes these resistors very stable, a matter of primary importance for the end in view. By a careful choice of composition and treatment of this ceramic material, its resistance at room temperature can be varied within wide limits.

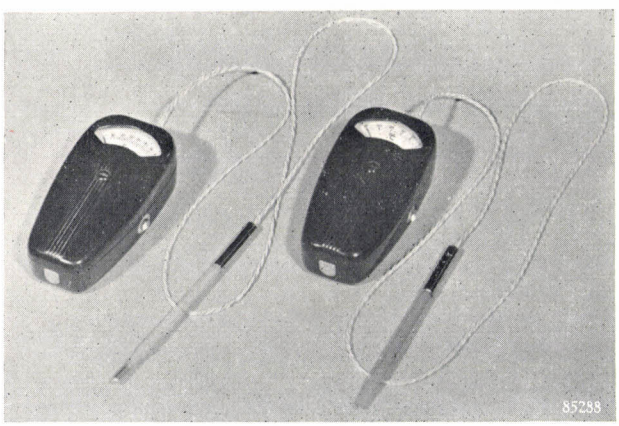


Fig. 1. Left, the electrical clinical thermometer, consisting of a probe and an indicating instrument, connected by a flex. The button visible on the right of the instrument is pressed in order to take a reading. The reading may be taken about 13 seconds after insertion of the probe. Right, the model for skin thermometry.

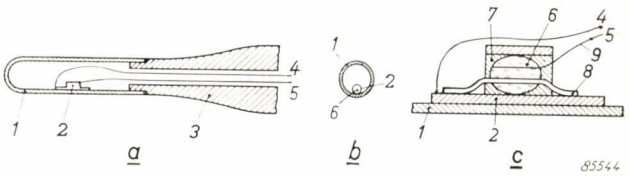


Fig. 2. *a*) Longitudinal cross-section of the probe of the clinical thermometer. *1* silver "bulb", a closed tube of the same form and dimensions as the bulb of a mercury clinical thermometer. *2* small silver cylinder soldered to the interior of the bulb and containing the NTC resistor. *3* hollow holder of nylon. *4* and *5* connecting wires.
b) Transverse cross-section of the bulb at the position of the small cylinder *2*. *6* is the NTC resistor.
c) Longitudinal cross-section of the silver cylinder *2*. *6* NTC resistor, *7* "Araldite" bonding resin. *8* thick wire of pure platinum, both ends being soldered to the lips provided on the cylinder *2*. *9* thin wire of platinum-iridium.

The end of the probe (*fig. 2*) consists of a small silver tube whose external dimensions are the same as those of the mercury bulb of a normal clinical thermometer. A small silver cylinder is soldered against the inner wall of this "bulb" and the closely fitting NTC resistor is cemented in it with "Araldite". There is no electrical contact between the cylinder and the outside of the NTC resistor but there is a good thermal contact; heat absorbed through the "bulb" from the body tissue is thus transferred to the NTC resistor via the circumference of the inner silver cylinder. In addition to this, the heat can flow directly into the interior of the NTC resistor via one of the wires (*8* in fig. 2*c*), which is therefore made comparatively thick (100 μ) and of pure platinum (like silver, an excellent conductor of heat). Both ends of this wire are soldered to the small lips on the cylinder surrounding the NTC resistor.

Clearly a current must flow through the NTC resistor in order that the measurement may be carried out. Owing to the good thermal contact between the NTC resistor and the silver "bulb", the heat generated in the former by this measuring current is rapidly distributed over the latter. This is most important: the heat generated causes a

[1]) The thermometers discussed in this article are at present used only for experimental purposes and are not generally available.

[2]) E. J. W. Verwey, P. W. Haayman and F. C. Romeyn, Semi-conductors with a high negative temperature coefficient of resistance, Philips tech. Rev. **9**, 239-248, 1947/48.

rise in temperature in the NTC resistor and consequently an error in measurement, which may be considerable unless the heat is rapidly dispersed. Tests have shown that in maintaining this error below an acceptable value, more heat may be dissipated in the NTC resistor with a thermometer of this construction then when the resistor is encased in a glass tube. This means that a higher current may be used, which makes the instrument more sensitive. The maximum permissible energy dissipation in the NTC resistor with the present design is 0.25 mW.

In order that the temperature of the object being measured shall be affected as little as possible by the measurement, conduction of heat from the object via the probe to the surroundings must be as small as possible. With this in mind, the holder to which the silver bulb is attached is made of nylon, which is a good heat insulator (much better than glass, for example). In addition, the surface of the bulb in contact with the nylon holder is made small with respect to the surface in contact with the body. Loss of heat via the copper connecting wires which link the NTC resistor to the meter is kept low by making these wires very thin. The error caused by heat loss via the probe and the error caused by heat generation in the NTC resistor as a result of the measuring current are opposed to each other and tend to cancel each other.

Yet a third source of error should be taken into consideration: this is operative even when the total dissipation of heat via the probe is so small that it does not noticeably affect the body temperature being measured. The silver bulb then has exactly the right temperature. Now the heat conducted away by the connecting lead *5* (fig. 2*c*) has to pass through the NTC resistor itself. Owing to the considerable thermal resistance of the latter, a temperature drop is set up across it, which leads to an error in measurement. This error can be reduced by increasing the thermal resistance of the conduction path beyond the NTC resistor. With this in mind, the wire *9* is very thin and is made of platinum-iridium whose thermal conductivity is about $^1/_3$ that of pure platinum. When these measures are taken, the heat resistance per cm length of the thin wire *9* is about 25 times as great as that of the thick wire *8*. Heat conduction along connecting wire *4* which is soldered to the small silver cylinder *2* (and thus electrically connected to wire *8*) and heat losses through the holder both bypass the NTC resistor, and do not therefore contribute errors of this nature.

The NTC resistor forms one of the branches of a Wheatstone bridge. The connecting leads run through the hollow holder and are joined via a flex to the measuring instrument which contains the other three resistance branches, the indicating meter and the supply battery. The bridge is not balanced for each reading, i.e. the instrument is not operated on the null principle: this would take up too much time. Instead, the resistances are so chosen that there is no current flowing through the meter (i.e. the bridge is in balance) when the temperature of the NTC resistor corresponds to the lowest value shown on the scale (35 °C). At higher temperatures the meter shows a deflection due to the out-of-balance current then flowing. The current flowing at temperratures lower than 35 °C causes the needle to deflect against a stop.

The battery is connected only during a measurement, by depressing a spring-button (fig. 1). Since the instrument can conveniently be read with the probe actually in position and almost immediately after its insertion, it is unnecessary to provide a maximum-reading device in the meter.

The meter is calibrated in °C, and its range runs from 35 to 42 °C, with a scale length of 3 cm.

The sensitivity of the meter circuit, i.e. the deflection of the meter per °C temperature variation in the NTC, depends upon the terminal voltage of the battery, which must necessarily be very stable in order to maintain a constant calibration. The mercury-oxide cell used meets this requirement provided that it is used only intermittently and briefly — as is the case for such temperature measurements — and provided that only a small current drain is involved [3]). The instrument includes a provision for a simple check on the battery voltage. With the voltage supplied by this type of cell (1.34 V), and the above-mentioned maximum permissible dissipation in the NTC resistor (0.25 mW) the values for the bridge-resistances are chosen so as to give maximum bridge sensitivity, i.e. maximum meter deflection per °C. Under these conditions, the NTC resistor has a value of the order of 1000 Ω and the total current drawn from the battery is small enough to be provided by a mercury-oxide cell of the smallest obtainable type. The indicating instrument is a moving-coil meter with a phosphor-bronze strip suspension; this gives a very rugged instrument which can withstand the shocks that a hand instrument may be expected to suffer. The temperature can be easily read to 0.1 °C which is the accuracy laid down in English and American standards.

In some cases it is important for the physician to be able to measure the temperature of the skin, e.g. in cases of disturbed blood-circulation conditions.

[3]) See, for example, G. W. Vinal, Primary batteries, J. Wiley, New York 1950.

For this application (measuring range 15 to 42 °C) it was necessary to change the form of the sensitive element to that shown in *fig. 3*. The skin thermometer is also shown in the photograph of fig. 1. With this larger measuring range, the fact that the resistance of the NTC resistor is not a linear function of the

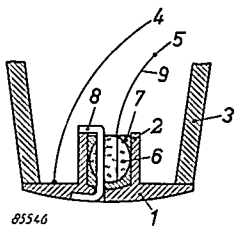


Fig. 3. Cross-section of the sensitive element of a skin thermometer. The element is designed so as to give good thermal contact with the skin. The numbers in the diagram have the same significance as in fig. 2. The components *1* and *2*, however, are now turned as a single piece from a rod of silver.

temperature becomes important; the resistance variation per degree temperature change decreases rapidly as the temperature rises (*fig. 4*). Fortunately, however, the sensitivity of the bridge for resistance variations of the NTC resistor increases with the increasing deviation from equilibrium of the bridge. A closer analysis of the properties of the bridge has shown that it is possible to design the bridge such that these two non-linear effects practically compensate each other, so that the relationship between the temperature and the bridge current is linear within an accuracy of 0.2 °C. The resistance values then used in the bridge for the skin thermometer (range 15-42 °C) happen at the same time to give maximum sensitivity.

A linear relationship offers considerable advantages. Not only is the absolute reading accuracy constant within the whole range but it is also possible to use a standard printed linear scale in the meter. If the meter is adjusted to give correct readings at two points, it will then be correct over the whole scale. For one of these temperatures, viz. the lowest scale reading, the bridge is balanced by adjusting the resistance in one of the arms; for the second temperature, the sensitivity of the bridge can be adjusted by connecting a small resistance in series with the meter.

The skin thermometer described here may also be used as a clinical thermometer if used per rectum (if the thermometer is used e.g. in the arm-pit, the design of fig. 3 is not such as to always ensure the direct contact of the sensitive element with the skin). Thus, by providing the instrument with two ranges, selected by a switch, it is possible to combine the skin thermometer and the clinical thermometer in one instrument. This is a doubly attractive proposition since the construction shown in fig. *3* is simpler and thus cheaper than that shown in fig. 2.

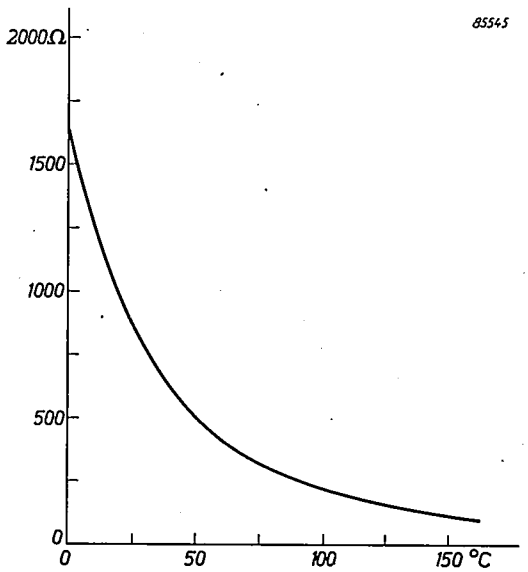


Fig. 4. The resistance of an NTC resistor as a function of the temperature.

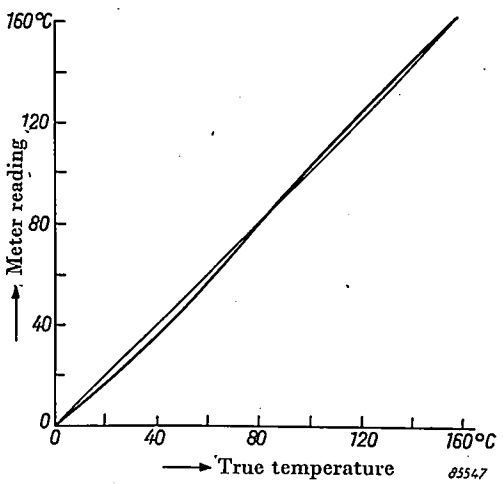


Fig. 5. Graph showing the slight deviation from a linear scale of an electrical thermometer with a range of 0-160° C. The temperature coefficient of the NTC resistor varies by a factor of 20 in this temperature range.

It is possible also to construct non-clinical thermometers with much greater ranges (e.g. 0 to 160 °C) based on the same principles. It is then no longer possible to keep the relationship between temperature and meter deflection exactly linear, so that a calibrated scale is necessary. It is possible, however, to choose the bridge resistances so that the deviation from linearity is reasonably small, so as to retain the advantage of an almost constant absolute reading accuracy (see *fig. 5*). Possible applications for such a thermometer include: rapid checking of the temperature of resistors in electrical apparatus during operation, and the temperatures of bearings, carbon brushes, transformer cores and electrode lead-in wires in electronic tubes.

C. C. J. ADDINK.

STEREO REVERBERATION

by R. VERMEULEN.

534.844:534.846.4:621.395.625.3

While too long a reverberation time makes speech unintelligible, a reverberation time which is too short makes music sound "dry" and brittle. Many varieties of acoustic materials are available for shortening the reverberation time and improving the intelligibility. Lengthening of the reverberation time and — perhaps more important — making the sound diffuse, can be achieved by electro-acoustic means. Tests have shown that in this way a good theatre hall can be made suitable for concerts.

Introduction

The acoustic properties of a hall are determined by the behaviour of the sound waves, in particular those which are reflected from the walls. The importance of the reflected sound can be well understood if one has listened to an open-air speech made without an amplifier installation, or if one imagines an open-air performance by a symphony orchestra without this aid.

A simple calculation will show that in a hall at even quite a small distance from the orchestra, the direct sound can sometimes be weaker than the reflected sound. At a distance r from a sound source of power P, the energy density of the direct sound is equal to $P/(4\pi r^2 c)$, where c is the velocity of sound. In a hall with a volume V and reverberation time T (defined as the time in which the sound intensity decreases by 60 dB, after the source has ceased radiating), the energy density of the indirect sound [1]) is $PT/(13.8\ V)$. These two energy densities are equal at a distance $r_0 = (1.1\ V/cT)^{\frac{1}{2}}$. Values of r_0 corresponding to various practical values of V and T are as follows:

$V =$	100	1000	10 000 m³
$T =$	0.7	1.0	1.5 s
$r_0 =$	0.7	1.8	4.5 m.

At distances greater than r_0 the indirect sound predominates. It is seen that this can occur at distances of only a few metres.

The qualities required of a hall for speech and for music are quite different. In a theatre, the intelligibility is of primary importance. If speech is to be clearly understood, the reflected sound must reach the audience with so little delay that it reinforces the direct sound but does not overlap the sounds which follow. For the latter, the persistence of the preceeding sound must be regarded as "background noise", which adversely affects intelligibility. As a rough guide, one can say that all sound which reaches the audience within 50 milliseconds can be regarded as useful sound [2]). Erwin Meyer has formulated the idea of "clearness" or "definition" [3]), which he defines as follows:

$$\frac{\int_0^{50\ \text{msec}} p^2(t)\ \mathrm{d}t}{\int_0^{t_1} p^2(t)\ dt},$$

where p is the sound pressure, t is the time (measured from the moment at which the source is silenced) and t_1 is a time much greater than 50 msec.

For a concert hall, on the other hand, the first requirement is not intelligibility, but a fine, full tone. Here it is much more difficult to specify the requirements. For speech, the reverberation must be accepted as an inevitable, disturbing accompaniment to the useful sound, because it simply is not possible to silence the sound suddenly after 50 msec. For music we know that the reverberation time not only may be, but must be, longer. The optimum value is clearly dependent on the nature of the music. It is often seen that a composer has consciously taken into account the acoustics of the space (church, concert hall, room) where he wanted his music to be played. The inclination to sing in the bathroom can probably be largely attributed to the long reverberation time of this acoustically "hard" room.

It is becoming increasingly clear that the reverberation time is not the only property governing the suitability of a hall for musical performances.

1) See for example A. Th. van Urk, Auditorium acoustics and reverberation, Philips tech. Rev. **3**, 65-73, 1938.

2) See the article referred to in [1]), pp. 72-73.

3) E. Meyer, Definition and diffusion in rooms, J. Acoust. Soc. Amer. **26**, 634, Sept. 1954.

One might even conjecture that here, too, reverberation is merely an inevitable subsidiary effect. Just as important, or perhaps even more so, is the "diffuse-ness" of the sound (and possibly also the nature of the fluctuations of the reverberation).

To study these phenomena more precisely, we attempted to produce an artificial diffuse reverberation in the laboratory by means of distributed loudspeakers which repeated the music played, with controllable intensity and lag. This experiment appeared to improve the acoustics to such an extent that we ventured to take the bold step of using this artificial reverberation to make a theatre suitable for concerts. We propose that this artificial diffuse reverberation be called "stereo reverberation".

Our first installation was in the Philips Theatre at Eindhoven, whose acoustical properties as a theatre were very satisfactory as a result of rebuilding in 1935, but which left much to be desired as a concert hall. In addition to this theatre, a hall known as the "Gebouw voor Kunsten en Wetenschappen" (Arts and Sciences Hall) in The Hague, is now fitted with a permanent installation for stereo reverberation [4]).

Principle of the installation for stereo reverberation

The delay wheel

The principle of the stereo reverberation installation may be explained with the help of *fig. 1*. Controllable time lags are obtained by means of magnetic recording and playback. Magnetic material such as is used for magnetic tape is coated on the rim of a wheel ("delay wheel") [5]). The music is recorded on this material via a microphone and a recording head. A number of play-back heads — in the final apparatus four, in experimental types six (*fig. 2*) or more (*fig. 3*) — are mounted around the circumference of the wheel and connected via separate channels to loudspeakers. These are installed in various places in the hall: on the ceiling, along the balustrade of the balcony, in a lighting cornice, in "dead" corners under the balcony, etc. Between the last play-back head and

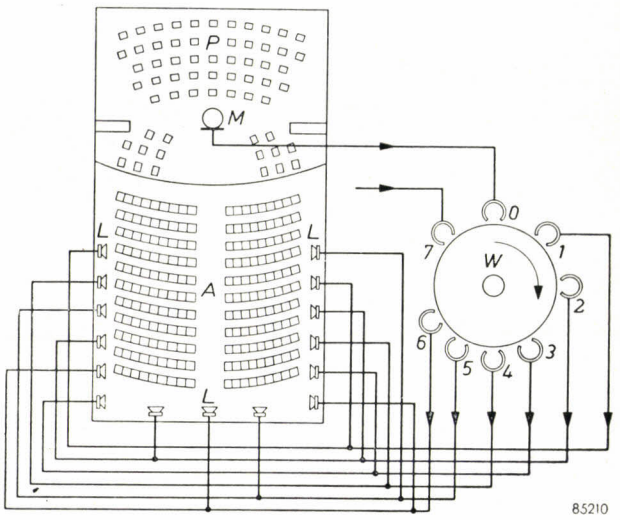


Fig. 1. Installation for simulating indirect sound with various time lags. *A* auditorium. *P* platform. *M* microphone. *W* delay wheel, coated on the edge with magnetic material suitable for magnetic sound recording. *0* recording head, *1...6* play-back heads. *7* erasing head. *L* loudspeakers.

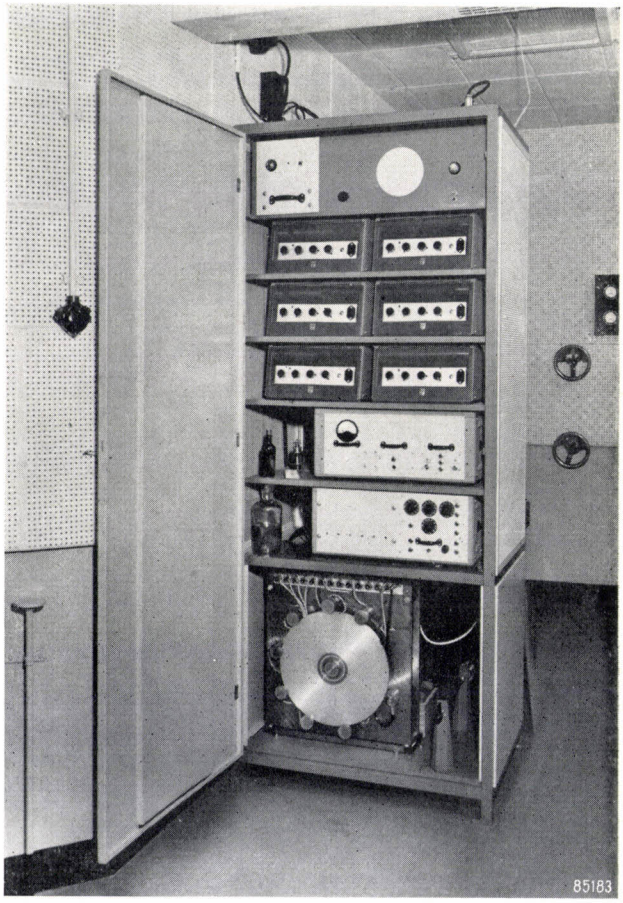


Fig. 2. Stereo reverberation installation in the Philips Theatre at Eindhoven. At the bottom of the cabinet is the delay wheel (cf. fig. 1); above it the amplifiers.

[4]) Demonstrations of stereo reverberation have also been given at the first I.C.A. Congress on Electro-acoustics, (Acustica **4**, 301, 1954), at Gravesano in Switzerland (at the invitation of the conductor W. Scherchen; see his book Musik, Raumgestaltung und Elektroakustik, Arsviva Verlag, Mainz 1955) and at the 3rd "Tonmeistertagung" of the Nordwestdeutsche Musikakademie Detmold (see D. Kleis, Elektron. Rdsch. **9**, 64-68, 1955 (No. 2)). Similar tests, but done in the open air, are described by H. S. Knowles, Acustica **4**, 80-82, 1954 (No. 1).

[5]) Others have also constructed a similar equipment, but have not created diffuse reverberation with it; see H. Schiesser, Einrichtungen zur Erzeugung künstlichen Nachhalls, Funk und Ton **8**, 361-368, 1954 (No. 7), and P. Axon and co-workers, Artificial reverberation, J. Instn. El. Engrs. **1**, 368-371, 1955 (No. 6).

the recording head is an erasing head, which ensures that the magnetic layer is blank on returning to the recording head.

For three reasons, however, the listener hears more than these seven (1 + 6) reports: firstly because to each play-back head, several loud-

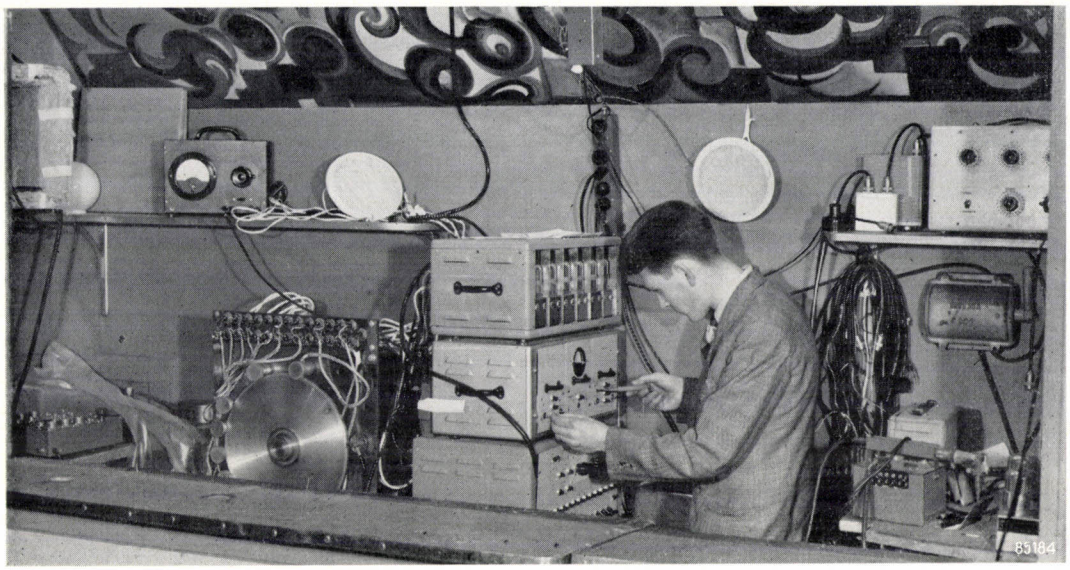


Fig. 3. Stereo reverberation installation in the "Gebouw voor Kunsten en Wetenschappen" (Arts and Sciences Hall), The Hague, in experimental form. The delay wheel was then fitted with ten play-back heads, of which six could operate simultaneously. It was possible to switch over rapidly from one set of play-back heads to another.

Let us consider the case of a sharp report produced in the hall (pulse I_0, *fig. 4*). After a certain transit time this reaches the microphone and then the artificial indirect sound begins. The latter, if we are using six play-back heads, consists in the first place of six successive reports from the loudspeakers ($I_1 \ldots I_6$). The intensity of each of these reports can be adjusted at will by the gain controls; the time intervals are also under control, by the spacing of the heads round the wheel and by its speed of revolution.

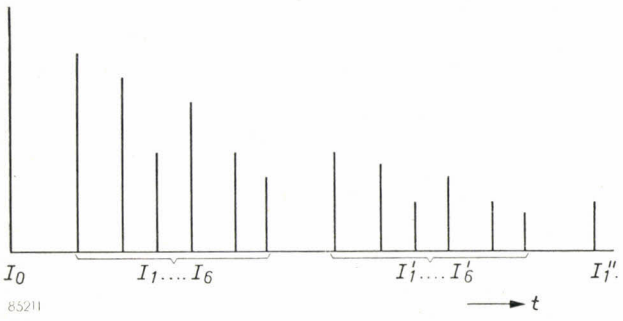


Fig. 4. If a report (pulse I_0) is produced in a hall with stereo reverberation, the six loudspeakers deliver during one revolution of the delay wheel the six reports $I_1 \ldots I_6$. If electrical feedback is applied from the sixth play-back head to the recording head, the loudspeakers give a second series of six reports ($I_1' \ldots I_6'$), a third, and so on. (Acoustic feedback is here neglected.)

speakers are connected, which are dispersed throughout the hall and are at different distances from the listener, so that the transit times are different; secondly because the walls reflect both the original sound and those coming from the loudspeakers, and thirdly because the report from each loudspeaker is picked up by the microphone, once more recorded on the wheel and, though attenuated, reproduced six-fold.

All these effects contribute to the fact that the original report is followed by a large number of others, so that the reverberation time can achieve considerable values. The third effect mentioned above — the feedback from the loudspeakers to the microphone — even involves the danger that one particular note continues to sound too long and in the extreme case howls back continuously; therefore this effect must be curtailed as much as possible (we shall return to this point presently). This acoustic feedback can be advantageously replaced by electrical feedback, in which a chosen fraction of the signal from the last play-back head is fed back to the recording head. This records the signal afresh: in our example, a second series of six reports then ensues ($I_1' \ldots I_6'$ in fig. 4); this series is once more recorded and a third series follows, and

so on. By adjusting the fraction of the output signal which is fed back, each series can be attenuated to any given degree, so that the reverberation time can be chosen at will.

Calculation of the effect of the stereo reverberation on the acoustic properties of the hall

The following calculation shows, for a simple case, the effect which stereo reverberation has on the acoustic properties of the hall. Suppose that the energy density in the hall is $E(t)$. Then $-V\mathrm{d}E/\mathrm{d}t$ is the rate at which acoustic energy in the hall diminishes. In a hall without stereo reverberation, this must be equal to the power absorbed by the walls [6]), which is proportional to E, say equal to aE. Stereo reverberation supplies extra power proportional to the energy density at some time τ ago: $\beta E(t-\tau)$. We can thus construct the equation:

$$-V\frac{\mathrm{d}E}{\mathrm{d}t} = aE(t) - \beta E(t-\tau).$$

A solution to this is:

$$E = E_0 \exp(-mt),$$

where

$$mV = a - \beta \exp(m\tau);$$

the quantitity m is inversely proportional to the reverberation time T, and equal to $13.8/T$. For small values of $m\tau$, an approximate value of $\exp(m\tau)$ is given by $1 + m\tau$. Thus we have:

$$m \approx \frac{a-\beta}{V+\beta\tau}.$$

From this it is seen that increasing the strength of the stereo reverberation (β) has the same effect as decreasing the absorption (a) or increasing the volume (V) of the hall. An increased lag τ also gives the effect of a hall of larger volume; we shall return later to this point.

The microphones

The above calculation was based on the assumption that one could speak of "the" energy density in the hall. In many cases the total energy (sum of potential and kinetic energy) is indeed fairly evenly distributed throughout the hall. With standing waves, however, as is well known, the potential and kinetic energy alternate with each other, and this means that the sound pressure at constant frequency changes sharply from place to place, and, conversely, that at a particular place the sound pressure is very dependent on the frequency. The microphone which picks up the signal which is fed back to the hall via the delay wheel and the loudspeakers, responds only to the pressure at the spot. The factor β in our calculation therefore varies sharply with the frequency, and the reverberation time, which is inversely proportional to $a-\beta$, varies also and to a much greater degree. If β as a function of the frequency shows a peak, then increasing the amplification will cause the note at which the peak occurs to go on sounding for a long time, while for most other frequencies, the reverberation time is not yet lengthened appreciably. With even greater amplification, the note fails to decay at all (howl-back: $a-\beta$ has become negative).

The obvious solution is to try and suppress the peak by a filter in the microphone channel. However, there are so many peaks in the frequency characteristic of a hall, and these peaks are so sharp, that the suppression of all of them would be a hopeless task, particularly since the peaks are modified by all changes made in the hall or on the platform. It is worthwhile, nevertheless, to attenuate those frequency ranges in which the highest peaks occur in such a way that increase of amplification causes a number of widely separated frequencies to howl back at the same time.

[6]) See for example the article referred to in [1]), where it is shown that the proportionality factor a is equal to $^1/_4 aAc$, where a is the average absorption coefficient and A the area of the walls.

In this connection it can be argued that electrical feedback via the delay wheel is a better means of obtaining a long reverberation time than acoustic feedback via the hall.

In a system with delayed feedback, the feedback signal and the input signal exhibit, in general, a relative phase difference which is proportional to the product of frequency and time lag. If there are a great number of feedback channels (as there invariably are in practice, with acoustical feedback), and if we suppose that they all make equal contributions to the input signal, then in general the contributions will show fairly random phase differences, so that the average resultant increases as the root of the number of feedback channels. There will be, however, one or more frequencies for which all the contributions are nearly in phase with each other. For these frequencies the resultant will be a maximum and equal to the algebraic sum of the contributions, and thus proportional to the number of feedback channels. This reasoning makes it clear that the ratio of the maximum value of the resultant (which occurs at a particular frequency and limits the maximum amplification) to the average value (at other frequencies) increases with the root of the number of feedback channels. In the case of a microphone and loudspeakers in the same hall, this number is very large and it is to be expected that a situation can easily arise in which one note continues to sound for a long time. With feedback via the delay wheel,

we are using only one feedback channel (from the last playback head to the recording head) and the above-mentioned danger is thus much less. It would be possible to introduce more feedback channels, e.g. from one or more of the preceeding playback heads to the recording head. Experiments have shown that this is undesirable, and this is partly explained by the above considerations.

Fig. 5. The stage in the "Gebouw voor Kunsten en Wetenschappen" in The Hague. M_1 is the line microphone (rod with ten condenser microphones). (Other microphones visible in the photograph were for broadcasting and had nothing to do with stereo reverberation.)

Instead of a microphone which responds to the average sound density in the hall, the other extreme would be one which responds exclusively to sound coming directly from the source and is insensitive to sound from the hall [7]). The danger of notes howling back would then be completely averted. This situation can be closely enough approximated to by using a microphone having a sharp directional characteristic. An arrangement for achieving this, a so-called "line microphone", consists of a group of ten condenser microphones mounted at equal intervals along a rod rather more than a metre in length (*fig. 5*). By suitably positioning this line microphone above the orchestra, we can ensure that the loudspeaker signal at the microphone, although still making an important contribution, no longer dominates. If it is not possible to cover the orchestra adequately with one line microphone, two may be used (*fig. 6*).

The use of one or more extra microphones can be desirable in order to strengthen weak instruments in the orchestra. Thus, for example, in the Philips Theatre at Eindhoven the organ was rather weak in comparison with the orchestra and choir in the annual performance of Bach's St Matthew Passion; it appeared to be an improvement when this instrument was boosted by a microphone of its own (*fig. 7*). With such cases in mind, the stereo reverberation system was provided with several microphone channels with mixing facilities. These should, however, be used with the utmost restraint, for the sound engineer must never encroach upon the conductors prerogative for the balance of the instruments in the orchestra.

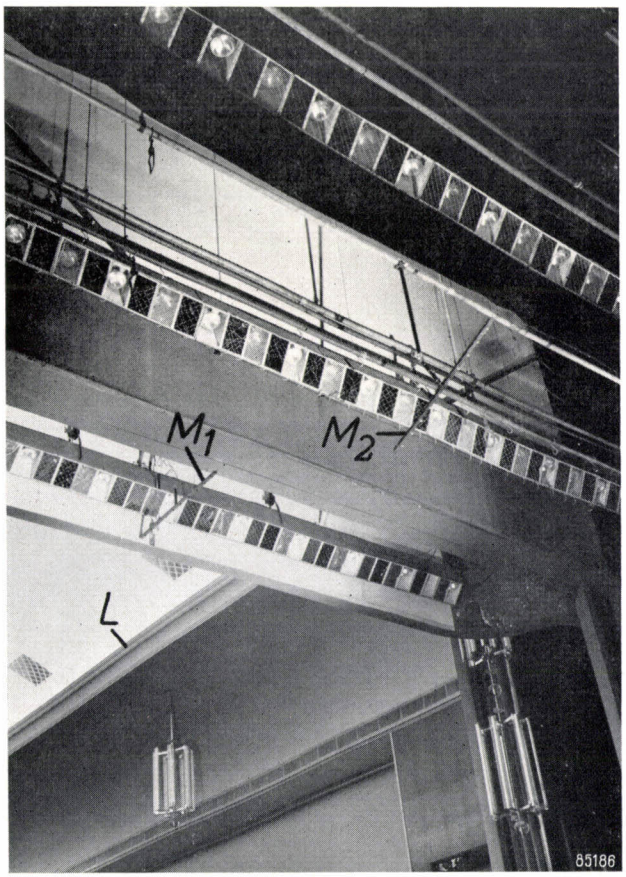


Fig. 6. In the Philips Theatre at Eindhoven two line microphones are used (M_1, M_2). The loudspeakers are mounted in a concealed position behind the lighting cornice L (compare fig. 9).

[7]) This occurs, for example, in the case of music reproduced in a hall, when it is (or was) actually performed elsewhere. We return to this point again at the conclusion.

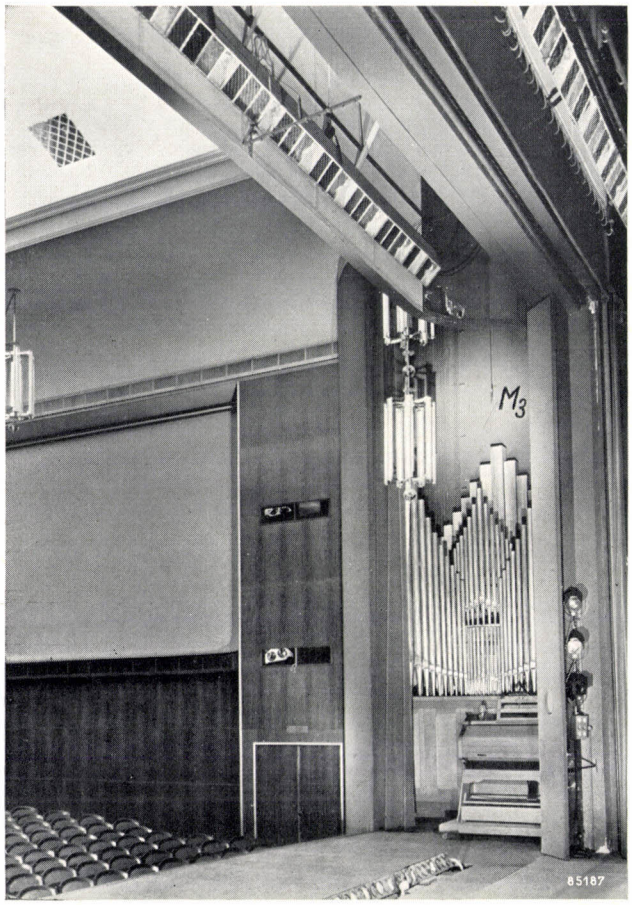


Fig. 7. The small organ in the Philips Theatre can be reinforced through an amplifier channel fed from a separate microphone M_3.

The loudspeakers

As already suggested, the diffuseness of the sound is perhaps even more important than the lengthening of the reverberation time. Diffuseness can be obtained by dispersing the loudspeakers over the hall (*fig. 8*) and connecting them to the various play-back heads. The wiring is simplified if all the loudspeakers belonging to one group (fed from the same play-back head) are connected in series. In the case of four play-back heads, one four-core cable is run around the hall, balconies, etc; where a loudspeaker is installed, the appropriate core is cut and the speaker connected in series (*fig. 9*).

The distribution of the loudspeakers between the various play-back heads should be done as randomly as possible. The only restriction is that the audience should never get the impression that the sound comes from the loudspeakers. We shall now try to explain further the general lines to follow in order to avoid this impression.

We shall use some results of the work of K. de Boer on stereophony [8]). In fact, we are here dealing

[8]) For a recapitulation of the principles of stereophony, with references to the literature, see R. Vermeulen, Philips tech. Rev. **17**, 171-177, 1955/56 (No. 6).

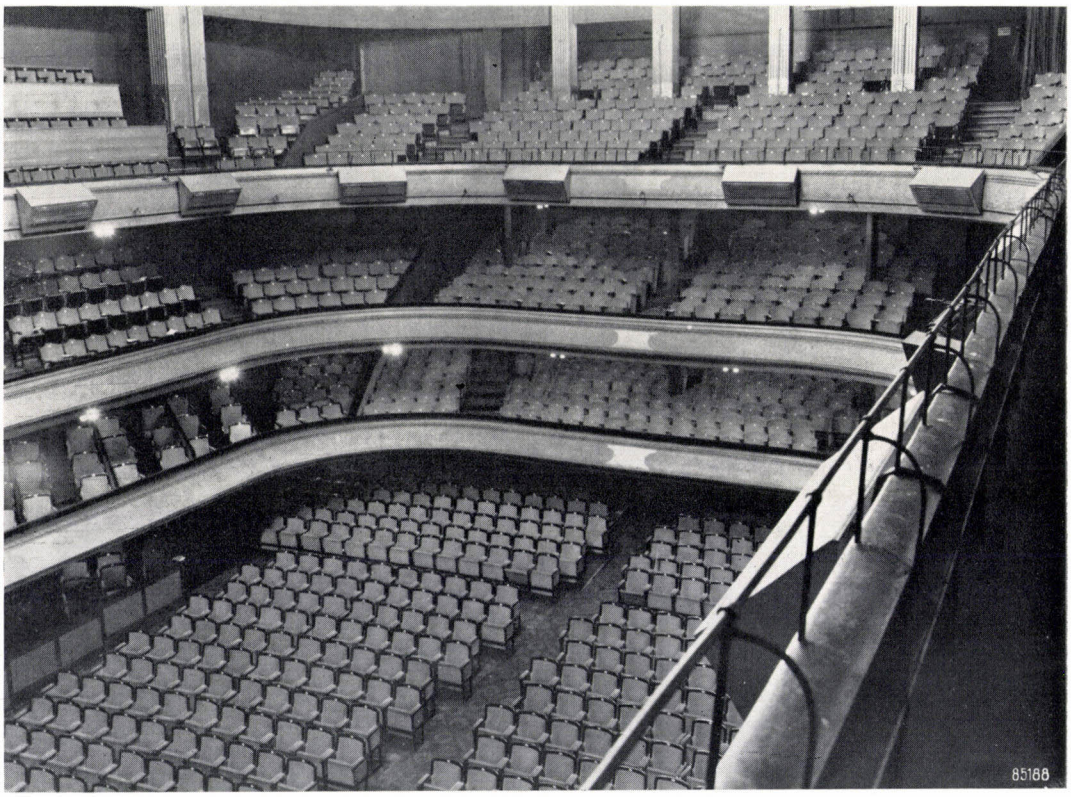


Fig. 8. In the "Gebouw voor Kunsten en Wetenschappen" the loudspeakers are mounted along the edge of the upper balcony and (not visible in the photograph) under the balconies.

with an analogous problem. The condition that nobody in the hall may consciously hear music coming from the loudspeakers, means that at all points in the hall the "sound image" must appear to be located in the orchestra. This latter can be

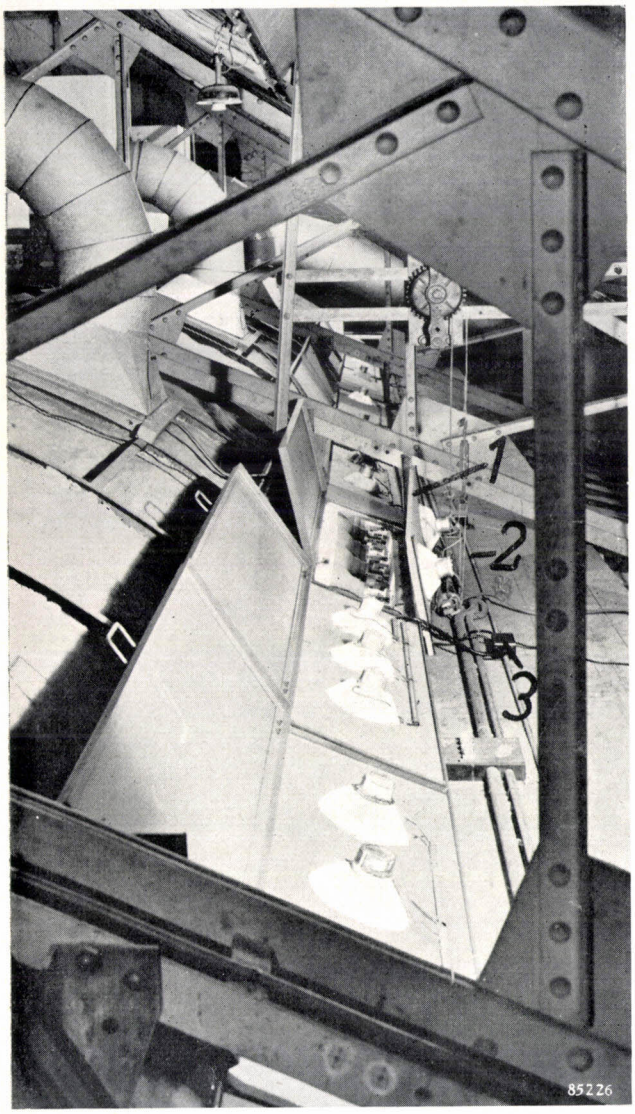


Fig. 9. The loudspeakers of the stereo reverberation installation in the Philips Theatre are mounted in groups of two or three in the panels which cover the lighting cornice (*L* in fig. 6). At *1* one of the panels has been lifted up. *2* is the (six-core) cable, *3* a junction box.

regarded as one of the two sound sources in a stereophonic installation, one of the loudspeakers being the other. For the sake of simplicity we assume that the observer is in the plane of symmetry of the two sources (if he is otherwise situated, the values to be mentioned presently must be modified by suitable amounts). We then know that he will locate the virtual sound source (the "sound image") in the orchestra if (a) the sound from the orchestra and that from the loudspeaker arrive at the same time but the intensity of the orchestra is at least 10 dB above that of the loudspeaker, or (b) orchestra and loudspeaker are equally loud but the sound from the orchestra arrives 2 msec earlier. These are the two extremes; for intermediate cases, as far as the position of the sound image is concerned, a lag of 1 msec in one sound is compensated by an increased intensity of 5 dB, according to the almost linear relationship [9]) shown in *fig. 10*. The sound image is still located in the orchestra even when the intensity of the orchestra is 5 dB less than that from the loudspeaker, provided the sound from the orchestra arrives 3 msec earlier.

Fig. 10. Differences in intensity (in dB) plotted against the phase differences Δt which produce the same angular displacement of the sound image.

Another condition is that for no listener may the sound from the loudspeakers arrive with so great a time lag that it is no longer experienced as reverberation of the orchestra, but as a separate echo. This means that for no observer may the first loudspeaker signal which reaches him arrive more than 50 msec after the direct sound from the orchestra. This value corresponds to that found earlier when investigating the maximum time interval during which, in speech, the indirect sound contributes to the intelligibility (see Introduction). Recent investigations [10]) have confirmed this value and also that the indirect sound may be stronger than the direct without disturbing the location of the sound image.

On the basis of these data we can set down the following rules for a stereo reverberation installation:

---

[9]) K. de Boer, Stereophonic sound reproduction, Philips tech. Rev. **5**, 107-114, 1940.

[10]) H. Wallach, E. B. Newman and M. R. Rosenzweig, The precedence effect in sound localization, Amer. J. Psych. **62**, 315-336, 1949; G. Meyer and G. R. Schodder, Über den Einfluss von Schallrückwürfen auf Richtungslokalisation und Lautstärke bei Sprache, Nachr. Akad. Wiss. Göttingen, IIa, 31-42, 1952.

1) No member of the audience may receive sound from any loudspeaker before the direct sound has reached him.
2) Nowhere may the first loudspeaker sound arrive more than 40 msec after the direct sound (a safety margin of 10 msec has been deducted from the limiting value of 50 msec).
3) The intensity of the loudspeakers may nowhere be more than 5 dB above that of the direct sound [11]) (apart from the diffuseness required, this is another reason why many dispersed loudspeakers must be used).

If the same hall is also to be used for plays or lectures and improvement of the intelligibility is desirable, the stereo reverberation installation can also make itself useful for this purpose. One must then ensure that no sound is repeated later than 40 msec after the direct sound, and no feedback should be applied.

Directions for operation of the installation

Complying with the rules listed above does not necessarily ensure that the installation works satisfactorily. Because the theory lacks sufficient experimental backing, one should beware of clinging to unfounded preconceived opinions.

An example will illustrate this. In one of our first experiments with stereo reverberation, it was thought that the artificial reverberation should be built up of as many repetitions as possible, in order to obtain the smoothest possible exponentially decreasing intensity. It appeared, in fact, that though this did give the impression of a long reverberation time, this was by no means accompanied by the feeling of being in a large hall — rather that of a small "hard" room such as a bathroom. To suggest a large space, it was necessary to increase the time interval between the echos and to make the reverberation not at all smooth. By careful adjustment in a laboratory room of about 1000 m^3 volume, we could create the acoustic impression of being in a cathedral.

One general piece of advice relating to electro-acoustical intervention in musical performances is that the engineer must show considerable retraint. He must suppress his desire to make the effect *striking* and take care that he does not exaggerate. His is a thankless task: whenever his work is recognized for what it is, he will be reproached, and the better he does his work, the more natural the result will appear and the less thanks he can expect. The highest praise he can receive is probably the simple verdict: "...the orchestra sounded much better".

As we have said, not only the Philips Theatre at Eindhoven, but also the Gebouw voor Kunsten en Wetenschappen in The Hague, is fitted with a stereo reverberation installation. On November 30th, 1954, this installation (then still provisional) had its public debut during a concert given by the Residentieorkest. After the concert, the effects which can be achieved were demonstrated more emphatically.

The verdict on the operation during the concert varied from "favourable" to "very favourable". Some people, however, found it difficult to observe the effect consciously. This is illustrated by a remark from a musician in the audience: "It was remarkable that one could not consciously hear that the installation was in operation; one only felt, or experienced it. Only during the demonstration after the concert did I become consciously aware of it, and convinced that my observations during the concert were not imaginary".

Though the improvement may be appreciated only unconsciously by some of the audience, it is another matter for the performing musicians, both members of the orchestra and soloists; they experience the stereo reverberation very clearly and consciously as making the hall more "playable". This undoubtedly contributes to the attainment of a higher artistic level.

We feel that it is an important milestone in the development of electro-acoustics that leading musicians not only permit microphones and loudspeakers and all that goes with them, in the concert hall, but actually welcome their help.

Other applications of stereo reverberation

Stereo reverberation will undoubtedly find other applications apart from the conversion of theatres, acoustically speaking, into concert halls. However, the apparatus is so complicated that these applications will be limited to the professional field for the time being.

One obvious application is in broadcasting studios. Here, stereo reverberation can be a means of adjusting the acoustics of the studio to the nature of the music or the play to be performed. This can, of course, be done afterwards by adding artificial reverberation to the microphone signal; the echo chamber is a device often used for this purpose. But this deprives the musicians of the stimulus of good acoustics and if there is an audience in the studio, they too miss the full effect. Stereo

---

[11]) See fig. 8 in the article by Meyer and Schodder, referred to in [10]).

reverberation can overcome these difficulties [12]).

Another application that springs to mind is in the cinema. Here stereo reverberation can be obtained in the manner described above, with the difference that the microphone is discarded and the recording head on the delay wheel is now fed with the signal derived from the sound track of the film. It is even simpler if the cinema is fitted for a system like "CinemaScope", i.e. fitted with loudspeakers around the hall and projectors suitable for films with more than one sound track; the direct and the delayed diffuse sound could then be recorded beforehand on these tracks, so that the cinema has no need to be equipped with the delay wheel.

Finally we may mention the "duplication of concerts" [13]), that is the stereophonic reproduction in one or more "overspill" halls of a concert given elsewhere. Clearly, diffuse reverberation in the "overspill" halls can considerably increase the musical value of the programme presented.

Summary. A shortage of good concert halls means that music is often presented in a hall which is less suitable for this purpose, for example, in a theatre. Such halls can be given better acoustics for music by providing artificial diffuse reverberation. An installation is described with which such "stereo reverberation" can be provided electro-acoustically. It is based on a so-called "delay-wheel", the rim of which is coated with a material suitable for magnetic sound recording. A recording head records the music performed. A number of play-back heads (say, four or six) around the circumference of the rotating wheel pick up the recorded sound with predetermined time lags, and separately feed a group of loudspeakers, mounted throughout the hall. The intensity of each group is independently controllable, and the time lags can be regulated by the speed of the wheel. The sound is picked up by one or two "line microphones", each consisting of ten condenser microphones with strongly directional characteristics, to reduce the possibility of continuous "howl-back".

It has been shown that such an installation can produce a great improvement in the musical acoustics of a hall. Some of the audience experience this only unconsciously, but the performing musicians are strongly aware of it as making the hall more "playable".

Finally, some other possible applications of stereo reverberation are discussed: in broadcasting studios, in the cinema and in the "duplication" of concerts.

[12]) According to a private communication from Dr. J. J. Geluk, Head of the Dutch Broadcasting Union Laboratories, the Union is planning to build two large broadcasting studios (7500 m³ each) in which the principles of stereo reverberation will be applied.

[13]) R. Vermeulen, Philips tech. Rev. **10**, 169-177, 1948/49 and R. Vermeulen and W. K. Westmijze, Philips tech. Rev. **11**, 281-290, 1949/50.

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

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

2217: A. Venema: Thermische emissie (T. Ned. Radiogenootschap **19**, 283-303, 1954). (Thermionic emission; in Dutch).

After an introduction to the subject, the author deals briefly with Schottky's method for calculating the work function, the influence of adsorbed layers on the work function, and its measurement on the basis of the Richardson formula. This is followed by a survey of emitter substances, with data on the temperature dependence of the current density. The thermionic efficiency is considered. The problem of the life of oxide cathodes is also dealt with. The shortcomings of the oxide cathode are mentioned, which have lead to the development of the so-called dispenser cathodes. Three types of the latter are discussed: the L-cathode, the impregnated cathode and the pressed cathode.

2218: D. Kleis: Dynamique de l'enregistrement magnetique (Onde électrique **34**, 753-760, 1954).

For good sound reproduction, an adequate dynamic range is essential. A large orchestra has a dynamic range of about 80 dB; with magnetic recording, a dynamic range of only 70 dB is attainable, so that it is important to ensure that this is used to the fullest extent. This puts high requirements on the recording apparatus as well as on the reproduction instrument. Unless special precautions are taken, there is the danger with such a large dynamic range that the noise level and distortion are unacceptably high; this is a result of the asymmetry of the erasing current but, more important, of the A.C. biassing current. With the usual type of play-back head, the signal voltage delivered to the amplifier is proportional not only to the amplitude of the flux in the play-back head but also to the frequency. This results in the signal voltage having a dynamic range considerably greater than that of the recording itself. To prevent the introduction of noise and distortion in the amplifier, special circuits must be used. This article discusses these factors and the means by which the desired dynamic range can be obtained in both recording and reproduction. The methods used include, in particular, the use of a negative feedback h.f. oscillator for erasing and

biassing during recording and an amplifier with frequency-dependent negative feedback during reproduction.

2219: G. W. Rathenau and G. Baas: Electron-optical observations of transformations in eutectoid steel (Acta Metallurg. **2**, 875-883, 1954).

More detailed account of investigations described in brief in Philips tech. Rev. **16**, 337-339, 1955/1956.

2220: J. M. Stevels: Networks in glasses and other polymers (Glass Ind. **35**, 657-662, 1954).

The dielectric losses, measured as tan δ, as a function of temperature are compared for a number of glasses, fused silica and crystalline quartz. It is shown that the sharp peak in the curve for quartz at low temperatures has its origin in a relaxation mechanism. For fused silica and the glasses examined, tan δ has a very broad maximum at low temperatures. It is reasonable to suppose that in the latter cases we are concerned with a movement of the oxygen, modified by the network in which it is bound. With decrease in Y (the average number of points of contact per tetrahedron), tan δ_{max} and also the corresponding decrease of the dielectric constant ε go through a maximum whereas the activation energy of the relaxation phenomenon goes through a minimum. This can be understood as due to the fact that although the network gets looser and looser, the network-modifying ions increase in number so that the oxygen becomes less mobile. Tan δ for silicones and other organic polymers also shows a broad maximum at very low temperatures. Here, however, sharp peaks are also present at about 200 °K. The possible origin of these peaks is discussed.

2221: F. A. Kröger and H. J. Vink: Physico-chemical properties of diatomic crystals in relation to the incorporation of foreign atoms with deviating valency (Physica **20**, 950-964, 1954).

If foreign ions of a valency deviating from that of ions of the base lattice are incorporated, electro-neutrality is maintained. This is effected in a number of ways, some of which are already known. For example, electro-neutrality may be maintained by the formation of vacancies, or by the occupation of interstitial sites. A second mechanism is that in which the incorporation of the foreign ions is accompanied by a reduction or an oxidation of the base lattice. It is shown that it is possible to consider the various ways of maintaining the electro-neutrality from a general point of view. There are a number of factors which govern the way the electro-neutrality can be maintained. Among these are the concentration of the foreign ions, the tendency of the base lattice to form lattice imperfections, the position of the energy levels associated with the various lattice imperfections, and the width of the forbidden zone. Another factor of importance is the atmosphere in equilibrium with the compound. In such a way, a third mechanism of maintaining the electro-neutrality is found. Application of these considerations to CdS as a base lattice gives a satisfactory agreement with experiment.

2222: F. van der Maesen and J. A. Brenkman: Acceptor activity of copper in N and P type germanium of different resistivity (Physica **20**, 1005-1007, 1954).

Experiments are carried out in which N and P type germanium samples with various resistivities or a bar with a resistivity gradient are saturated with copper at temperatures of 750 and 810 °C. The number of introduced acceptors calculated from the change of resistivity after quenching, is dependent on the position of the Fermi level. In material that remains N type after saturation, there is a considerable increase of the acceptor activity. It is therefore possible that copper produces extra acceptor levels with a rather high activation energy.

2223: H. J. G. Meyer: On the theory of transitions of F-centre electrons (Physica **20**, 1016-1020, 1954).

In the expressions for the probability of a radiationless transition in F centres, as given by Huang and Rhys and more recently by the present author, certain parameters occur, the numerical values of which have to be found from a comparison of the experimental and the theoretical absorption spectrum due to the corresponding transition. If certain small but systematic deviations between theory and experiment are neglected, values for the radiationless transition probability are found which are of a reasonable order of magnitude. It is furthermore shown in a qualitative way that the deviations may be explained by the fact that the theoretical spectrum is determined under the assumption that the Condon approximation is valid. For alkali-halide F-centres this is probably not allowed.

2224: B. H. Schultz: Surface recombination as a function of the concentration of charge carriers in the interior (Physica, **20**, 1031-1033, 1954).

Measured values of the surface recombination rate s of electrons and holes of P and N germanium are compared with what is to be expected from the considerations of Brattain and Bardeen on surface states. There is qualitative agreement: s is indeed large for materials of small resistivity, but it depends on the resistivity to a lesser extent than the theory predicts. For intrinsic germanium (i.e. germanium in which neither donors nor acceptors are dominant) which has been etch-polished, the value of s measured was 20 cm/sec.

2225: F. A. Kröger, H. J. Vink and J. Volger: Resistivity, Hall effect and thermo-electric power of conducting and photo-conducting single crystals of CdS, from 20—700 °K (Physica **20**, 1095-1099, 1954).

Short report on investigations which were described at length in Philips Res. Rep. **10**, 39-76, 1955: see these abstracts No. **R260**.

2226: G. H. Jonker: Semiconductor properties of mixed crystals with perovskite structure (Physica **20**, 1118-1122, 1954).

Mixed crystals of $La\,Me^{3+}O_3$ and $Sr\,Me^{4+}O_3$ (Me = Ti, Cr, Mn, Fe or Co) with perovskite structure have been prepared in ceramic form. The compounds with the general formula $[La_{1-x}Sr_x][Me^{3+}_{1-x}Me^{4+}_x]O_3$ show a high electric conductivity as the Me-ions are present in two valency states. Interesting properties are met in mixed crystals containing two kinds of metal ions of the transition group. Because of the difference in ionisation energies, one may expect that only one kind of the metals in these mixed crystals exists in two valency states. It is possible to prepare samples with high restivity, from which it can be concluded that one metal is present in the trivalent state and the other one in the tetravalent state, e.g. in $[La_{0.75}\,Sr_{0.25}][Fe^{3+}_{0.75}Mn^{4+}_{0.25}]O_3$. By the determination of the maxima of the resistivity in various series of mixed crystals, the following sequence of preference for the tetravalent state is found: Mn > Cr > Fe. Some of the measurements are complicated by the ferromagnetic properties of the materials, which have a strong influence on the conductivity. It was therefore difficult to find the place of Co in the sequence.

R 271: G. Klein: Rejection factor of difference amplifiers (Philips Res. Rep. **10**, 241-259, 1955, No. 4).

By analyzing the ordinary triode difference amplifier it is shown that the rejection factor can be made arbitrarily large, without the need for preselection of valves or stringent mutual equality of corresponding circuit elements. Some circuits are given which guarantee a high rejection factor, even with 10% difference of the corresponding components of the two halves. The theory is verified by a number of measurements.

R 272: J. Volger, J. M. Stevels and C. van Amerongen: Dielectric losses of various monocrystals of quartz at very low temperatures (Philips Res. Rep. **10**, 260-280, 1955, No. 4).

Between 14 and 150 °K the $\tan\delta$ vs T curve of quartz, measured at frequencies of 1 or 32 kc/s, may show a variety of maxima. The relaxation phenomena involved are correlated with both primary and radiation-induced lattice defects. Results of experiments with clear quartz, artificially irradiated quartz, natural smoky quartz and amethyst are reported. A discussion related to the nature of a number of lattice imperfections is given.

R 273: A. van Weel: Phase-linear television receivers (Philips Res. Rep. **10**, 281-298, 1955, No. 4).

A television receiver with phase-linear intermediate frequency amplifier is described. Normal selectivity demands are satisfied and conventional circuits are used. Performance compares very favouraby with receivers of which the I.F. phase errors are compensated for in the video-frequency part of either transmitter or receiver. The picture quality, apart from being optimum for the given bandwidth, is almost independent of tuning variations.

VOL. 17 No. 10, pp. 269-304 APRIL 1956

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

X-RAY SPECTROCHEMICAL ANALYSIS

by W. PARRISH *).

545.824:621.387.4

Qualitative and quantitative spectrochemical analysis, using the emission spectra of elements in the optical wavelength region, has been successfully applied since the days of Bunsen and Kirchhoff. In the past decade a basically similar method of analysis, using the emission spectra in the X-ray wavelength region, usually excited by way of fluorescence, has been introduced into laboratory practice. This method, which of course employs completely different techniques, has already proved very useful in a large variety of cases.

The specific character of the spectrum of the light emitted by each chemical element when heated or burned, has been put to use for the identification of elements for nearly a century. To-day this method ranks among the most sensitive means for the detection of minute traces of impurities. Techniques have been established which make it an easy and fairly accurate method of *quantitative* chemical analysis, supplementing and sometimes supplanting classical wet chemical methods [1]).

The optical spectra of most elements are very complicated: they usually contain hundreds or even thousands of lines (*fig. 1a*). Although this is not a limitation in principle, the usefulness of spectrochemical analysis is often limited in practice by difficulties in the identification of components in a mixture when the spectrum is very crowded. Other limitations of the method will be mentioned at the end of this article.

In the past decade another method of spectrochemical analysis, based on the X-ray instead of the optical spectra emitted by the elements, has acquired considerable importance. This method has similar characteristics to those of the optical method and it has additional features, e.g. that of being non-destructive — certainly an important asset for some cases. An essential advantage, however, as compared with the optical method, is the great simplicity of X-ray spectra (fig. 1*b*). It would perhaps seem that such an advantage — which is of a practical nature — would be completely out-

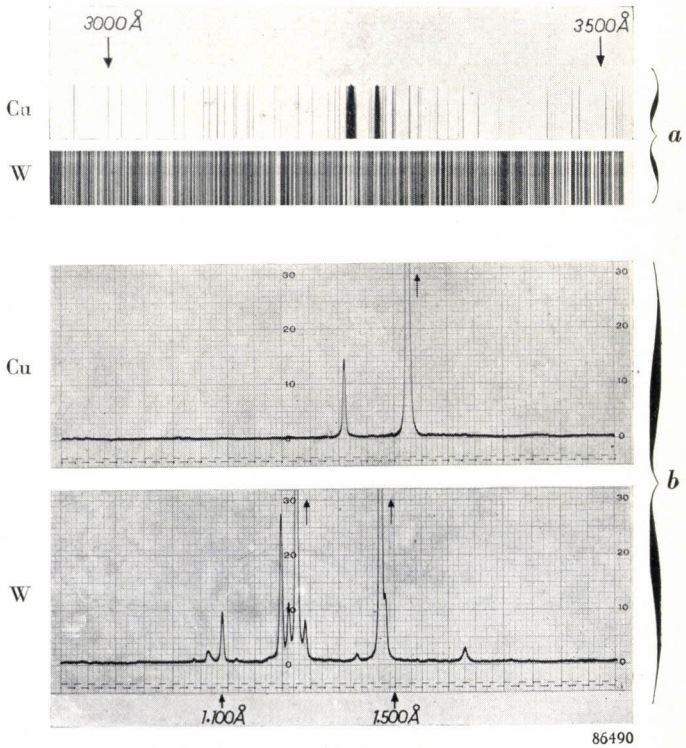


Fig. 1. *a*) Optical spectra of copper and tungsten in the region 3000 to 3500 angstroms. *b*) X-ray spectra of same elements.

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

1) See, for example, N.H. Nachtrieb, Principles and practice of spectrochemical analysis, McGraw-Hill, New York 1954; N. W. H. Addink and W. de Groot, Spectrochemical analysis, Philips tech. Rev. **12**, 337-348, 1950/1951.

weighed by the more complicated technique involved in handling X-rays instead of light. However, this is not the case: equipment is now available which renders the X-ray spectrochemical analysis about as easy in routine application as the optical one. This situation is a consequence of the fact that X-ray techniques as applied to another method of chemical and physical analysis, viz. X-ray diffraction, have attained a high degree of perfection and automation during the past decade and have thereby gained a firm footing in many laboratories. It will be pointed out below that the measurement of radiation intensities by counter tubes in these techniques, instead of the earlier prevailing recording by photographic methods, has played an essential part in the development of X-ray spectrochemical analysis.

Foundations of the method

The origin of the characteristic X-ray spectra may be briefly summarized as follows [2]). According to the Bohr model, the electron orbits of each atom are arranged in shells, viz. the K, L, M, . . . shells. When the atom is given sufficient energy, an electron may be ejected from one of the inner shells. The place of this electron will then promptly be filled by an electron from an outer shell, whose place will in turn be taken by an electron still farther out, etc. The atom thus returns to its normal state in a series of steps, in each of which a photon is emitted. The energy differences between the electron orbits in different shells are such that these photons have wavelengths in the X-ray region, ranging from a few tenths of an angstrom to some dozens of angstroms.

The characteristic emission spectra arising in this way from different elements consist of comparatively few lines owing to the small number of electron orbits participating. Moreover they are very much alike in general structure, revealing a simple variation in wavelengths which is governed by the atomic number of the element. These facts were first established by the experiments of the English physicist Moseley in 1913/14. *Fig. 2* shows a series of X-ray spectra as recorded by Moseley. The very regular trend of the wavelength for each spectral line as a function of the atomic number is shown by *fig. 3*. These curves are the basis of X-ray spectrochemical qualitative analysis. A comparison of the *intensities* of the spectral lines of the elements in a mixture or compound will permit a *quantitative* analysis of their concentrations.

In the practical application of X-ray spectrochemical analysis there are two major problems

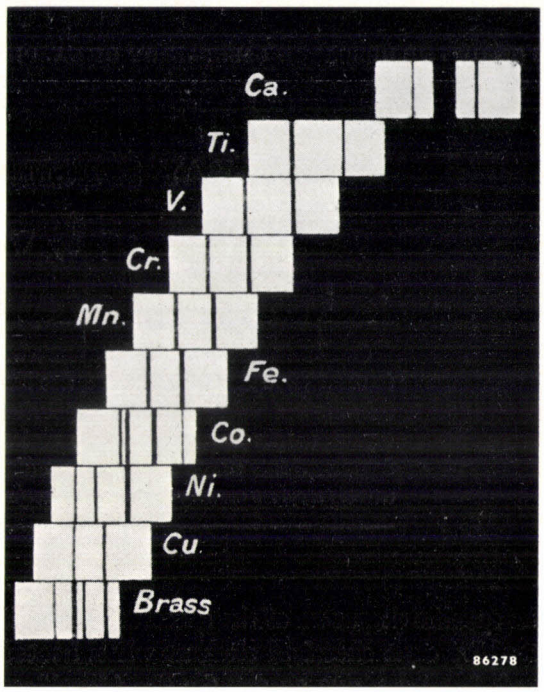


Fig. 2. X-ray spectra of a number of elements photographed by Moseley. (From Phil. Mag. **26**, 1024, 1913.)

which should be briefly considered here before going into the details of the present-day technique, viz. the *excitation* of the X-ray spectra and their *analysis*.

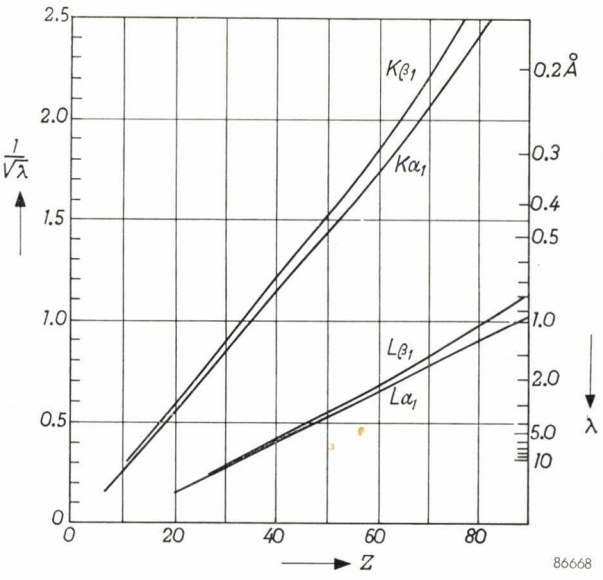


Fig. 3. Wavelengths λ of characteristic X-ray spectral lines of elements, plotted against atomic number Z. According to Moseley's law, there should be a linear relationship between $1/\sqrt{\lambda}$ and Z.

[2]) For a more complete account and a general treatment of the physics of X-ray spectra see for example: M. Siegbahn, Spektroskopie der Röntgenstrahlen, Springer, Berlin 1931. A. H. Compton and S. K. Allison, X-rays in theory and experiment, Van Nostrand, New York 1935.

An atom can receive the energy required for the ejection of an electron from an inner shell either by collision with a high-energy free electron or by absorption of a high-energy photon (X-ray quantum). Accordingly there are two ways of exciting the characteristic X-ray spectrum of each element, viz. either by bombarding the specimen with electrons accelerated by a high voltage or by irradiating it with a beam of X-rays. The latter process, called *fluorescence* because of the analogy with the optical case, is now used and in fact has given its name to the whole method of X-ray spectrochemical analysis, which is usually referred to as "X-ray fluorescence analysis".

The first mentioned process of exciting characteristic spectra — by electron bombardment — is that used in X-ray tubes for generating the monochromatic X-rays applied in diffraction analysis (e.g. the K-radiation of copper). In order to utilize this process for spectrochemical analysis each specimen has to be made in the form of a target and introduced as the anode into a type of demountable X-ray tube. The method was applied for a long time in the investigation of characteristic spectra by Siegbahn and others. The element hafnium was discovered by Von Hevesy and Coster in 1923 using this method, and the method was again used when the element was first prepared in a ductile state in the Philips Laboratories at Eindhoven [3]). It is clear, however, that the difficulties of the technique with demountable X-ray tubes virtually prohibited its large-scale application to routine analysis. Moreover, serious quantitative errors were liable to occur owing to the fact that the specimen exposed to the electron bombardment might volatilize or melt in such a way that the elemental composition at its surface would be altered. For this reason, Von Hevesy and others prior to 1930 had used fluorescence excitation but in order to get sufficient intensity they had to place the specimen very near to the primary X-ray target, inside the tube; the difficulties of vacuum technique etc. thus remained [3]).

In addition to the characteristic line spectrum, a continuous, nearly structureless X-ray spectrum is generated in a target bombarded by electrons. (This continuous spectrum is due to deceleration of the impinging electrons and is used in medical and other X-ray applications where the polychromatic character of the radiation is not objectionable — or is even desired.) It is important to note that a specimen excited by fluorescence does not emit a continuous spectrum. On the other hand, the primary X-ray beam used for exciting fluorescence always contains the continuum. This and other details of the fluorescence excitation method actually used will be discussed later.

Spectrochemical analysis by means of spectra excited by electron bombardment is still applied in one special instance, viz. in the testing for spectral purity of X-ray tubes used for diffraction. Impurities in the target contribute X-rays of undesirable wavelengths. Spectroscopic investigation of these radiations yields the necessary information concerning the nature of the impurities. At the same time, fluorescent X-rays emitted by portions of the tube other than the target can be detected. An example of this application of spectrochemical analysis will be seen later in fig. 7*b*.

The *analysis* of the X-ray spectra is usually effected by applying the principle of diffraction. Mechanically ruled diffraction gratings, which are commonly used for optical spectrographs, can only with great difficulty be made fine enough for the much shorter X-ray wavelengths. A single crystal of a suitable material is used instead. It was Von Laue's discovery of the diffraction of X-rays by crystals which enabled Moseley to carry out his basic experiments on X-ray spectra. The diffraction by a crystal is governed by the well-known Bragg condition:

$$n\lambda = 2d \sin \Theta; \quad n = 1, 2, 3, \ldots \quad (1)$$

A set of atomic planes of the crystal lattice spaced apart a distance d will "reflect" the X-ray beam of wavelength λ only (if at all) at incident angles Θ meeting this condition. (A reflected ray will then make an angle 2Θ with the incident ray.) If for the moment the possibility of reflections of different orders (n) is disregarded, reflection for a given wavelength will occur at only one sharply defined angular setting of the crystal. In order to obtain the complete emission spectrum of the specimen, the crystal has to be rotated to scan the desired angular region.

The resulting arrangement for the analysis of the fluorescent X-ray spectrum is schematically shown in *fig. 4a*. This arrangement is quite similar to that used in X-ray diffractometry, which is similarly based upon the Bragg equation (fig. 4*b*). In diffractometry, however, *a single known wavelength is used to analyze a diffracting specimen for all d-spacings present*, whereas in spectrochemical analysis *a single known d-spacing of a diffracting crystal is used to analyze the spectrum of the specimen for all wavelengths present*. Thus, in diffractometry the specimen is rotated and the different angles 2Θ at which diffracted energy is detected are related to

[3]) G. von Hevesy, Chemical analysis by X-rays and its applications, McGraw Hill, New York, 1932. J. H. de Boer and J. D. Fast, Z. an. allg. Chem. **187**, 193-208, 1930.

different d-values. In spectrochemical analysis the crystal is rotated and the different angles 2Θ at which diffracted energy is detected are related to different wavelengths.

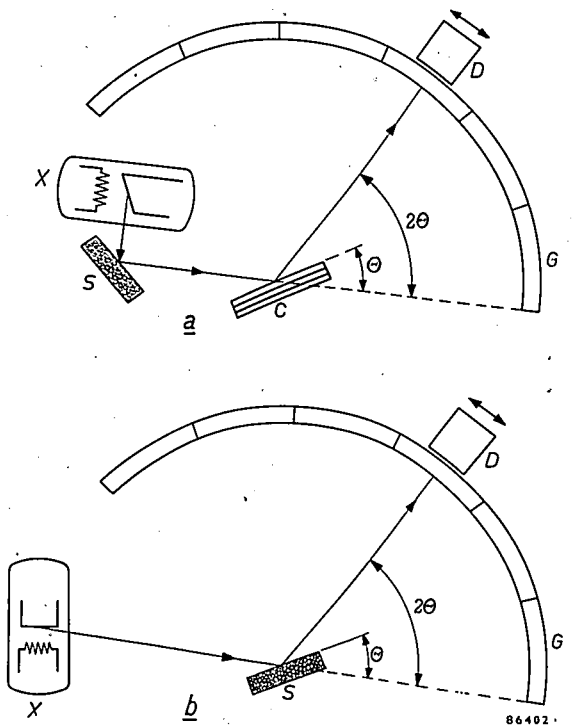


Fig. 4. Schematic arrangement a) for X-ray spectrochemical analysis, b) for powder diffractometry. X X-ray tube, S specimen, C analyzing crystal. D detector, G goniometer.

Because of the evident similarity in instrumentation between X-ray spectrochemical analysis and X-ray diffractometry, it will be useful to emphasize the difference in the information that can be obtained from these methods. X-ray diffraction patterns reveal differences in the character, dimensions and orientations of crystals. Hence, diffraction analysis identifies compounds by their specific crystal structure and lattice spacings; it distinguishes between different crystalline modifications of elements (e.g. graphite and diamond); it permits quantitative analysis of mixtures of crystallized substances; it distinguishes between mixtures and solid solutions; and it gives information on the orientation of crystallites in a sample, on strain and on crystallite size. Clearly, diffraction methods are mainly applied to matter in the crystalline state.

The characteristic X-ray spectra, on the other hand, are emitted by the atoms of an element regardless of the crystal structure or chemical composition of the sample in which the element occurs and independent of the physical conditions of the sample. The X-ray spectrographic analysis therefore supplements the diffraction analysis in giving information on the elemental constitution of a sample.

Details of the technique

In embarking on a more detailed description of the equipment and technique, it is convenient to continue the discussion of the actual measurement and recording of the X-ray spectra begun in the preceding section, and to return later to the excitation and other details of the technique. It should be pointed out that in this description it will not be possible to mention all the procedures now in use.

Recording of the spectra

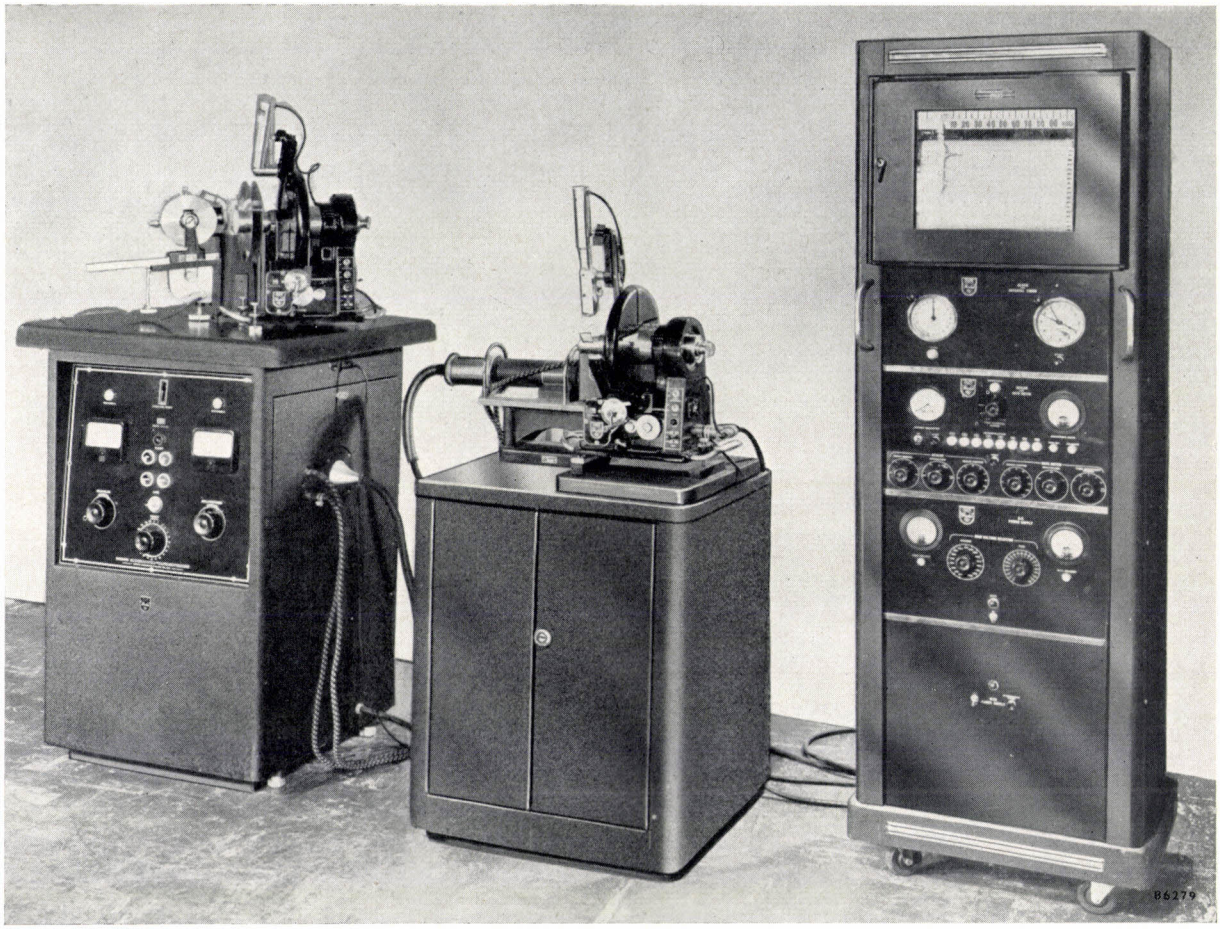
The X-ray energy reflected by the analyzing crystal at different angles can be recorded photographically (which was the original method used by Moseley and others), but the modern technique of using Geiger-Müller tubes or other counter tubes is much better adapted to the purpose. These give an immediate evaluation of the line intensities, which are important for quantitative chemical analysis and which are much more difficult to determine from photographically recorded spectra [4] [5]).

To make this point clearer, we should turn for a moment to the field of X-ray diffraction analysis proper. The method most widely adopted for routine analyses in this field is the Debye-Scherrer method, in which a finely powdered specimen is used: such a specimen, in principle, does not have to be rotated to yield the complete diffraction pattern, since in each angular position of the specimen all possible orientations (Bragg angles Θ) with respect to the fixed incident beam are represented among the numerous crystallites. All the diffraction lines belonging to different d-spacings are thus recorded simultaneously on a photographic film laid around the specimen. With a counter tube substituted for the film, on the contrary, scanning of the pattern is evidently necessary, and it must be performed at a relatively low rate, in order to accumulate enough counts at each Bragg angle [5]). This apparent drawback of "diffractometry" with a counter tube detector has not prevented it from acquiring considerable importance: the reason for this must be attributed to the advantage of easy and accurate intensity evaluation and to the development of a highly

[4]) W. Parrish, E. A. Hamacher and K. Lowitzsch, The "Norelco" X-ray diffractometer, Philips tech. Rev. **16**, 123-133, 1954/1955.

[5]) W. Parrish, X-ray intensity measurements with counter tubes, Philips tech. Rev. **17**, 206-221, 1955/1956 (No. 7-8).

a



b

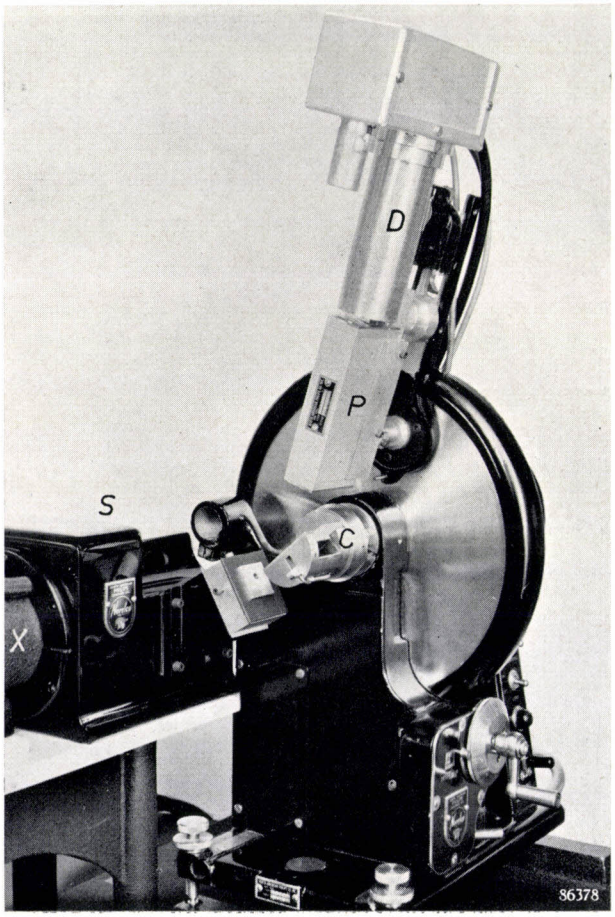


Fig. 5. *a*) "Norelco" equipment both for X-ray diffraction and X-ray spectrochemical analysis. The X-ray generator, with powder cameras and a goniometer for diffractometry on top of it, is shown in the left; the X-ray spectrograph for spectrochemical analysis is in the middle; the electronic circuits rack including the counting-rate meter and recorder is on the right. The same X-ray generator and electronic circuits are used for both methods but the methods cannot be used simultaneously. *b*) Goniometer arranged for spectrochemical analysis: *X* horizontally mounted X-ray tube, *S* specimen chamber, *C* analyzing crystal. *P* parallel slit system, *D* radiation detector (scintillation counter).

perfected instrumentation permitting very good angular resolution and accuracy [4]) [5]) [6]).

The determination of intensities from a photographically recorded diffraction pattern being a rather time-consuming process, it is found on balance that diffractrometry gives a noticeable saving in in time. It is clear that this advantage will be even more pronounced in those cases where only a few lines have to be measured instead of a complete pattern. This is exactly what is usually encountered in spectrochemical analysis: only one characteristic spectral line need be measured for each element, and in many cases the elements whose concentration is to be measured will be known, so that only a very limited angular region need be scanned. The simul-

[6]) E. A. Hamacher and K. Lowitzsch, The "Norelco" counting-rate computer, Philips tech. Rev. **17**, 249-254, 1955/56 (No. 9).

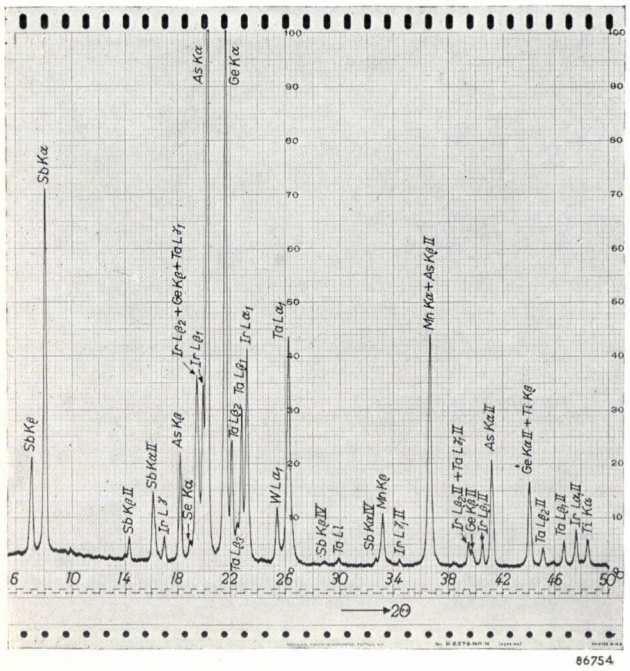


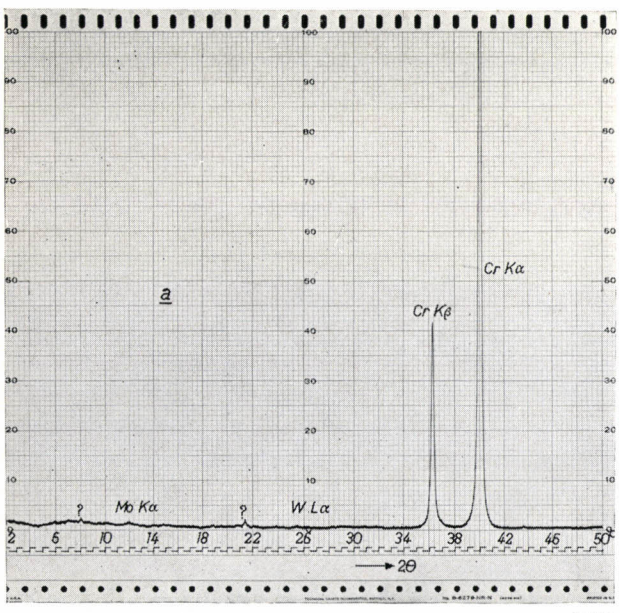
Fig. 6. Qualitative analysis of a mixture of several elements using a quartz analyzing crystal ($10\bar{1}1$-plane), scintillation counter, scanning speed $^1/_2{}^\circ$ per minute. The K-spectra of 22 Ti, 25 Mn, 32 Ge, 33 As, 34 Se, 51 Sb and the L-spectra of 73 Ta, 74 W and 77 Ir can be seen. (The higher order reflections are indicated by Roman numerals.)

taneous rapid development of diffractometry and of X-ray spectrochemical analysis is, therefore, not to be regarded as being a mere coincidence [7]).

The close relationship between X-ray spectrochemical analysis and diffractometry is clearly seen in *fig. 5*, which shows the "Norelco" equipment for X-ray spectrochemical analysis. Except for the fluorescent specimen holder and the X-ray optics, it comprises the same elements as the "Norelco" diffractometer [4]).

The usual procedure in the analysis of a specimen of unknown composition is first to make a qualitative analysis with the recording counting-rate meter by scanning the entire spectrum at a moderate speed, say $^1/_2{}^\circ$ or $^1/_4{}^\circ$ (2Θ) per minute. This permits identification of the elements present and allows an approximate measure of the concentrations; see *fig. 6*. Exact measurements are then made by setting the counter tube on the peak of one line selected for each element, and the intensities measured using either fixed-time or fixed-count methods [5]) [6]). The background is measured separately and subtracted from the peak intensities.

[7]) The advantage of the counter tube detector when looking for a specific spectral line is so marked that techniques of this nature have been commonly applied also in optical spectrochemical analysis. An early application of this method for monitoring purposes was described in this Review several years ago: O. G. Koppius, An application of Geiger counter tubes for spectrochemical analysis, Philips tech. Rev. **11**, 215-222, 1949/50.



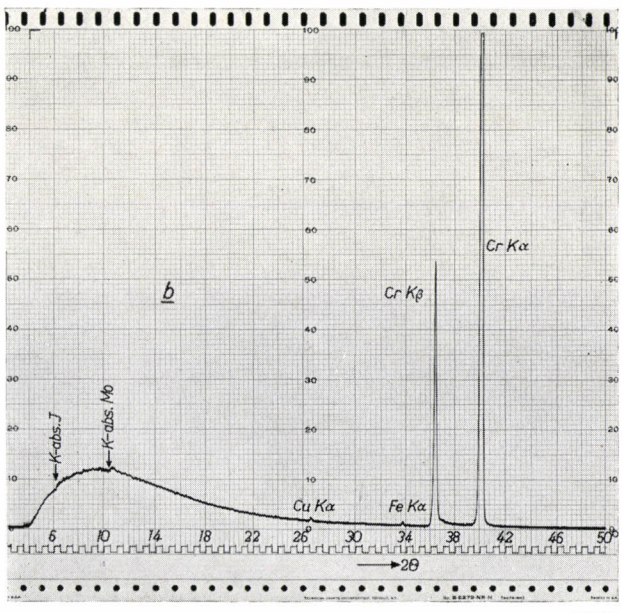


Fig. 7. *a*) Chromium spectrum produced by X-rays from a tungsten target tube (fluorescence) operated at 50 kV peak. No continuous spectrum is excited in this case.

The weak W Lα line is due to scattering of the primary X-ray beam by the specimen and crystal. (The lines at 8° and 21.5° marked "?" have not been identified).

b) The characteristic K-spectrum of chromium produced by direct electron excitation at 50 kV peak. The spectrum is superimposed upon the continuous spectrum. The X-ray tube current was made smaller than in (*a*) by a factor such that approximately the same intensity of the Cr K-radiation was obtained.

The two absorption edges are due to iodine (in the scintillation crystal) and molybdenum (in the slits). The copper and iron lines are due to impurities in the target.

Excitation of the spectra

In the present equipment, for an electron beam of given power, the intensity of the X-ray spectral lines when excited by fluorescence will be smaller by a factor of roughly 1000 than when direct excitation by the electron beam is used. The convenient method of fluorescence excitation has therefore become possible in practice only by the development of high power X-ray tubes, very sensitive detectors and suitable X-ray optics. It would, however, be misleading to compare the two methods of excitation merely on the basis of the intensity difference. The intensity is important with a view to the *time* that will be necessary to measure it: a certain number of quanta (counts) has to be accumulated in order to sufficiently reduce the statistical error of the measurement (see [5]). The *sensitivity* of the analysis, i.e. the lowest detectable concentration of admixtures in a specimen, however, will depend on the *peak-to-background ratio* of the spectral lines. This ratio is generally much higher in the fluorescent spectra than in the directly excited spectra, owing to the absence of the continuous X-ray spectrum. The difference in the spectra is clearly brought out by *fig.* 7.

The question arises, what kind of primary X-radiation will be most suitable for exciting fluorescence?

It was mentioned above that the primary X-ray target will always emit a continuous spectrum. This spectrum for a given accelerating voltage V of the electrons begins abruptly at a wavelength λ_{min}, rises to a broad maximum and falls off gradually with increasing wavelength (*fig.* 8); the limiting wavelength (in Å) is related to the voltage (in kV) by the equation:

$$V = 12.35/\lambda_{min} \quad \ldots\ldots\ldots \quad (2)$$

As the voltage is increased (which will increase the amount of energy eV carried by each electron) the limit of the spectrum shifts towards shorter wavelengths. As soon as the energy eV exceeds the energy required to remove an electron from an orbit in, say, the M-shell of a target atom, all the M-lines of the characteristic target spectrum terminating in this orbit will be emitted in addition to the continuum. On further increasing the voltage, the intensities of the M-lines will increase rapidly, and again at a certain voltage groups of lines of the L-series will appear. At a still higher voltage the K-series will appear.

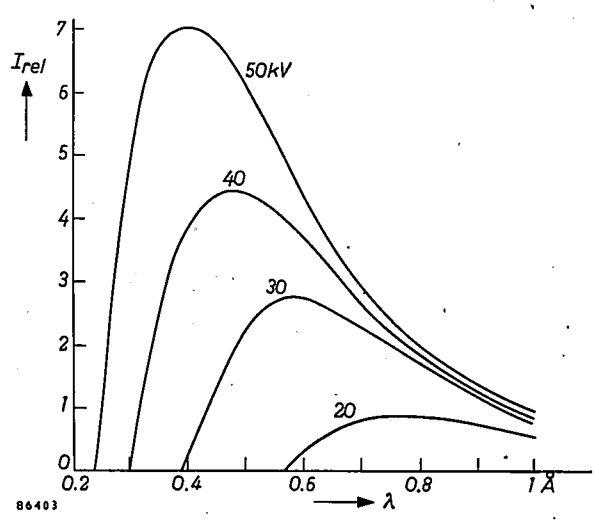


Fig. 8. The continuous spectrum of a tungsten target X-ray tube (relative intensity I_{rel} vs. wavelength λ) obtained with various voltages (peak values) and the same tube current. The curves were redrawn from separate rate meter recordings. Experimental conditions: full-wave rectification, silicon crystal analyzer (111-plane), scintillation counter.

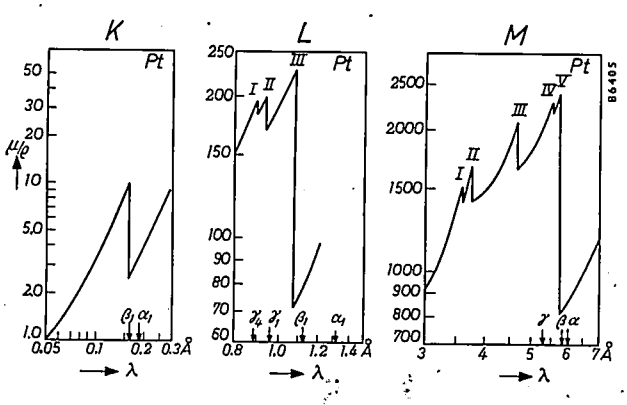


Fig. 9. Mass absorption coefficient (μ/ϱ) of platinum as a function of wavelength λ in the regions of the K, L and M absorption edges. Some characteristic emission lines of the K, L and M series of platinum are shown with arrows.

Similar considerations apply to the secondary target, i.e. the specimen, onto which the primary X-ray beam will be directed. Primary radiation in a broad wavelength region will be more or less absorbed by the specimen. With decreasing wavelength, the impinging photons will carry more energy and when a certain wavelength λ_{min}, characteristic of the element is reached, this energy will be sufficient to knock an electron out of one of the orbits, say of the M-shell. A group of M-lines will then simultaneously be excited and at this wavelength (or rather at a slightly shorter one) the absorption vs. wavelength function of the element shows a steep rise (absorption edge). At still shorter wavelengths, other absorption edges corresponding to other orbits occur: each element has five absorption edges for the M-spectrum, three for the L-spectrum and one for the K-spectrum (*fig.* 9).

For exciting the fluorescence, the primary radiation must evidently have a wavelength shorter than the absorption edge of the spectral lines

selected for the analysis in hand. Continuous as well as characteristic radiation of the primary target can serve the purpose. It should be noted that in order to get a continuous spectrum of short enough wavelengths, the voltage of the X-ray tube must comply with eq. (2), λ_{min} now representing the selected absorption edge of the element to be analyzed. Thus, to just excite the K-series in an aluminum specimen, a voltage $V = 1.6$ kV on the X-ray tube will suffice, whilst for a uranium specimen a minimum voltage of 115 kV would be necessary. Since the X-ray equipment generally used in X-ray spectrochemical analysis does not provide for tube voltages exceeding about 50 kV, elements with atomic number higher than about 60 (neodymium) can not be analyzed by a K-line. The voltage necessary to excite the L-series, however, is lower, reaching only 21.7 kV for uranium, so that L-lines can be used for the elements of higher atomic number.

A closer consideration of limitations of the method due to the voltage and other factors will be given below. A number of typical wavelength data are compiled in *Table I* [8]).

Table I. Characteristic X-ray wavelengths (in Å) of some elements.

Element	K-abs. edge	$K\beta_1$	$K\alpha_1$	L_{III} abs. edge	$L\gamma_1$	$L\beta_1$	$L\alpha_1$
13 Al	7.951	7.981	8.338				
24 Cr	2.070	2.085	2.290	20.7		21.32	21.71
29 Cu	1.381	1.392	1.540	13.29		13.08	13.36
40 Zr	0.688	0.701	0.786	5.58	5.38	5.84	6.07
60 Nd	0.285	0.294	0.332	1.995	1.878	2.166	2.370
74 W	0.178	0.184	0.209	1.215	1.098	1.282	1.476
92 U	0.107	0.111	0.126	0.722	0.615	0.720	0.911

The continuous spectrum emitted by the primary target may have several times more total energy than its characteristic spectrum (cf. fig. 7*b*). The ratio depends mainly on the voltage; for a copper target at 50 kV, the energy of the continuum is about an order of magnitude greater than that of the $K\alpha$ lines. Since, however, the fluorescence of the specimen is mainly excited by the radiation closest to its absorption edge, the characteristic lines in the primary radiation can play an important role. The L-spectrum of tungsten, (which is seen in fig. 1*b*) for example, is very important in exciting K-fluorescence of the elements from about 34 Se to 25 Mn. In one specific case, viz., nickel excited by a tungsten target tube at 40 kV (with full wave rectification), only about 10% of the Ni $K\alpha$ intensity was found to be due to the continuum of the tungsten target: 40% was due to the $L\alpha_1$, α_2 lines and 50% to the $L\beta$ and $L\gamma$ lines of the tungsten. Of course, it will not generally be possible to pick a primary target with lines close to the absorption edges of the elements to be analyzed; in such cases one has to be satisfied with using the continuum for the excitation. Tungsten as a primary target can thus be used for a large group of elements, and fluorescence intensities of up to 4×10^6 counts per second may then be obtained from large pure specimens. Tungsten radiation should, however, not be used for the analysis of tungsten in a specimen, and in general any specimen element should not be excited by the identical element as the primary target. The reason for this is that the primary radiation — the continuum as well as the line spectrum — is scattered by the specimen and the analyzing crystal to some extent and will therefore appear in the spectrogram with a certain intensity (cf. fig. 7*a*), usually less than 1% of the principal fluorescent lines of the specimen. Clearly, the contribution due to a scattered primary W L-line would not be distinguishable from the fluorescent intensity of the same W L-line.

It should be mentioned that owing to the regular dependence of the absorption edges on the atomic number, a method of chemical analysis can also be based on locating these edges for a specimen. Measurement of the absorption coefficient with monochromatic or in some cases polychromatic X-ray beams allows of quantitative analysis. This method, which is much more restricted in its performance than the analysis based on the emission spectra, will not be discussed here [9]).

The X-ray tube

The X-ray tube providing the primary beam is similar to that used in X-ray diffraction work, with the main exception that the focus is of much larger size, viz. about 10×5 mm, and that it is not viewed at a very small angle (the viewing angle is 20°). Indeed, contrary to applications where the X-ray source is used for obtaining *images* (either shadowgraphs as in diagnostics or "slit-images" as in diffraction), there is no point in using a point-source for the irradiation of the fluorescent specimen. A larger focal spot has the advantage of permitting higher loadings. The permissible loading, of course,

[8]) A complete list of X-ray wavelengths has been published by Y. Cauchois and H. Hulubei, Longueurs d'onde des émissions X et des discontinuités d'absorption X, Hermann & Cie, Paris 1947. The excitation potentials are given by S. Fine and C. F. Hendee, Nucleonics **13**, 36-37, 1955 (No. 3).

[9]) A. Engström, Acta Radiologica, Supp. **63**, 1946. H. A; Liebhafsky, Anal. Chem. **21**, 17-34, 1949; **22**, 15-16, 1950. **23**, 14-16, 1951; **24**, 16-20, 1952; **25**, 689-692, 1953. B. Lindström, Acta Radiologica, Supp. 125, 1955.

also depends on the thermal conductivity and melting point of the selected primary target metal. For tungsten in the water-cooled "Norelco" tube, a 10×5 mm focal spot permits continuous loading up to 3.5 kW as compared to 1 kW with a 10×1.6 mm focal spot used for diffraction.

The choice of the voltage for operating the X-ray tube is an important point. When the selection of a primary target with a suitable line spectrum for the fluorescence excitation is not possible and the continuous spectrum of a tungsten target has to be used, it will usually be desirable to run the tube at the highest permissible voltage. This will ensure that the largest possible number of unknown elements in the specimen will be excited to fluoresce. It will also ensure the greatest possible intensity of fluorescence of each element, since with increasing tube voltage the portion of the continuum having wavelengths shorter than a given absorption edge will increase (cf. fig. 8).

In two cases the X-ray tube voltage should be made lower than the available maximum, viz. *a*) when it is desirable not to excite fluorescence of *all* elements in the specimen (selective excitation; this case will be considered more closely in the Appendix); *b*) when very long wavelength fluorescence spectra are to be excited. The reason in this latter case is the scattering of the primary radiation through the system. The peak-to-background ratio will be smaller at higher voltages since, for equal X-ray tube power, the height of the maximum (and hence the integrated intensity) of the continuum will increase more than the intensity in a longer wavelength region (fig. 8).

It should be pointed out that for selective excitation and in general when using relatively low tube voltages, it will be advantageous to operate the X-ray tube at constant potential instead of with the more usual full-wave rectified supply. With the latter mode of operation, the continuous X-ray spectrum produced will shift back and forth during each half cycle and will attain its extreme short wavelength position only during the peak of the operating voltage. A specimen element having its excitation potential not far below the peak voltage of the tube will therefore be excited only during a short part of the cycle. The gain in intensity obtained for the Kα line of a number of target elements when operating the X-ray tube at a constant potential of 30 kV rather than at a full-wave rectified voltage of 30 kV peak, is plotted in *fig. 10*.

In comparing the modes of operation it is of course necessary to take into account the power rating of the apparatus available. The "Norelco" diffraction unit is full-wave rectified with a rating of 2.5 kW, while the PW1010 unit made by Philips in Europe gives constant potential with a rating of 1 kW. Similar considerations determine the feasibility of lowering the X-ray tube voltage, since this will necessitate higher tube currents to maintain the power at the same level. With the sealed-off X-ray tubes available for this application and the "Norelco" power unit, the maximum tube currents are about 50 mA; full use of the power available from the "Norelco" unit therefore cannot be achieved at voltages below 50 kV.

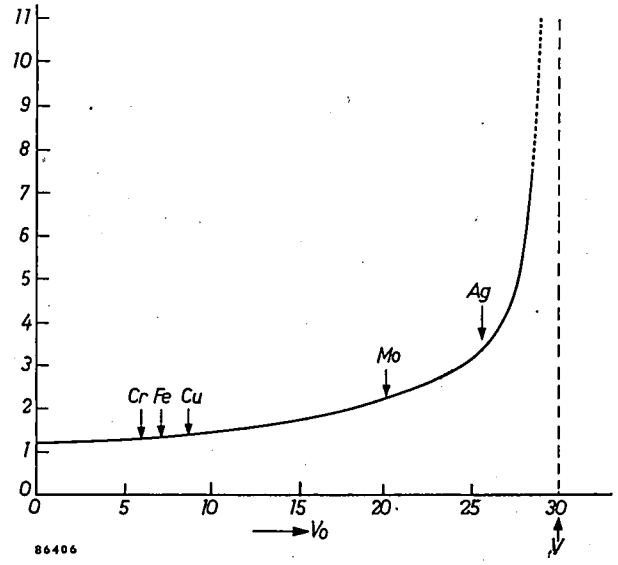


Fig. 10. The gain in line intensity of the primary beam obtained by operating the X-ray tube at constant potential V instead of with full-wave rectified voltage of peak value V, is plotted against the critical excitation potential V_0. The curve shown was calculated on the basis of $V = 30$ kV, assuming the intensity increases with $(V - V_0)^2$ and neglecting self-absorption in the target. The gain factor increases with critical excitation potential V_0 of the elements; a few typical elements are indicated. The gain factor also increases as the operating voltage V is decreased.

At low voltages it is particularly important to have an X-ray tube window transmitting the primary X-ray beam without much loss. The beryllium windows normally used vary in thickness from about 1.5 to 1 mm and will transmit X-rays up to 3 or 4 Å as can be seen in *fig. 11*.

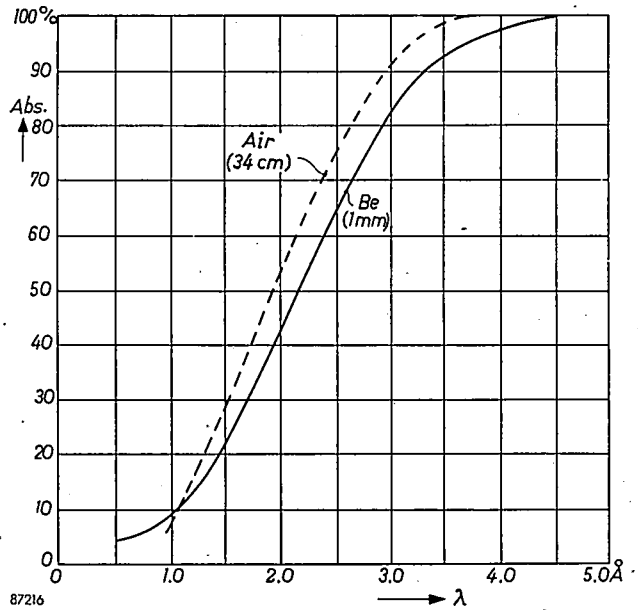


Fig. 11. The percentage absorption of a 1 mm thick beryllium window and of an air path 34 cm long (i.e. about the length used in the "Norelco" spectrograph) as a function of wavelength λ of the X-rays.

These soft X-rays are appreciably absorbed in air so that large intensity losses would occur. The air absorption in a 34 cm path (as used in the "Norelco" equipment) is also shown in fig. 11. The air absorption problem in the goniometer for fluorescent radiation of these and even longer wavelengths can be solved in a simple way by introducing helium in the place of air into the X-ray optical path. This is done by attaching a flexible rubber bag or polyethylene sleeve to the goniometer and passing helium through the system, see *fig. 12*; a transmission of about 85% for aluminium Kα radiation of $\lambda = 8.3$ Å is thus obtained in the "Norelco" goniometer. Where helium is not readily available, a vacuum path can be used. In this case two more windows are practically unavoidable; thin "Mylar" polyester film has been used for the purpose, a sheet of 0.06 mm absorbing about 45% of Al Kα.

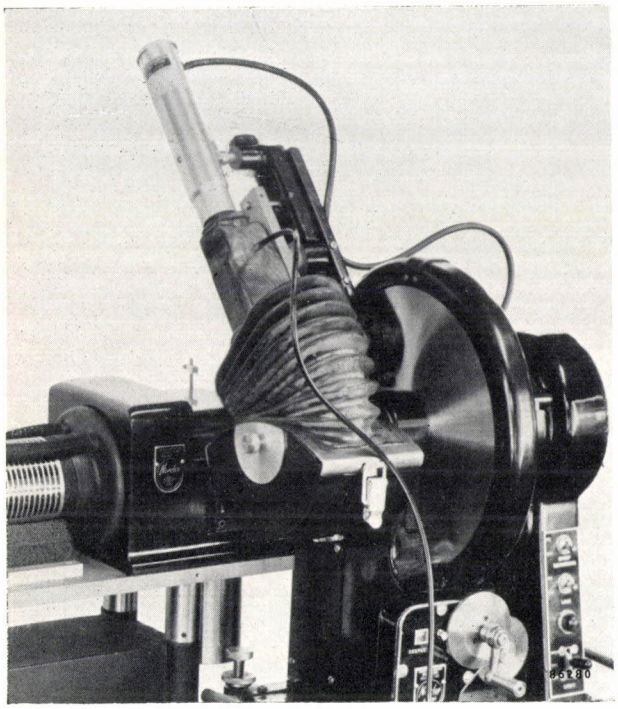


Fig. 12. "Norelco" helium attachment. By passing helium through the rubber-enclosed X-ray path, the serious loss of intensity due to absorption of soft X-rays in air is reduced. This avoids the necessity of using vacuum techniques.

X-ray optics

In order to collect sufficient radiation for the analysis of minor constituents, it is desirable to have a large specimen surface acting as a source. It is *not* possible (without reducing the radiation collected) to render the *apparent* fluorescent source small by viewing the specimen at a very small angle to the surface.

The well-known expedient of decreasing the apparent size, which is employed in nearly every application of X-rays in order to obtain a high brilliance point or line source (cf. above), is only applicable to electron-excited targets. The electrons are stopped at a relatively small depth, usually a fraction of a micron under the surface, so that only a thin layer of the target acts as an X-ray source. This layer is practically transparent to the emitted X-rays, even to those travelling obliquely. In the case of fluorescence, on the other hand, the primary X-rays which excite the fluorescent X-rays penetrate to a considerable depth in the specimen before being absorbed and the latter will equally absorb the emitted rays. Consequently no gain in brilliance is obtained by using small glancing angles (the specimen is a Lambert type source).

A fluorescent source of large apparent size precludes the possibility of using a focussing arrangement like that used in diffractometry [4]). The arrangement applied in most cases and also in the "Norelco" equipment is shown in *fig. 13* [10]). The analyzer is a large flat single-crystal plate, of width about 1 in. and length L about 3 in. In a position of the crystal corresponding to the Bragg angle Θ for a particular fluorescent line (cf. eq. 1) each part of the large specimen will have its radiation reflected by a particular zone of the crystal, as indicated in the figure. The receiving window of the detector must be wide enough to take in all the (nearly) parallel reflected rays. The whole system can be visualized as consisting of a great number of systems acting in parallel narrow beams, separated by the set of parallel thin metal foils (parallel slit

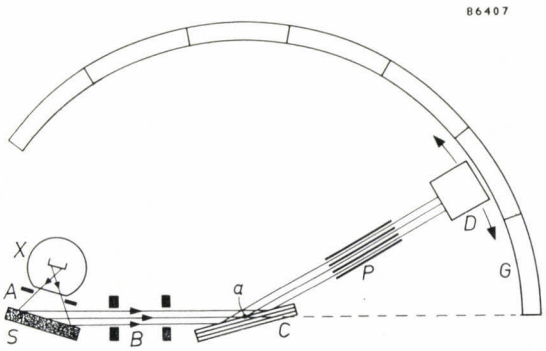


Fig. 13. Non-focussing X-ray optics used for fluorescence analysis of large specimens. S specimen excited by primary radiation from X-ray tube X; C analyzing crystal which is rotated around axis a in order to scan different wavelengths. The detector D is rotated on the goniometer G at twice the angular speed of the crystal. The parallel slit system P ensures the required angular resolution; the latter is approximately determined by the angular aperture 2α of the slits ($\tan \alpha = s/l$, where l is the length and s the spacing of the foils). The slits A and B reduce scattered radiation. (It should be noted that the parallel slit system has a function here different from its function in the diffractometer optics; its orientation is perpendicular to that in the diffractometer [4]). Friedman and Birks [10]) employed tubes in place of flat foils.)

[10]) An arrangement employing similar principles was first described by H. Friedman and L. S. Birks, A Geiger counter spectrometer for X-ray fluorescence analysis, Rev. sci. Instr. **19**, 323-330, 1948.

system). which are positioned between crystal and detector. The role of these metal partitions can be explained as follows (*fig. 14*). Imagine the incident and reflected rays for one value of angle Θ as material lines rigidly fixed to each part of the analyzing

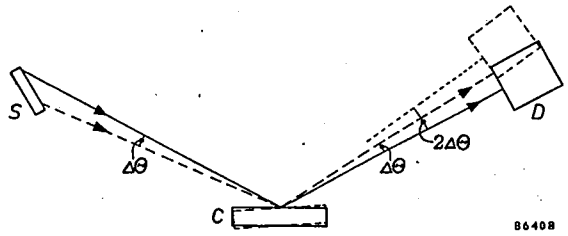


Fig. 14. When the parallel slit system is omitted, the analyzing crystal when being rotated will continue to contribute rays reflected at a given angle Θ, within a certain angular region $\Delta\Theta$ determined by the size of the specimen. The region is limited by the introduction of the parallel slit system.

crystal. When the latter is rotated, the lines representing the incident rays will "sweep" over the specimen. These rays will continue to "see" the source within a certain angular region $\Delta\Theta$ of the crystal. This region would be the approximate width of the recorded fluorescent line if the partitions were not present, since the wide-window detector, although it is rotated through double the angle $\Delta\Theta$, would continue to receive part of the reflected rays. The parallel slit system, however, limits the recorded line width to a reasonable value and therefore provides an acceptable angular resolution, dependent on the angular aperture 2α of the slits. For a slit system with $2\alpha = 0.58°$ a line width of $0.69°(2\Theta)$ was obtained in a typical case. Reducing 2α by a factor of 4 (to 0.15°) improves the resolution by a factor of 3 (line width 0.23°), which is sufficient for most analytical purposes. It also improves the peak-to-background ratio by a factor of 3, but it decreases the intensity by a factor of 2.

If only very small specimens are available, the fluorescent intensity will be much smaller because a much smaller part of the primary radiation is intercepted; it is then necessary to use a larger segment of radiation from the specimen in order to obtain enough counts in a reasonable time. A focussing arrangement as shown in *fig. 15* can then be used. The analyzing crystal has a curved reflecting surface. The specimen (which is a "point" or "line" source in this case), the reflecting surface of the crystal and the receiving slit of the detector all lie on a fixed circle of radius r. To obtain the best focussing [11]) the crystal is first bent so that the reflecting atomic planes conform to a circle of radius $2r$, and then the reflecting surface of the bent crystal is ground to the radius r. Scanning of the spectrum is effected by moving both the crystal and the detector around the focussing circle, the former at half the speed of the latter. The detector with its scatter slit is swivelled around the center of the receiving slit by means of a wire-and-pulley device during the movement, so that the detector keeps seeing the whole crystal. This type of X-ray optics has been used for fluorescence analysis of samples in milligram quantities [12]).

The focussing arrangement just described can also be used with large specimens. The tiny specimen of fig. 15 is then replaced by a fine slit which transmits a segment of radiation from the large specimen placed behind it. This, however, does not give a higher intensity or better resolution than the nonfocussing arrangement of fig. 13 using a flat crystal.

The detector

The detectors in current use for X-ray spectrochemical analysis are basically the same as used in diffractometry and will be described in a separate article in this Review, so that a few notes will be sufficient here.

The Geiger counter tube has been the detector most commonly used in X-ray diffraction and spectrochemical analysis [4]). In its modern form, using halogen as a quenching gas, it has a sensitive volume sufficiently wide to detect nearly the entire large beam used in the X-ray optics of fig. 13. Its principal limitation lies in its long dead time (about 270 microseconds) giving rise to counting losses at

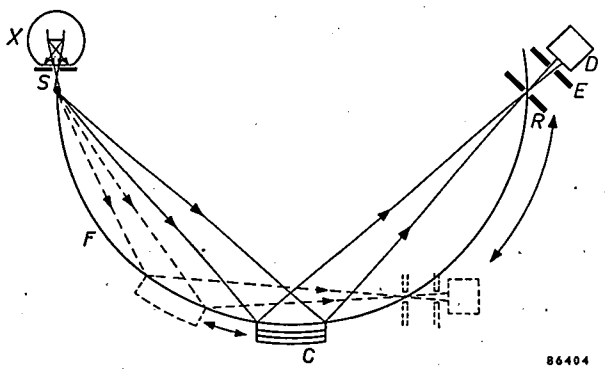


Fig. 15. Focussing X-ray optics used for very small specimens. A *curved* analyzing crystal C is employed. The specimen S forms a line source (perpendicular to the plane of the drawing). S, C and the receiving slit R of the detector, are situated on the focussing circle F. To measure different wavelengths, the crystal C is moved along F at one half the angular speed of R as shown by the dotted lines, and the detector, with its receiving slit R and anti-scatter slit E is swivelled to point at C [12]).

[11]) See, for example, Yvette Cauchois, Recent developments in bent-crystal technique, Trans. Instr. Meas. Conf. Stockholm 1949, 41-45.

[12]) L. S. Birks and E. J. Brooks, Applications of curved-crystal X-ray spectrometers, Anal. Chem. 27, 437-440, 1955 (No. 3).

high intensities, and in its low sensitivity (due to low absorption) for the shorter wavelengths. The proportional counter as developed for X-ray measurements in recent years has about the same spectral sensitivity characteristics as the Geiger counter but a very short dead time (< 1 μsec). The most generally useful detector for fluorescence analysis is the scintillation counter. It has the same advantage as the proportional counter in eliminating the dead time problem, and moreover when using an NaI scintillating crystal it has a nearly uniform and high spectral sensitivity throughout the important wavelength region (*fig. 16*).

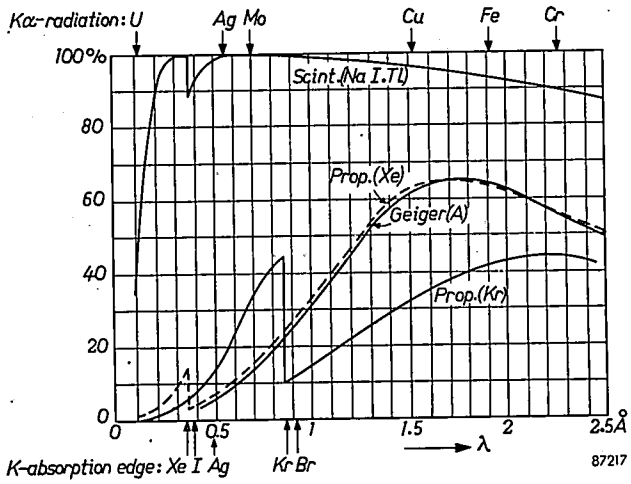


Fig. 16. Calculated quantum counting efficiency (%) as a function of wavelength for several X-ray detectors. Scintillation counter with NaI-Tl crystal 1 mm thick; Geiger counter, "Norelco" 62019, and mica window 13 μ, argon filling at 55 cm Hg, 10 cm in length; proportional counter, side window, beryllium 130 μ + mica 13 μ, xenon at 30 cm Hg, 2.7 cm length; proportional counter, side window, beryllium 130 μ + mica 13 μ, krypton at 50 cm Hg, 2.7 cm length. (From J. Taylor and W. Parrish, Rev. sci. Inst. **26**, 367, 1955.)

For the proportional and the scintillation counter, the pulse height produced by an X-ray quantum is proportional to the quantum energy, i.e. inversely proportional to its wavelength. With suitable circuits for pulse height discrimination, this property can be used to great advantage for increasing the peak-to-background ratio and for separating lines in the case of overlapping spectral lines of different elements (see Appendix).

The analyzing crystal

The diffracting crystal used for analyzing the fluorescent spectra has to meet the following requirements.

It must have a chemical composition such that its own fluorescence caused by the impinging fluorescent radiation will not disturb the measurements. When analyzing at wavelengths less than 3 Å, the analyzing crystal should therefore not contain elements with atomic number above 20; the K-radiation of the crystal elements in the analyzer will then have large wavelengths and will be completely absorbed in the air before reaching the detector. When analyzing at longer wavelengths, higher atomic number elements may be used in the analyzing crystal provided that their L-spectra do not occur in the region which is being analyzed (their K-spectra are not excited in this case). The analyzing crystal must be large enough to allow the use of wide beams as shown in fig. 13. Moreover, the reflecting surface must be quite perfect, i.e. free of distortions caused by grinding or cleaving, and the crystal should not have a pronounced mosaic structure, in which several portions have somewhat different angular orientations. Such a structure would spread the angular width of a line corresponding to the spread of orientation of the mosaic portions, it would also reduce the peak intensity and it might even cause the reflections to split up into multiple peaks, thus seriously complicating the interpretation of the spectrum [13]).

Besides these basic requirements, the analyzing crystal should have a suitable lattice spacing d for the diffraction. For a given wavelength λ, the reflection angle 2Θ will be larger the smaller d, according to eq. (1). The d-value should be small enough to make the angle 2Θ greater than approximately 10° or 15° even at the shortest wavelength used; otherwise excessively long analyzing crystals would be needed in order to prevent the direct fluorescent beam from entering the detector (see fig. 13). A small d-value is also favorable for producing a large dispersion $\mathrm{d}\Theta/\mathrm{d}\lambda$ of the spectrum, to give good separation of adjacent lines; this is seen by differentiating eq. (1):

$$\frac{\mathrm{d}\Theta}{\mathrm{d}\lambda} = \frac{n}{2d\cos\Theta}. \quad \ldots\ldots \quad (3)$$

On the other hand, a small d-value will set an upper limit to the range of wavelengths to be analyzed, since at $\lambda = 2d$, the angle 2Θ will become 180°. The λ-limit is in fact even lower than $2d$, because the reflection angle range of the goniometer is not 180° but is mechanically limited to about 150°. It will therefore be necessary to select a larger d-value for longer wavelengths.

[13]) The analyzing crystal should not — and in practice never will — be perfect on an atomic scale: slight imperfections in the arrangement of the atoms are required in order to avoid extinction effects which diminish the intensity of the reflection; see for example R. W. James, The optical principles of the diffraction of X-rays, G. Bell and Sons, London 1948.

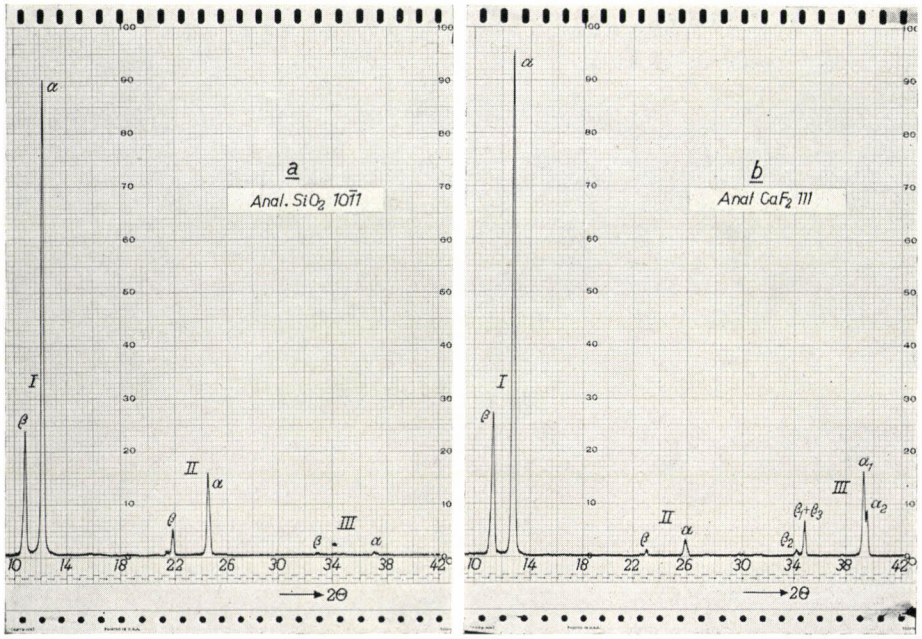


Fig. 17. Molybdenum K-spectrum excited by a tungsten target X-ray tube and reflected by different analyzing crystals. (Nickel filter placed on window of X-ray tube to eliminate scattered L-spectrum of tungsten.) *a*) Reflection from quartz cut parallel to $10\bar{1}1$. *b*) Reflection from fluorite, cleavage plane 111. The first, second and third orders are indicated by Roman numerals.

A final important requirement refers to the higher order reflections caused by the analyzing crystal. The second order reflection of a radiation with wavelength of $\lambda/2$ and the third order of $\lambda/3$ will occur at the same angle 2Θ as the first order reflection of λ. This may result in overlapping of lines originating from different elements and also in a lower peak-to-background ratio when a strong continuum is scattered by the fluorescent specimen. It is desirable in these cases to select a crystal which has very weak or zero second order reflection. An example of this would be silicon or fluorite (CaF_2), both cut parallel to the 111 plane. Both crystals have a third order reflection which is approximately 1/10 as strong as the first order. *Fig. 17* shows the difference in the spectra obtained with crystals having different relative intensities of higher order reflections. *Table II* gives a list of crystals commonly used for X-ray spectrochemical analysis.

Specimens

In the discussion of the X-ray optics it was stated that a large fluorescent specimen surface is desirable for obtaining high intensity reflections with the flat analyzing crystal arrangement. This arrangement can be used, however, also for smaller size samples, down to a few mm diameter. In this case the primary X-ray beam should be stopped down to the actual specimen size in order to avoid an excess of scattered radiation arising in the vicinity of the specimen holder. If only very small amounts of samples are available, say a few milligrams, it may be necessary to use the focussing crystal arrangement (fig. 15). The specimen may in that case be coated on a thin glass fibre mounted on the focussing circle (or rather, the "focussing cylinder").

The specimen preparation methods for large samples depend on the nature of the specimen, which may vary widely: liquids, powders, solids such as minerals or metals, or thin films can be analyzed. The surface should be flat and clean and the distribution of constituents in it should be

Table II. Some analyzer crystals for X-ray spectrochemical analysis. The first four crystals are most frequently used. For the crystal planes marked * the second order reflection is weak or absent.

Crystal	Reflecting plane used (*hkl*)	Lattice spacing *d* in Å	Reflection angle 2Θ for CuKα radiation	Relative intensity
LiF	200	2.01	45.1°	High
SiO_2 (quartz)	$10\bar{1}1$	3.343	26.7	High
NaCl	200	2.820	31.7	High
CaF_2	111*	3.16	28.2	High
Al	111	2.338	38.5	High
,,	200	2.026	44.8	Med.
SiO_2 (quartz)	$10\bar{1}0$	4.255	20.9	Med.
,,	$20\bar{2}3$	1.375	68.2	Low
,,	$50\bar{5}2$	0.812	143.5	Low
Si	111*	3.135	28.5	High
Mica	002	9.96	8.9	Low

representative of the composition of the whole sample. Adequate polishing and etching are required for metal specimens in order to avoid smearing of one component over another. When making casts of molten metals it is important to avoid contamination by the mould. For liquid specimens it is important to prevent volatilization, e.g., by placing an X-ray transparent lid on the container. Equal particle size is desirable in a mixed powder in order to minimize errors due to differences of absorption of fluorescent X-rays by the particles. Another precaution is to minimize the whole inter-particle absorption effect by preparing the powder specimen in the form of a flat solid wafer in a hydraulic press [14]).

The latter factor, i.e., the attenuation of fluorescent X-rays by absorption within the specimen, is of general importance in the analysis. This effect and the subsidiary one that fluorescent rays emitted by one element may contribute to generating fluorescence of another element, cause the relationship between measured fluorescent intensity and concentration of an element in the specimen to be non-linear. For semi-quantitative analyses it is possible to allow roughly for these effects, and in fact no other analytical method is so useful for approximately analyzing wide variations in composition in so short a time. For accurate quantitative analysis by absolute measurements, however, it would be necessary to calculate precisely the effects mentioned, but this is difficult to do except in the simplest cases [15]). The usual way for accurate quantitative analysis, therefore, as in optical spectrochemical analysis, is to prepare a certain number of reference standards, i.e. specimens of known composition in different concentration regions. The standards are used to prepare working calibration curves of X-ray fluorescent intensity as a function of the concentration of each constituent. Only for combinations of elements with nearly equal absorption are the curves relating X-ray intensity to concentration nearly linear [10]). The greater the differences in the absorption, the closer together in composition should the standards be chosen. "Internal" standards prepared by adding known amounts of an element to the specimen are also used in some analyses [16]).

[14]) Some of the effects of absorption and particle size of powders are discussed by P. M. de Wolff, Part F in "X-Ray Crystallography", Ed. J. Bouman, North-Holland Publ. Co., Amsterdam 1951.

[15]) See, for example, J. Sherman, Amer. Soc. Test. Mat., Spec. tech. publ. No. 157, Philadelphia 1954.

[16]) J. Adler and J. M. Axelrod, Spectrochim. Acta 7, 91-99, 1955 (No. 2).

Performance and limitations of the method

The accuracy of the concentrations thus determined will depend not only on the specimen preparation and the calibration curves, but also on the stability of the X-ray source and on the counting statistics involved in the intensity measurements: for good accuracy a large number of counts must be accumulated (this factor and the "counting strategy" were discussed at some length in [5])). In the largest specimens used for the flat-crystal method (fig. 13), the major constituents usually produce a large enough line peak intensity to accumulate in one minute (of time) the number of counts required to give an accuracy of 1% or better in the analysis. The lower limit of concentrations that can be determined by the method is set by the peak-to-background ratio; it will be obvious from the preceding sections that this ratio will be very different for different cases. With some preliminary wet chemical or physical extraction methods to enhance the peak-to-background ratio, analyses of minor constituents in the region below 10^{-3}% may be made with fair accuracy. Successful analyses of constituents of a few parts per million have been reported, and with the focussing arrangement of fig. 15, microgram quantities have been determined in milligram samples with standard deviations of the order of 10 to 15%.

A comparison with optical spectrochemical analysis is useful for an assessment of the value of X-ray spectrochemical analysis. Such a comparison, of course, will reveal different aspects for each specific analytical problem for which the alternative of both methods is offered. In general, it may be said that the X-ray method cannot compete with the optical one in the analysis for trace elements: the optical method will in some cases go down to concentrations of one part in 10^7. On the other hand, the optical method is less sensitive to differences in the concentration of a major constituent, which are most easily measured by the X-ray method. Another general aspect is the range of elements for which the methods are useful: The optical method is generally used for about 65 or 70 elements scattered all over the periodic system; it is not readily applicable to electronegative elements such as the halogens, sulphur, etc., which have spectra in the far ultraviolet. The X-ray method is useful for all elements above atomic number about 13 but cannot at the moment be applied to the lower end of the periodic table for reasons discussed more fully below. A general advantage of the X-ray method is its non-destructive nature and its ability to cope with all kinds of specimens. It can be applied

to crystalline as well as to amorphous materials, and the specimen can be in the liquid or solid state, in powder or glass or metal form, with constituents dispersed or dissolved. This compares favorably with the optical method on which the specimen imposes certain limitations. From a methodological point of view, finally, it is interesting to compare the ways of handling overlapping spectra. The only way of handling this problem in the optical case is the use of high dispersion, and owing to the complicated character of most optical spectra, this might necessitate the use of grating spectrographs of very great length. In the X-ray case, the spectra, being much simpler, cause much less overlapping, and moreover there are several methods other than increased dispersion for separating adjacent lines. These methods, which in themselves form an interesting survey of the physical phenomena involved in X-ray spectroscopy, are described in the Appendix to this article.

The main *limitations* of X-ray spectrochemical analysis lie in the excitation and handling of radiation of very long or very short wavelengths. Summarizing and amplifying what has been said about these limits in the above, it can be stated that the X-ray fluorescence method is useful for all elements in the periodic system from about 13 Al to 92 U. Instrumentation is simplest and work easiest for the wavelength region between about 2.2 and 0.3 Å, comprising the K-spectra from 25 Mn to 60 Nd and the L-spectra from 62 Sm to 92 U (see *Table III*).

The K-spectra for elements 61 to 92, although even simpler than the L-spectra and therefore desirable, cannot at the moment be used, since for their excitation high primary target voltages (45 kV to 116 kV) would be necessary, for which the available X-ray generators are not suitable. Moreover, the radiation protection problems would be much more serious than at voltages below 50 kV. Finally, even with the smallest d-values for the analyzing crystal that can be selected, the reflection angles would be very small (see Table III).

For the elements 13 Al to 24 Cr, the K-spectra lying between about 9.0 and 2.3 Å have to be utilized. Radiation of this wavelength region (which also contains some L-spectra and M-spectra of higher atomic number elements) is strongly absorbed in air and will therefore necessitate the helium or vacuum path technique described above.

The K-spectra of elements of atomic number below 13 lie in the ultra-soft X-ray region, which is hardly accessible at the moment. Unfortunately, this group comprises the elements carbon, nitrogen, and oxygen, which are so important for organic and biological studies.

Applications of X-ray spectrochemical analysis

It is impossible to list here the many successful applications of the method to analytical chemical problems in industry. The method has been applied to a wide variety of metallurgical problems, e.g., for the analysis of steels and many types of alloys; in mineralogy for the analysis of ores, minerals and concentrates; and in the petroleum industry for the analysis of sulphur in oil and of bromine in liquid hydrocarbons. A few typical examples are mentioned below.

Birks and others, in 1950, applied the X-ray fluorescence method to the quantitative determination of tetraethyl lead and ethylene bromide in aviation gasoline [17]). A one-minute count on the Pb Lα line in a sample containing about 4 ml tetraethyl lead per gal. gave the tetraethyl lead content with a probable error of $\pm$ 0.06 ml per gal. Similarly, a one-minute count on the Br Kα line gave the ethylene bromide content with a probable error of $\pm$ 0.16 ml per gal. in a 1.8 ml per gal. concentration. It was thus possible to make 10 to 20 determinations per hour for both Pb and Br with no special specimen preparation of the liquid and

[17]) Anal. Chem. 22, 1258-1261, 1950.

Table III. Major analytical regions determined by wavelengths and instrumentation.

Wavelength region (Å)	Lowest exciting voltages (KV)	Elements	Spectra (α-lines)	Path	Lattice spacing of analyzer	Detector
9-2.3	1.4-6.5	13 Al - 24 Cr 34 Se - 61 Pm 69 Tu - 92 U	K L M	Helium, or vacuum with thin windows	d = 4.3 Å reflects Al Kα at 150°	Prop. counter with thin window; scint. counter below 4Å
2.2-0.3	6.5-44	25 Mn - 60 Nd 62 Sm - 92 U	K L	Air	d = 1.5 Å reflects Ba Kα at 15°	Scint. counter; prop. and Geiger counter above 1 Å.
0.3-0.1	45-116	61 Pm - 92 U	K	Air	d = 1.5 Å reflects U Kα at 4.9°	Scintillation counter

achieve the same accuracy as conventional chemical methods [18]). X-ray tubes, detectors and techniques recently developed would give even higher accuracy in less time. The method is now widely used for automatic plant control because it is so rapid, accurate and simple.

Small amounts of Ni, Fe and V have a deleterious effect in cracking catalysts. Standard chemical and optical spectrographic methods were found to lack the precision to distinguish between good and poor catalysts. Dyroff and Skiba [19]) developed an X-ray fluorescence method for determining these contaminants in silica-alumina catalysts which took about 15 minutes for a complete analysis on samples as small as 2 grams. They achieved a precision of 3% for Fe present in the 0.1 to 1.0 weight % range, and 0.002% (absolute) for Ni and V in the 0.002 to 0.10 weight % range.

Davis and Van Nordstrand [20]) used the fluorescence method for rapid and accurate control analysis for lubrication oil blending. They obtained an accuracy of 2% to 3% of Ba and Ca in concentrations about 0.05%, and 1% to 2% accuracy of Zn in concentrations greater than 0.005%. It was necessary to prepare several calibration curves for Zn in the various Ba concentrations due to the absorption effects discussed earlier [14]). The time for the analysis of each element was 3 to 12 minutes; this is only a fraction of the time required for standard chemical methods, which give no better accuracy.

X-ray fluorescence has proved to be a reliable method for the analysis of Nb and Ta in ore minerals and for following the increase in concentrations during mineral dressing processes [21]). It takes a competent chemist 5 to 15 days to make a chemical analysis for these elements, while 10 to 50 samples can be analyzed with comparable accuracy in a day by the X-ray method.

The method has been applied to small samples, trace element analysis and disperse systems. Thin evaporated films of Ni, Fe and Cr (about 100 Å thick, corresponding to surface densities of 1 to 100×10^{-6} gm/cm^2 have been analyzed by Rhodin [22]) with an accuracy of 2% or better, the results being in excellent agreement with those obtained by microcolorimetric methods.

These are only a few of the applications of the method. The reader may consult recent volumes of Analytical Chemistry (1950 et seq.) where many specific applications have been published.

To conclude this article, a threefold development of the application of X-ray spectrochemical analysis can be anticipated: a widening of the field within the wavelength regions 0.3 to 2.2 and 2.3 to 9 Å, to include even more analytical problems; a possible extension of the wavelength limit beyond 9 Å, a region where instrumental difficulties become very great, but which is important because it embraces elements involved in organic and bio-chemistry; lastly an intensification in the routine application of the method to specific cases. In fact, the method is capable of a considerable degree of automation, as has been shown by the design of several completely automatic instruments. One of these, the North American Philips "Autrometer", is shown in *fig. 18.*

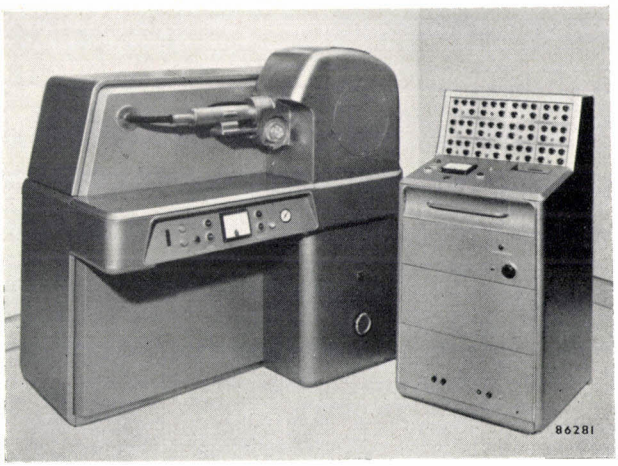


Fig. 18. The "Norelco" "Autrometer", developed for automatic X-ray spectrochemical analysis in industrial plant control. The X-ray unit is on the left and the electronic circuits unit on the right.

In this instrument, the crystal and detector, arranged according to fig. 13, are automatically stepped in sequence to a number of preselected Bragg angle positions and the intensities recorded. Once such an instrument has been calibrated, the operation is simple and analyses are performed rapidly so that a very effective solution to industrial production control problems is offered.

Appendix: Methods of handling overlapping lines

Table IV shows some cases of closely spaced X-ray spectral lines. The methods of handling such cases will be illustrated using niobium and tungsten as an example.

The Nb Kα line first order reflection and the W Lα line, obtained with a flat analyzing quartz crystal ($10\bar{1}1$ plane, $d = 3.34$ Å) and using a parallel slit assembly of angular

[18]) In contrast to X-ray *absorption* analysis, other additives and impurities were found to have little effect on the accuracy of the measurements; see Anal. Chem. **22**, 1238-1248 and 1248-1258, 1950.
[19]) Anal. Chem. **26**, 1774-1778, 1954.
[20]) Anal. Chem. **26**, 973-977, 1954.
[21]) Anal. Chem. **26**, 800-805, 1954.
[22]) Anal. Chem. **27**, 1857-1861, 1955.

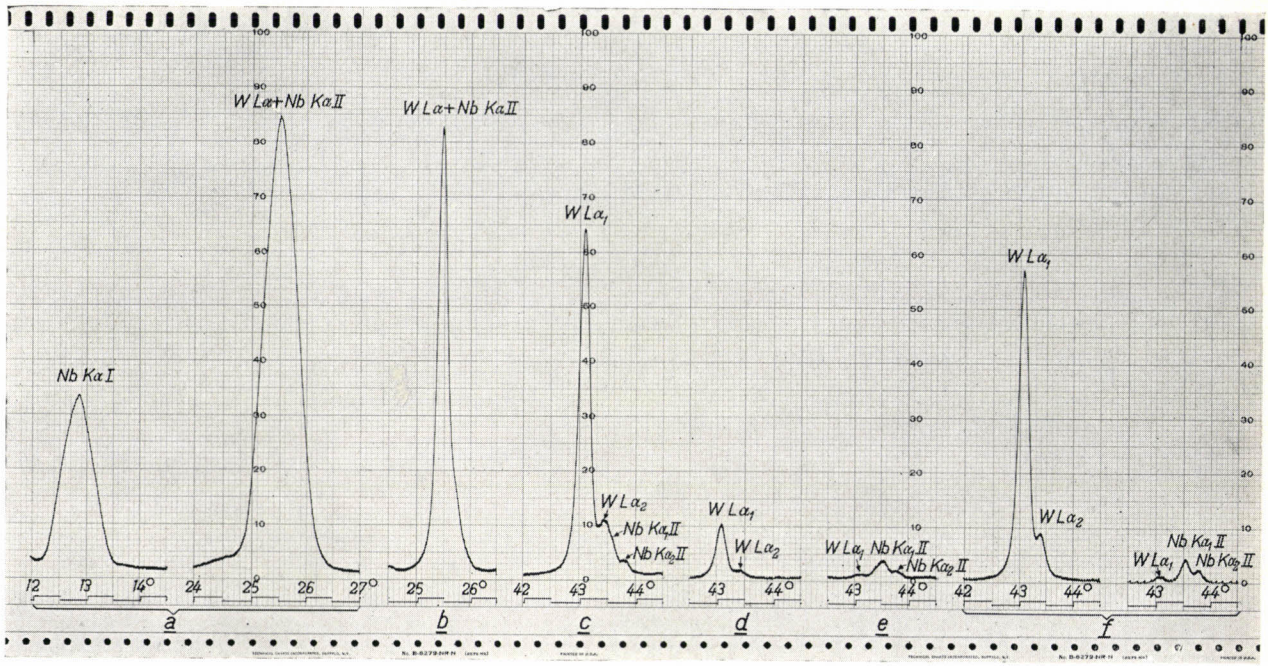


Fig. 19. Methods of separating overlapping spectral lines. The specimen contains WC 72%, TiC 10%, NbC 10%, Co 8%; i.e. approx. W 67.6% and Nb 8.9%. Tungsten target X-ray tube, 50 kV peak, 45 mA (except for method *d*, see below). Full scale intensity = 6400 counts/sec for (a) and 3200 counts/sec for (*b* - *f*). Scanning speed $\frac{1}{2}°$ per min, time constant 4 sec, scintillation counter.
(*a*) Nb Kα and W Lα + Nb Kα second order reflected from quartz $10\bar{1}1$ ($d = 3.34$ Å) with parallel slit system aperture $2\alpha = 0.58°$.
(*b*) W Lα + Nb Kα second order as in (*a*) but $2\alpha = 0.15°$.
(*c*) W Lα and Nb Kα second order reflection from lithium fluoride, 200 ($d = 2.01$ Å), $2\alpha = 0.15°$.
(*d*) Same as (*c*) but X-ray tube operated at 19 kV peak, 33 mA.
(*e*) Same as (*c*) but 0.0007″ nickel filter placed between specimen and crystal.
(*f*) Same as (*c*) but pulse height analyzer set for W Lα and then for Nb Kα. The W Lα line is nearly but not completely eliminated.

aperture 0.58°, are shown in *fig. 19a*. The second order reflection of Nb Kα lies in the right hand side of the foot of the W Lα line, thus prohibiting a direct comparison of the relative line intensities of Nb and W.

If we increase the angular resolution of the instrument by reducing the aperture of the parallel slits to 0.15°, we obtain the line profile in fig. 19*b*. The second order Nb Kα line now shows as a slight hump on the lower right hand side of the WLα line, but the separation is still not complete.

If we now replace the quartz crystal with a LiF crystal, using the cleavage surface (200 plane, $d = 2.01$ Å), the larger angular dispersion separates the NbK$\alpha_1\alpha_2$ lines in the second order from the WLα line, fig. 19*c*; relative intensity measurements can now be made with a fair accuracy although the overlapping is not yet completely eliminated.

Increasing the resolution and dispersion is a familiar expedient in optical spectrographic analysis, although larger instrumental efforts for its realization are required than in the X-ray method. The following methods of handling overlapping lines are unique for the X-ray case.

The excitation of the Nb K radiation can be avoided by lowering the X-ray tube voltage: the excitation potential of Nb Kα is approximately 19 kV, whereas the W Lα line appears at 10 kV. The effect is shown in fig. 19*d*. Owing to the full-wave rectified operation of the X-ray tube (cf. fig. 10) and its limited current capacity, the intensity of the W Lα line has been reduced by about a factor of 8 by the lowering of the voltage.

An overlapping line can in some cases be eliminated by using a selective filter on the X-ray detector. This possibility is due to the fact that the absorption edges of elements are located at wavelengths which vary systematically with the atomic

Table IV. Sources of closely spaced spectral lines. Roman numeral indicates order of reflection.

Source	Lines	Wavelengths in Å
Elements close together in atomic number	25 Mn Kα_1	2.103
	24 Cr Kα_1	2.085
	45 Rh Lα_1	4.374
	46 Pd Lα_1	4.368
K-spectra of low atomic number overlap L-spectra of high atomic number	30 Zn Kα_1	1.435
	75 Re Lα_1	1.433
	21 Sc Kα_1	2.780
	56 Ba Lα_1	2.776
Higher orders of K-spectra of high atomic number overlap K-spectra of low atomic number	28 Ni Kα_1	1.500 (I)
	41 Nb Kα_1	1.494 (II)
	42 Mo Kα_1	2.127 (III)
	33 As Kα_1	2.114 (II)
Higher orders of K-spectra overlap L-spectra	47 Ag Kα_1	1.677 (III)
	70 Yb Lα_1	1.672 (I)
	41 Nb Kα_1	1.494 (II)
	74 W Lα_1	1.476 (I)

number, much in the same way as the emission lines. It is therefore possible in many cases to pick an element having an absorption edge which will affect one spectral line of the specimen much more than another line close to it. For the Nb + W sample, a nickel filter can be used: the nickel, having its K-absorption edge at 1.48 Å, selectively absorbs the W Lα radiation ($\approx$ 1.43 Å) while hardly affecting the harder Nb K radiation. With a Ni-filter approximately 0.02 mm thick, the W Lα-line is almost completely eliminated, the NbKα lines being reduced in intensity by only a factor of about 2; see fig. 19*e*.

The final and most effective solution to the problem of overlapping lines is the pulse height discrimination technique. It is based on the fact that the pulse height generated in a proportional or scintillation counter by the absorbed X-ray quantum is proportional to the quantum energy, subject to a certain spread resulting in a finite "energy resolution" of the detector [23]). Whenever the wavelength difference of two overlapping lines is larger than corresponds to this energy resolution, the lines can be separated by means of the electronic "window" of the pulse height analyzer. The two curves of fig. 19*f* illustrate this point: In the left curve the pulse height analyzer was set for the W Lα-radiation and it is seen that the W L$\alpha_1\alpha_2$-lines are obtained while no Nb radiation is recorded; in the right curve the analyzer was set for the Nb Kα radiation — only very little W Lα-radiation is recorded in this case. It should be noted that this method not only permits complete separation but also gives the highest intensities in both lines and is very easy to perform.

The three latter methods of handling overlapping lines — which were stated to be unique for the X-ray case — can effectively be applied to the case where the whole spectrum "overlaps", i.e., where the spectral lines are not dispersed by an analyzing crystal at all [24]). Since the detector can be placed very near to the specimen, the intensities are enormously greater than in the dispersive (diffraction) type of analysis, but on the other hand, it is not possible to eliminate scattered radiation so that the background of the measured intensities is relatively high. Non-dispersive analysis, therefore, is restricted to specimens with only a few constituents present in not too small concentrations.

Summary. In X-ray spectrochemical analysis, elements are identified and their concentrations determined by means of the wavelengths and intensities of lines in their characteristic X-ray spectra. This method, which in principle is similar to the well-known optical spectrochemical analysis, has several important advantages: It is non-destructive; it can readily be applied to all kinds of specimens, solid, powder or liquid, crystalline or amorphous; the X-ray spectra are very simple, cases of overlapping lines of different elements are not frequent, and a number of methods are available for separating or distinguishing them. The instruments developed for the method are simple to operate, so that it is quite possible that in the future X-ray spectrochemical analysis will be used for routine analytical work as extensively as the optical spectrochemical method.

The X-ray spectra are usually excited by way of fluorescence, induced by the continuous or characteristic radiation of an X-ray tube with tungsten or other target. A number of details concerning the primary target, the voltage and mode of operation of the X-ray tube and the nature of the fluorescing specimen are discussed in the article. The X-ray spectra are analyzed by diffraction from a single crystal; a Geiger counter, proportional counter or scintillation counter is used for measuring the reflected rays. The instrumentation thus is very similar to that used in X-ray diffractometry. A discussion of the X-ray optics for the fluorescence analysis and of the desirable characteristics of the analyzing crystal, the specimens and the standards used is also given.

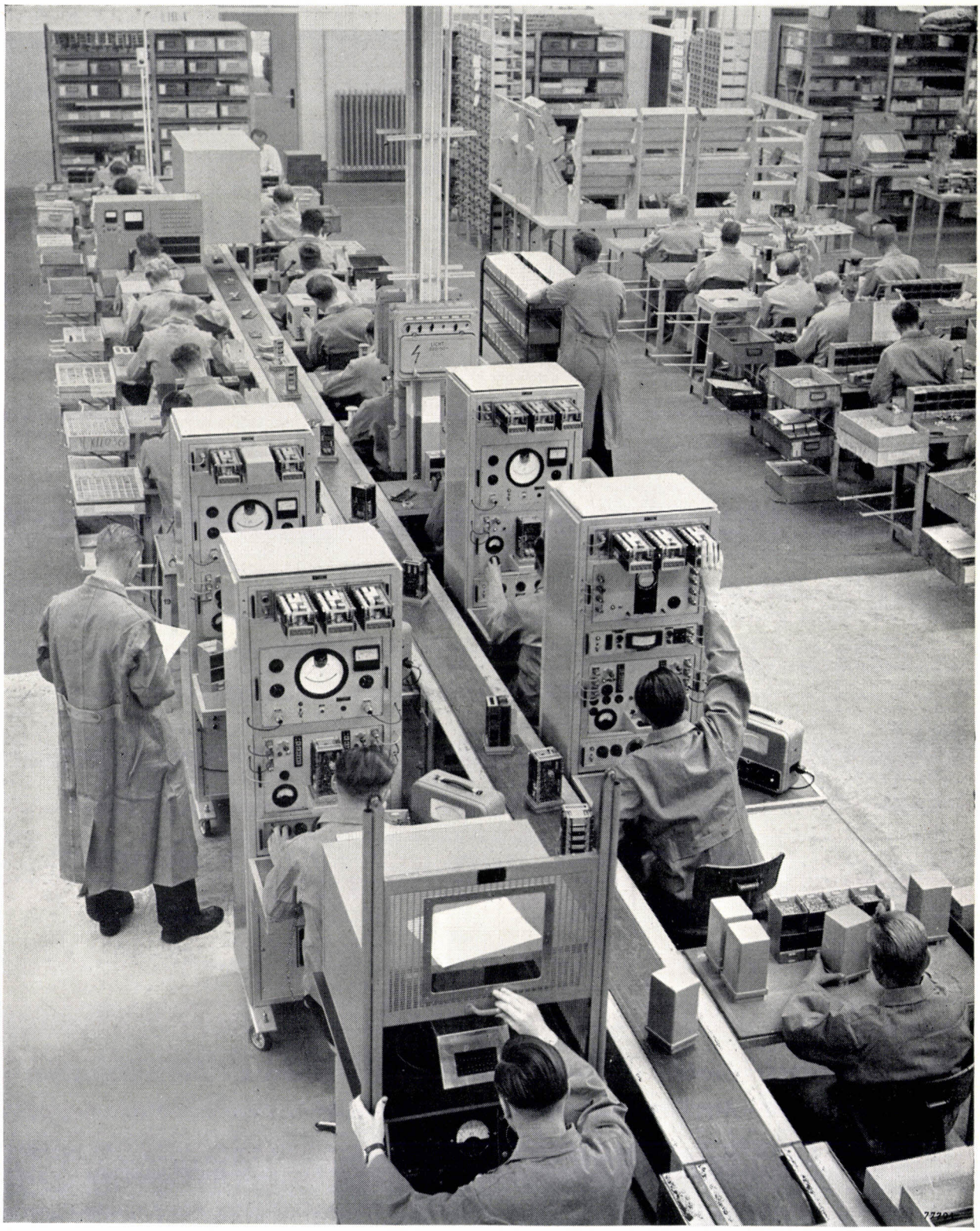
The method is capable of good precision within reasonable time: quantitative analyses to within 1% or less can often be performed in a few minutes. Minor constituents down to 10^{-3}% or less have in some cases been determined. Specimens of a few milligrams can be handled when special focussing arrangements are used. The method is useful for all elements of atomic number between 13 (aluminum) and 92 (uranium); using a 50 kV X-ray tube, the K-radiation of elements 13 to 60 and the L-radiation of elements 34 to 92 can be excited. For the elements 13 - 24 the K-spectra lie between about 9 and 2.3 Å, and owing to the absorption in air, special techniques are required for these very soft X-rays.

---

[23]) A. R. Lang, Wavelength resolution of X-ray proportional counters, Proc. Phys. Soc. **A65**, 372-373, 1952; C. F. Hendee and S. Fine, Moseley's law applied to proportional counter resolution of adjacent elements, Phys. Rev. **95**, 281-282, 1954 (1 July).

[24]) See, for example, H. Friedman, L. S. Birks and E. J. Brooks, Amer. Soc. Test. Materials, Spec. tech. Publ. No. 157, Philadelphia 1954.

SERIES PRODUCTION OF CARRIER WAVE TELEPHONY EQUIPMENT



In the short-haul carrier telephony system STR 112, developed by Philips Telecommunication Industries at Hilversum *), an entirely new design is now used for the speech channel units. All the components and circuitry for each channel (including signalling circuits) are built into an airtight box (so-called "conclave" construction). This photograph shows the manufacture of these speech channel units on a moving belt. Along the belt (moving towards foreground of picture) the operations are, successively: wiring of the chassis, mounting of resistances and coils, adjustment of filters and coils and the assembly of the unit. In the middle foreground are four testing racks for measurements and adjustment of the units. At the end of the moving belt (foreground) are the high tension test rack and the final inspection.

*) Comm. News **14**, 78-127, 1954, The photograph is taken from this publication.

A NEW FLUORESCENT LAMP IN A STARTERLESS CIRCUIT

by W. ELENBAAS and T. HEHENKAMP.

621.327.534.15

The fact that most tubular fluorescent lamps require a starter switch, and consequently flicker slightly during the starting period, is a disadvantage of these lamps, which otherwise offer important advantages. This objection is overcome in a new lamp ("TL"M lamp) and the associated ballast arrangement. Lamp and ballast are so designed that ignition is ensured even at low temperatures.

Introduction

In recent years there has been a general effort to go over to circuits for fluorescent lamps in which there is no need for a starter. This is a consequence of a desire for quicker and steadier ignition and, at the same time the elimination of a component that increases the risk of breakdowns. It is true that modern glow discharge starters are reliable and durable but, after all, their lifetime is not unlimited. Moreover it is not always easy, without measuring instruments, to find out whether a breakdown is due to the lamp or to the starter.

Even in the early years of the fluorescent lamp, attempts were being made to devise starterless circuits. Starterless outfits that were marketed at that time clearly possessed too many disadvantages, so they were adopted only on a limited scale and some of them disappeared soon afterwards. One example was the "Instant-Start" equipment, which was employed on a limited scale in America; this supplied a voltage of about 450 V for igniting the special "Instant-Start" 40 W lamps [1]). Here, the elimination of the starter brought with it the drawbacks that the ballast equipment had to be of greater size and higher price and caused bigger electrical losses, and that, since the lamp electrodes were not heated before ignition, electrodes of a special type had to be used.

Another example is the resonance device (*fig. 1a*) developed by Philips in 1939 for a 25 W lamp [2]). On switching on, the electrodes are heated by a relatively high current and at the same time a voltage of about 350 V (r.m.s.) develops across the lamp, thus ensuring quick ignition Here we shall give a brief explanation of the way the device works, in order to make clear the disadvantages of the circuit and the simplifications to it which will be discussed presently.

The circuit is to be regarded as a variant of the well-known Steinmetz circuit shown in fig. 1*b*. The self-inductance L_1 and the capacitance C_1 are in resonance at the frequency of the AC supply. For a supply voltage $E\sqrt{2}\sin\omega t$, the circuit has the following properties: the current through the impedance Z is $E/j\omega L_1$ and is therefore independent of Z; if, moreover, the reactive part of Z is equal to ωL_1, the input current is in phase with the supply voltage (power factor = 1). The first property has the consequence that, whenever the circuit is open at Z, the voltage across the break becomes very high. All this makes the circuit very attractive for employment with a fluorescent lamp, the lamp (in series with a choke coil) taking the place of the impedance Z.

In explaining the circuit of fig. 1*a* we may ignore the bimetal relay B, which will be dealt with presently, and likewise the capacitor C_2, the sole purpose of which is to suppress radio interference.

On switching on, the position is as illustrated in fig. 1*c*. The electrodes are fed in series by a heating current I_h, which passes through the two coils S_1 and S_2 wound on the same core (each coil separately having the self-inductance L_2) in such a way that the resultant self-inductance is zero. Thus, if r represents the total resistance of the system and L_1 and C_1 are in resonance with the supply frequency, the current I_h is given by

$$I_h = \frac{E}{r}.$$

After ignition the position is that of fig. 1*d*. Connections may now be considered to have been made as indicated by dotted lines in the diagram, with the result that the circuit is transformed into that of fig. 1*b*; the lamp current I_{la} thus becomes

$$I_{la} = \frac{E}{j\omega L_1}.$$

[1]) R. N. Thayer and D. D. Hinman, Requirements for reliable instant-starting fluorescent lamps, Illum. Eng. **40**, 640-658, 1945.

[2]) Netherlands patent No. 55 200.

The circuit parameters are such that $|I_h| \gg |I_{la}|$ (for rapid heating of the electrodes) [3]) and at the same time the voltage across the lamp is high (350 V). This results in quick and reliable ignition of the lamp.

On account of the war it was not until 1945 that the circuit of fig. 1*a* was adopted on a fairly large scale, but it was unable to maintain its position for more than a few years. Here too, large size and high price were the objections and, these apart, the excess current safety device constituted a troublesome complication. This device was necessary because the state of resonance continued to exist when, at the end of the lamp's life, it failed to ignite; the heavy current I_h then persisting would have overloaded the equipment had it not brought the bimetal relay (*B* in fig. 1*a*) into operation. The contact of this relay, normally open, then short-circuits the choke coil L_1, reducing the current to a low value once again.

Two trends are discernible in the recent development of the fluorescent lamp: one towards special electrodes that are not pre-heated, the other towards more normal electrodes that are. An example of the first type is the Philips "TL"S 40 W lamp [4]), in which an incandescent lamp alone constitutes the whole of the ballast equipment. The second type has come to the fore in various countries [5,6,7]) and is also represented in the new Philips "TL"M lamp, which will be discussed in this article. In both cases the ignition voltage of the lamp has been brought down by various means to a value considerably

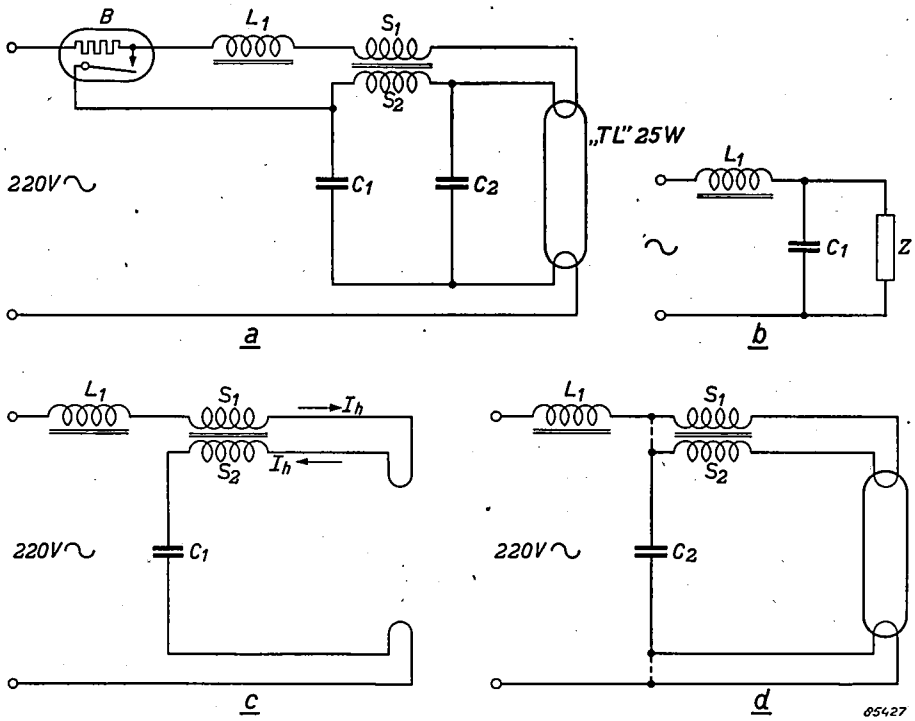


Fig. 1. *a*) Circuit diagram of a ballast for "TL" 25 W lamp, which was on the market in 1945. L_1 choke coil and C_1 capacitor slightly off resonance with the supply frequency. S_1/S_2 double choke. C_2 capacitor for interference suppression. *B* bimetal relay (excess current safety device).
b) Steinmetz resonant circuit. With L_1 and C_1 in resonance with the supply frequency, the current through the impedance *Z* is independent of *Z*.
c) Equivalent circuit of (*a*) before the lamp is ignited (*B* and C_2 omitted). The net self-inductance of S_1 and S_2 together is zero; there is a heavy heating current I_h flowing.
d) Equivalent circuit of (*a*) when lamp is ignited. Connections may be considered to have been made along the dotted lines; the circuit is then effectively that of (*b*), in which *Z* represents the lamp in series with the shunt-connected chokes S_1 and S_2.

[3]) In general it is desirable to keep the resistance *r* low, with an eye to losses. However, since I_h would then take on excessively high values, it is necessary to deviate from resonance somewhat and thus to keep $|I_h| < |E/r|$ (but in any case $\gg |I_{la}|$). Out of resonance, I_{la} is no longer completely independent of *Z*; nevertheless the reasoning remains valid in essentials.

[4]) W. Elenbaas and T. Holmes, Philips tech. Rev. **12**, 129-135, 1950/'51.

[5]) S. Anderson, Control gear for fluorescent lamps, G.E.C. Journ. **17**, 159-177, 1950.

[6]) A. R. Lemmers and W. W. Brooks, New fluorescent lamp and ballast design for rapid starting, Illum. Eng. **47**, 589-594, 1952.

[7]) J. Cates, Fluorescent discharge tube circuits and operating problems, Proc. Inst. El. Engrs., Part II, **100**, 389-400, 1953.

lower than in older types. This has not only opened the way for starterless circuits; it also has the consequence that the lamp is subjected to less damage when switched on.

Means of lowering the ignition voltage

It has long been known that the ignition voltage of fluorescent lamps can be considerably lowered by various means other than the pre-heating of the electrodes [8]). Three methods have actually been adopted in practice:

1) a conductive strip or coating on the inside of the glass wall,
2) an earthed conductive strip on the outside of the glass wall, and
3) earthed metal surfaces in the immediate neighbourhood of the lamp.

The objection to the two last-named methods is that it is a condition for proper functioning that one of the lamp electrodes should have a sufficiently high potential difference with respect to the strip or plate. Now this is usually the case when the mains supply is earthed on one side, but not with non-earthed supplies nor with two phases of a three-phase supply earthed at the star-point. In method (3) this difficulty is sometimes overcome by applying a suitable potential to the metal plate via a resistor, instead of earthing it. This resistor must have a value so high that there is no danger when the plate is touched; this signifies, in accordance with the safety regulations in force in many countries, that the current through the body when contact is made may not exceed 0.5 mA. With one individual lamp it is easy to satisfy this regulation, but it becomes difficult when fittings are mounted in long rows: then the metal parts are interconnected so that the resistors lie in parallel, with the result that it could be dangerous to touch a fitting.

The "TL"M 40 W lamp

For the reason just mentioned, when a new lamp (the "TL"M) was developed the choice lay between an internal and an external strip (the latter connected via a resistor with one of the electrodes). The external strip was chosen, because this (unlike the internal strip) involved no extra energy loss (see the article cited in footnote [4]), pp. 131-132). As we shall presently see, the highest voltage developing across the lamp is about 250 V, so that a resistance as low as 0.5 MΩ gives satisfactory protection against shock. A value of 1 MΩ has actually been chosen, to keep well within the safety limit in all circumstances.

A well-known phenomenon in fluorescent lamps with external ignition aids is that the ignition voltage depends on the humidity of the air [1]). This is also the case to some extent in the lamp with an external strip connected via a resistor to one of the electrodes. In the "TL"M lamp, this phenomenon is completely eliminated by applying a silicone coating to the outside of the glass, which makes the lamp quite insensitive to humidity.

Choice of heating current and no-load voltage

In *fig. 2* the ignition voltage of the "TL"M 40 W lamp is plotted against the heating current of the electrodes for various ambient temperatures. It is possible with the help of this figure to determine,

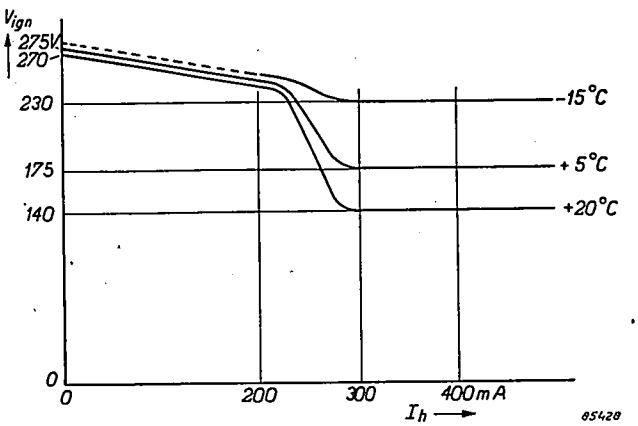


Fig. 2. Relation of ignition voltage V_{ign} of the "TL"M 40 W lamp to the electrode heating current I_h, for various ambient temperatures.

for a given electrode heating current, what no-load voltage is required from the ballast to give reliable ignition of the lamp in all circumstances. To this end, the curve is selected which relates to the lowest temperature to be expected in the field of application of the lamp; this curve is then corrected by a safety factor which allows for both the variations between one lamp and another and the deviations from the nominal supply voltage. The result is then as shown schematically in *fig. 3*. The region above curve *1* is that of reliable ignition.

In order now to narrow down the field of selection, regard must also be paid to other considerations. The most important criterion is the desired performance of the lamp under repeated switching conditions. This means the number of times that a lamp can be ignited before it fails, the burning time being extremely short. In normal practice a lamp will not be switched on more than a few

[8]) W. Uyterhoeven, Elektrische Gasentladungen, J. Springer, Berlin 1938, pp. 92-93.

thousand times during its lifetime; however, the lifetime should not be limited by the criterion of the permissible number of switchings, so that several times this figure should be taken for testing purposes: 10 000 times, say, may be regarded as amply sufficient.

Extensive tests have made it clear that the no-load voltage has a big influence on the switching endurance, while the heating current has not, provided it remains above a certain minimum value. The result of these tests is also shown in fig. 3: the permissible region is that below curve *2*. The hatched-in region between curves *1* and *2* is thus suitable for realising the aim in view.

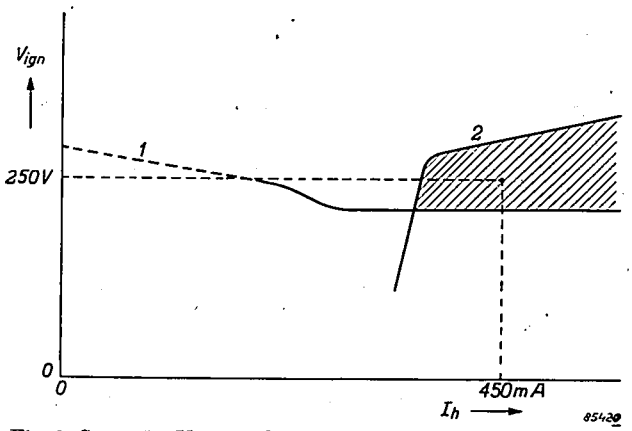


Fig. 3. Curve *1* : V_{ign} as a function of I_h for the lowest ambient temperature at which the lamp must still ignite, modified by a safety factor which takes into account the spread between lamps and mains voltage fluctuations.
Curve *2*: Limit below which the ignition voltage must be kept in order to obtain a sufficient switching endurance. The hatched in area represents the region of dependable ignition with high switching endurance.

In order that the ballast should have small dimensions it is generally desirable to choose the lowest permissible values for the heating current and no-load voltage. However, a higher heating current reduces the ignition time, so that in this connection a compromise must be sought. For the "TL"M lamp the operating point indicated in fig. 3 has been chosen. It is clear that for this point the switching endurance goes well beyond the minimum requirement laid down, but after switching on some thousands of times a certain blackening was found to occur at the ends of the lamp. The fitting of shields around each electrode reduced this blackening effect to normal proportions.

Circuit for heating the electrodes

The current for heating the electrodes may be supplied in two basically different ways: maintaining a constant *voltage* or maintaining a constant *current*.

For constant voltage supply, the electrodes are connected to secondary windings of a transformer which supply them with the correct filament voltage. With this method, the filament current is dependent on the resistance between each electrode and the transformer, in other words, on the leads and contact resistances in the lamp, lampholders and ballast [9]). It is above all the contact resistance in the lampholders which can be of importance, for it frequently occurs as a result of faulty installation (lampholders not properly spaced or angularly displaced with respect to each other) that one or more of the lamp pins make bad contact or none at all with the springs of the lampholders. One of the electrodes is then not heated, or not properly heated, and the consequences are once again blackening of the lamp ends and a sharp reduction in life. Another objection to constant voltage supply is that if a short-circuit occurs between the leads to the lampholder, the filament current transformer becomes overloaded — usually dangerously so.

For constant current supply, as is usual in circuits employing a starter, and also used in some starterless circuits adopted in the past (fig. 1*a*), the two filaments, in series with each other and with a high impedance, are connected to a source of fairly high voltage. The disadvantages referred to above are then absent: the current through the electrodes is practically independent of possible contact resistances, whilst if contact is broken altogether both electrodes remain dead; in such a case the lamp fails to start but is not exposed to damage.

The ballast equipment for the "TL"M lamp

The foregoing will have made clear some of the important requirements to be satisfied by the ballast equipment for the "TL"M lamp: at the time of ignition, a current of about 0.45 A must pass through the electrodes in series, while at the same time there must be a voltage of about 250 V across the lamp.

It was obviously worthwhile to investigate the extent to which the resonant circuit device (fig. 1*a*) could be adapted so as to satisfy these requirements. After all, the device combines several highly desirable properties: constant current filament supply, high power factor, little distortion of the current taken from the mains, and high impedance to frequencies which are adopted for AF control signals [10]) in mains supplies.

---

[9]) W. Calvin Gungle, Electrical characteristics of lamps and ballasts, Illum. Eng. **48**, 579-584, 1953.
[10]) This refers to techniques for the telemetering of information and for remote control via the mains.

A first simplification is that the fairly low heating current and the low ignition voltage of the "TL"M lamp make the excess current safety device superfluous; the bimetal relay can therefore be dispensed with. Further, the choke coil L_1 can be obviated by providing the double choke coil S_1/S_2 with a suitable leakage. The circuit diagram is thus simplified into that of *fig. 4.*

That the properties desired are actually obtained by the means outlined above may be demonstrated with the aid of the equivalent circuits in *fig. 5,* in which the values indicated for the various impedances are those selected for an equipment for a mains supply of 220 V at 50 c/s.

During the starting period the current taken from the mains, which serves to heat the electrodes, is determined by the leakage reactances and the capacitance (fig. 5*a*). This current amounts to 220/(640—80—80) = 0.45 A. The voltage between

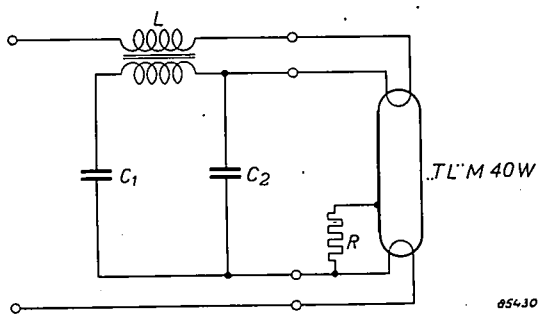


Fig. 4. Circuit diagram of ballast for "TL"M 40 W lamp. The two coils L on one core possess a certain leakage. C_1 resonance capacitor. C_2 capacitor for interference suppression. R resistor of 1 MΩ connecting the conductive strip to one of the electrodes.

the ends of the lamp is thus 0.45 × (640—80) = 250 V. Immediately the lamp ignites, the equivalent circuit of fig. 5*b* becomes applicable. It follows from the vector diagram (*fig. 6*) that the current taken from the mains is practically in phase with the mains voltage, so that the power factor is nearly 1. It may be seen from fig. 5*c* that the higher harmonics set up by the fluorescent lamp in the lamp current arrive at the mains much attenuated, for the circuit acts as a low-pass filter. The current taken from the mains is in consequence practically sinusoidal — an unusual property for a device with a high power factor. The impedance of the whole circuit to audio frequencies such as those used in mains networks for control signals is obviously high, as a consequence of the leakage reactances.

Performance in practice

Practical tests have been carried out on a large number of "TL"M lamps with their ballast equipment in order to find out whether the results came up to expectations.

At 20 °C and the rated mains voltage, ignition is shown to take place evenly and without flickering in about 1.5 seconds. At lower temperatures and

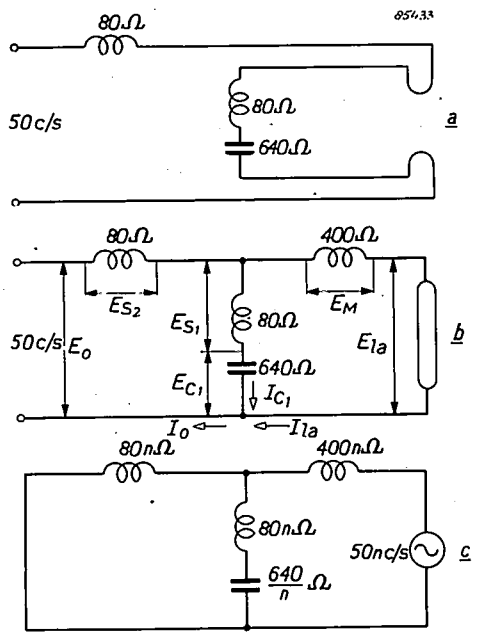


Fig. 5. Impedances occuring in the circuit of fig. 4:
a) before lamp ignites, frequency 50 c/s;
b) with lamp ignited, frequency 50 c/s;
c) with lamp ignited, frequency 50n c/s. Owing to the impedance of 80n Ω in series with the supply, the mains current is distorted only very slightly.

mains voltages this time becomes only a little longer; down to —15 °C, ignition is still reliable at 90% of the nominal supply voltage. The number of switchings before failure occurs amounts on the average to about 20 000. With 3 hours burning per start, which corresponds more to working conditions, the average lifetime is 7500 hours. The life of the lamp usually comes to an end as a result of one of the electrodes melting through; the flow of current through the equipment is thereby interrupted.

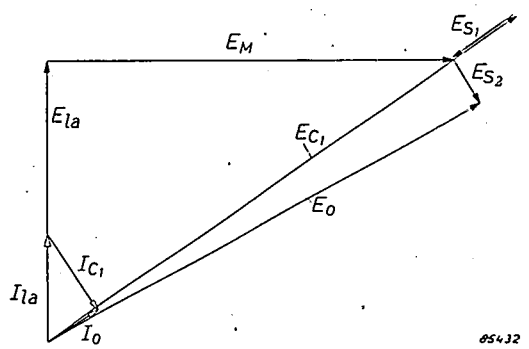


Fig. 6. Vector diagram relating to fig. 5*b* (see this figure for the meaning of the symbols).

The "TL" S 40 W lamp with the ballast equipment of the "TL"M lamp

We should like to mention in conclusion that the ballast equipment described above is also very well suited to the "TL"S 40 W lamp. For this lamp, in series with the associated incandescent lamp as ballast resistor, a mains voltage of 220 V is sufficient for ignition. On the other hand, if a ballast choke coil is employed, a higher voltage is necessary. The reason for this is that the low resistance of the incandescent lamp in the cold state allows a higher glow-discharge current to pass through the "TL"S lamp than does the high impedance of the choke coil; this heavier glow-discharge current helps towards ignition [4]). As we saw above, the ballast of the "TL"M lamp supplies a no-load voltage of 250 V, and this proves to be sufficient to ignite the "TL"S lamp at ambient temperatures not lower than 5 °C. Since the "TL"S lamp requires no heating current, it is possible to reduce the size of the equipment somewhat and also the number of terminals; a special version for the "TL"S lamp originated in this way.

Comparing the "TL"S with the "TL"M lamp, the former is found to ignite more quickly — almost immediately after switching on in fact. Against this, the "TL"M lamp continues to ignite at ambient temperatures as low as —15 °C, exhibits a greater switching endurance and, since the internal strip of the "TL"S lamp is lacking, has a rather higher light efficiency.

Summary. The "TL"M 40 W lamp is a new tubular fluorescent lamp which functions without a starter. The ignition voltage lies below 250 V (r.m.s.) owing to two features: heating of the electrodes when the lamp is switched on and the presence of a conductive strip on the outside of the glass. The strip is connected to one fo the electrodes via a 1 MΩ resistor. The ignition voltage is made independent of the humidity of the air by covering the lamp with a silicone layer. The ballast equipment is a simplified version of a device that was on the market in 1945. It is based on a Steinmetz resonance circuit. On switching on, it supplies a filament current of about 0.45 A to the series-connected electrodes, while simultaneously a voltage of approximately 250 V develops between them. The current taken from the mains is distorted only very slightly, the power factor is nearly 1 and the input impedance for audio-frequencies is high. At an ambient temperature of 20 °C and the rated mains voltage, the lamp ignites without flickering in about 1.5 seconds. At —15 °C ignition is still reliable at 90% of the rated supply voltage. The switching endurance amounts on the average to about 20 000 times. Allowing 3 hours burning for each time the lamp is switched on, the average lifetime is 7 500 hours. Protection against excess current is not necessary.

A similar, somewhat smaller, ballast equipment has been developed for the "TL"S 40 W lamp, which was originally designed to operate in series with an incandescent lamp.

GROWTH SUBSTANCES IN PLANTS

by R. van der VEEN.

581.143:631.811.98

Growth substances are compounds of a hormonal nature which, although present in the plant in extremely low concentrations, govern the life of the plant. Much regarding these substances is still obscure. Until recently it was thought that the chemical structure of all compounds with growth substance activity (both natural and synthetic) conformed to two essential requirements, but the members of a newly discovered group of compounds appear to be active as growth substances, yet do not comply with these requirements. Paradoxically enough, in certain concentrations growth substances are lethal to some plants. This has found practical application in the control of weeds.

The concept of "growth substances" is intimately associated with the name of F. W. Went, who in 1926 while working in the Botanical Laboratory at Utrecht, discovered that the elongation of the stems of young plants was essentially caused by one specific substance.

It soon became apparent that this growth substance brought about many other reactions in the plant. A survey of the phenomena induced by growth substance is given below.

2. *Geotropism.* Growth substance is transported in a vertical downwards direction by the living cells. If a plant be placed at an oblique angle or horizontal, its lower side will receive more growth substance than its upper side. Consequently its lower side wille elongate more, as a result of which the plant will again begin to grow upwards (*fig. 1a*).

The roots of a plant are even more sensitive to growth substance than its aerial parts. If a tap

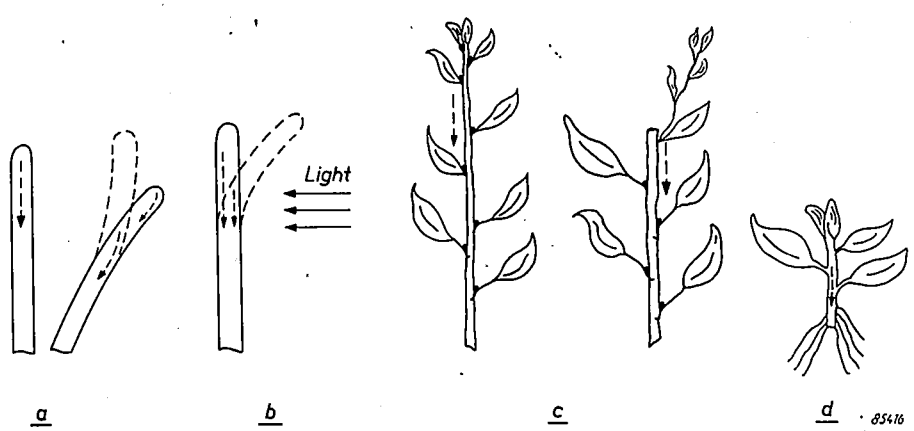


Fig. 1. *a*) *Geotropism.* In a stem set at an oblique angle, the lower side receives more growth substance than the upper side. The lower side thus grows more rapidly, and the stem curves towards the vertical.
b) *Phototropism.* On the lighted side of the stem there is less growth substance than on the shaded side; consequently the latter grows more rapidly, and the stem bends towards the light.
c) *Apical dominance.* Left: growth substance from the apex inhibits the development of the lateral buds. Right: on removal of the apex, the uppermost lateral buds sprout.
d) *Root formation.* The accumulation of growth substance in the lower extremity of a tap root brings about the development of roots.
In the figures, the dotted arrows indicate the movement of growth substance.

Reactions brought about by growth substance

1. *Elongation.* The cell walls of young cells elongate, often to many times their original length, under the influence of very low concentrations of growth substance. Greater concentrations of growth substance, however, inhibit this elongation.

root be laid horizontally, then here also more growth substance will be present on its lower side. This will lead to too high a concentration for the root growth, however, so that the elongation is not promoted, but inhibited, in such a way that the lower side grows slower than the

upper side, and the root will thus grow downwards.

3. *Phototropism.* It appears that growth substance is broken down by enzymes in the plant, and that the breakdown is powerfully stimulated by light. On the lighted side, therefore, the breakdown will be far more rapid than on the shaded side. In consequence the shaded side of the stem will grow more quickly, causing the plant to bend towards the light (fig. 1*b*). (This representation, though broadly accurate, is somewhat over-simplified).
4. *Apical dominance.* Much growth substance is formed in the rapidly growing apex of a plant. This growth substance is continuously transported down the stem. Now in each leaf axil is a lateral bud which can grow out into a lateral branch. Growth substance prevents the lateral buds from sprouting; in the presence of growth substance the buds thus remain dormant. If, however, the stem is decapitated (i.e. if its apex is cut off), the flow of growth substance is interrupted, and the lateral buds are thus able to sprout [1]). The uppermost lateral buds sprout first (fig. 1*c*). Since the latter themselves produce growth substances the lower buds are arrested in their development and remain dormant.
5. *Root formation.* It has been found that a high concentration of growth substance stimulates the formation of roots. When a cutting is planted all the growth substance present in it, will be transported downwards and will collect in the lower end. For this reason the roots will form at that end and not higher on the cutting. By applying artificial growth substance to cuttings, the rooting process can be accelerated.
6. *Abscission.* The spontaneous shedding of leaves and fruits is usually caused by a reduction of the growth substance concentration in the plants, and can be readily prevented by the application of artificial growth substance to the plants, e.g. by spraying.
7. *The withering of certain blooms.* Many orchids, such as the popular Cattleyas, will often retain their freshness for more than 14 days on the plant. If they are pollinated, however, they wither in one day. Immediately after pollination, a large amount of growth substance is formed. The same effect is obtained by applying a little growth substance to the stigma, instead of pollen.

[1]) For the correlation between this phenomenon and the length of day see R. van der Veen, Philips tech. Rev. **14**, 179, 1952/53.

It is possible to go on enumerating phenomena induced by growth substance, but the above will suffice to give an impression of the highly important and varied role played by the growth substance in the life of plants.

Growth substances as weed-killers

As a consequence of its powerful hormonal nature, growth substance can ultimately give rise to a profound disturbance of various physiological equilibria in the plant, unless it is broken down in the latter. This disturbance may even be severe enough to bring about the death of the plant within one or twow eeks, and has led to the practical application of synthetic growth substances as weed-killers (herbicides). Growth substances for use in this role must be very active and must not be rendered inocuous by the plant. Since the monocotyledons such as grasses and cereals are far less sensitive to growth substances than the dicotyledons, dicotyledonous weeds in grass and cereal fields can be eradicated by spraying with a specific amount of growth substance [2]). The best known of these so-called hormone weed-killers based on growth substance activity are 2,4-D[1] [3]), 2,4,5-T [2] and MCPA [3].

The chemical structure of growth substances

The chemical structure of the growth substance that plays the principle role in the plant kingdom is known. It is a compound of fairly simple structure and can be prepared synthetically, viz. 3 - indolylacetic acid [4].

It is produced in the plant by the breakdown (in several steps) of the amino acid tryptophane, and is itself broken down further in turn, especially under the influence of light (see above). Indolylacetic acid is a link in a relatively long degradation chain. The concentration of the growth substance in a plant depends upon the rates of the various reactions in this chain. Since the indolylacetic acid breakdown reaction is dependent upon the illumination, a low light intensity will cause the concentration to be high and consequently the plant will proceed to elongate rapidly.

Following the discovery of the natural growth substance, numerous synthetic substances have been found that have a more or less analogous influence on the plant. Their number runs into hundreds. Some of them have but a weak growth substance action, others an action scarcely exceeded

[2]) This possibility has already been mentioned in this journal, see R. van der Veen, Philips tech. Rev. **16**, 356-357, 1954/55.
[3]) The numbers in square brackets refer to the structural formulae in the appendix.

by that of the natural growth substance. Some of the better known are summarized in the appendix.

It is truly remarkable that so many compounds, at first sight so widely different, exert an analogous action on plants. Attempts to find some correlation between the structures of the various compounds has occupied many workers.

In 1938 Koepfli, Thimann and Went [4]) established that all substances with growth substance activity have the following characteristics in common:

a) a molecule with a "nucleus" consisting of a ring system,
b) at least one double bond in this ring,
c) one side chain on the ring,
d) a carboxyl group at the end of the side chain, and separated from the ring by at least one carbon atom, and
e) a special spatial relationship between the ring and the carboxyl group.

In 1949 Veldstra and Booy [5]) reduced these five characteristics to two, namely the possession of

a) a ring system displaying a high surface activity, and
b) a group of acidic character, which, on adsorption of the molecule at an interface, can come to lie outside the plane of the ring.

At the last held botanical conference (Paris 1954) both these characteristics were re-confirmed; no single exception to the characteristics of Veldstra and Booy had yet been found among the great number of compounds with growth substance activity.

A new group of growth substances

Co-operation between Philips-Roxane and the Organo-Chemical Institute of the Dutch National Council for Industrial Research, T.N.O., has recently lead to the discovery of a new group of compounds of fairly high growth substance activity [6]). There is no ring system in these compounds however, so that they do not comply with condition *a*)

[4]) J. physiol. Chem. **122**, 763-783, 1938.
[5]) Biochem. biophys. Acta **3**, 278-312, 1949.

[6]) This work was conducted under the direction of Dr. Van der Kerk in the Organic Chemistry Laboratory at Utrecht. The discovery was reported in: G. J. M. van der Kerk, M. H. van Raalte, A. Kaars Sijpesteijn and R. van der Veen, Nature **176**, 308-310, 13 Aug. 1955.

Fig. 2. The reaction of tomato plants to spraying with solutions of N,N-dimethylthiocarbamyl glycollic acid [5] (code name KD 31) of various concentrations. Photograph (*a*) was taken two days, photograph (*b*) ten days after spraying. Even after ten days the plants treated with 0.05% and 0.1% solutions have still not recovered.

laid down by Veldstra and Booy, and thus do not conform to their formulation of growth substance characteristics. Examples of the new growth substances are N,N-dimethylthiocarbamylglycollic acid [5] and N,N-dimethyl-S-carboxymethyl-dithiocarbaminate [5a].

In this new group of growth substances there is an atomic grouping that may be regarded as an intermediate form between

$$(CH_3)_2N—\overset{\overset{\|}{S}}{C}— \quad \text{and} \quad (CH_3)_2\overset{(+)}{N}=\overset{\overset{|}{S^{(-)}}}{C}—.$$

in place of the ring.

It is known from organic chemistry, that in the second structure, with a double bond between N and C, the atoms lie preferentially in one plane, and that this plane structure is imparted to a greater or lesser degree to all the molecules in which this group occurs. The new growth substances, therefore, have in their molecules an atomic group of plane structure, which takes the place of the ring system of the previously known growth substances.

In addition to this plane atomic group, the molecule contains a side chain, that is attached to the right-hand carbon atom (in the above formulae). The side chain conforms to condition (*b*) of Veldstra and Booy, i.e. it is acidic in character, and will lie out of the plane containing the flat atomic group when the molecule is adsorbed at an interface.

Regarding the side chain, the following can be noted. Whether or not a compound of the recently-discovered type is active, depends upon the structure of the side chain in a manner that is completely analogous to that found for the side chains of the 2,4-dichlorophenoxy group. Thus the side chain —O—CH_2—COOH for example, occurring in N,N-dimethylthiocarbamylglycollic acid [5] is very active; the action of solutions with different concentrations of this compound on young tomato plants may be seen in *fig. 2a* and *b*. If the side chain in question is replaced by —S—CH_2—COOH, a barely less active compound is obtained, with regard to its effect on the plant within two days (*fig. 3a*). After a longer period, the plant recovers from the effects of a high concentration of the sulphur-bridged growth substance (fig. 3*b*) but not from the effects of the oxygen-bridged growth substance (fig. 2*b*). The same is found for 2,4-dichlorophenoxy-

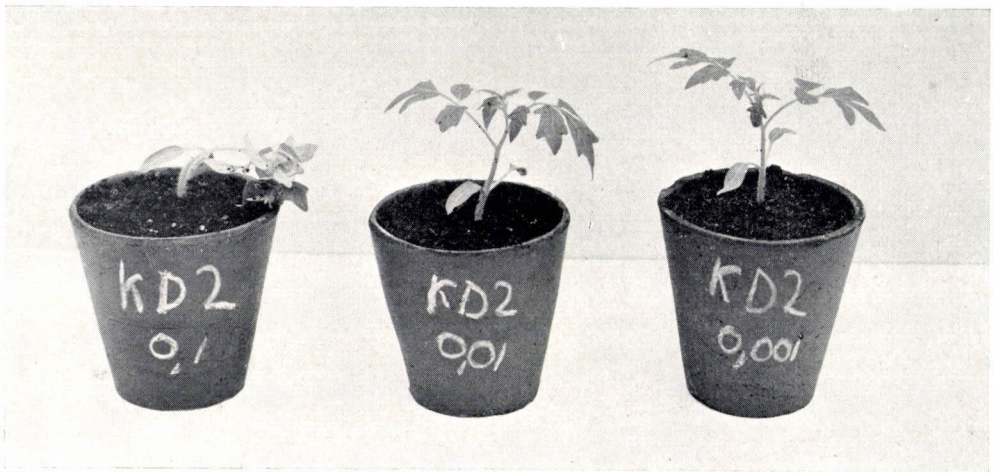


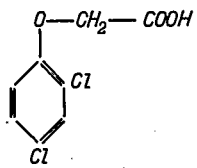
Fig. 3. The reaction of tomato plants to spraying with solutions of N,N-dimethyl-S-carboxymethyl dithiocarbaminate [5a] (code name KD 2) of various concentrations, (*a*) two days (*b*) ten days after spraying. After ten days all the plants have completely recovered.

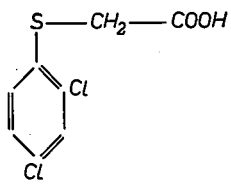
acetic acid (a growth substance of the old type having a ring system) if the bridging oxygen atom (cf. [1] and [1*a*]) in the side chain is replaced by a sulphur atom. Plants thus appear to be better able to break down a sulphur bridge than an oxygen bridge in this type of compound.

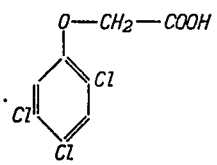
This is also illustrated by the following. α-naphthoic acid [6] displays growth substance action. On replacement of the carboxyl carbon of this acid by sulphur, α-naphthyl-sulphinic acid [6*a*] is formed; this compound is likewise active as a growth substance, but is broken down very rapidly by the plant. Tomato plants, sprayed with an 0.1% solution, show a very pronounced reaction after one hour (consisting in a downward curving of the leaves) which reaches a maximum in eight hours. Recovery then sets in and after 24 hours there is no longer anything unusual to be seen. If the plants are sprayed with indolylacetic acid (the growth substance most widely occurring in nature), they also recover in the course of time, but the recovery is much slower.

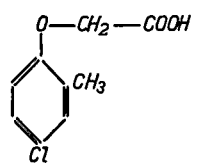
APPENDIX: STRUCTURAL FORMULAE

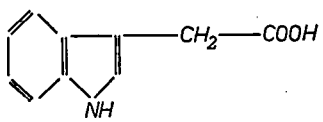
Below are given the structural formulae of the compounds mentioned in the text.

[1] 2,4-dicholorophenoxyacetic acid (2,4-D) . . . . . 

[1*a*] Like [1], but the bridging oxygen atom has been replaced by a sulphur atom . . . . . 

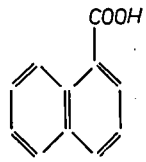
[2] 2,4,5-trichlorophenoxyacetic acid (2,4,5-T) . . . . . 

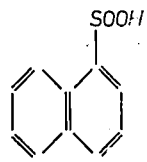
[3] 2-methyl-4-chlorophenoxyacetic acid (MCPA) . . . . . 

[4] 3-indolylacetic acid . . 

[5] N,N-dimethylthiocarbamyl glycollic acid . . . . . $(CH_3)_2N-\overset{\overset{S}{\|}}{C}-O-CH_2-COOH$

[5*a*] N,N-dimethyl-S-carboxymethyl dithiocarbaminate . . . . . $(CH_3)_2N-\overset{\overset{S}{\|}}{C}-S-CH_2-COOH$

[6] α-naphthoic acid . . . . . 

[6*a*] α-naphthylsulphinic acid . . . . . 

Summary. After a survey of the principal reactions brought about by growth substances in plants, the use of growth substances in the combating of weeds is dealt with in brief. The author then passes on to the chemical structure of growth substances. The main growth substance is chemically fairly simple, viz. 3-indolylacetic acid, and can be prepared synthetically. The five common characteristics of substances with growth substance activity, framed by Koepfli, Thimann and Went in 1938, are then quoted. These were reduced to two in 1949 by Veldstra and Booy. Philips-Roxane and the Organo-chemical Institute of the Dutch National Council for Industrial Research T.N.O., working in co-operation, however, have recently discovered a group of compounds of fairly high growth substance activity which have no ring system and do not therefore conform to the first condition of Veldstra and Booy. In the new group of growth substances, a similarly plane atomic group occurs in place of the ring system. A side chain attached to the atomic group does comply with the second condition of Veldstra and Booy.

A CINEMA PROJECTOR FOR 70 mm AND 35 mm FILMS

778.554.1

In the last few years various new systems of making and projecting films have been used. Philips is indirectly involved in one of these, viz. the Todd-A.O. [1]) system. The American Optical Company, the firm that undertook the development of the "Todd-A.O." system turned to Philips in Eindhoven for the development and manufacture of the necessary special projectors. We shall deal briefly with the "Todd-A.O. system" and its position in regard to two other well-known new systems, namely "Cinerama" and "CinemaScope", before turning to the projector itself.

The aim of these new systems is to strengthen the impression of "realness" created by the film. The method used in the above three systems is to photograph a wider field of view by using wide-angle taking lenses and to take up more of the field of vision of the audience with the projected picture by widening the screen. It is not sufficient to fill the bigger screen by projecting a normal film that has been enlarged still further; this would only produce the same effect as if the audience were to sit nearer to the screen.

It is most desirable to combine the use of a wide screen with stereophonic sound. The positions on the screen where the actions take place now vary so much that it would be distracting were the accompanying sound to come from just one fixed direction.

Increase in the picture angle at the filming stage has been taken furthest in the "Cinerama" system. In this system three cameras are used which take adjoining pictures, and this gives a total picture angle of 146° in filming. When these are projected, a wide curved screen is used on which the three films are projected next to each other (*fig. 1*) by three synchronized projectors. It is true, as can be seen from fig. 1, that the great majority of the audience will see the picture at an angle considerably less than 146°, but experience shows that this is not objectionable.

The success of "Cinerama", which in Europe has been seen in London, Paris, Milan and Rome, was so complete that audiences accepted the shortcomings of the system. From the audience's point of view the greatest objection is that the two dividing lines on the screen where the pictures join, have not yet been successfully eliminated. The brightness and colour of the three pictures are never absolutely alike. Moreover, every picture projected dances a little: the position of the frames on the film and the transport mechanisms of the projectors are never quite perfect and this means that consecutive frames never occupy exactly the same position in the film gate. The dancing effect is not the same for each of the three films so that they can be seen to quiver with respect to each other at the picture boundaries. Possibly even more distracting is the fact that at a picture boundary the faults in the picture, though slight, change in a discontinuous manner. This can cause, for example, a long object situated across the width of the screen to show a kink at a picture boundary. This is especially annoying when an object of this kind (e.g. a boat) moving right across the screen, passes a picture boundary.

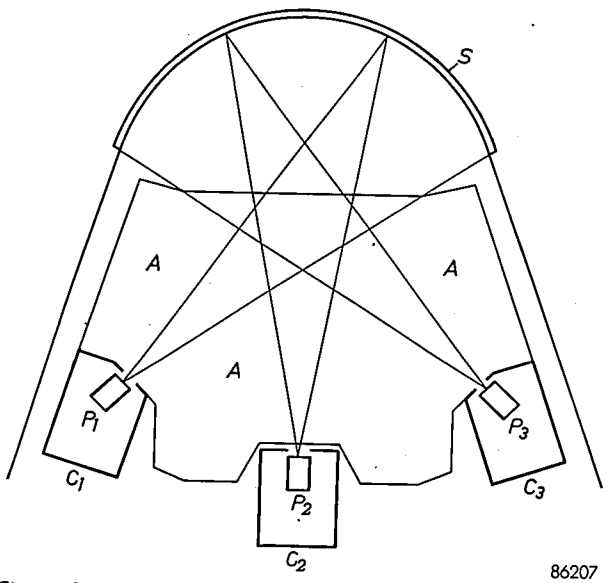


Fig. 1. Schematic diagram of the Cinerama system, showing the arrangement of the screen S, the three projectors P_1, P_2 and P_3 in their projection rooms C_1, C_2 and C_3, and the audience A.

From the point of view of the cinema management, "Cinerama" has the further disadvantage that the three projection rooms have to be placed in the auditorium itself, so that a considerable number of seats have to be given up to make room for them (fig. 1). Projection is not possible from rooms situated high up as the light beams would strike the screen obliquely and horizontal lines for instance would be reproduced as curved ones because of the curvature of the screen. Moreover, the whole arrangement is very complicated and several technicians are necessary to operate it. It is understandable that other systems have been tried with a view to achieving the desired effect while avoiding the difficulties mentioned.

An elegant approach to the problem was made possible by the development of a special wide angle

[1]) Todd is the name of the man who advocated the system; A.O. stands for "American Optical Company".

objective by Professor Brian O'Brien of the American Optical Company. This lens has a picture angle of 128° (but nevertheless has a comparatively large numerical aperture). A film taken with this lens consequently covers an area only slightly less than that of the three "Cinerama" cameras combined. Such a film can be projected on to a wide, curved screen and this produces, with a single projector, a similar effect to that of "Cinerama". Incidently, the "picture angle" of the projector does not need to be made specially large.

This special objective forms the nucleus of the "Todd-A.O." system, but the system has other characteristic features. One of these is that it has completely broken with tradition by using 70 mm film in place of 35 mm and at the same time the frame frequency has been increased from 24 to 30 frames per second. The higher frame frequency produces smoother motion and lessens flickering in the picture reproduced on the screen. This is very desirable, for the eye is most sensitive to flickering at the periphery of the field of vision and, as a result, flickering on a wide screen might become noticeable.

The problem of obtaining a sufficiently bright picture on the large screen is greatly simplified by the wide film, since the area of the frame on 70 mm film can be made 3.5× as large as that on normal 35 mm. For the same luminous intensity on the film, 3.5 times as much light strikes the screen. In addition, the higher frame frequency contributes to less heating of the film.

Another important advantage is that the big screen can be filled without abnormally large magnification. With the small frame of 35 mm film, the grain of the film emulsion would be visible and would tend to blur the image.

In *fig. 2* we have a strip of 70 mm film and in *fig. 3* strips of normal 35 film and of 35 mm film for "CinemaScope". In all the films the perforations have the same spacing. With the 70 mm film, however, the height of the frame is 5 times the perforation spacing whereas it is only 4 times with 35 mm film; this means that the frame on the wide film is not only wider but also higher than that on normal film. In contrast with "CinemaScope", the scene is photographed on the film in natural proportions. The shape of the projected picture therefore corresponds to that of the frame (the film mask is 48.5 by 22 mm) and a so-called anamorphic projection lens (as in "CinemaScope") is not used.

Fig. 2. Strip of 70 mm film. It has six magnetic sound-tracks, three on either side of the film. The wide black bands on the outside of the perforations carry two sound-tracks each; the narrower bands inside the perforations carry only one each. The height of the frame is equal to 5 perforations.

a *b* *c*

Fig. 3. Strips of 35 mm film (*a*) normal and (*b* and *c*) "CinemaScope". The normal 35 mm film has one optical sound-track; the 35 mm "CinemaScope" film has either four magnetic sound-tracks, two on either side of the film (*b*), or one optical sound-track (*c*). Only in the second case (*b*) is stereophonic reproduction possible.

In "CinemaScope", for which many cinemas have already been adapted, 35 mm film is used. By means of the anamorphic lens used in this system, the picture is compressed in a horizontal sense (fig. 3*b* and *c*) when being taken. When the film is being projected the picture is expanded horizontally by a similar lens, so that the natural relationships are restored. The relationship of width to height of the projected picture can be larger by these means than it actually is on the film. The picture angle during shooting is about 90° as compared to 146° for "Cinerama" and 128° with "Todd-A.O.".

On the wide film there is room for the six magnetic sound-tracks which give well-nigh perfect stereophonic reproduction. (A magnetic sound-track can give a higher quality of reproduction than an optical sound-track). Behind the screen there are five groups of loudspeakers, whilst in the auditorium there is a further number of loudspeakers used for special sound effects. Each group is supplied by its corresponding sound-track. It is now not necessary to have the six sound-tracks on a separate synchronously driven film as is the case with the "Cinerama" system. The greater frame-height and the higher frame-frequency combined, increase the film speed by a factor $^{5}/_{4} \times {}^{30}/_{24} =$ more than 1.5 times over that with normal film. This improves the quality of the sound: high frequencies can now be better recorded and reproduced.

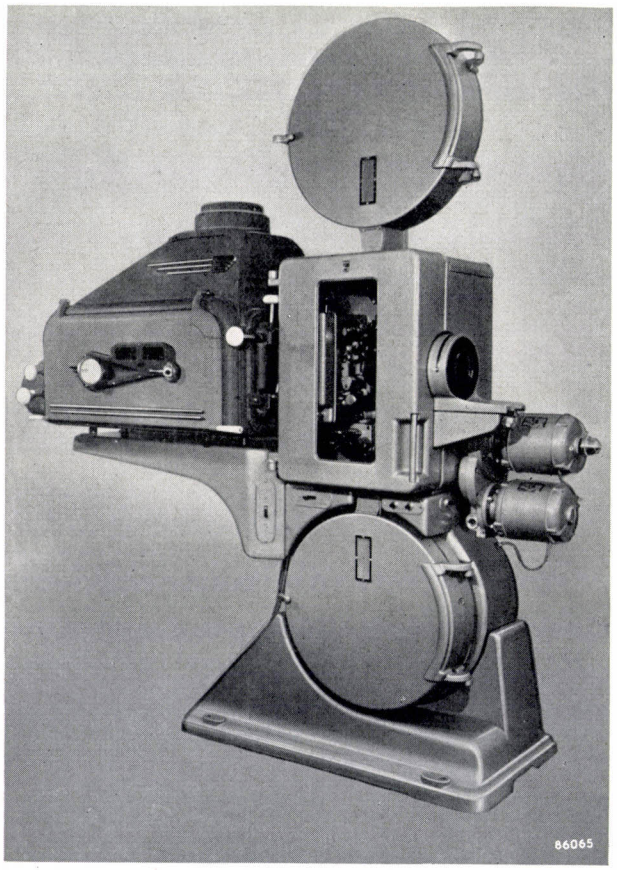


Fig. 4. The 70 mm film projector (type EL 4000) which at the same time is suitable for projecting normal 35 mm and "CinemaScope" films (with 4 magnetic sound-tracks or with one optical sound-track).

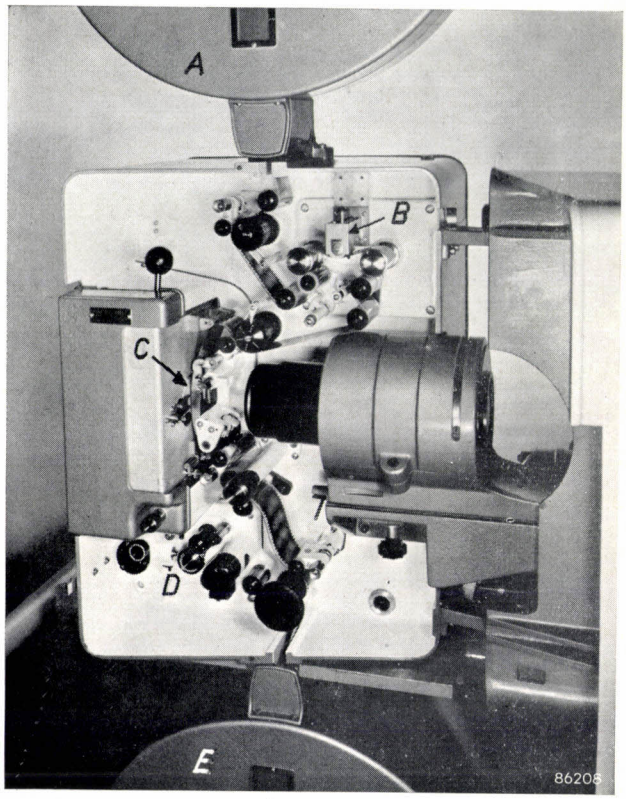


Fig. 5. Path of film when projecting either 70 mm or "CinemaScope" (35 mm) with magnetic sound-tracks. The projector in the figure is threaded with 70 mm film. After leaving the upper spool *A*, the film is fed past the magnetic sound head *B*. The film then passes to the take-up spool *E* via the curved film guide *C*, by-passing the optical sound head *D*. If normal 35 mm film or "CinemaScope" film with an optical sound-track is being projected, the film by-passes the magnetic sound head and is fed past the optical sound head.

The projector

As already mentioned, the special projector for the "Todd-A.O." system has been developed in Eindhoven and is now being manufactured there (Type EL 4000). The projector is so made that it is suitable not only for the new system but also for normal and for "CinemaScope" films. A cinema adapted for "Todd-A.O." is, with one type of projector, equipped for most other current projection systems at the same time (not, of course, for "Cinerama".) *Fig. 4* is a photograph of the projector.

The projector has an optical sound head (for the optical sound-tracks on 35 mm film) and a magnetic sound head for 35 mm "CinemaScope" film and for 70 mm film. These heads are so arranged that the film can by-pass the heads not in use (*fig. 5*). The magnetic sound head has a total of 10 pick-ups (*fig. 6*); 6 of these correspond to the sound-tracks of 70 mm film and the other 4 to the sound-tracks on "CinemaScope" film (35 mm).

The sprockets have 4 toothed rims; the inner two are used for transporting 35 mm film, the outer

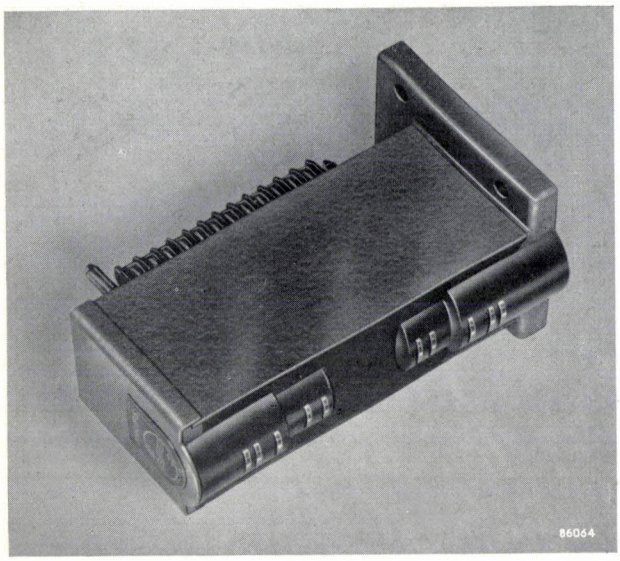


Fig. 6. The magnetic sound head of the EL 4000 cinema projector. 10 pick-ups can be seen as narrow ridges projecting from the mount. The four in the middle, which are those corresponding to the sound-tracks on "CinemaScope" film (35 mm), do not project quite as far as the others so that when 70 mm film is being projected they cannot cause any damage.

two for 70 mm film (*fig. 7*). The film is drawn forward by a 4-slot Maltese cross mechanism and the intermittent sprocket must therefore have a periphery equal to 4 times the size of the frame [2]). As already stated, the height of the frame in 70 mm film is $^5/_4$ times that of 35 mm film. The periphery, and consequently also the diameter of the intermittent sprocket must differ by this same factor for the two sizes of film. Hence the 70 mm film is clear of the sprocket teeth for 35 mm film.

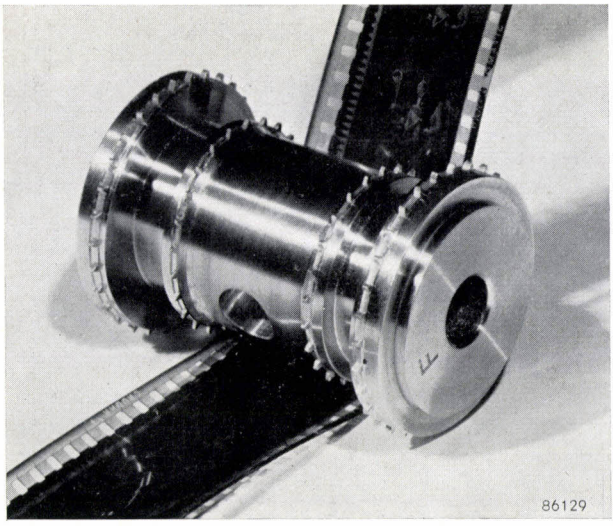


Fig. 7. Sprocket for 70 mm and 35 mm film. The toothed rims in the centre (for transporting 35 mm film) are set deeper so that they clear the 70 mm film.

[2]) See Philips tech. Rev. **16**, 158-171, 1954/55.

The sprocket shown in fig. 7 is the type used both for the take-off sprocket and for the take-up sprocket, which respectively draw the film at uniform speed from the feed spool and lead it on to the take-up spool. The diameters of the toothed rims for 70 mm and 35 mm film also differ by a factor of $^5/_4$ but the peripheries are six times the size of the frame.

When transferring from one film size to the other, some re-adjustments have to be made: e.g. the film guide (which guides the film at the gate, see fig. 5) and the pad rollers (which keep the film on the teeth of the sprockets) have to be changed. In addition the speed of the projector has to be changed and for this purpose the apparatus has 2 motors (visible in fig. 4) and a selector switch. The objective too must of course be the correct one for the projection system in operation.

70 mm film has a greater tendency than 35 film to buckle in the film gate under the influence of the heat generated by the arc lamp. This would cause the projection to become blurred and so to prevent this, the film guide has been slightly curved (fig. 5) which gives the film more rigidity at that point. The normal flat pressure springs need now to be replaced by thin steel strips, whose tension can be adjusted and which press the film tightly to the gate.

The projector is fitted with a single-blade rotating shutter and the two interruptions per frame that are necessary are achieved by allowing the shutter to make two revolutions in this period. The effective region of the shutter blade is at the outer edge and at a large distance (about 12 cm) from the point of rotation. This blade interrupts the light-beam only 12 mm from the film at a point where the diameter of the beam is small. With these measures shutter losses are reduced to 46% [3]).

The high number of revolutions of the shutter (3600 revs/min for 70 mm film) cause the projector to run with more noise than is usual. Also the flapping of the wide film is noisier than that of normal film. This is due to the increased width, the greater frame-shift and the higher frame-frequency. However, when the door of the projector is closed, the noise is not annoying for the operator.

The "Todd-A.O." system had its public debut on 13th October 1955 with the film "Oklahoma" shown in the Rivoli Theatre on Broadway, New York. The strip of 70 mm film reproduced in fig. 2 contains a scene from this film.

J. J. KOTTE.

[3]) The high speed of the effective part of the shutter blade produced by the doubled rotation speed and the large radius of the shutter, lessen the effect of the so-called "covering angle". A small beam diameter at the point of interception is also favourable in this connection. See article quoted in [2]).

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

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

2227: M. E. Wise: A quickly convergent expansion for cumulative hypergeometric probabilities, direct and inverse (Biometrika **41**, 317-329, 1954).

Sampling without replacement from finite populations is treated. If n units are taken from a population of N containing Nx defectives, we may want to know the probability P that (say) c defectives or less are found in the sample of N; or, conversely, what is the value of x corresponding to given values of c and P. It was generally believed that these results were far more difficult to work out than for the corresponding problem with "infinite batches" in which we do not have to allow for the units removed in the sample as affecting the composition of the rest of the batch; this case only involves binomial probabilities, whilst the finite batch case involves hypergeometric ones. This paper derives simple but accurate approximations for these probabilities and for the inverse case. The mathematical method is an unusual one and is given in full, as it might be usable in quite different problems leading to hypergeometric series. Two numerical examples are given, in each of which, for given c, points in a P-x curve are calculated and compared with the corresponding exact values.

2228: P. F. J. Bogers: Organische stoffen bij demineralisatie (Water **38**, 229-302, 1954). (Organic substances in demineralization; in Dutch).

Report of investigations into the poisoning of ion-exchange resins by organic substances present in water.

2229*: J. Smit and H. P. J. Wijn: Physical properties of ferrites (Advances in Electronics and Electron Physics **6**, 69-136, 1954).

Survey of theory and experimental results concerning the properties of Ferroxcube, up to the end of 1953. See also Philips tech. Rev. **8**, 353-360, 1946; **9**, 185-190, 1947/48; **11**, 313-322, 1949/50; **13**, 181-193, 1951/52; **16**, 49-58, 1954/55; and these abstracts Nos. **2092** and **2109**.

2230: F. A. Kröger and H. J. G. Meyer: The edge emission of ZnS, CdS, and ZnO and its relation to the lattice vibrations of these solids (Physica **20**, 1149-1155, 1954).

It is shown that the vibrational structure appearing in the luminescent emission close to the fundamental absorption bands of ZnS, CdS and ZnO corresponds to the frequency of the longitudinal lattice vibrations of the optical branch. This proves that the edge emission is a property of the pure, unperturbed lattice. In connection with observations by Klick it is concluded that the fluorescence is due to excitons. The appearance of double peaks in the infra-red absorption spectrum of ZnO and BeO is explained by a two-phonon process, involving phonons of the optical and acoustical branch according to a theory given by Born and Blackman.

2231: C. Z. van Doorn: Temperature dependence of the energy-gap in ZnS (Physica **20**, 1155-1156, 1954).

The temperature dependence of the energy-gap in ZnS was determined by direct measurement of the shift of the absorption-edge as a function of temperature. The results were in good agreement with calculations by W. W. Piper and measurements by J. H. Gisolf.

2232: T. Tol: Fundamentale Probleme bei der Röntgen-Kinematographie mit Bildverstärker (Research Film, No. 5, pp. 1-6, Dec. 1954). (Fundamental problems in X-ray cinematography with the image intensifier; in German).

Brief survey of selected topics presented in previous article in Philips tech. Rev. **17**, 69-97, 1955/1956 (No. 3).

2233: L. A. Æ. Sluyterman and H. J. Veenendaal: The rate of condensation of some tripeptide methyl esters in solution (Rec. Trav. chim. Pays-Bas **73**, 1001-1008, 1954).

The condensation rate of a few tripeptide methyl esters in methanol has been studied at 65 °C. The peptide esters with a glycyl residue at both ends have been found to be equally reactive, regardless of the nature of the amino acid residue in the middle. A side chain situated on one of the terminal amino acid residues decreases the reaction rate.

2234: C. G. J. Jansen, R. Loosjes and K. Compaan: Distribution anormale des vitesses des électrons émis par une cathode à oxydesen, régime d'impulsions (Le Vide **9**, 129-134 1954).

French version of paper published in Philips Res. Rep. **9**, 241-258, 1954 (see these abstracts No. **R246**).

2235: C. G. J. Jansen, R. Loosjes, J. J. Zaalberg van Zelst and G. A. Elings: Une méthode oscillographique pour determiner le Potential d'Extraction en fonction de la température (Le Vide **9**, 135-138, 1954).

An oscillograph has been developed which describes the saturation emission current i_s of a diode as a function of the temperature T of the cathode. The curve is described in about ten seconds, during which the anode-cathode voltage remains practically constant. The temperature T is measured by a Pt-PtRh thermocouple connected to the metal of the indirectly heated cathode. As the $i_s = f(T)$ curve is determined in such a short time, the effects of poisoning of the cathode surface are kept to a minimum. Moreover, it is possible to determine very quickly not only the work function φ of the cathode but also its temperature coefficient $d\varphi/dT$ from a graph of saturation current density j_s as a function of absolute temperature, with the help of Richardson's formula.

2236: A. Venema: Applications de la cathode L (Le Vide **9**, 269-272, 1954).

The construction and working of the L-cathode are described and some of the special precautions necessary in its fabrication are mentioned. The special properties of this type of cathode are then dealt with in connection with the metallic emitting surface (no oxides) and the replenishment of the activating material (barium). The application of the L-cathode in the EC56 disc-seal triode (power amplifier, up to about 4000 Mc/s) is described. Another example of its use is in the 4J50 magnetron, which gives a 250 kW output at a wavelength of 3 cm. The use of the L-cathode in klystrons and in X-ray tubes is also described.

2237: J. Meltzer: Morphopogical notes on Bryobia forms of fruit trees and ivy (Entomologische Berichten **15**, 337-339, 1955).

Bryobia mites (Bryobia praetiosa Koch) found on apple and pear trees appear to differ morphologically and in size from those found on ivy. The author believes that the two types represent different species. There are indications that those living in fruit trees cannot live in ivy, and vice versa. Owing to unfavourable weather, experiments to test this hypothesis have not given definite results.

2238: J. G. van Pelt and H. Keuker: De electrometrische Karl Fischer titratie (Chem. Weekblad. **51**, 97-99, 1955, No. 7). (The electrometric Karl Fischer titration; In Dutch).

A simple and practical apparatus for electrotitration for the determination of water and a method for quick preparation of the necessary reagents.

2239: N. W. H. Addink, J. A. M. Dikhof, C. Schipper, A. Witmer and T. Groot: Quantitative spectrochemische Analyse mittels des Gleichstrohmkohlebogens (Spectrochimica Acta **7**, 45-59, 1955, No. 1). (Quantitative spectrochemical analysis by means of the D.C. carbon arc; in German).

Detailed account of the "Constant Temperature Arc" method of analysis developed in Eindhoven (see N. H. W. Addink and W. de Groot, Philips tech. Rev. **12**, 337-348, 1950/1951. The method consists of completely volatilizing 5 mg of a powdered sample in a shallow anodic crater of a carbon arc, with the addition of materials to modify the rate of volatilization if required. The line intensities are calibrated and corrections are made by comparison with selected Fe-lines, originating from a "standard light source" so as to get comparable analytical results; the calculations required are illustrated by several examples which also indicate the accuracy of the method. Tables are given of the empirically determined K-values, so that they can be checked in other laboratories.

2240: A. Kats and J. M. Stevels: Colour centres in alkali silicate glasses (Z. physik. Chem. **3**, 255-260, 1955, No. 3-4).

It appears that the theory of colour centres, caused by X-ray and U.V. irradiation, as developed in recent years for alkali halides, also holds for alkali silicate glasses. The effect of such irradiation has been studied by the changes in the absorption spectrum between 2000 Å and 10 000 Å for various glasses with different alkali oxide content. The theoretical background of the results is briefly discussed; a full account of this work will appear in Philips Research Reports.

VOL. 17 No. 11, pp. 305-340 MAY 1956

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 RECEIVER FOR THE RADIO WAVES FROM INTERSTELLAR HYDROGEN

I. THE INVESTIGATION OF THE HYDROGEN RADIATION

by C. A. MULLER *).

522.6:621.396.9

Radio astronomy, born just before the second world war, has developed with amazing rapidity. Radio astronomical research can now boast of results of far-reaching importance, which have considerably improved and sharpened our picture of the universe. The instruments employed in radio astronomy are for the greater part electronic. In the Netherlands, the author, member of the Netherlands Foundation for Radio Astronomy, has headed the development of a receiving installation for research on the radio spectral line of wavelength of 21 cm, which is radiated by the hydrogen atoms in interstellar space. This receiver has been used in conjunction with a 7.5 m diameter parabolic reflector at Kootwijk. A similar receiver is destined for a "radio observatory" at Dwingelo. The receiver will be described in two articles; the first, printed below, outlines the place of radio astronomy in modern astronomical research and discusses the nature and properties of the radiation and the requirements which the receiver must fulfil. The second article will give a detailed description of the receiver circuits.

Radio waves from space

One of the most important discoveries in astronomy in this century was the observation of radio waves originating outside the Earth. This radiation, discovered by chance in 1931 by Jansky in the United States, was first systematically investigated a number of years later by the amateur Reber. Not until during the second world war was work in this field begun in earnest, first in Australia and in England. At present, radio astronomy is carried on in a number of institutes spread over the whole world.

The reception of the radio waves requires, of course, instruments which are utterly different from the usual optical instruments used in conventional astronomy. Fortuitously, after the second world war, many radar installations for which there was now no further use could be found in various countries. These were very well suited for use as receivers for the V.H.F. radio waves from space. In the Netherlands, an abandoned German radar antenna of 7.5 m diameter was moved by the P.T.T. (Dutch Post Office) to Kootwijk Radio Station and later placed temporarily at the disposal of the Netherlands Foundation for Radio Astronomy. This body includes representatives from the Dutch Post Office, the observatories of Leiden, Utrecht and Groningen, and those of Belgium, the K.N.M.I. (Royal Dutch Meteorological Institute) and the Research Laboratory of N.V. Philips. The work of the Foundation is subsidised by the Netherlands Organisation for Pure Research (Z.W.O.). A group of investigators of the Foundation has been making continuous observations at Kootwijk, the results being worked out at the Leiden observatory. A new parabolic reflector of 25 m diameter has been built in the neighbourhood of Dwingelo and will shortly be brought into use. This antenna is movable in all directions, and has an automatic steering mechanism. In addition, two more parabolic antennae 7.5 m in diameter are being set up in Dwingelo.

Not all the radiation from space can reach the Earth: the atmosphere is transparent for wavelengths between about 1 cm and 20 m. Radiation

*) Leader of the Observations Group of the Netherlands Foundation for Radio Astronomy, Dwingelo.

with a wavelength shorter than 1 cm or longer than 20 m is almost completely absorbed in the atmosphere or reflected from the ionosphere. The wavelength range of this "radio window" is, however, very wide (8 to 10 octaves) compared with the narrow "optical window": the atmosphere of the Earth transmits "by chance" just the visible range of the spectrum, but this covers less than one octave (3700 Å — 8000 Å). Here we see one of the reasons for the importance of radio astronomy.

It was soon found that the intensity of the radiation differed for the various wavelengths within the transmitted range: we can thus speak of a radio spectrum. The optical spectra of stars, e.g. that of the sun, exhibit a great number of lines (corresponding to atomic transitions) against a background of the continuous emission spectrum (caused mainly by the capture of electrons by hydrogen atoms). The radio spectrum is probably mainly a continuous emission spectrum; discrete atomic or molecular transitions corresponding to frequencies in the radio range have a very small probability. In general, therefore, no spectral lines can be observed here on Earth. However, in 1944 the Dutch astronomer Van de Hulst [1]) drew attention to the possibility that at least one emission line in the radio range, viz. at wavelength of 21 cm, should be observable. Radiation of this wavelength is emitted by the highly rarefied clouds of hydrogen between the stars and which, together with the equally rarefied dust clouds, constitute the matter of interstellar space.

As a result of the extreme rarefaction, the hydrogen in interstellar space can exist in atomic or ionized form. Until recently, one could observe hydrogen gas only in the neighbourhood of very hot stars, where it is almost completely ionized. One of the effects of the capture of electrons by hydrogen ions is the emission of visible spectral lines and a continuum. Only indirect methods had been available by which anything could be concluded about the physical condition of the more extended regions of the neutral hydrogen. Van de Hulst argued that for the transition in the hydrogen atom giving rise to a quantum of wavelength 21 cm, there was a good chance that it would be observable on Earth. In fact, it is possible to observe this spectral line. The calculation involved in this argument is a good illustration of the magnitudes juggled with by astronomers. Any arbitrary H-atom exhibits this transition only about once in 11 million years; furthermore the density of the interstellar hydrogen is extremely low, of the order of 1 atom per cm^3 (1 cm^3 hydrogen at 1 atm pressure contains at room temperature about 2.5×10^{19} atoms). However, since the hydrogen clouds stretch out over enormous regions — they are spread throughout the whole of the Galactic system, with an equatorial diameter of 10^{23} cm the power intercepted on Earth from particular directions is still just enough to observe.

An extensive search was started for this spectral line, and in 1951 it was indeed found, almost simultaneously by workers in the United States, in Holland and in Australia [2]).

The full importance of this discovery can hardly be overestimated: a new tool was now available for investigating the structure of the Galactic System, a tool which was much more direct than those hitherto available. We will return to this in the next section.

The intensity of the hydrogen radiation is very weak; the power received is many times (10 to 1000 ×) less than the noise level which is produced in the receiver itself. The intercepted radiation also has the character of noise, like the receiver noise. The value of the intensity is a function of the frequency, with a marked maximum around 1420 Mc/s, corresponding to the wavelength 21.1 cm.

The noise character of the interstellar radio waves, i.e. the statistical variations in their intensity with time, is not to be unexpected. It is caused by the fact that the sources of radiation, i.e. the hydrogen atoms, emit their radiation independently of each other. The number of atoms in which the transition occurs in unit time, and which therefore send out a quantum of energy, fluctuates statistically with time: the power intercepted on Earth therefore fluctuates too.

In optical astronomy, the receiving apparatus (e.g. a photocell) generally produces comparatively little noise, at most about as much as the noise power of the radiation itself. Use is often made of a photographic plate; the long exposure completely eliminates the effect of the noise. In radio astronomy the intensity of the received radiation, particularly for short waves, is small with respect to the receiver noise. This is also true for the 21.1 cm radiation; hence the receiver noise has a critical effect on the accuracy of the measurement. In principle, the effect of the receiver noise can be reduced by means of long integration times, which imply longer periods

[1]) H. C. van de Hulst, Ned. T. Natuurk. **11**, 201, 1945.

[2]) H. I. Ewen and G. M. Purcell; C. A. Muller and J. H. Oort; J. L. Pawsey; all in Nature **168**, 356-358, 1951. The amplifier then used at Kootwijk was developed in co-operation with Dr. F. L. H. M. Stumpers of the Philips Research Laboratory at Eindhoven.

for a given measurement. In practice, lengthening integration times is restricted by the unavoidable instability of the receiver. It is therefore necessary to take more sophisticated measures to reduce or eliminate its effect.

The Netherlands Foundation for Radio Astronomy, and in particular the Kootwijk group, has developed a receiving and amplifying installation which aims at squeezing out the last drop of available information. This is the receiver to be described in these pages. Before embarking on the description it may be useful to say a few words about some of the results so far obtained in the study of the hydrogen radiation, to give an impression of the many achievements of this new method and its potentialities.

Radio-astronomical observations at a wavelength of 21 cm

As already stated, observations of the interstellar hydrogen line yield information on the structure of the Galactic System. The latter consists of a great number of stars (about 10^{11}), distributed in space roughly in the form of a flattened disc with a diameter of about 30 000 parsec (1 parsec $\approx$ 3 light years $= 3 \times 10^{18}$ cm), and an average thickness of about 1000 parsec (*fig. 1*). This disc rotates on an axis through its centre, perpendicular to the galactic plane, the rotation time in the neighbourhood of the sun being about 200 million years. The galaxy is not composed only of stars: there is also interstellar matter; the two are present in about equal quantities as far as mass is concerned. The density of the material in the galaxy is not homogeneous. Stars frequently occur in groups, whose members can be distinguished from other stars in that they all perform nearly enough the same motion; also the interstellar matter is concentrated in clouds. Certain galactic bodies have preferred positions near the equatorial plane; others, on the other hand, are only to be observed at high "galactic latitude". Gas and dust clouds, for example, are always to be found in the region of the equator. This is a rather a pity, since these clouds with their light-scattering properties obscure a large section of the galaxy from us. The solar system is also situated near the equatorial plane and at a great distance from the centre. The absorbing matter makes it quite impossible to observe the region around the centre optically; it is only possible to make indirect deductions about this region from observations of objects in the neighbourhood of the Sun and from observations at higher galactic latitude, in which we look past the centre. It is therefore not possible to determine accurately the direction and distance of the centre. Clearly then, the information obtainable about the Galactic System by optical means is necessarily vague. The first detailed results come from very recent work and of necessity concern only the immediate neighbourhood of the Sun. They lead to some indications of the existence of a spiral structure, such as occurs in many galaxies outside our own, but whose existence in our own galaxy has not hitherto been demonstrable.

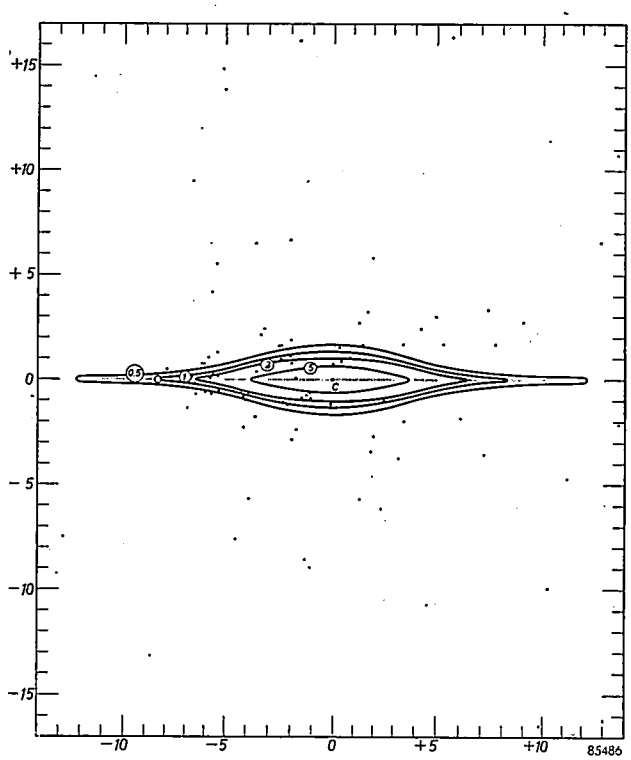


Fig. 1. Schematic section of the Galactic System, perpendicular to the equatorial plane. The Sun is indicated by the open circle (between 0.5 and 1), the centre is the point *C*. The interstellar matter is shown as a narrow hatched band; the full lines are contours of equal density (relative values indicated by the encircled numbers; the contour containing the Sun is arbitrarily chosen as of unit density). The galaxy rotates about an axis through *C*, perpendicular to the equatorial plane. Scale: 1 unit = 1000 parsec.

Radio waves suffer much less from absorption by the interstellar matter than light. Interstellar gas, in general, absorbs very little but the interstellar dust, though having little effect on radio waves strongly scatters visible light. The cause of this lies in the fact that the diameter of the dust particles is of the same order as the wavelength of light. The radio waves are much longer and therefore suffer little scattering. Thus it is possible to observe sources of radio waves which are situated almost on the opposite side of the galaxy from the Earth.

From observations of other galaxies such as, for example, the spiral nebula in Andromeda, it seems that the interstellar matter occurs preferentially in the arms of the spiral. In our own galaxy, this is probably also the case. The concentration of material near the equatorial plane suggests this. Owing to the small absorption, which makes it possible to observe the hydrogen radiation from great distances, it seems possible to obtain more exact information about the structure of the galaxy as a whole from study of this radiation. To this end, an extensive mapping of the intensity and of the other characteristic properties of the 21 cm radiation over the whole sky is required. The results so far obtained at Kootwijk, give an impression of the distribution of the hydrogen clouds over the different directions in the galaxy as seen from the Sun. This, however, is not sufficient; it is also required to know the distance of the source of radiation coming from a given direction. Oort and coworkers [3]) have used an elegant method for determining this distance. This is based on the differential galactic rotation, the phenomenon that the angular velocity of rotation of the galaxy decreases from the centre outwards. This means that a hydrogen cloud viewed in a given direction, possesses a velocity component towards or away from the Earth purely as a result of the differential rotation. In *fig. 2* the

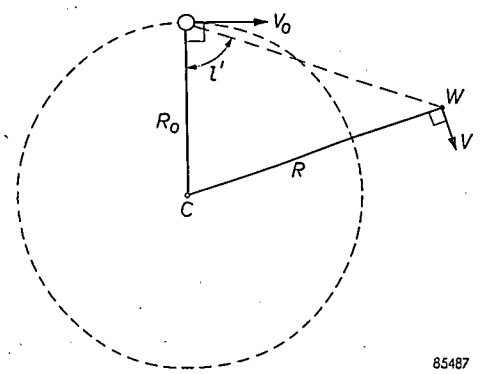


Fig. 2. Illustrating the relation between differential galactic rotation and radial velocity relative to the earth.

situation in the equatorial plane of the Galactic System has been sketched. Let V_0 be the translational velocity with which the Sun rotates around the centre C. The distance from the Sun to the centre is R_0. A hydrogen cloud W is situated at a distance R from the centre; the angle at which it is observed from the Sun (or the Earth) is l' (the direction of the centre corresponds to $l' = 0$; l' is related to the galactic longitude l according to the equation $l' = l - 327.5°$). The translational velocity of the cloud is V. The component of this velocity in the direction of sight is then $VR_0 \sin l'/R$, the component of the velocity of the Sun in the direction of the cloud is $V_0 \sin l'$, and the radial component of the velocity of the cloud relative to the earth as a result of the differential rotation is thus:

$$V_{\text{rad}} = \left(\frac{V}{R} - \frac{V_0}{R_0}\right) R_0 \sin l'.$$

Writing this in terms of angular velocities ω and ω_0, we get

$$V_{\text{rad}} = (\omega - \omega_0)\, R_0 \sin l', \quad \ldots \quad (1)$$

(A radial velocity away from the Earth is denoted positive). In the direction of the centre and in the opposite direction (anticentre) V_{rad} is zero; for all other directions however it has a finite value because $\omega \neq \omega_0$ if $R \neq R_0$. The relationship between ω and R is known approximately from statistical investigations and thus, if V_{rad} can be determined for a cloud, the distance from the centre and thus also from the Sun can be estimated.

The radial velocity can be found by studying the profile of the received 21 cm line in detail. The line is not sharp, that is to say, radiation of neighbouring wavelengths is also received. The centre of the line in general does not lie exactly at 21.1 cm, owing to the Doppler effect: the existence of a radial velocity between source and observer gives rise to an observed wavelength λ' of a spectral line which differs from the real wavelength λ in the ratio

$$\frac{\lambda'}{\lambda} = 1 + \frac{V_{\text{rad}}}{c}, \quad \ldots \ldots \quad (2)$$

where c is the velocity of light. If V_{rad} is positive (object moving away), the line appears to lie at a greater wave length, and vice versa.

If we observed only one hydrogen cloud in our line of sight, and if this cloud as a whole participated only in the rotation of the galaxy and possessed no motion of its own, the spectral line would remain sharp and merely be displaced. V_{rad} could then be calculated directly from the displacement. Actually, the clouds have their own motions, with an average speed of 8 km/sec in arbitrary directions, and therefore the line is not sharp, but exhibits a broad profile. *Fig. 3a* and fig. 3*b* show profiles of the 21 cm line observed in various directions. In the direction of the centre ($l' = 0$) and the anticentre ($l' = 180°$), corresponding to the galactic longitudes $l = 327.5°$ and $l = 147.5°$, the profile is almost symmetrical; differential rotation gives rise to no radial velocity

[3]) H. C. van de Hulst, C. A. Muller and J. H. Oort, Bill. Astr. Inst. Neth. **12**, 117, 1954. For a simpler treatment, see H. C. van de Hulst, Observatory **73**, 129, 1953.

in this case. In all intermediate directions, the profile is asymmetrical. The centre of gravity of the "line" is displaced and sometimes the line shows two or three definite maxima. Correction of the asymmetrical profiles for the individual cloud motions deduced from the profiles at $l' = 0$ and $l' = 180°$, makes it possible to derive the effect of differential rotation and, at the same time the

The interstellar hydrogen is concentrated in the hatched regions, and close examination shows that these regions have the character of spiral arms. Unfortunately the measurements give less unambiguous results as soon as one is dealing with that region of the system which lies closer to the centre than the Sun; there are then two points along the line of sight where the cloud can be situated and it

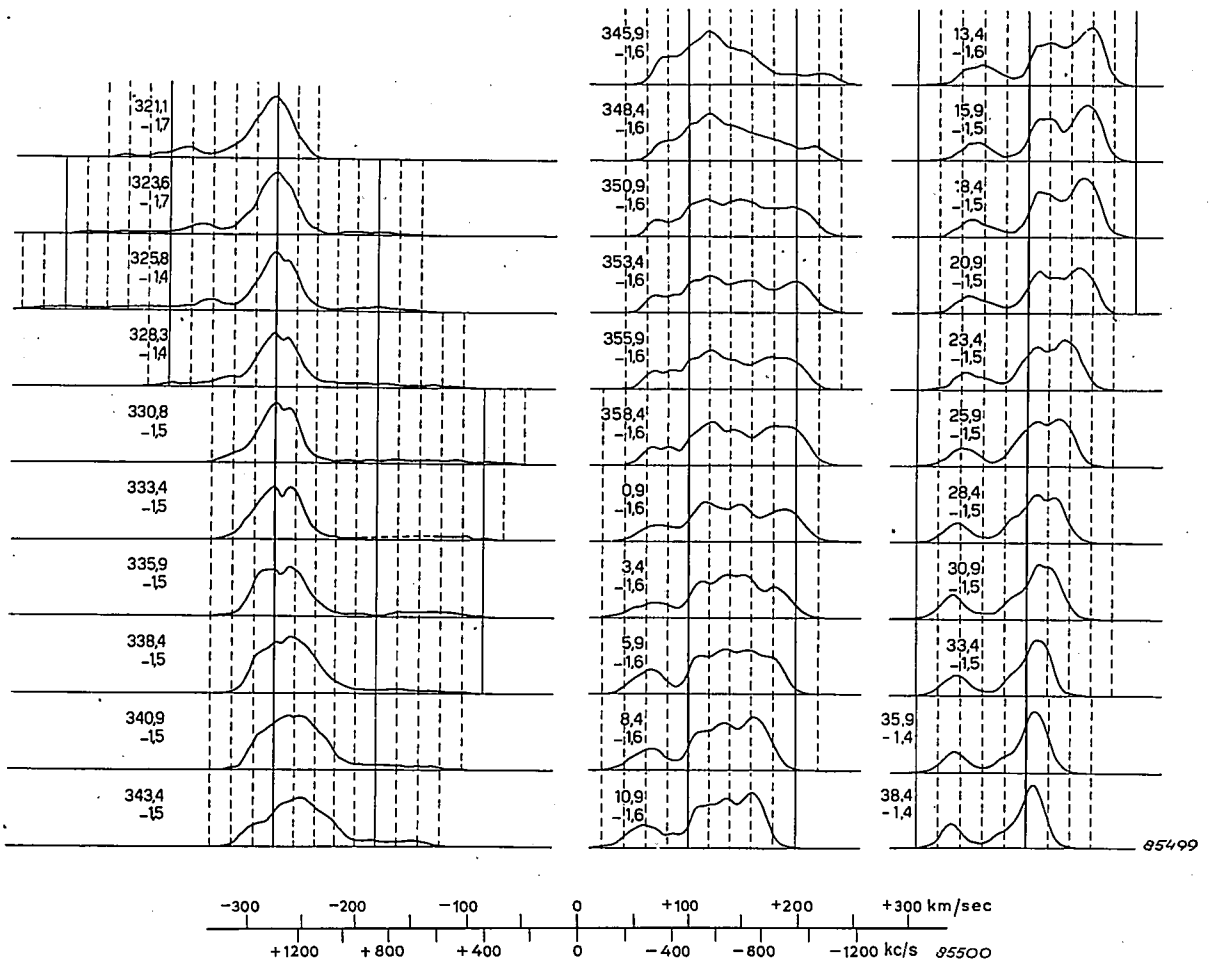


Fig. 3*a*. Line profiles of the hydrogen spectral line at 21 cm, for directions between 320° and 40° galactic longitude, i.e. near the galactic centre ($l = 327.5°$). (The negative numbers —1.4.., —1.7 represent the galactic latitude.) The deviations from the ideal profile (as found in the direction of the centre itself) depend on the Doppler effect resulting from differential rotation. The profiles can be interpreted with the help of the scale underneath the figure; this gives the relationship between relative velocity and frequency displacement.

distances of the clouds or cloud complexes from which the radiation orginates. The occurrence of more than one maximum indicates the existence of various regions of clouds situated at various distances from the Sun along the line of sight under consideration.

The distribution of clouds found in this way can be plotted on a chart. This gives a picture such as that in *fig. 4* [3]).

is very difficult to distinguish between them. Even so, indications have been found here too of a spiral structure, which is shown with dotted lines in fig. 4.

Radio-astronomical research thus lends support to the hypothesis of a spiral structure suggested by conventional astronomical work.

This is not a suitable place to go further into the many interesting results which radio astronomy has obtained in the few years of its existence. Apart

from the work on hydrogen radiation there is, for example, work proceding on continuous radiation which, amongst other things, gives us an impression of the distribution of matter generally in the Galactic System. There are also the investigations on the radiation from the Sun (chiefly originating in the corona). We shall, instead, return to our theme and discuss the basis of the Kootwijk receiver.

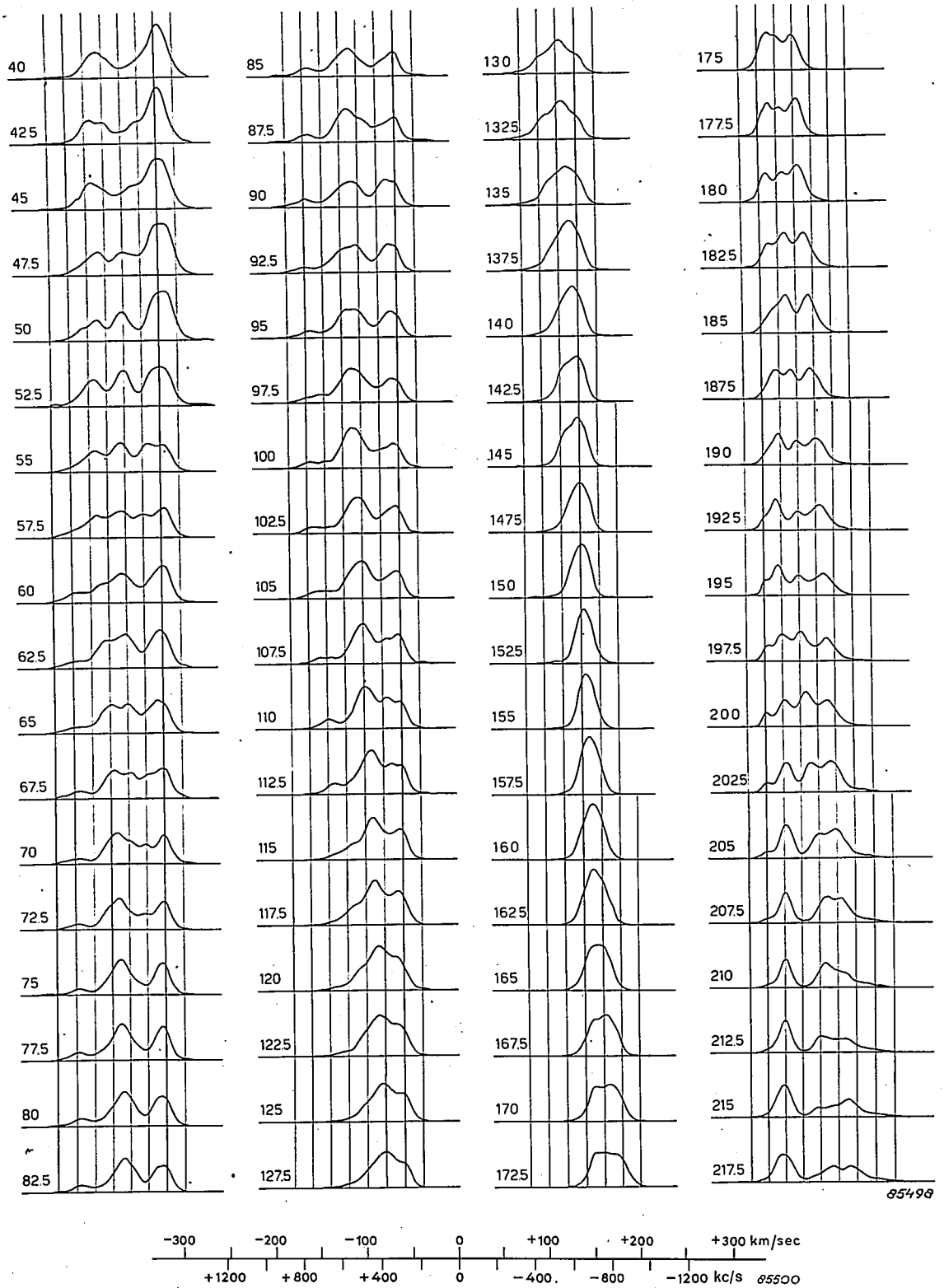


Fig. 3*b*. As for fig. 3a, for other directions in the equatorial plane of the Milky Way. The ideal profile occurs in the direction of the anticentre ($l = 147.5°$).

Properties of the radiation and design of the antenna

The aim of the investigations at Kootwijk was to measure the intensity of the hydrogen radiation as a function of position in the sky and of the frequency. In radio-astronomy, the intensity of the radiation received is usually expressed as a temperature. The energy received is imagined as coming from a closed envelope surrounding the antenna, and the temperature given is that which this envelope would have to have in order that, within the narrow bandwidth of the receiver, it should supply the same radiation power to the antenna as the latter in fact receives from space. In such a system, there is radiation equilibrium between antenna and envelope, and the

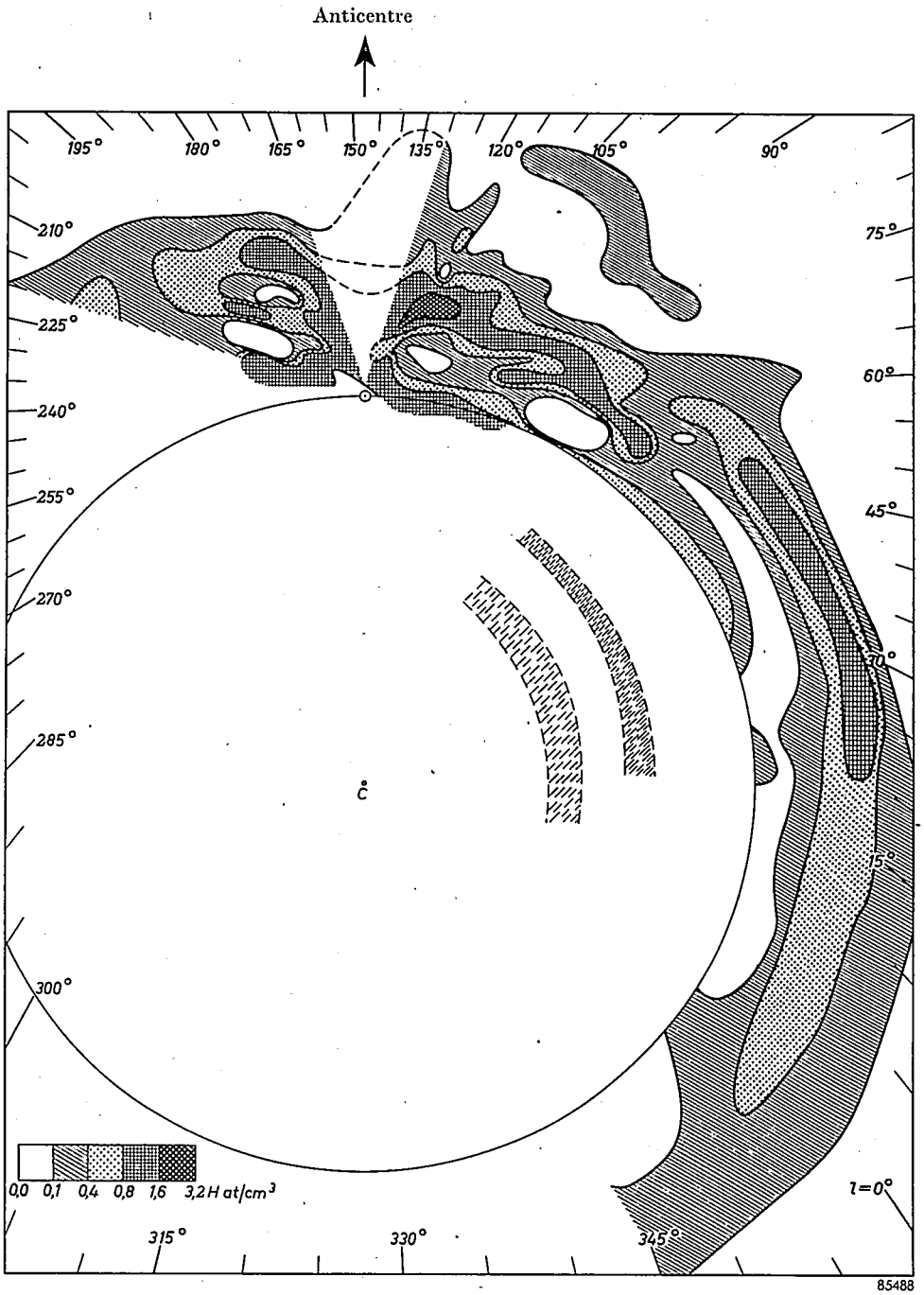


Fig. 4. Distribution of interstellar hydrogen in the equatorial plane of the Galactic System (schematic). The numbers around the edges give the galactic longitude l; the centre is indicated by C and the Sun by a small circle vertically above it. The hydrogen density is given by the shading (see code bottom left). The spiral character of the arms can be detected. Observations of spiral arms which lie closer to the centre than the Sun, are drawn with broken lines. (Taken from articles [3]) and [4])). Between $l = 130°$ and $l = 165°$ the effect of the differential rotation is so small that the measurements are not sufficiently accurate.

two temperatures are equal. It is thus usual to speak of "antenna temperature". This temperature would correspond to the kinetic gas temperature of the hydrogen clouds if they radiated as a black body (which depends partly on the thickness of the gas layer) and the antenna received radiation only from the clouds. According to various measurements, the kinetic gas temperature is about 125 °K. The highest measured "antenna temperature" is about 118°K. This corresponds, for a receiver bandwidth of 40 kc/s, to a noise power of 6×10^{-17} watts. (In the R.F. region the Rayleigh-Jeans law applies, which states that the radiation intensity of a black body is proportional to the temperature and the square of the frequency.) In the receiver to be described, temperature differences of 1 °K representing differences in power of 5×10^{-19} watts, can be measured.

The connection between "observed" intensity distributions (i.e. antenna temperature) and the "real" intensity distribution depends on the direction characteristics of the antenna used. The antenna pattern shows the sensitivity of the antenna for reception from various directions, and consists generally of a "major lobe" and a few smaller "side lobes" (*fig. 5*). The observed radiation intensity must be corrected for this antenna pattern. In practice the observed antenna temperature is taken as an average of the real temperature distribution across the half-width of the major lobe (c.f. fig. 5).

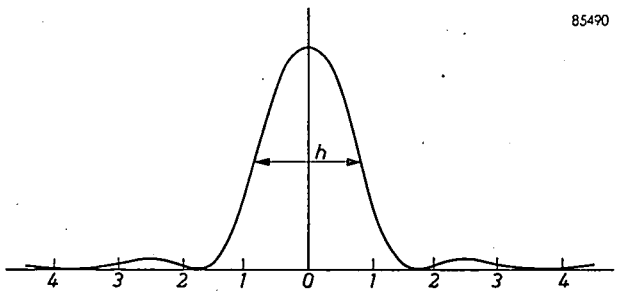


Fig. 5. Schematic directional diagram of a parabolic antenna. The ordinate represents the relative sensitivity, whilst the abscissa represents the direction relative to the axis, plotted in arbitrary units of angular measure. h is the half width of the "major lobe".

The finest details of the directional distribution which can be measured, are therefore fixed wholly by this half width.

The directional diagram of a parabolic reflector is almost entirely determined by diffraction phenomena. According to the theory, the half width of the major lobe (in radians) is given by:

$$\varphi \approx \lambda/D$$

in which λ is the wavelength of the radiation and D the diameter of the mirror. In the case of optical mirrors, the wavelength concerned is about $10^5 \times$ smaller than the hydrogen radio waves; and since the mirrors used in both cases differ in size by a factor no greater than 10, this means that the resolving power of a radio telescope is 10 000 to 100 000 times worse than that of an optical telescope. This is one of the disadvantages of radio astronomy.

The Kootwijk reflector (now at Dwingelo) is a paraboloid of revolution constructed of punched aluminium plate. It has a diameter of 7.5 m and a focal length of 1.70 m (*fig. 6*). Mounted at the focus is the antenna, a half-wave dipole. The observation cabin, which contains the receiver is mounted so that it turns together with the reflector in the horizontal plane. The reflector can also be moved in the vertical plane. The accuracy of adjustment is 0.1° in both directions. This is more than adequate since the half-width of the major lobe of the antenna pattern in the two main directions (along the axis of the dipole and perpendicular to it) is 2.7° and 1.9° respectively.

The frequency of the hydrogen line is accurately known from laboratory measurements [4]), viz.

$$f_0 = 1420.4056 \text{ Mc/s}.$$

The line width (half-width of the line profile) which is theoretically to be expected as a result of the thermal motion of the hydrogen atoms in a cloud, is about 5 kc/s. The Doppler effect mentioned above gives rise, in practice, to complicated profiles whose width varies between 0.05 and 1 Mc/s (cf fig. 3). Furthermore, there is in general also a weak continuous radiation from the galaxy, the temperature of which is usually much less than 10 °K, but which can rise in some directions to 60 °K. This radiation can be regarded as independent of frequency in the small frequency range under consideration. The spectrum of the galactic radiation (in this frequency range) thus looks like that shown schematically in *fig. 7*. The finest details of this profile which can be measured, are determined by the bandwidth of the receiver.

Principle of the method of measurement

In principle, the measurement of the profile of the spectral line, as sketched in fig. 7, could be done by traversing the frequency range with a receiver of very narrow bandwidth and recording the amplified power. An important factor which makes it necessary to use a more complicated measuring technique, has already been mentioned above: the intensity of the radiation received is at most only

[4]) E. G. Prodell and P. Kusch, Phys. Rev. 79, 1009, 1950.

a few percent of the noise intensity generated in the first stages of the amplifier. In addition, the statistically fluctuating character of the received radiation makes it necessary to take each measurement over a considerable period, in order to average out the radiation intensity at a particular frequency sufficiently ($1°K \approx 0.1\%$ of the noise level in the receiver). To traverse completely the frequency range of interest takes at least an hour. However it is scarcely possible to keep the properties of the receiver constant enough for so long a time. It is therefore preferable to choose a method which avoids the effects of small changes in amplification. Since the receiver characteristics change only slowly, if a measurement is carried out within, say, 0.01 seconds, the receiver characteristics can be considered constant during this interval. Of course, the accuracy of such a short measurement is small. It can be improved, however, by taking the average result of a large number of measurements made in quick succession.

In the Kootwijk receiver, shown in *fig. 8*, each measurement consists of determining the difference between the output voltage at a frequency f_1 on the spectral line and that at an adjacent frequency f_2 (see fig. 7). Each measurement occupies a period of 1/400 sec; the receiver is switched to and fro between the two frequencies f_1 and f_2 at a switching rate of 400 c/s. To traverse the complete line profile, the two frequencies are displaced together slowly while maintaining

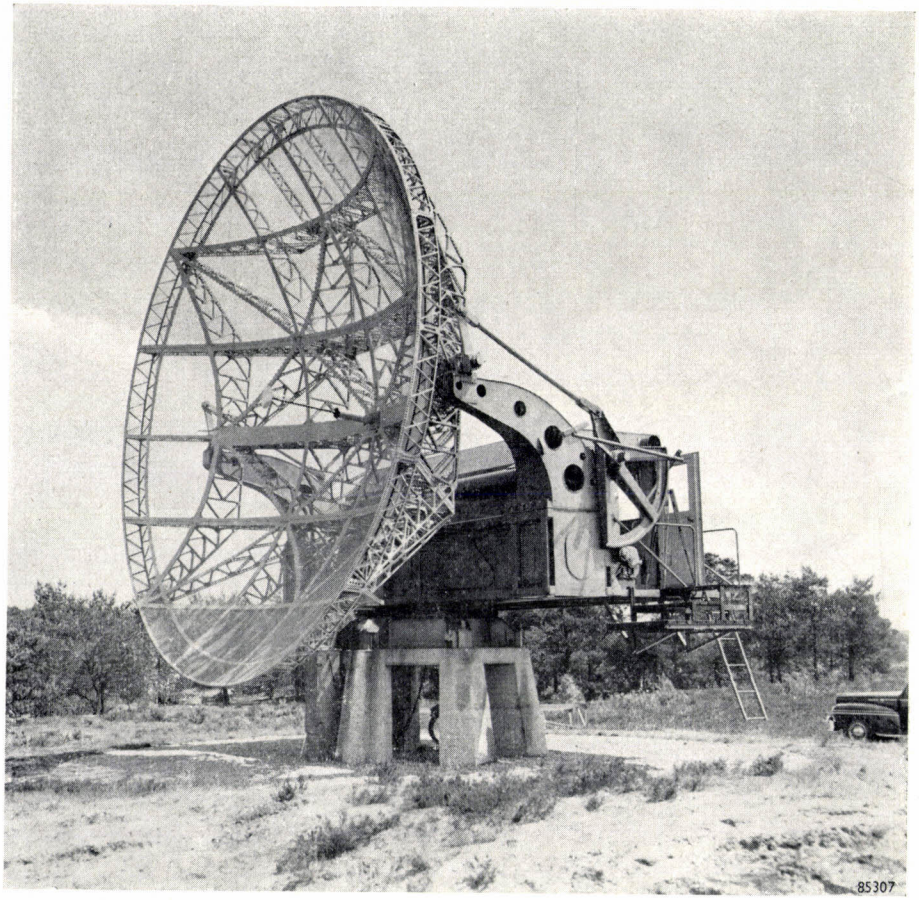


Fig. 6. The parabolic antenna at Kootwijk (now at Dwingelo). The reflecting surface is formed of punched aluminium sheet which is not easily visible in the photograph.

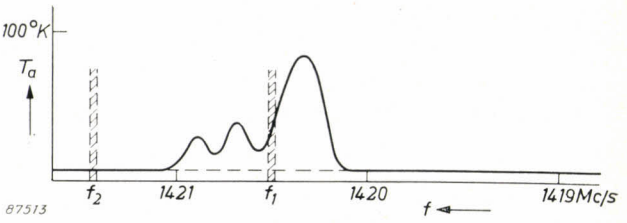


Fig. 7. Schematic representation of a line profile, i.e. of the average intensity (expressed as antenna temperature T_a) of the Galactic radiation as a function of the frequency near the 21 cm line.

a constant frequency difference. The frequency difference is so chosen ($f = 1080$ kc/s) that the comparison frequency f_2 always falls outside the spectral line. Since the (noise) signal at the frequency f_1 is always, on the average, somewhat greater than at the frequency f_2, the output voltage after rectification displays a square wave form with a frequency of 400 c/s, the amplitude of which is a measure of the difference in noise levels at the two frequencies. This method not only eliminates the effect of changes in the receiver characteristics, since each measurement lasts only a short time, and but also, since it is a difference method, suppresses the contribution of the continuous background. By measuring only the fundamental of the 400 c/s square wave voltage, with an extremely narrow frequency range on either side, the effect of the receiver noise is considerably reduced.

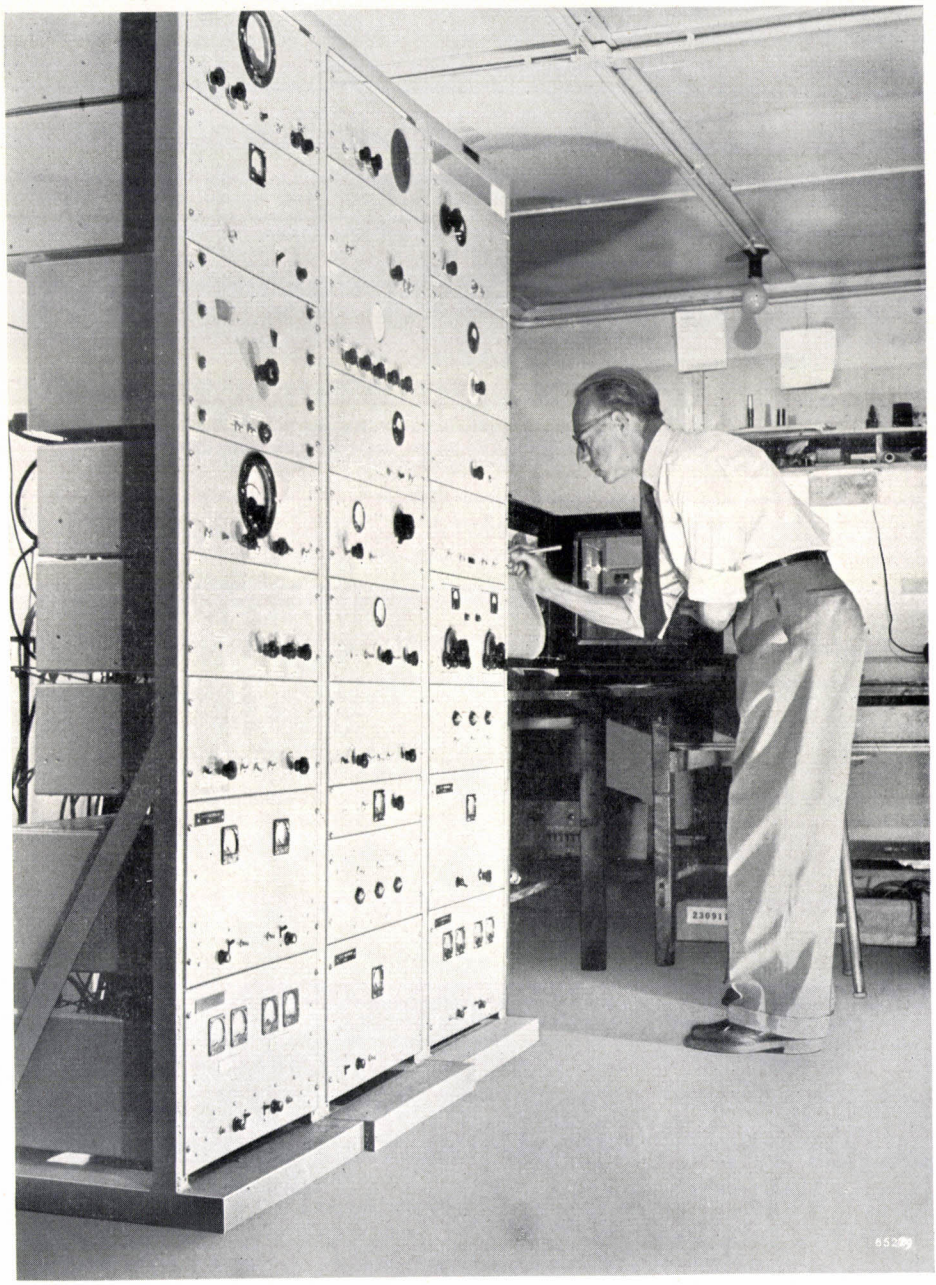


Fig. 8. The radio receiver as installed at Kootwijk.

Since during each half period of the measurement the receiver is tuned to a frequency outside the spectral line, the available information time is only half used, and the accuracy of the measurement is thus reduced by a factor 2. To improve on the usage of the available time, two comparison frequencies are used in practice. A more detailed explanation and some of the circuit details are given in the second part of this article.

The amplitude of the 400 c/s output signal is recorded on a strip-chart recorder in which the chart movement is correlated with the frequency. *Fig. 9* shown a typical recording obtained on the Kootwijk equipment.

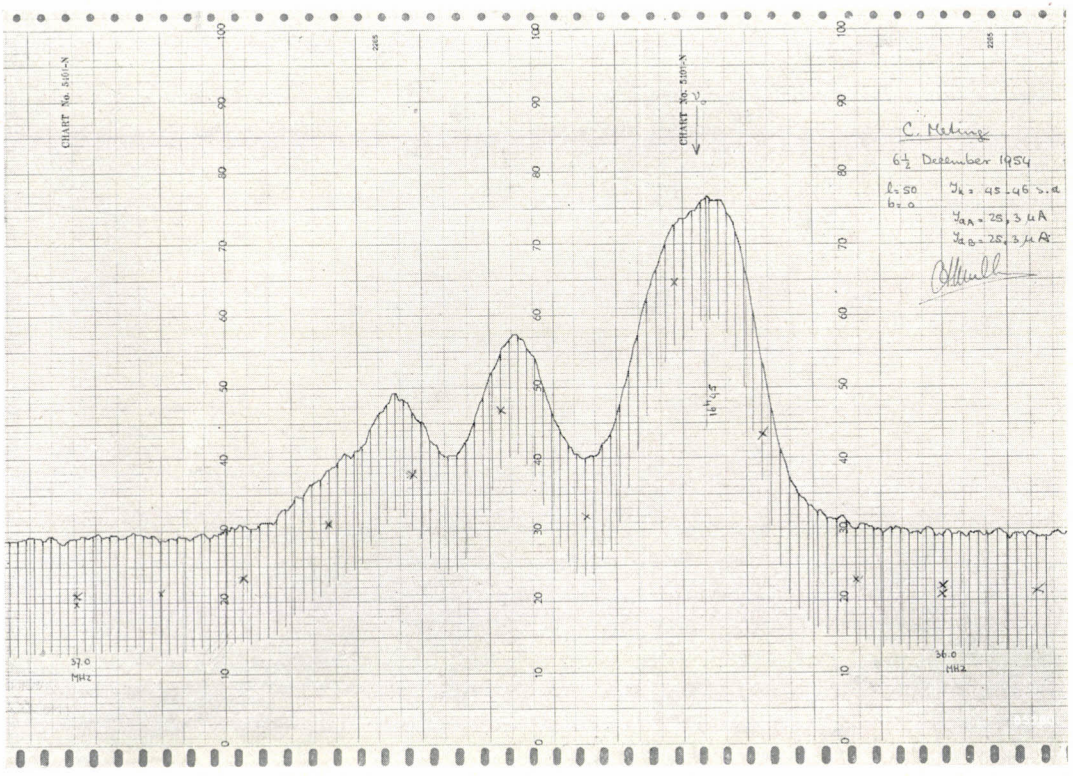


Fig. 8. Example of a record. The vertical lines at 10 kc/s intervals mark the frequency.

Summary. Radio astronomy has grown in a very short time to be a powerful tool among astronomical techniques. This article deals briefly with the structure of the Galactic System and the role played by interstellar hydrogen as source of radio waves of sharply defined frequency; this is followed by a short description of the principles of operation of a receiver used to detect these waves and measure their intensity set up at Kootwijk (and now removed to Dwingelo) by the Netherlands Foundation for Radio Astronomy. The hydrogen radiation from space is picked up by a dipole antenna in the focus of a parabolic reflector 7.5 m in diameter. The radiation has the character of noise and is very weak, many times weaker than the noise level in the receiver itself; this necessitates special methods of measurement in order to achieve a reasonable accuracy. If we express the received noise power as a temperature defined by means of the Rayleigh Jeans law, the maximum received intensity corresponds to a temperature of about 110 °K. The accuracy of measurement reached is 1 °K, i.e. 0.1% of the receiver noise.

A TRANSISTOR HEARING-AID

621.395.92:621.375.4

The most obvious advantages of the transistor over the thermionic valve are that the dimensions and the power consumption are considerably smaller. It requires, moreover, only one source of energy of very low voltage (a few volts only); and in comparison with electronic tubes there is not only the advantage (where dry batteries form the source) of lower power consumption but also the fact that the energy (the number of watt hours) is supplied more cheaply from a battery of a few volts than from a battery of some tens of volts. Where hearing-aids are concerned, small dimensions and low battery costs are of course two of the most important characteristics. It is hardly surprising therefore, that one of the first applications of Philips transistors has been in hearing-aids.

The first Philips transistor hearing-aid (type KL 5400) has meanwhile been superseded by a second (type KL 5500, *fig. 1*) in which the specific properties of transistors are even better employed. This apparatus, now on the market for about a year, weighs only half as much as the hearing-aid using

valves, earlier described in this Review [1]), whilst running costs (batteries) have been reduced to $^1/_{10}$. Another appreciable advantage of the new apparatus is that it provides a greater acoustical gain and produces a higher sound level than the valve hearing-aid; a wider circle of those who are hard of hearing may therefore profit from it.

The hearing-aid KL 5500 consists of an electromagnetic microphone, an amplifier with four transistors, a telephone inset and a battery supply of one, two or three 1.2 V cells. The nominal acoustical gain is 58 dB with one cell and 69 dB with three cells, corresponding to a power output of the amplifier of 1 and 10 mVA respectively.

The transistors (three of type OC 70 and one, in the output stage, of type OC 71) are resistance-coupled and connected according to the common emitter type of ciccuit [2]). The output transistor supplies its power directly to the telephone without the intermediary of an output transformer as required with a tube. The omission of an output transformer is possible because of the inherently low output impedance of the transistor circuit in question [2]). As a rule all impedances of transistor circuits are lower than those of circuits with electronic tubes. A consequence of this is the great advantage of transistors that undesired couplings due to stray capacitances are negligible, so that a compact construction can be used. Since transistors, unlike valves, exhibit no microphony, no special provisions are required to counteract this effect; this also contributes towards a compact construction.

For a hearing-aid, a good limiter in the output stage is essential; this is to prevent the threshold of pain of the wearer being exceeded in case of shouting or other loud noises; patients with reduced span are particularly susceptible to this trouble. In the hearing-aid KL 5500 limiting is obtained by selecting appropriate values for the current and the impedance in the collector circuit of the OC 71. For further particulars we refer the reader to a more detailed article to appear in this Review.

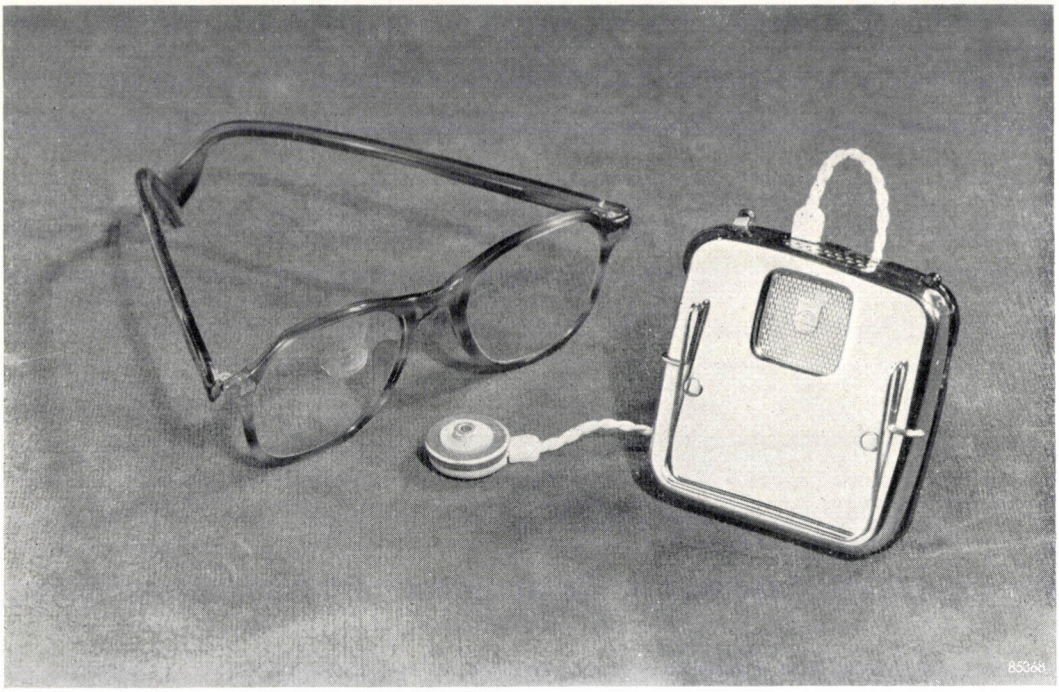


Fig. 1. The hearing-aid KL 5500 which works on transistors.

In order to provide a sound that is agreeable to the wearer, one of the essential requirements is a smooth and adjustable response curve. An uneven response curve is mainly due to resonances of microphone and telephone. The response curves of both have been improved by providing acoustical cavities with damping. The sensitivity is impaired by this, but a higher sound level than that of a valve hearing-aid can nevertheless be readily obtained. If necessary, the power output can be increased to the relatively high value of 10 mVA, by using three cells (3.6 V).

As far as the adjustment of the response curve is concerned, the wearer has first of all the choice of

[1]) P. Blom, An electronic hearing-aid, Philips tech. Rev. **15**, 37-48, 1953/54.

[2]) See, for example, Philips tech. Rev. **17**, 242, 1955/56 (No.9).

three telephones with different characteristics; moreover, the response curve of the amplifier can be adjusted by means of a three-position selector knob. A choice of nine response curves is thus available. As in all fields of communications, the signal-to-noise ratio is an important quantity. The first transistors were, as regards noise, somewhat inferior to electronic tubes, but the modern types have been so much improved that at present a transistor apparatus can be made having less noise than a good valve hearing-aid. The equivalent input signal for noise in a valve hearing-aid is approximately 25 dB above zero level (10^{-16} W/cm^2), whereas in the hearing-aid KL 5500 this quantity is less than 20 dB.

Not only the electronic noise but also the case noise has to be considered. This latter is caused by the fact that the ear of the patient lies, as it were, on his clothes. This noise shows a fairly uniform distribution throughout the spectrum, and the level of interference thus caused corresponds roughly to the level of normal speech at a distance of 1 metre. In the KL 5500 this noise has been considerably reduced by acoustical insulation of the microphone cavity from the rest of the apparatus.

P. BLOM.

BEAM TRANSMITTERS WITH DOUBLE FREQUENCY MODULATION

by C. DUCOT *).

621.396.43:621.376.32:621.396.5

The development of the very-high-frequency techniques (wavelengths of a few centimetres) has made it possible to establish carrier-telephony communication by means of accurately directed radio beams. Equipment developed for this purpose will be described in a number of articles in this Review. It will be shown that the application of double frequency modulation results in a favourable signal-to-noise ratio, and also provides some practical advantages important to the telephone service.

Beam transmitters operating on wavelengths between a few centimetres and a few metres are being increasingly used for the transmission of telephone calls and TV-signals. Two methods of modulation may be adopted for this, viz. frequency modulation [1]) and pulse modulation [2]). For transmitting a large number of telephone channels or one or more TV-signals, only frequency modulation can at present be employed. In this system a group of speech channels is transmitted as a single signal, in the same way as in carrier telephony via cables. A beam-transmitter link using frequency modulation may therefore take the place of a normal cable circuit without any complications. At large distances between transmitter and receiver the input signal of the receiver is necessarily attenuated, so that the ratio between the desired signal and the inevitable noise deteriorates. This noise, which originates mainly in resistors and electronic tubes, will be referred to here as *fluctuation noise*. Apart from this, in carrier telephony, other interference voltages are added to every channel, owing to the fact that the transmission of a signal always causes some distortion.

This distortion is mainly due to the fact that the modulation characteristic of the frequency modulator in the transmitter as well as the demodulation characteristic of the discriminator in the receiver are never perfectly straight. Furthermore, a certain distortion arises because the phase-characteristic of the networks through which the signal passes in transmitter and receiver, are not perfectly linear. Finally, modulation distortion may arise from the signal being transmitted via a long cable (e.g. an antenna feeder) at least one extremity of which does not have a reflection-free termination, i.e. a resistance equal to the characteristic impedance of the line. The distortion due to the two latter causes is often termed phase distortion. When the signal is transmitted through a wave guide, even a perfectly reflection-free termination cannot entirely prevent phase distortion. Finally it should be mentioned that modulation distortion can also arise when the electromagnetic waves from the transmitter reach the receiver along different paths.

In the case of a large number of channels, which carry widely varying loads in the course of normal telephonic traffic, interfering alternating voltages are manifested as noise in every channel. In accordance with the usual terminology this noise will be referred to as *intermodulation noise* [3]).

In this article we shall deal with a method which makes it possible in many cases to reduce the noise

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

1) Cf. A. van Weel, An experimental transmitter for ultra-short-wave radio telephony with frequency modulation, Philips tech. Rev. **8**, 121-128, 1946.

2) Cf. C. J. H. A. Staal, An installation for multiplex-pulse modulation, Philips tech. Rev. **11**, 133-144, 1949.

3) See: J. te Winkel, Feedback amplifiers for carrier telephone systems, Philips tech. Rev. **16**, 287-296, 1954/55.

in a radio link, thus enabling either greater distances to be bridged or more channels to be transmitted.

The method consists of so-called double frequency modulation, in the sense that first an auxiliary carrier wave is modulated with the signal, after which the main carrier wave, generated by an oscillating transmitting valve, is modulated with the modulated auxiliary carrier.

By way of introduction, we shall first consider the requirements to be met by a radio link for carrier telephony. After discussing the method of double modulation we shall finally describe an experimental installation operating on this principle.

Criteria for a radio link

During a congress held in Florence in October 1951, the C.C.I.F. (Comité Consultatif International de Téléphonie) recommended that beam-transmission links for carrier telephony should conform, whenever possible, to the standards applying to international cable telephone circuits. These standards are defined with the aid of a "standard circuit". This consists of a transmission path of total length 2500 km, split up at nine places where demodulation into the supergroups (each containing 5 groups of 12 channels) and renewed modulation takes place.

Owing to attenuation and the interposition of line amplifiers (repeaters) the signal power varies considerably along a telephone line (e.g. the power is low at the input side of a line amplifier and high at the output side). At certain places the power is equal to that at the place of origin [4]); these places are then said to be at *relative zero level*. The strength of a signal is often expressed in decibels relative to a signal having a power of 1 mW at a point at relative zero level, and denoted by dBm_0 (m standing for mW and 0 for zero level).

One of the criteria for the above-mentioned standard circuit is that the level of the total noise in each of the channels, in so far as it originates from that particular line section, and measured psophometrically [5]) at a place with relative zero level, must not exceed a value of 7500 pW for more than 1% of the time. In this connection it must be observed that with radio links, it is necessary to take account of a special condition not found with cable circuits, viz. the occurrence of fading. When frequency modulation is used, the effect of fading is not that the damping of the circuit increases, since the magnitude of the alternating voltages obtained in the receiver after frequency demodulation is determined by the frequency swing, and this swing is not affected by fading. Fading, on the other hand does cause an increased fluctuation noise, thus reducing the signal-to-noise ratio.

When assessing the fluctuation noise in circuits incorporating one of more radio links, therefore, the incidence of fading should be reckoned with. The effect of this, however, is the same for the two systems which we shall compare here (viz. single and double modulation).

When a telephone circuit includes a radio link, there is a fixed relation between the signal level at any point of the circuit and the frequency swing of the beam transmitter. A quantity characteristic of the link is, therefore, that particular value of the frequency swing which produces a signal strength of 1 mW in the points at relative zero level. The r.m.s.-value of this frequency swing will be denoted d.

There is considerable freedom in the choice of d, since by varying the amplification before the modulator stage in the transmitter and that after the demodulator in the receiver in opposite directions, the damping of the whole circuit can be kept constant at a variable value of the frequency swing. By increasing the frequency swing, the ratio of signal to fluctuation noise can be improved. This involves, however, a greater distortion, which, as we explained above, manifests itself in an increase in the intermodulation noise. A certain value of d may be found where the sum of these two noise components is a minimum, and the frequency swing should preferably be adjusted to this optimum value. It would, therefore, be illogical to compare the properties of two systems solely on the basis of either the fluctuation noise or the intermodulation noise. The properties of a link can, however, be expressed unambiguously by the noise occuring at the optimum frequency swing.

As stated, intermodulation noise is caused by the fact that the signals are distorted in transmission. A measure of this noise, therefore, is the relative strength of the higher harmonics when a purely sinusoidal signal is transmitted instead of the composite signal from all telephone channels. In many cases it will be sufficient to consider only the second harmonic [6]). In that event the total noise

[4]) See e.g. fig. 1 in J. de Jong, Maintenance measurements on carrier telephony equipment, Philips tech. Rev. 8, 249-256, 1946.

[5]) The psophometrically measured power (which takes into account the frequency-dependent sensitivity of the human ear) can be found with an accuracy sufficient for our pupose by measuring the power in the channel in question (frequency band 4 kc/s) with a frequency-independent instrument, and reducing the result by 3 dB.

[6]) Only in the rare event that the channels are at frequencies so high that the whole frequency band of all channels is contained in one octave will the part played by the second harmonics and the corresponding combination frequencies be of little importance.

power W_b, measured psophometrically in one channel, at a point at relative zero level, can be expressed by the formula:

$$W_b = \frac{A}{d^2} + Bd^2, \quad \ldots\ldots \quad (1)$$

where A and B are constants [7]).

The fluctuation-noise level in the frequency band of one channel may be expressed by a specific frequency swing. The mean square value of this frequency swing is represented by the constant A. This quantity is not dependent upon the earlier introduced quantity d (which determines the effective signal), but does depend upon the power of the transmitter, the directional properties of the antennae, the conditions of propagation of the electromagnetic waves, and furthermore upon the noise factor [8]) of the receiver, the number of transmission links and upon the central frequency of the frequency band available for the channel in question.

The intermodulation noise, on the other hand, does depend upon d; its contribution to the total noise power in each channel is in fact proportional to d^4. The proportionality constant B is dependent upon the various factors causing the distortion, upon the value of the total signal of all channels, and (like A) upon the central frequency of the frequency band available for the channel in question.

The total noise level in milliwatts in a point at relative zero level is then obtained by dividing the sum of A and Bd^4 by d^2, which corresponds to formula (1).

This expression is a minimum for a value of d given by:

$$d_0 = \sqrt[4]{\frac{A}{B}}, \quad \ldots\ldots \quad (2)$$

and the minimum value of the noise power amounts to:

$$W_{b\,min} = 2\sqrt{AB}, \quad \ldots\ldots \quad (3)$$

This quantity determines the properties of the transmission link. (Communication is better according as the above quantity is smaller).

It should be observed that, where minimum noise is concerned, i.e. where $d = d_0$, the contributions of fluctuation noise and of intermodulation noise towards the total noise are the same. A great difference between these quantities is, therefore, an indication that the frequency swing d has not been appropriately chosen.

Double frequency modulation

One of the major difficulties in the development of a radio link is maintaining the linear relation between the signal applied to the transmitter and the signal obtained in the receiver after demodulation. Any deviation from this linearity gives rise to distortion and thus, as we explained above, to intermodulation noise. A first requirement, therefore, is that modulation in the transmitter is effected linearly, i.e. that the difference between the momentary frequency and the central frequency is exactly proportional to the voltage applied to the modulator. Circuits to realize this, however, can only operate at relatively low frequencies (up to approx. 150 Mc/s) with present facilities. The far higher frequencies that are, as a rule, required for the signal to be transmitted, may then be obtained by a system of frequency transformation. A block diagram of a transmitter operating on this principle is shown in *fig. 1a*. The signal to be transmitted is indicated here by S (frequency f_s). The oscillator O_1 (frequency f_1) is modulated by the frequency modulator FM. The modulated voltage is applied to the mixing tube M, in which, with the aid of the oscillator O_2 (frequency f_2) a voltage with the far higher central frequency f_3 is formed ($f_3 = f_2 + f_1$) this voltage, after being amplified in A_H, is applied to the transmitting antenna A_T. The output tube of the transmitter must be an amplifying tube, and

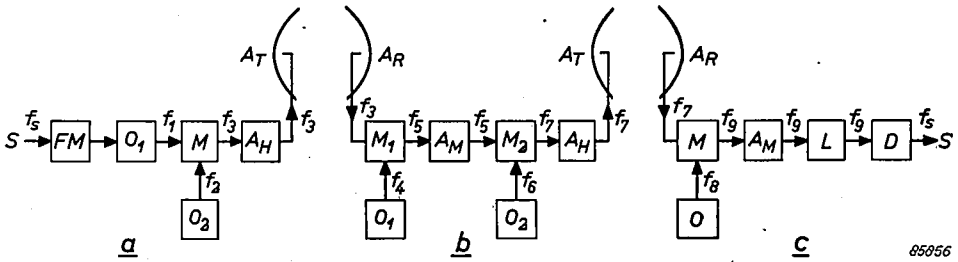


Fig. 1. Block diagrams of a transmitter (*a*), a relay station (*b*), and a terminal station (*c*), using an amplifying tube in transmitter and relay station. The significance of the letters is given in the text.

the tubes now available for the high frequencies involved (e.g. 4000 Mc/s) can supply only a low power (1 W or less). This limited output power has an unfavourable effect on the ratio of signal to fluctuation noise.

[7]) See also: A. T. Starr and T. H. Walker, Microwave radio links, Proc. Instn. El. Engrs. **99**, 241-255, 1952 (partIII).

[8]) For the significance of this term, see e.g.: G. Diemer and K. S. Knol, The noise of electronic valves at very high frequencies, II, The triode, Philips tech. Rev. **14**, 236-244, 1952/53.

When an oscillator is used as transmitting tube, a far higher output can be obtained, viz. of the order of 10 W. The signal to be transmitted is in this case directly applied to the transmitting tube O (*fig. 2a*). In this tube, frequency modulation with a very good linearity should be effected. It has been mentioned earlier, however, that this requirement cannot be satisfied at high frequencies. (Strong negative feedback could be applied, but in that case it is very difficult to obtain and maintain a stable adjustment of the transmitter). The block-diagrams *b* in figs. 1 and 2 represent relay stations, which relay the signals if more than one radio link is used. In the two cases (1*b* and 2*b*) totally different circuits have to be applied. If an amplifying transmitting tube is used (fig. 1*b*), an I.F. voltage (frequency f_5) can be obtained by combining the H.F. signal received in antenna A_r with the output of oscillator O_1 (frequency f_4) in the mixing stage M_1. This I.F. voltage, after being amplified in the I.F. amplifier A_m, is applied to a second mixing stage M_2. In this stage, by means of the oscillator O_2 (frequency f_6), an H.F. signal (frequency f_7) is again obtained which, after being amplified by amplifier A_H, is applied to the transmitting antenna A_T. As a rule f_7 is made different from f_3 so as to prevent many undesirable coupling effects between input and output. In relay stations it is therefore not always necessary to return to the A.F. signal by means of a frequency demodulator (as *is* obviously necessary when either telephones or a telephone circuit are connected to the relay station). If, however, an oscillator is used as the transmitting tube (fig. 2*b*), demodulation then has to be effected in every relay station. This is done by applying the I.F. amplified signal, via a limiter L, to the frequency demodulator D. The A.F. signal thus obtained is then again applied, via the amplifier A_S, to the oscillating transmitting tube O_2. Since the process of demodulation and modulation is always accompanied by some distortion, this can be regarded as a serious drawback of using an oscillator as the transmitting tube.

The block diagrams of figs. 1*c* and 2*c* represent terminal stations. It will be clear that these may be identical for both systems. In the mixing tube M, the I.F. signal with frequency f_9 is obtained from the antenna signal with frequency f_7 and the oscillator signal with frequency f_8. This I.F. signal is applied, via the I.F. amplifier A_M and the limiter L, to the frequency demodulator D, which produces the A.F. signal S.

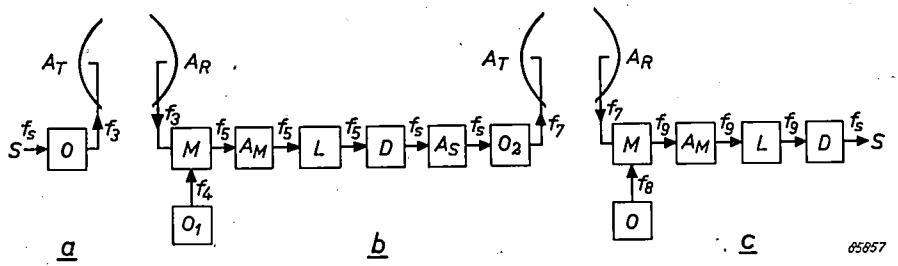


Fig. 2. Block diagrams of a transmitter (*a*), a relay station (*b*), and a terminal station (*c*), using an oscillating transmitting tube and a system of single modulation. The significance of the letters is given in the text.

In *figs. 3* and *4* the various frequencies mentioned in connection with figs. 1 and 2 are represented on a frequency scale. The sections *a*, *b* and *c* are correspondingly related to the transmitter, the relay station and the terminal station.

Owing to the high power available when an oscillating transmitting tube is used, a radio link

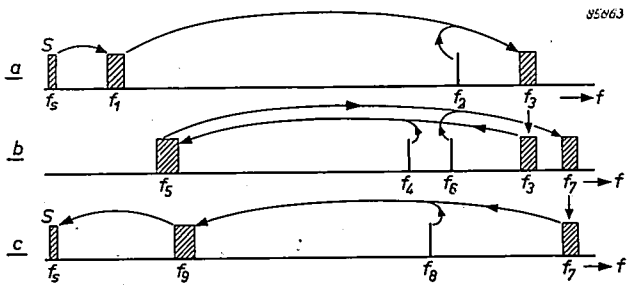


Fig. 3. Schematic representation of the frequencies of the signal voltages in a transmitter (*a*), a relay station (*b*), and a terminal station (*c*), for a circuit according to fig. 1.

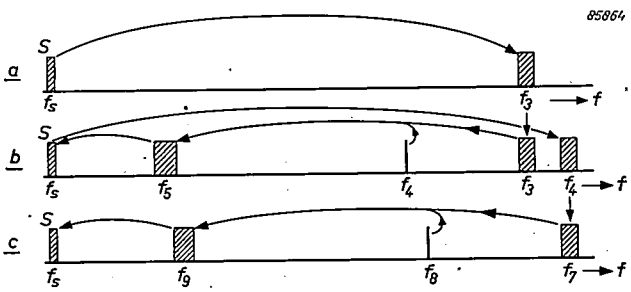


Fig. 4. Schematic representation of the frequencies of the signal voltages in a transmitter (*a*), a relay station (*b*) and a terminal station (*c*), for a circuit according to fig. 2.

operating on this principle is less likely to be affected by interference due to fading. A certain degree of fading may affect transmission by a system of oscillating transmitting tubes (fig. 2) only as a reduction of the signal-to-noise ratio, whereas under similar conditions, communication via a system with amplifying transmitting tubes (fig. 1) may be completely interrupted.

A system of modulation using an oscillator as transmitting tube and yet permitting modulation at a far lower frequency is the system of double modulation, invented by Armstrong. The application of this system to beam-transmitters was, to our knowledge, first suggested by L. E. Thompson [9]) and put into practice by the Radio Corporation of America (RCA) for an experimental communication link between New York and Philadelphia [10]). This system was subsequently applied by Western Union Telegraph for a communication chain linking New York, Washington and Pittsburgh [11]). *Fig. 5* shows the block diagram of a transmitter, a relay station and a terminal station operating on this principle, whilst *fig. 6* again represents the frequencies of the currents and voltages occuring at various points of the circuit. In the transmitter, an auxiliary carrier-wave with frequency f_1 is modulated with the signal S to be transmitted by means of the frequency modulator FM. The auxiliary carrier-wave is generated in the oscillator O_1. After being modulated it is applied, via the amplifier A_{SP}, to the oscillating transmitting tube O_2 (frequency f_3), in which f_3 is frequency-modulated with the modulated auxiliary carrier frequency. In a relay-station, the first stage of which is again a normal FM-receiver (cf. fig. 2*b*) the auxiliary carrier frequency is obtained by means of the frequency demodulator D and applied, via amplifier A_{SP}, to the oscillating transmitting tube O_2. Demodulation of the auxiliary carrier, therefore, is not necessary in a relay station. The first stage of a terminal station is likewise a normal FM-receiver (cf. fig. 2*c*). The auxiliary carrier obtained from the frequency

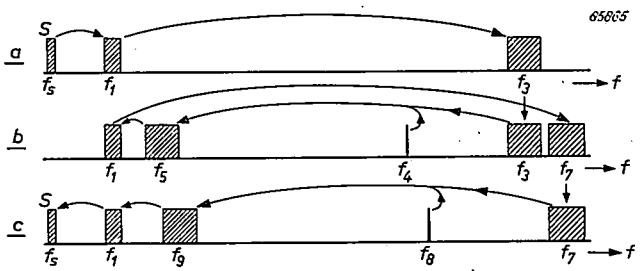


Fig. 6. Schematic representation of the frequencies of the signal voltages in a transmitter (*a*), a relay station (*b*), and a terminal station (*c*), for a circuit according to fig. 5.

demodulator D_1 is applied, via the amplifier A_{SP} and the limiter L_2, to a second frequency demodulator D_2, which finally produces the A.F. signal S.

Owing to the fact that the frequency f_1 of the auxiliary carrier must always be several times higher than the highest frequency of the A.F. signal S to be transmitted, the ratio of signal to fluctuation noise is unavoidably impaired when double modulation is applied. In the previous section, however, it was explained that it is not fair to compare two modulation systems solely with respect to the fluctuation noise. The application of double modulation unquestionably offers certain substantial advantages, viz:

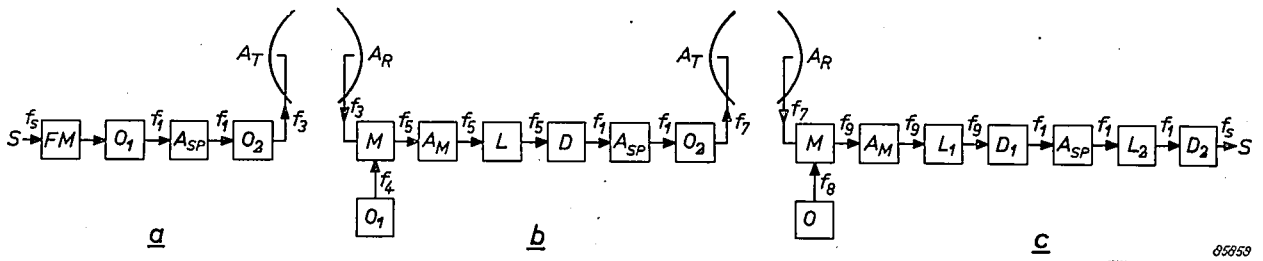


Fig. 5. Block diagram of a transmitter (*a*), a relay station (*b*), and a terminal station (*c*), using an oscillating transmitting tube and a system of double modulation.

1) Good linearity is required only in the first modulator and in the final frequency demodulator, both operating at a relatively low frequency. The modulation distortion affecting the carrier frequency or I.F. signals produces after demodulation only higher harmonics of the auxiliary carrier frequency. Since the carrier has to pass a filter before being applied to the second demodulator this distortion will affect the intermodulation noise to a very small degree only. For this reason no

[9]) L. E. Thompson, A microwave relay system, Proc. Inst. Rad. Engrs. **34**, 936, 1946.
[10]) G. G. Gerlach, Microwave relay communication system, R.C.A. Review **7**, 576-600, 1946.
[11]) J. J. Lenehan, A radio relay system employing a 4000 Mc/s 3-cavity klystron, Western Union tech. Rev. **6**, 111-116, 1952.

stringent demands as regards linearity are made on the phase characteristics of H.F. and I.F. circuits in transmitters, relay stations and receivers. The same applies to the linearity of the modulation characteristics of the oscillating transmitting tubes of transmitter and relay stations, and to the characteristics of the frequency demodulator in the relay stations and of the first frequency demodulator in a terminal station. Also the phase distortion arising from the H.F. signals being transmitted via long antenna feeders will ultimately not play an important role. A relay station in which the auxiliary carrier is not demodulated, will therefore contribute very little to the intermodulation noise, a property which is not shared by either of the other systems discussed here.

2) A consequence of the points mentioned under 1) is that the frequency swing effecting modulation of the carrier with the auxiliary carrier is not greatly restricted by considerations of linearity. A wide frequency swing (e.g. 5 Mc/s or even wider) is thus permissible, which again favourably influences the ratio of the signal voltage to the fluctuation noise, whereby the adverse influence of the high frequency of the auxiliary carrier on this ratio is largely compensated.

3) With double modulation the requirements put on the limiters are less exacting than those with single modulation [7]).

4) Modulation and demodulation of the auxiliary carrier is necessary only in those places where this is required by the telephone service. The distortion, and hence the intermodulation noise is thus virtually solely determined by the number of relay stations where this *has* to take place, and not by the total number of relay stations (the latter often being five times more numerous than the former).

5) Double modulation offers the possibility of transmitting, apart from the "normal" speech channels, also a number of signals with single modulation, e.g. a service channel or a signal for measuring and inspection purposes, which require less exacting demands as regards freedom from noise. (The advantages of double modulation obviously do not apply to these channels.)

We should mention here that the application of double modulation was particularly attractive to us, as we have at our disposal an oscillating transmitting tube of extremely simple and sturdy construction, viz. the multi-reflex klystron, designed by F. Coeterier. Modulation with the wide frequency sweep in question presents no difficulties with this tube, because its "electronic bandwidth" amounts to about 30 Mc/s. A multi-reflex klystron was described some years ago in this Review; in this issue an article appears in which later versions of this type of tube are discussed [12]).

Comparison of the signal-to-noise ratios for systems with single and with double modulation

We shall now make a quantitative comparison of the signal-to-noise ratios for beam-transmission systems with single and with double modulation. It will be clear from the foregoing that in the former system the transmitting tube is assumed to be of the amplifying type (fig. 1).

For making the comparison, let us suppose that with either of these beam-transmission systems a standard circuit is formed similar to that referred to in the C.C.I.F. specifications regarding the permissible noise. This circuit has, therefore, a length of 2500 km, and consists, as will be assumed here, entirely of beam-transmission links. In compliance with C.C.I.F. specifications, we shall assume that in nine of the relay stations, demodulation into the super-groups and renewed modulation takes place.

In order to study the two systems as far as possible under equivalent conditions, all characteristic quantities not depending upon the choice of system, will be assumed to be equal in either case. These quantities are: number and length of the links, the carrier-wave frequency, the gain of the antenna system, the conditions of propagation of the electromagnetic waves between transmitting and receiving antenna, and the noise factor of the receivers. It was mentioned earlier that in view of the available tubes, the power of the transmitters of the two systems cannot be the same. Also the frequency swing will be different for either case. As regards the latter, we shall assume that in the system according to fig. 1 this is adjusted to the optimum value with a view to the noise, whilst in the system according to fig. 5 the same applies to the frequency swing upon the auxiliary carrier. In the latter case the frequency swing on the transmitted carrier will be adjusted to the maximum value permissible for the transmitting tube in question.

When a sinusoidal signal is transmitted instead of a signal composed of telephone channels, it will likewise undergo a certain distortion, or, in other words, higher harmonics will be formed. We shall now introduce the symbol μ to designate that

[12]) F. Coeterier, The multireflection tube, a new oscillator for very short waves, Philips tech. Rev. 8, 257-266, 1946. The multi-reflex klystron as a transmitting valve in beam transmitters, Philips tech. Rev, this issue, p. 328.

frequency at which, for single modulation, a specific ratio of the second harmonic to the fundamental occurs (we may take this ratio as, say, 10^{-6}, i.e. — 60 dB). With double modulation the frequency swing on the auxiliary carrier which causes the same distortion will be denoted μ'. We shall further introduce the following notation: P is the output power of the transmitter when an amplifying transmitting tube is used (fig. 1), and P' is the output power radiated by an oscillating transmitting tube (fig. 5). f_1 represents the central frequency of the auxiliary carrier and Δ the frequency swing with which the auxiliary carrier is modulated upon the transmitted carrier.

It can now be demonstrated [13]) that the ratio of the total noise power of a system according to fig. 1 to that of a system according to fig. 5 can be expressed by:

$$\eta = \frac{1}{\sqrt{2}} \frac{\mu'}{\mu} \frac{\Delta}{F} \sqrt{\frac{P'}{P}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

If the value of η turns out to be greater than 1, a system with double modulation is clearly preferable, but even if the value of η were unity or slightly less, double modulation would still be preferable as a rule, in view of some of the advantages enumerated above, which find no expression in formula (4).

If, for instance, 60 telephone channels are to be transmitted then the following values are valid for the quantities in (4):

$$\Delta = 5 \quad \text{Mc/s},$$
$$F = 1.5 \text{ Mc/s}.$$

If, further, $P'/P = 10$, which is about right for the tubes available at present, it follows from (4) that

$$\eta = 7.5 \frac{\mu'}{\mu} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

On the assumption that the intermediate frequency of a system with single modulation is 100 Mc/s, we may put the value of μ at about 0.35 Mc/s. From (5) it follows now that if η is to be greater than 1, μ' must be greater than 46.5 kc/s. This can easily be realized [14]) with a frequency demodulator tuned to the frequency of the auxiliary carrier (e.g. 1.5 Mc/s).

[13]) See: C. Ducot, Procédé technique pour l'amélioration des performances des faisceaux hertziens en téléphonie, Onde Electrique **35**, 41-54, Jan. 1955 (No. 334).

[14]) It would be possible to shift the frequency of the auxiliary carrier to, say 13 Mc/s by frequency transformation in the receiver, as was effected by Thompson, thus easily satisfying the above requirement for μ', but in most cases this provision is not necessary.

Experimental installation with double modulation

In order to test the advantages of double frequency modulation for beam transmitters employing the multi-reflex klystron transmitting tube, the Laboratoires d'Electronique at de Physique appliquées, Paris, have developed an experimental installation, consisting of a transmitter and a receiver.

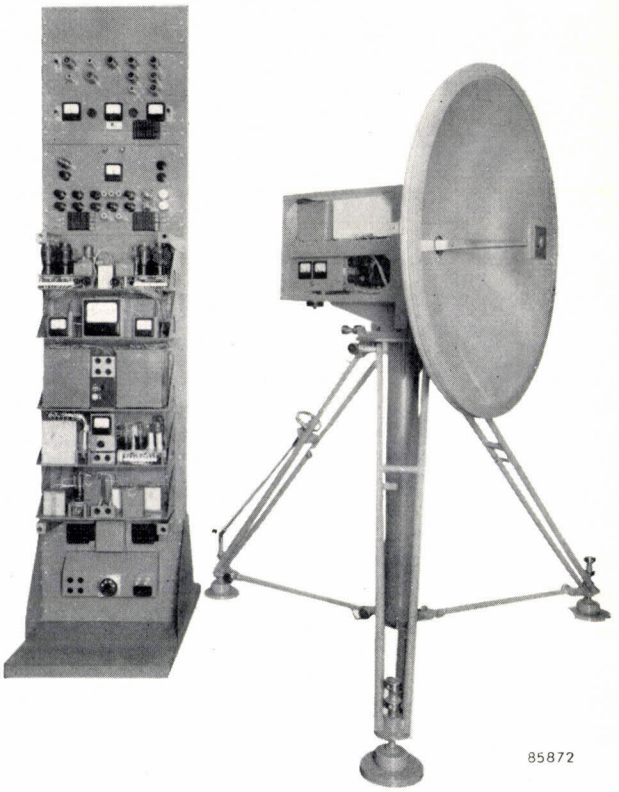


Fig. 7. The experimental transmitter for double frequency modulation. To the left is the rack, to the right, mounted on a tripod stand, the reflector assembly, containing the output stages of the transmitter.

Each is made up of two parts, one built into a paraboloid reflector assembly, and the other built into a rack and connected via cables to the former.

Fig. 7 shows the complete transmitter, which is represented by the block diagram of *fig. 8*. The frequency of the auxiliary carrier is 2 Mc/s. Since linear modulation with the required frequency swing will present difficulties at such a low frequency the auxiliary carrier is obtained by applying the output voltages of two oscillators O_1 and O_2, having frequencies of 9 and 11 Mc/s, to a mixing tube M. The 9 Mc/s oscillator is modulated with the aid of a reactance tube FM. The auxiliary carrier is fed, via filter F and amplifier A_{SP1}, to the part of the transmitter built into the reflector assembly. Here the auxiliary carrier is once more amplified, by amplifier A_{SP2}, and subsequently fed to the oscillating transmitting tube O_3. This is a multi-reflex klystron,

operating at a frequency of about 3500 Mc/swith an output of 10 W. Two stabilizers St_1 and St_2 respectively ensure a sufficient stability of the auxiliary carrier frequency and of the master carrier frequency. The stabilizer is dealt with in the third article of this series [15]).

which is applied via the cable to the other section of the receiver. This contains firstly an I.F. amplifier with a low-noise input stage. The bandwidth of this amplifier (the frequency band in which the amplification does not drop more than 3 dB below that at the central frequency) amounts to 14 Mc/s. In this

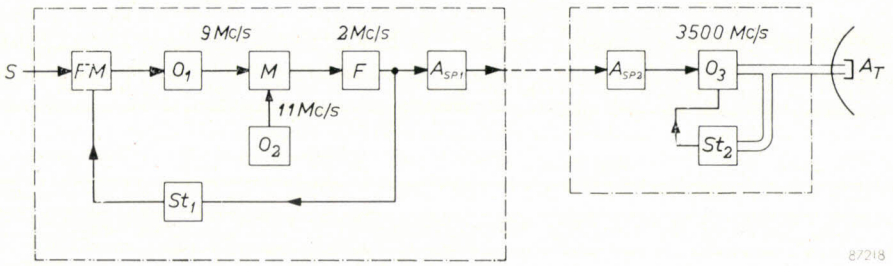


Fig. 8. Block diagram of the transmitter of fig. 7. The left-hand part, surrounded by the dotted line, is built into the rack; the right-hand part, is built into the reflector assembly.

The transmitting antenna A_T consists of a reflector in the shape of a paraboloid of revolution containing, at the focus, a radiator of a type designed by Cutler.

Fig. 9 shows the complete receiver, which is represented by the block diagram of *fig. 10*. The antenna A_R is identical to that of the transmitter. The part of the circuit built into the reflector assembly contains an oscillator O consisting of a triode, the frequency of which is determined by a cavity resonator. In the mixing stage M the antenna signal together with the oscillator signal form an I.F. voltage with a frequency of 58 Mc/s,

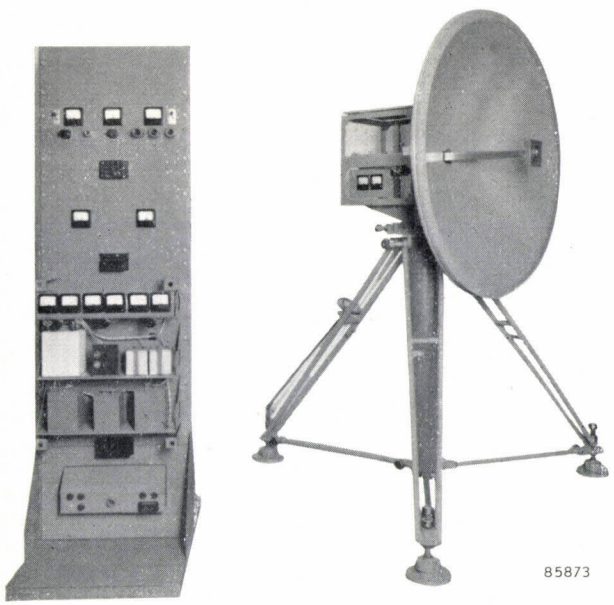


Fig. 9. The experimental receiver for double frequency modulation. To the left is the rack, to the right, mounted on a tripod stand, the reflector assembly containing the first stage of the receiver.

[15]) J. Cayzac, Automatic frequency stabilization for a beam transmitter working on centimetric waves, Philips tech. Rev. this issue, p. 334.

amplifier, staggered circuits are employed. The I.F. signal is applied, via limiter L_M, to the first frequency demodulator D_1 which reproduces the auxiliary carrier with a frequency of 2 Mc/s. This is applied, via amplifier A_{SP}, limiter L_{SP}, and filter F, to the second frequency demodulator D_2, which produces the transmitted signal S.

Measurements on the installation

With the installation in question an experimental telephony link bridging a distance of 12.5 km was established. In view of this fairly short distance, small paraboloid reflectors were used, with a diameter of only 1 m. In order to approximate more closely to the conditions of a link of longer distance, the transmitter output was reduced to 4 W, and an attenuation of 5 dB was incorporated before the I.F. amplifier of the receiver, so that the noise factor of the receiver increased by 4 dB. Because of an obstacle between transmitter and receiver, the signal voltage ultimately received underwent an additional attenuation by 10 dB. The frequency swing with which the carrier was modulated by the auxiliary carrier was 5 Mc/s.

It can be shown that in these experiments, the fluctuation noise was equal to that which would occur with a 50 km link when the received signal is attenuated 24 dB by fading.

If the distance between transmitter and receiver were 50 km instead of 12.5 km, the voltage obtained at the receiver antenna would be four times lower, and therefore reduced 12 dB with respect to the actual conditions used. It should be kept in mind, however, that for such a long-distance link reflectors of larger diameter are used, e.g. 3 m, giving a received signal greater by 18 dB. A further amplification (4 dB) will be obtained from the fact that the transmitter power will be adjusted at 10 W and also from the fact that as a rule there is

no obstacle between transmitter and receiver. Finally, no attenuation will be incorporated before the I.F. amplifier, so that the noise factor of the receiver decreases 4 dB. The ratio of signal to fluctuation noise will thus altogether be $(18 + 4 + 10 + 4) - 12 = 24$ dB higher than the one occurring in the experiment. A fading of 24 dB on the 50 km link therefore makes the fluctuation noise equal to that in the experimental link.

this way a curve as plotted in *fig. 11* was obtained. At small values of W_s, W_b consists mainly of the fluctuation noise, which is independent of W_s. At greater values of W_s, however, the intermodulation noise begins to play a part, and W_b increases accordingly.

We have already mentioned in the introduction

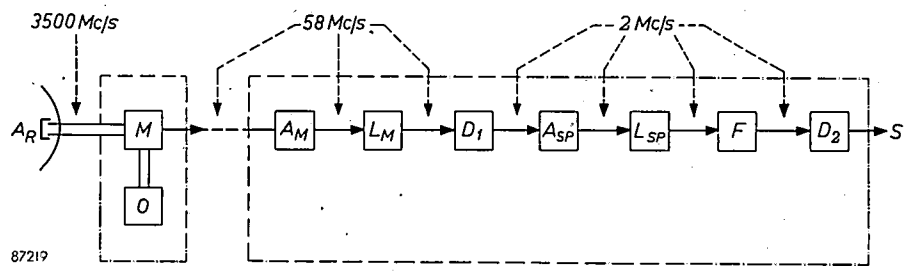


Fig. 10. Block diagram of the receiver of fig. 9. The left-hand section (surrounded by dotted line) is built into the reflector assembly; the right-hand part is incorporated in the rack.

The installation was connected to a carrier-telephony system for 48 channels, of the type described in an earlier issue of this Review [16]). In this system the 48 channels are divided into 4 groups, situated in the frequency bands 12-60, 60-108, 108-156 and 156-204 kc/s. For measuring purposes, incoherent signals supplied by a noise generator were substituted for the speech signal voltages. As a noise generator, use was made of an amplifier of the type applied in the above-mentioned carrier-telephony system for the amplification of a group of 12 channels. The noise voltage supplied by this amplifier was fed, after frequency transformation, to the four groups of channels, 8 channels being kept free. Of each of these "free channels" the noise power W_b was now measured as a function of the power W_s applied to the other channels. In

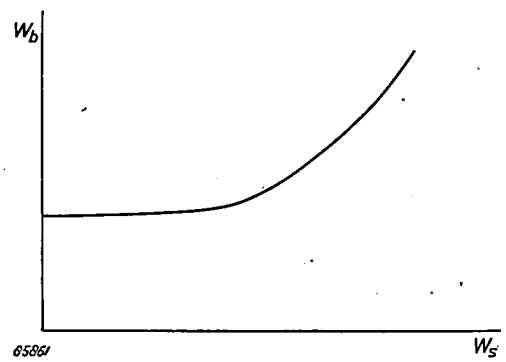


Fig. 11. Noise power W_b in one of the channels, plotted as a function of the power W_s supplied to the other channels. The curve applies to one specific ratio between the level of the signal to be transmitted and the frequency swing.

[16]) G. H. Bast, D. Goedhart and J. F. Schouten, A 48-channel telephone system, Philips tech. Rev. **9**, 161-170, 1947 and **10**, 353-362, 1948.

that, without essentially altering the system, the frequency swing on the auxiliary carrier can be adjusted to various values by incorporating attenuators before the first modulator in the transmitter and after the second demodulator in the receiver, the attenuation of which is varied in the opposite sense. If the modulating voltage is reduced and (consequently also the frequency swing) the fluctuation noise will increase, but the intermodulation noise will become smaller.

When the frequency swing is widened, the reverse will obviously take place. In *fig. 12*, W_b is plotted

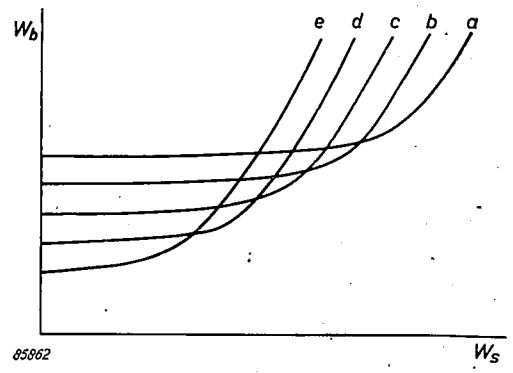


Fig. 12. Noise power W_b in one of the channels, plotted as a function of the power W_s supplied to the other channels, for various values of the frequency swing. For a given signal level, the frequency swing increases from *a* to *e*.

as a function of W_s, for various values of the frequency swing, increasing from *a* to *e*. From this we see that each value of W_s has an appropriate frequency swing producing a minimum value of W_b. In order to adjust the frequency swing to its optimum value, we must know what value of W_s corresponds to the normal power in a 48-channel system. This

power is of course subject to considerable fluctuations. Statistical studies have revealed that with a 48-channel system, the power at points at relative zero level exceeds 1.2 dBm_0 for 1% of the time at the most.

This figure has been derived from an article by Holbrook and Dixon [17]). More recent research, however, has shown that the power values found by them as the normal load of a carrier-wave system can be reduced by 5 dB. This latter consideration has already been allowed for in the above.

We must therefore adjust the frequency swing, by means of the 2 attenuators, so that the minimum value of W_b occurs at this value of W_s. This position was determined experimentally and retained during the subsequent measurements. Curve *a* in *fig. 13*

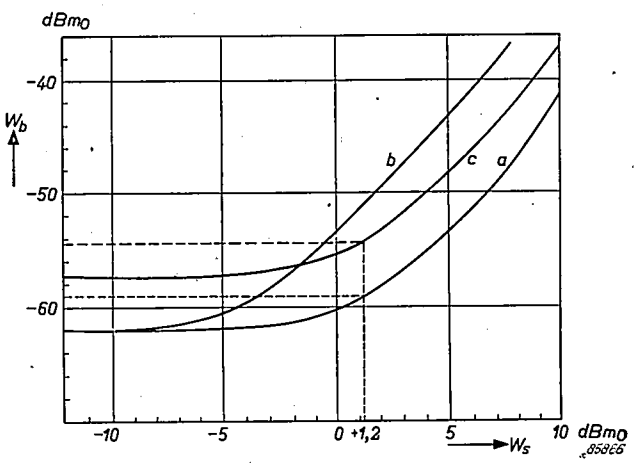


Fig. 13. *a*) Experimental curve, showing the noise power W_b in one of the 8 free channels as a function of the power W_s supplied to the remaining 40 channels.
b) Curve computed from *a* which would apply if, using the same type of transmitters and receivers as in *a*, transmission were over a standard circuit of length 2500 km, the frequency swing being adjusted to the optimum value for each individual link. *c*) Curve which applies if the frequency swing in each of the links were adjusted to the value most favourable to the whole. The power values are plotted in dBm_0, i.e. in dB relative to a power of 1 mW at points at relative zero level.

shows W_b plotted as a function of W_s for one of the channels, viz. that of the highest frequency band (200—204 kc/s). The scale is in dB relative to 1 mW at points at relative zero level, i.e. in dBm_0. The particular frequency swing, earlier designated by *d*, which is representative of a power of 1 mW at points at relative zero level, will consequently occur at $W_s = 0$ dBm_0. From measurements in various channels an average value of 40 kc/s was found for *d*.

Curve *a* in fig. 13 shows that $W_b = -59$ dBm_0 when $W_s = 1.2$ dBm_0 in the channel in question. When W_s is very small, so that we are concerned only with the fluctuation noise, $W_b = -62$ dBm_0. Both as regards fluctuation noise and intermodulation noise, this channel, having the highest frequency band, shows the least favourable properties. (The curves measured in the other channels, however, do not differ a great deal from fig. 13.) Our further discussion can, therefore, be safely restricted to this channel.

Conclusions from the measurements

From the results of the measurements described above, certain conclusions may be drawn concerning the practical value of transmitters and receivers of the type examined in the specified standard circuit, in particular as regards the question whether the C.C.I.F. noise specifications can be complied with. The length of 2500 km of the standard circuit will obviously have to be bridged by means of a number of relay stations. Let us assume that fifty links, each of 50 km length, will be used.

A reduction of the distance between the relay stations and a correspondingly increased number of links, has a favourable effect on the fluctuation noise, since in each link the level of this noise is proportional to the square of the distance, whereas the number of links is inversely proportional to this distance. An increase in the number of relay stations, on the other hand, incurs great expense. For this reason a length of 50 km per link was considered to be a good compromise.

Let us further assume that in two of these links fading will reduce the signal strength to a hundredth (— 20 dB) of its normal value [18]). This means that the fluctuation noise for the total circuit will be $2 \times 100 + 48 \times 1 = 248$ times (24 dB) greater than that of a single link in which no fading occurs. Earlier in this article it was calculated that, in the experiments, the fluctuation noise relative to the signal is 24 dB greater than would occur in a normal 50 km link, so that, as regards fluctuation noise, the experiments are representative of a standard circuit under the conditions mentioned above.

In the standard circuit as laid down by the C.C.I.F., the auxiliary carrier is to be demodulated and then generated and modulated again at nine places. If we presume that the intermodulation noise is formed exclusively in the first modulators of the transmitters and in the second demodulators of the receivers (disregarding the phase distortion in networks and transmission lines), this means that the intermodulation noise will be nine times (9.5 dB) greater than in the experimental link.

[17]) B. D. Holbrook and J. T. Dixon, Load rating theory for multichannel amplifiers, Bell System tech. J. **18**, 624-644, 1939.

[18]) Such a strong fading occurring on two links simultaneously is most unlikely. Measurements on an American-built 107-link beam transmitter chain have shown that a fading of 20 dB occurred on the least favourable of the links during 1% of the time only.

With these data and with the aid of curve *a* of fig. 13, a curve can be plotted, which applies to the whole circuit. This curve, indicated by *b*, would apply if the measurements in question were carried out on the whole standard circuit, in which the frequency swing *d* for each link is adjusted to the optimum value found for a single link. Compared with the experimental link, however, the fluctuation noise and the intermodulation noise have not increased in the same proportion, so that in fact another value of *d* will give better results. In equation (1) *A* will have retained the same value, but *B* will be nine times greater. The optimum frequency swing d_{opt} will consequently, according to (2), have to be decreased by a factor of

$$\sqrt[4]{\frac{9}{1}} = 1.73 \ (4.8 \text{ dB}).$$

The result is that curve *b* of fig. 13 is shifted upwards and to the right over distances of 4.8 dB, thus forming curve *c*. It can be seen that at $W_s = 1.2$ dBm_0, the value of W_b becomes -54.5 dBm_0. When the measurement is carried out psophometrically, this value can again be reduced 3 dB, so that we arrive at -57.5 dBm_0. The permissible value, as mentioned in the introduction, is 7500 pW, i.e. -51.2 dBm_0. It may, therefore, be concluded that even under conditions of severe fading, such as have been presumed for this case, the noise will remain a good 6 dB below the permissible level.

From the measurements of the experimental link it may also be deduced that this level is not exceeded even if the number of channels transmitted is raised to 60; the total noise then increases only by 3.7 dB.

Experiences with the experimental link have furthermore shown that various quantities could be allowed to deviate considerably from their optimum values before the signal-to-noise ratio deteriorated appreciably [19]). The local oscillator of the receiver, for example, can be detuned by as much as 5 Mc/s before the noise rises by 3 dB. (The frequency stability of the oscillator is such that deviations greater then 1 Mc/s are most unlikely.) The central frequency of the auxiliary carrier can be altered by 50 kc/s without more than a 3 dB rise in noise (the maximum deviation from this frequency is restricted to 10 kc/s by the stabilizing circuit). As a general conclusion it may be claimed that the system of double modulation, in spite of its apparent complexity, is characterized by a very high degree of stability.

[19]) These experiments were carried out with the installation connected to a 24-channel carrier telephony system.

In conclusion, the author wishes to convey his thanks to Prof. G. A. Boutry, director of the Laboratoires d'Electronique et de Physique appliquées, for permission to publish this article, and to acknowledge the valuable contribution to this work of G. Andrieux, J. Cayzac, R. Roy and their co-workers, who built the equipment for this experimental link.

Summary. This article describes an experimental beam transmitter and receiver operating on the principle of double modulation. Some of the general criteria for a beam-transmission system for multi-channel carrier telephony are first discussed. In the transmitter, either an amplifying or an oscillating transmitting tube can be employed. Some of the drawbacks of each system are mentioned. When an oscillating transmitting tube is used, the signal to be transmitted has to be directly modulated on this tube. It is very difficult to attain a sufficiently linear frequency modulation at the very high oscillation frequency of this tube. Insufficiently linear modulation causes an excessive intermodulation noise. With *double* frequency modulation, the signal to be transmitted is modulated upon an auxiliary carrier of moderate frequency, so that the requirement of linearity can be easily satisfied. The carrier to be transmitted is then modulated with the auxiliary carrier, a process for which the linearity requirement can be considerably relaxed. Demodulation of the auxiliary carrier need take place only at relay stations where this is required by the telephone service. Many of the relay stations, therefore, contribute hardly at all to the intermodulation noise. An experimental transmitter and receiver was built with the aid of which a telephone circuit with 48 channels was established. Measurements were made of the signal-to-noise ratio. From the results it can be shown that with this equipment, a standard circuit with a length of 2500 km can be built up which satisfies the signal-to-noise requirements laid down by the C.C.I.F.

THE MULTI-REFLEX KLYSTRON AS A TRANSMITTING VALVE IN BEAM TRANSMITTERS

by F. COETERIER.

621.373.423

In recent years there has been a marked increase in the number of types of microwave valves whose operation is based on the transit time effects of electrons. Klystrons, reflex klystrons, travelling-wave tubes and magnetrons are already widely used, while other types are still in a more experimental stage. Each valve has its own particular field of application. The multi-reflex klystron, which is the simplest in design of all the members of the family of velocity-modulated valves, has fully proved its value as a transmitting valve in beam transmission links.

The preceding article [1]) discusses a beam transmitter in which a multi-reflex klystron is used as a transmitting valve. The following characteristics of the valve are important in this application:

a) Relatively high power output at short wavelengths — 10 W or more at 8.5 cm.

b) Possibility of linear frequency modulation over wide bandwidths, without appreciably influencing the power output. A favourable characteristic is that the modulation may be effected by varying the voltage of one or more electrodes which consume hardly any power.

A paper on the multi-reflex klystron has already appeared in this Review [2]); it will be referred to in the following as I. The present paper will report on certain new aspects and deal with the electronic tuning of the valve, a subject which was not considered in I. In conclusion, a description will be given of the construction of the valve.

Principle of the multi-reflex klystron

We shall first recapitulate the operation of the multi-reflex klystron, which was dealt with in detail in I.

The construction is represented diagrammatically in *fig. 1a*. K is the cathode. A and A' are grids forming the anode system. M and M' are grids which are connected to a resonant system and between which a high-frequency alternating voltage is set up.

We shall call the space between M and M' the resonator gap. R and R' are repeller electrodes, which will be dealt with in more detail later. The potential distribution in the absence of a high frequency alternating voltage, is shown in fig. 1*b*; to a first approximation it can be represented by a parabola.

The electrons emitted by K are accelerated and modulated in velocity between M and M' by the high-frequency voltage. Since R is at cathode potential, the retarded electrons turn back before they reach R, while the accelerated electrons penetrate beyond it. The parabolic field between K and R has the property of making the period of oscillation of the electrons constant and hence independent of their velocity. If the field were to remain parabolic between R and R', all electrons would return after the same time interval; in other words, there would be no density modulation of the beam. By arranging for the field between R and R' to deviate from the parabolic form, the transit time of the accelerated electrons is made to differ from that of the retarded electrons, so that, in the returning beam, density modulation is obtained which, by induction between M and M', maintains the high frequency oscillation. The beam then proceeds to the cathode into which all those electrons vanish which have drawn energy

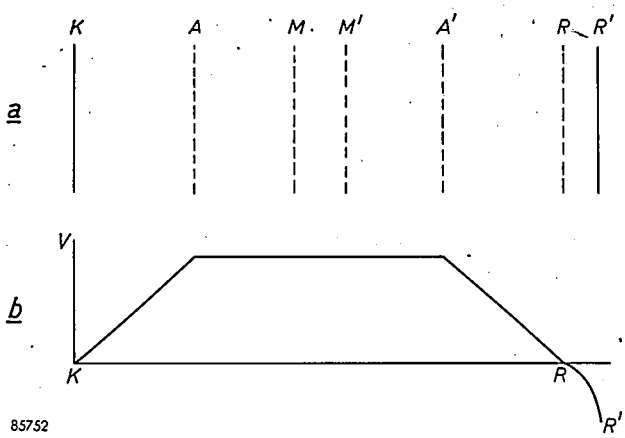


Fig. 1. *a*) Diagram showing the principle of the multi-reflex klystron. K cathode; A and A' grids forming the anode; M and M' grids of the modulator system; R and R' repeller electrodes.
b) Static potential variation between the electrodes.

[1]) C. Ducot, Beam transmitters with double frequency modulation, Philips tech. Rev., this issue, p. 317.

[2]) F. Coeterier, The multi-reflection tube, a new oscillator for very short waves, Philips tech. Rev. 8, 257-266, 1946, further referred to here as I.

from the alternating field during their outward and return excursion, while the remainder continue to swing to and fro for a certain time between K and R. As the periodic time of the electrons is independent of their velocity, it is possible to adjust the voltage such that the electrons always arrive at the resonator gap (MM') at the right moment, where they repeatedly give up energy. As a result of this periodic surrender of energy, the multi-reflex klystron has a high efficiency as compared with an ordinary reflex klystron (see I). In the latter tube the modulated electron beam returns only once to the resonator gap.

After this introduction we shall now proceed to consider in more detail the operation of the repeller electrodes.

The field between the repeller electrodes

The manner in which the most efficient field might be chosen between the repeller electrodes R and R' was discussed in I. In the meantime, new theoretical considerations have led to a modification which has an important bearing on the construction of the tubes. We shall discuss this theory with reference to *fig. 2*.

It should be noted in the first place that a high-frequency field is present between the grids A and A'. The transit time of an electron between A and A' is not small with respect to one cycle of this field, so that it is difficult to describe the behaviour of the electrons. It appears, however, that the operation of this system is analogous to a similar system in which a high-frequency field is present only in the resonator gap, the remaining space being field-free. We therefore assume that a high-frequency field is present only between M and M', and that this space is so narrow that we can regard the moment of an electron's arrival and departure as a single instant of time t_0. The field outside the resonator gap is independent of the time. The transit time in this field can therefore depend only upon the velocity of the electron.

In fig. 2*a* the high-frequency alternating voltage $v = V_0 \cos \omega t_0$ is shown as a function of ωt_0. According to our assumptions, the change in the energy of an electron which passes the gap at the time t_0 is equal to ev, where e is the electronic charge[3]). It must be remembered, however, that the electrons come from both sides; we assume that an electron coming from the cathode undergoes a change in energy according to fig. 2*a*, while the change in energy of an electron coming simultaneously from the other side will be the inverse of this. If the electrons come from the cathode, we see that

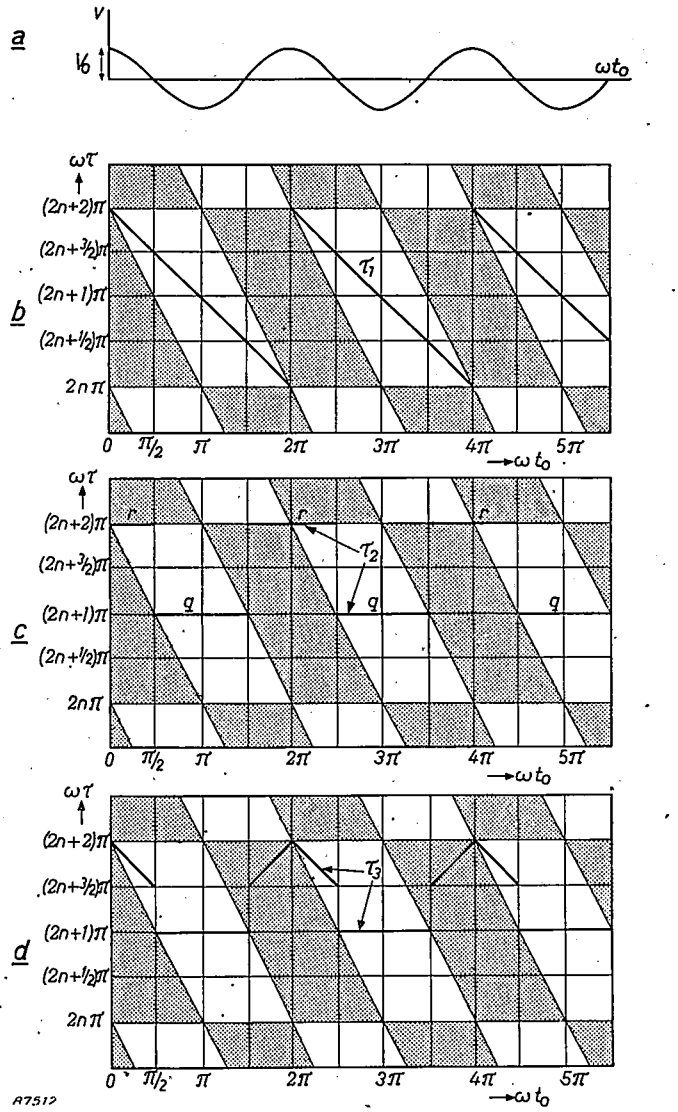


Fig. 2. *a*) The high-frequency voltage $v = V_0 \cos \omega t_0$. *b*), *c*) and *d*) Electron transit-time diagrams; t_0 is the moment at which the electrons first pass through the resonator gap; τ is the transit time, i.e. the time between two outward excursions by the electrons. The various transit-time functions τ_1 (*b*), τ_2 (*c*) and τ_3 (*d*) are described in the text.

they are retarded if $\pi/2 < \omega t_0 < 3\pi/2$, $5\pi/2 < \omega t_0 < 7\pi/2$, and so on. During the intervals these electrons are accelerated. Now we wish to know how great the change is in the energy of an electron after it has travelled once to and fro. If τ is the transit time, i.e. the time between the two successive flights through the resonator gap, the change in energy is evidently

$$\Delta E = eV_0 \cos \omega t_0 - eV_0 \cos \omega (t_0 + \tau)$$

[3]) As a consequence of certain effects such as transit time, space-charge and finite distance of the electron to the boundaries of the grid structure, the change in energy is actually somewhat less than ev, viz. βev, where $\beta < 1$. For the sake of simplicity this β-factor is omitted from the argument.

(the second term has a minus sign, because the electron returns in the same field). This expression may be re-written:

$$\Delta E = 2eV_0 \sin\frac{\omega\tau}{2} \sin\omega\left(t_0+\frac{\tau}{2}\right).$$

If $\Delta E > 0$, the electron has drawn energy from the system. With the aid of the above expression, the areas where $\Delta E > 0$, i.e. the "unfavourable" areas, are shown shaded in the (t_0, τ) plane in fig. 2*b*, 2*c* and 2*d*. The question now arises, how must we choose the transit time τ as a function of t_0 in order to cause the electrons to surrender as much of their energy as possible? For this optimum transit time τ_1 the following expression is valid:

$$\omega\tau_1 + \omega t_0 = 2n\pi \quad (n \text{ an integer})$$

since all these electrons return at the instant $2n\pi$, and in fig. 2*a* we see that the change in energy for returning electrons is then at a negative maximum. The function τ_1 (t_0) is represented in fig. 2*b* by a sawtooth line, and the whole of this line passes through the unshaded region: hence all electrons give up the maximum amount of energy.

The static field needed to produce a transit time function of this nature cannot, however, be realized in practice, since all electrons with the same velocity must have the same transit time and, as fig. 2*a* shows, therefore for all integral values of n we must have for any arbitrary quantity a:

$$\tau\,(n\pi + a) = \tau\,(n\pi - a).$$

The transit-time curve must therefore in any case be symmetrical about the points $t_0 = n\pi$, which is not the case with the sawtooth line. However, this is not the only condition which the transit-time function should satisfy. The transit time of the electrons which are delayed during the first outward movement and which we wish to cause to travel to and fro as frequently as possible, must, moreover, be independent of the velocity, as otherwise their phase would be shifted during repeated reflection. The favourable value for this transit time is thus an uneven number of times π, since every time the electrons return to the resonator they are subjected to an opposing field. Hence.

$$\omega\tau = (2k+1)\pi \text{ for } \frac{(4n-3)\pi}{2} < \omega t_0 < \frac{(4n-1)\pi}{2}.$$

In fig. 2*c* this is represented by the lines q. We still have a certain freedom in our choice of τ in the interlying t_0-regions, (i.e. the times during which the electrons are accelerated when first passing through the resonator gap). The electrons may, owing to their surplus of energy, pass beyond the repeller R (fig. 1), which is at cathode potential, and thus arrive in the space between R and R'. The transit time can be adjusted by the choice of the field in this space. This has no influence on the retarded electrons, because, of course, these turn back before they reach R.

In article I, an optimum transit time was given for the accelerated electrons, which was a constant and was exactly half a period longer than for the retarded electrons (lines r in fig. 2*c*), so that the total transit-time curve τ_2 was a square-wave function. As the figure shows, these electrons after travelling once to and fro, have an energy surplus of exactly zero and were therefore retarded during their return. In so far as they do not arrive at the cathode, they thereafter remain in the same field as the retarded electrons, where their transit time is such that each time they pass through they surrender energy to the resonator circuit.

Efforts have been made in practice to approximate to the transit-time curve τ_2 as closely as possible by means of the static field. The calculation shows that in view of the discontinuities of slope in this curve, there must be a sudden irregularity in the field strength, and this, of course, can be approached only very roughly. It was found, however, that endeavours to improve the slope of this field had the effect of slightly lowering the efficiency of the tubes.

This may be explained as follows. Consider a transit-time curve τ_3 which differs from τ_2 only in the region of the accelerated electrons and also has the necessary symmetry. This curve is drawn in fig. 2*d*. In this case the accelerated electrons fall into 2 groups, depending on t_0. For $\omega t_0 < 2n\pi$ they have a surplus of energy after travelling once to and fro, so that they return to the cathode with finite velocity and thus no longer return to the resonator. For $\omega t_0 > 2n\pi$ they have surrendered energy to the system so that they reverse their direction before reaching the cathode. Owing to the fact that, in this region, curve τ_3 coincides with τ_1, these electrons henceforth always arrive at the favourable moment, that is to say at the moment of maximum inverse voltage in the resonator gap. Upon repeated reflection, a transit time according to curve τ_3 will cause the electrons to give up a greater amount of energy and thus increase the efficiency over that of curve τ_2.

The potential slope corresponding to a given transit-time curve can be found by graphical integration. If we approximate to τ_3 by a somewhat

smoothed curve (which will occur in practice), the potential curve will be as shown in *fig. 3* (curve *c*). This field can be realized fairly well by giving the electrode *R* a specific thickness, thereby realizing the discontinuity of slope in the curve. On the other

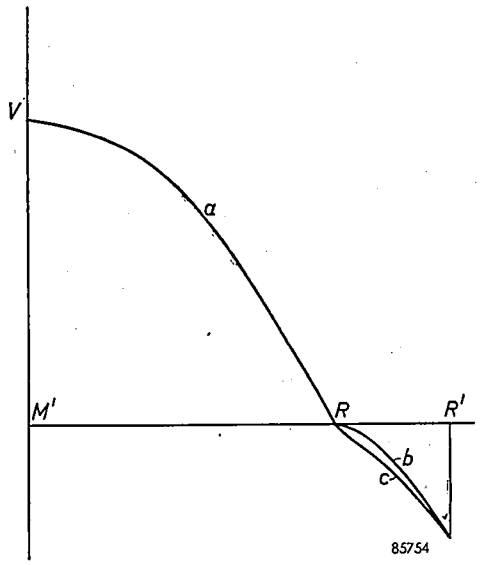


Fig. 3. The static potential between the planes M' and R' of fig. 1. *a*) the ideal potential between M' and R: a parabolic curve. *b*) and *c*) The potential slopes between R and R' corresponding to the transit time curves τ_2 and τ_3 respectively.

hand, the field corresponding to τ_2 (curve *b*) is much more difficult to realize. Thus, a simpler construction results here in better efficiency.

The electronic bandwidth

As stated in the introduction, the frequency can be varied by varying the voltage on one of the electrodes. As the transit time plays an essential part, the valve reaches optimum oscillation and hence delivers maximum power at a specific voltage. The

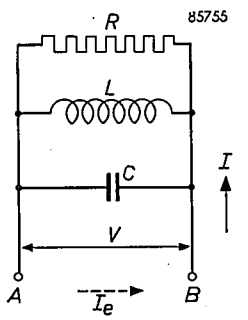


Fig. 4. Equivalent circuit of the high frequency section. I high-frequency current in the resonator. I_e high-frequency component of the total electron stream.

difference between the frequency settings at which the output power falls to half its maximum value is called the electronic bandwidth. The equivalent circuit in *fig. 4* may be used for determining this width. Points A and B represent the openings in the grids M and M' of fig. 1. Between these points the following elements are shunted:

1) The resonator, represented by an LC circuit. The resonant frequency ω_0 is such that

$$\omega^2_0 LC = 1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

2) The load, formed by the transformed impedance of the aerial and by the losses in the system. Together, they can be represented by a damping resistor R between A and B. If V represents the peak value of the alternating voltage between A and B, the total high-frequency output W is

$$W = \tfrac{1}{2}\frac{V^2}{R} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

The useful output is a certain constant fraction of this.

3) The electron beam, which itself also represents a certain impedance R_e between A and B.

If, in the circuit formed by these elements, an oscillation is present with a frequency ω and a constant amplitude, the total admittance between A and B must be zero. We can therefore write:

$$\frac{1}{R} + \frac{1}{R_e} + j\left(\omega C - \frac{1}{\omega L}\right) = 0.$$

Assuming that $\omega = \omega_0 + \Delta\omega$, then for $\Delta\omega/\omega \ll 1$ we may write:

$$\omega C - \frac{1}{\omega L} = 2C\Delta\omega$$

so that:

$$\frac{1}{R_e} = -\frac{1}{R} - j\,2C\,\Delta\omega \quad . \quad . \quad . \quad . \quad (3)$$

If I_e represents the beam current component of angular frequency ω, then $1/R_e = I_e/V$. This quantity is called the electronic admittance. We see that the angular frequency ω_0 is generated ($\Delta\omega = 0$) when I_e and V are in phase, and that the frequency is changed when there is a phase shift between them. This phase shift may be brought about by varying the transit time of the electrons in the static field, which is done in practice by changing the potential on electrode R (fig. 1). The relation between this potential and the frequency is called the modulation characteristic.

In principle, it would be more correct to modulate the electrodes A, A', M and M' (fig. 1) together, since by modulating R, the transit time is changed on only one side of the resonator. In practice, however, it is important for the modul-

ation to consume as little power as possible, so that it is preferably carried out on those electrodes over which hardly any current flows. Owing to the asymmetrical transit time, the bandwidth is slightly reduced, an effect which is tolerated for the reasons given.

The question is now how large $\Delta\omega$ can be without causing a severe depreciation of the high-frequency output. To determine this exactly, it would be necessary to know the relation between I_e and V explicitly at a given transit time. This rather elaborate calculation exceeds the scope of this paper, but we can say something in general terms about $\Delta\omega$. It is evident that when I_e is shifted 90° in phase with respect to V, there can be no more energy supplied. To a rough approximation, therefore, a 45° phase shift means that there will still be half the output left, and this in fact roughly agrees with the theoretical results [4]).

Taking $\Delta\omega_r$ as the frequency variation occurring at a 45° phase shift, then according to equation (3):

$$\frac{1}{R} \approx 2C\ \Delta\omega_r$$

and the total electronic bandwidth $= 2\Delta\omega_r \approx 1/RC$. The value of C is exclusively determined by the construction of the resonator. As regards $1/R$, we find from (3) that

$$\frac{1}{R} = \left(\left|\frac{I_e}{V}\right|\right)_{\omega=\omega_0}.$$

This shows the advantage of the multi-reflex klystron over an ordinary klystron; owing to the repeated reflections, the alternating current component I_e can take on much higher values, so that the bandwidth is many times larger.

It is true that bandwidths of the same order of magnitude can be realized with ordinary reflex klystrons, but this is done by keeping the direct voltage very low, as a result of which V remains much smaller. This in turn greatly reduces both the high-frequency output and the efficiency (to no more than a few watts and a few % respectively). Reflex klystrons with this power output can indeed be used as transmitting valves in a link transmitter, but experience shows that the greater power of the multi-reflex klystron is a considerable advantage.

Certain types of high-frequency valves exist, viz. travelling-wave tubes [5]), with which an even greater electronic bandwidth is possible. However, their efficiency is lower than that of the multi-reflex klystron, and their construction presents greater technological difficulties.

We shall now briefly review the average characteristics of the series-manufactured valve used in the link transmitter discussed in the preceding article [6]).

Wavelength	8.5 cm
Anode direct voltage	3000 V
Anode direct current	20 mA
Useful power output	12 W
Efficiency	20 %
Electronic bandwidth	20 Mc/s.

The high frequency voltage V' is measured as follows. The extra energy of the maximally accelerated electrons amounts to eV'; thus V' is the exact potential difference that must exist between R' and K (fig. 1) in order to suppress all current flowing to R'. The potential difference in this case is:

$$V' = 600 \text{ V}.$$

At $\Delta\omega = 0$ the resultant high-frequency output is equal to $\frac{1}{2}I_eV'$. This value, owing to the losses, is greater than the useful output, and in the present example can be taken at 24 W. It therefore follows that:

$$I_e = 80 \text{ mA}.$$

Since the theory shows that I_e would be approximately equal to the anode direct current I_0 if the electrons were to pass through only once, we must have a fourfold reflection. The maximally retarded electrons lose $4 \times 600 = 2400$ eV, that is to say most of their energy. This fourfold reflection is in good agreement with conclusions derived from the modulation characteristic (see [4])).

Construction of the tube

The construction of the multi-reflex klystron, a photograph of which is shown in *fig. 5*, includes elements of conventional transmitting-valve design and is very simple for a high-frequency valve.

The oscillating system M and the anode system A, with which the desired potential slope is adjusted, are connected. (M is not visible in fig. 5). Both are made of molybdenum sheet and the seams are brazed together with copper or nickel. The electron beam passes simply through the holes in the electrodes. By using grids in these holes it would be

[4]) Since I_e is the result of a number of reflections, the change of the transit angle $\omega\tau$ for one movement to and fro is much less than 45°. The amount by which the voltage must be changed in order to bring the valve back to optimum generation accurately indicates the number of reflections occurring.

[5]) J. R. Pierce, Traveling-wave tubes, D. van Nostrand Co. (1950).

[6]) The same valve is also being used in the channel link transmitter for relaying Eurovision television programmes. See: A. Laurens, Onde él. **34**, 999-1005, Dec. 1954.

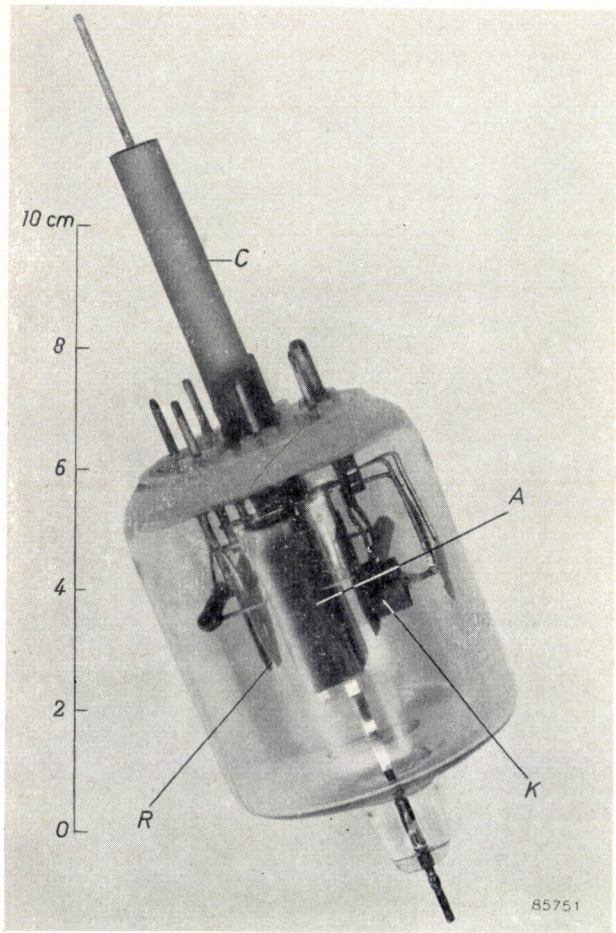


Fig. 5. The multi-reflex klystron used in the beam transmitter. *A* anode system, *C* coaxial output system. *K* cathode. *R* repeller electrodes.

possible to obtain a β factor closer to 1 (see [3])) but the grids would get too hot owing to bombardment by electrons. The correct frequency is pre-set by means of a correction screw in the side-wall of *M*. The connection for the H.T. electrode is arranged at the top of the valve to avoid flash-over and electrolysis of the glass [7]). The glass envelope is made fairly wide to minimize the influence of ambient temperature.

[7]) These and other improvements in design were introduced by E. G. Dorgelo & H. G. Gerlach of the Electronic Tubes Development Laboratory, Eindhoven.

The other components are mounted on a sintered glass base [8]). The external lead of the coaxial output system is formed of molybdenum sheet rounded into a cylinder; a small gap is left between the edges which facilitates the sealing-in. Furthermore, the structure is now more resilient: the valve can now readily be inserted in a coaxial plug-socket. The output system is not connected with the anode, so that it need not be at high tension and the outer sleeve can be earthed.

Since the holes through which the electron beam passes have to be fairly small in order to obtain good interaction of beam and electrodes, the cathode must be able to deliver a current density of approximately 2 A/cm^2 at the given output. The cathode is, moreover, exposed to bombardment by accelerated electrons. No oxide cathode would be able to stand up for long to such conditions. The obvious solution for this valve was therefore to use the L-cathode, already described in this Review [9]), which has a high maximum emission and is proof against electron bombardment.

For concentrating the electrons, an axial magnetic field with an induction of about 700 gauss is needed; this is provided by a "Ticonal" permanent magnet. Valve and magnet together weigh approximately 4 lbs.

[8]) E. G. Dorgelo, Philips tech. Rev. **8**, 2-7, 1946.

[9]) H. J. Lemmens, R. Loosjes and M. J. Jansen, "A new thermionic cathode for heavy loads", Philips tech. Rev., **11**, 341-350, 1949/50.

Summary. The multi-reflex klystron, which was first described in this Review in 1946, is eminently suitable for use as a transmitting valve in beam transmitters, owing to its large electronic bandwidth and its relatively high output power at centimetre wavelengths. The first part of this article develops the theory of the valve. It is shown that the optimum transit time of the electrons as a function of the moment of time at which they pass through the resonator enables the requisite potential curve to be easily established between the repeller electrodes. It is then demonstrated that the large bandwidth of the valve is due to its low electronic admittance, which in its turn is due to the repeated reflection of electrons. From tests made on the valve, it appears that the electrons which surrender energy are reflected four times. As compared with other high-frequency valves, the multi-reflex klystron is very simple in design. Most of its components are made from molybdenum sheet mounted on a sintered glass base.

AUTOMATIC FREQUENCY STABILIZATION FOR A BEAM TRANSMITTER WORKING ON CENTIMETRIC WAVES

by J. CAYZAC *).

621.316.726:621.396.61.029.64

This article, the third relating to a centimetric telecommunication transmitter, describes a device with which the oscillation frequency of the transmitting valve (a multi-reflex klystron) can automatically be kept constant to within 1 in 10 000.

The carrier frequency of the radiation emitted by a beam transmitter working on centimetric waves is required to be extremely stable. To avoid interference between different relay stations, an accuracy of 1 in 10 000 may be necessary at a carrier frequency of about 3400 Mc/s (8.5 cm wavelength). This is especially the case in long-distance communication via a succession of relay stations, the carrier frequencies of which have various values within a certain frequency band.

The preceding article in this Review describes a transmitting valve, the multi-reflex klystron [1]), which is particularly suitable for use in beam transmitters owing to its ability to supply an output of at least 10 W at centimetric wavelengths and because of the fact that it can readily be frequency modulated. The frequency at which this valve oscillates is dependent upon various quantities, especially upon the supply voltages (which, incidentally, is why it can be so easily modulated). At a nominal frequency of 3400 Mc/s, the anode voltage of the valve is 3000 V, and for every volt variation on the anode a frequency shift results of 100 to 200 kc/s. Careful stabilization of the supply voltages would therefore seem to be called for. However, since the frequency may vary for other reasons, because of thermal effects, for instance, it is necessary to find a means of automatically compensating any frequency change that might occur, irrespective of the cause. The power used for this purpose must of course be only a small fraction of the useful power.

The carrier frequency of a transmitter can be kept constant within accurate limits by comparing it with the frequency (many times multiplied) of a crystal oscillator, and by using any difference between them for corrective regulation. This article, however, is concerned with a different procedure which, although less accurate, nevertheless satisfies the requirements and offers, moreover, a number of practical advantages.

The frequency correction device to be described consists of two sections. The H.F. section, which must be introduced between the transmitter valve and the antenna, contains no electronic valves or other elements requiring maintenance. Since relay transmitters are often mounted on high towers, this can be an important consideration. A cable leads to the second (low-frequency) section, in which exclusively conventional receiver valves are used.

Principle of the frequency correction

Consider a symmetrically modulated carrier-wave with a frequency f_c. The spectrum of the radiated power may be represented by a band lying symmetrically about f_c and of width equal to twice the frequency swing, $2f_z$ (*fig. 1a*). A small part of this power is fed to two cavity resonators whose resonant frequencies, f_1 and f_2, lie just outside the radiated spectrum, i.e. the frequency difference $f_1 - f_2$ is taken somewhat larger than $2f_z$. The Q factor of both cavity resonators is so adjusted as to cause their resonance curves to overlap somewhat, as shown in fig. 1*b*. At a given carrier frequency f_c, the average power in each resonator can now be measured with the aid of a crystal detector.

Assuming that the crystals, and their coupling with the cavities, are identical and not selective, the same signal will be obtained from both crystals only when the carrier frequency f_c lies midway between the frequencies f_1 and f_2 (fig. 1c) i.e. when it is equal to f_0, or $f_0 = \frac{1}{2}(f_1 + f_2)$. The signals in the two cavity resonators are represented by the hatched areas. If, as in fig. 1*d*, f_c lies closer to f_2 than to f_1, the average signal in cavity *2* will be stronger than in cavity *1*, and vice versa. Alternate measurement of the signal from both cavities results in a current with an A.C. component, the amplitude of which varies according to the difference $|f_c - f_0|$. At $f_c = f_0$, the alternating current is at zero. The sign of $f_c - f_0$ determines the phase. After amplification,

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

[1]) F. Coeterier, The multi-reflex klystron as a transmitting valve in beam transmitters, Philips tech. Rev., this issue, p. 328

the alternating current is fed to a phase-sensitive detector circuit to produce a direct voltage, the magnitude and sign of which are proportional to $f_c - f_0$. If this output voltage, after amplification, is applied to an electrode of the valve which controls

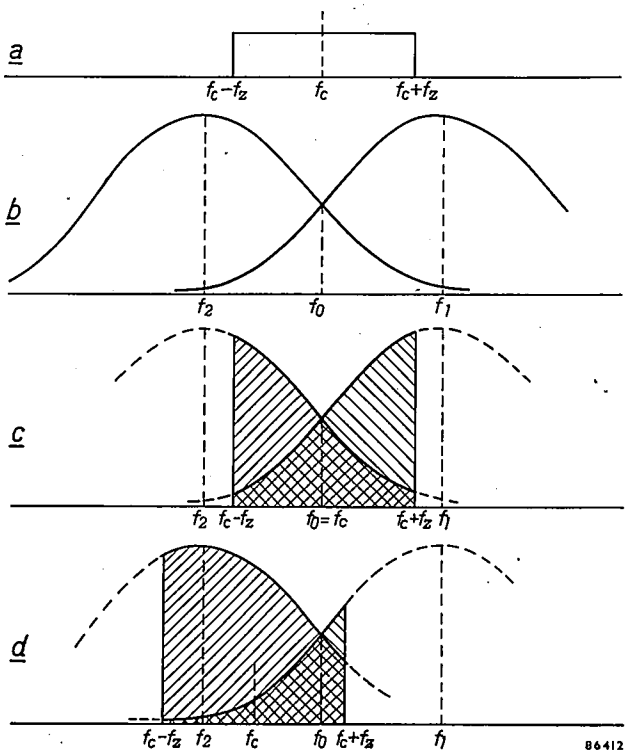


Fig. 1. *a*) The transmitted frequency band, showing the central frequency f_c and frequency swing f_z. The signal strength is plotted vertically.
b) The band-pass curves of cavity resonators *1* and *2*, with resonant frequencies f_1 and f_2.
c) The signals (hatched areas) produced by the frequency band in (*a*) in cavity resonators *1* and *2*, when $f_c = f_0$.
d) The same when $f_c \neq f_0$.

the frequency, it will automatically bring the carrier frequency f_c back to the frequency f_0.

It is essential that the characteristics of both crystals remain constant within very narrow limits over the course of time. Since this condition is difficult to fulfil in practice, we have evolved a method which entirely eliminates differences of this kind. The method involves the use of the same crystal for measuring the fields in the two cavity resonators alternately, in a manner to be described below.

The above arguments are strictly applicable only if the modulation causes the spectral components of the transmitted frequency band to be symmetrically distributed about a fixed frequency. However, the system can still be used if the distribution is asymmetric (as, for instance, in modulation with television signals, when one side band is suppressed). In that case it is possible to design an arrangement with which uses only one signal at a fixed place in the transmitted band. For this purpose the signal can be that which fixes the zero brightness level (black level) of the television picture.

It should be added that the system described is especially suitable for use when the transmitted frequency band is wider than the permissible frequency shift. That is generally the case with telecommunication systems working on decimetric or centimetric waves.

Design of the stabilizer

The system is used on a beam transmitter with a multi-reflex klystron [2]), the output power of which is 10 W at a frequency of 3400 Mc/s. The frequency band produced by modulation amounts to a maximum of 6·5 Mc/s above and below the carrier frequency. By varying the voltage on the repeller electrode, the central frequency can be varied by 50 kc/s per volt in a range from − 100 to + 100 V with respect to the cathode potential.

The high frequency section

The bandwidth required is 13 Mc/s. In our case the resonant frequencies of the two cavity resonators are set at

$$f_1 = f_0 - 7.5 \text{ Mc/s and } f_2 = f_0 + 7.5 \text{ Mc/s}.$$

The whole radiated band thus falls between f_1 and f_2 (as in fig. 1). The relation between the difference of the signals in both resonators and the frequency f_c should be made as linear as possible. This is done by adjusting the Q factor of the cavities to a certain value, which in the present instance was found to be 280.

This value may be derived as follows. As a function of the frequency f, the high-frequency voltage V measured in a cavity resonator is

$$V = \frac{V_0}{1 + 4Q^2\left(\frac{f-f_r}{f_r}\right)^2}.$$

in which f_r is the resonant frequency of the cavity, Q the quality factor and V_0 the voltage for $f = f_r$. The characteristic of the discriminator $y(f)$ is given by the difference of the voltages in the two cavities. Assuming that both V_0 and Q are equal in both cavities, we obtain:

$$y = \frac{V_0}{1 + 4Q^2\left(\frac{f-f_1}{f_1}\right)^2} - \frac{V_0}{1 + 4Q^2\left(\frac{f-f_2}{f_2}\right)^2}.$$

If we put

$$2Q\frac{f-f_0}{f_0} = x,$$

and

$$2Q\frac{f_0-f_1}{f_0} = 2Q\frac{f_2-f_0}{f_0} = a,$$

then, since $f_1 \approx f_2 \approx f_0$, it follows that

$$y = V_0\left[\frac{1}{1+(x+a)^2} - \frac{1}{1+(x-a)^2}\right].$$

[2]) C. Ducot, Philips tech. Rev., this issue, p. 317.

At $x = 0$, $d^2y/dx^2 = 0$, so that the characteristic is already fairly linear in the vicinity of $x = 0$. The linearity can be further improved if

$$\left(\frac{d^3y}{dx^3}\right)_{x=0} = 0,$$

which is the case when $a = \frac{1}{2}\sqrt{1.5}$, so that

$$Q = \frac{f_0\sqrt{1.5}}{2(f_0 - f_1)}.$$

With the values given for f_0 and f_1, we find $Q = 280$.

At the frequencies concerned, a Q factor of this order can easily be obtained. It is, however, essential that both cavity resonators have the same Q and that both are excited in the same way. For this purpose we have adopted the construction shown in *fig. 2*. The two cavities are milled from a single block of copper and are weakly coupled via identical coaxial conductors to the wave guide through which the high-frequency energy to be radiated passes. Both coupling systems are arranged symmetrically in a plane perpendicular to the direction of propagation, so that the coupling is identical in amplitude and phase. The total power consumed by the systems is about 0.5 mW, which is a mere fraction of the output power of about 10 W.

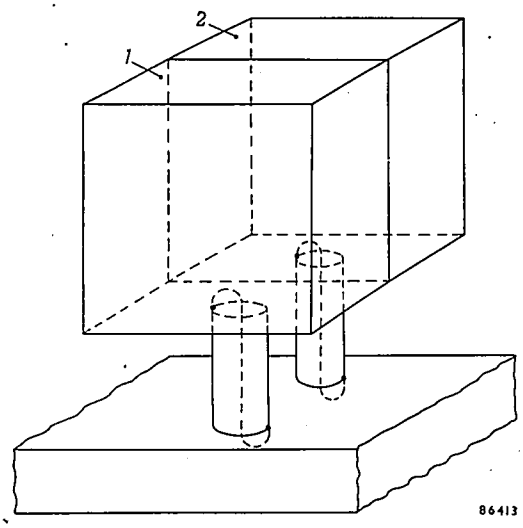


Fig. 2. The two cavity resonators *1* and *2* are symmetrically coupled to the wave guide in a plane perpendicular to the direction of propagation.

Fig 3 illustrates what would happen if each resonator were provided with a crystal detector and if one crystal had a higher sensitivity than the other. The frequency at which the output signals are identical is shifted from f_0 to f_0'. Moreover, even if f_c were now to coincide with f_0', unequal signals would be formed by the modulation in both crystals. As may be seen from fig. 3, the signal in cavity *1* is larger, so that f_c would be further shifted towards the left. The result would therefore be an impermissible frequency shift. For these reasons, the voltage in both cavities is measured alternately with one and the same crystal as stated earlier. This alternation of measurement can be done at a low frequency, because the frequency shifts concerned take place very slowly. The control apparatus can therefore have a time constant of about one second.

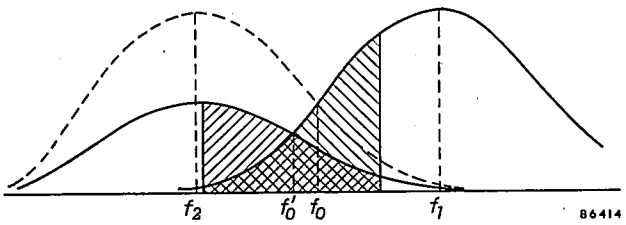


Fig. 3. If the detector in cavity resonator *2* is less sensitive than that in *1* the band-pass characteristic obtained will be that shown by the full-curve instead of the broken curve. Point f_0 is shifted to f_0'.

The coupling of the two resonators with the crystal is shown schematically in *fig. 4*. The crystal *4* is arranged in complete symmetry with respect to the resonators *1* and *2*, and is coupled identically with the resonators through a small hole in the partition wall *3*. The commutation of the crystal is effected by two "absorption modulators" in anti-phase, one in each resonator. In operation, the energy in one resonator is almost completely absorbed, while the other resonator oscillates unattenuated. In the next half cycle, the absorption modulator operates in the other resonator.

These absorption modulators make use of the fact that, at the frequencies concerned, the losses of some ferrites are highly dependent upon a weak, external magnetic field [3]).

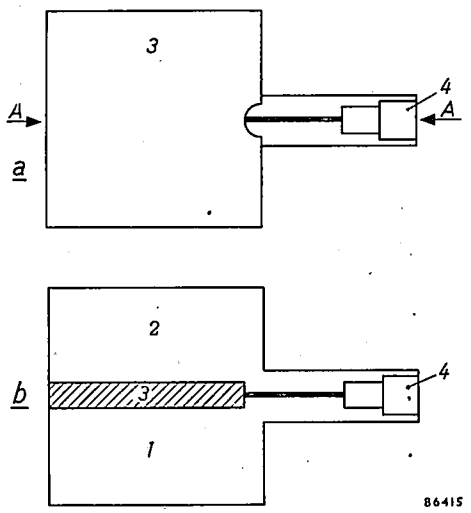


Fig. 4. Positioning of the crystal *4* in the plane of the partition wall *3* between resonators *1* and *2*. *a*) Side view, omitting resonator *1*, *b*) cross-section at *AA*.

[3]) See H. G. Beljers, W. J. van de Lindt and J. J. Went, J. appl. Phys. **22**, 1506, 1951. The phenomenon in question, and its application for modulating the amplitude of microwaves, will shortly be dealt with in this Review.

The system used here is shown in *fig. 5*. Two coaxial lines, K_1 and K_2, which are coupled at one end with the cavity resonators, are terminated at the other end by a loop through which rods of ferroxcube, F_1 and F_2, are passed. Around each rod is a coil, a varying current through which produces a varying magnetic field in the rods. The high-frequency losses in each of the rods vary correspondingly and thus vary the Q factor of the cavity resonators via the coaxial lines. The rods are matched to the resonators by impedance transformers T.

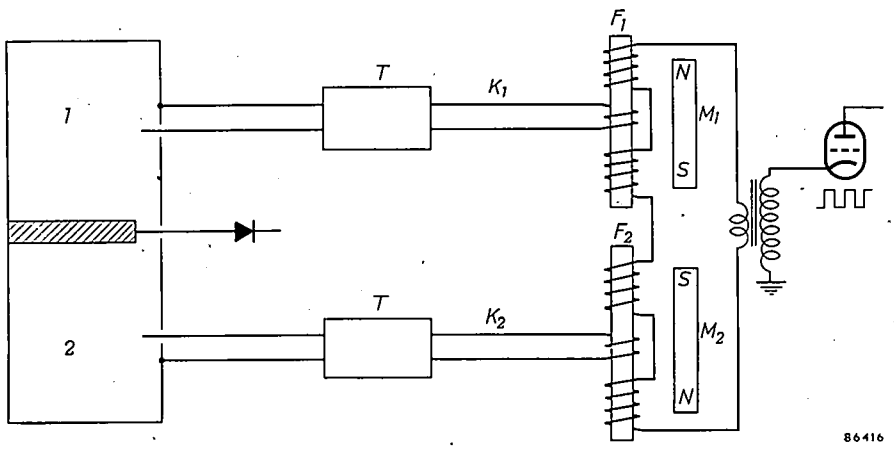


Fig. 5. Diagram of modulation system. The coaxial lines K_1 and K_2 couple the resonators *1* and *2* to the ferroxcube rods F_1 and F_2, which are successively magnetized by an alternating current of rectangular waveform. The permanent magnets M_1 and M_2 pre-magnetize the ferroxcube in opposite directions; T = matching transformers.

Fig. 6 shows how the Q factor of a cavity resonator depends upon the magnetic field in which the ferroxcube is placed. It can be seen that the value of Q is independent of the direction of the field. If, as in our case, we wish to modulate with a rectangular waveform, it will mean that the current fed to the coils will have a DC component. This raises difficulties, since the coils are fed by a matching transformer. It is simpler to premagnetize the ferroxcube rods with the aid of permanent magnets (M_1 and M_2 in fig. 5) to a value corresponding to an "average" Q (points A and B in fig. 6). By oppositely premagnetizing the two rods the same current can be made to flow in both coils. As appears from fig. 6 the Q factors of the two resonators are then modulated in opposite senses.

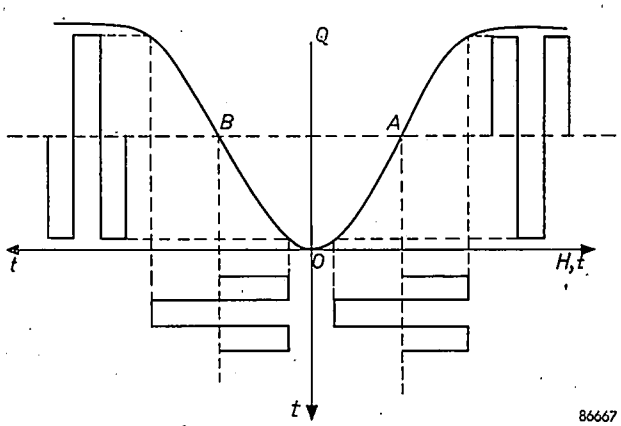


Fig. 6. The quality factor Q of the cavity resonator as a function of the magnetic field H surrounding the ferroxcube. The Q factor of the two cavity resonators is oppositely modulated by oppositely pre-magnetizing the ferroxcube rods, the direction of H being the same.

The chief problem here is that the Q factor of each resonator in the active phase must always have the same value. The problem is simplified in this case by the fact that, with a magnetic field above a certain strength, the losses in the ferroxcube are fairly constant and very low. Hence by choosing the amplitude of the alternating current on the large side, small changes in amplitude have hardly any effect on the value of Q.

It has been found that variations in the temperature of the different circuit elements coupled to the cavity resonators (fig. 5) give rise to slight variations in the resonant frequencies. For this reason, the whole circuit is placed in a thermostat controlled enclosure.

The low-frequency section

As we have seen, a device is needed that can distinguish from which cavity the stronger signal originates, and which enables the sign of the correction voltage to be determined. This device, a phase discriminator, functions as follows. Two signals, S_1 and S_2, which have the same waveform but are in anti-phase, are applied to the screen grids of two pentodes, connected as shown in *fig. 7*. These signals are derived from the square waveform current S_0, flowing in the coils around the ferroxcube rods. The duration of the (positive) pulse is, however, reduced by one half, as shown in fig. 7. The two control grids are negatively biased such that

both valves are normally cut off, irrespective of the potential on the screen grids. The square-wave control signal from the crystal is applied, after amplification, to both control grids, but only that valve conducts which receives positive pulses simultaneously on its control and screen grid. The amplitude of the anode current is proportional to the amplitude of the alternating voltage on the grids. According to whether resonator *1* or *2* gives the stronger signal, anode current will flow in valve B_1 or B_2.

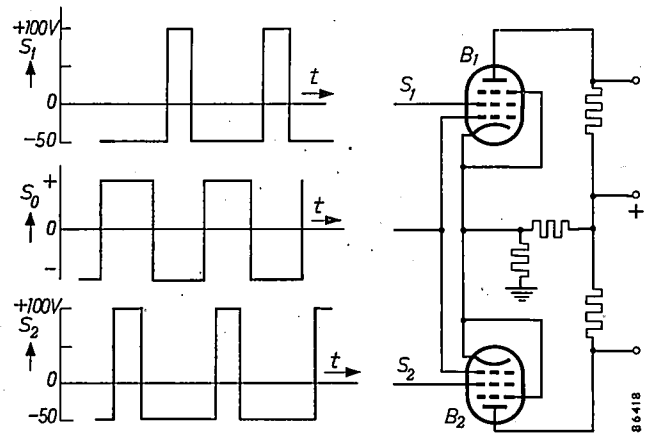


Fig. 7. The phase discriminator. Signals S_1 and S_2, derived from a signal S_0 proportional to the current through the coils, are applied to the screen grids of pentodes B_1 and B_2.

The remaining part of the circuit is shown in *fig. 8*. The output voltages from the pentode are rectified by diodes D_1 and D_2, and the rectified voltages are fed to the grids of two triodes connected as cathode followers. The cathode of one is earthed (C) and that of the other (A) is connected to the repeller electrode of the multi-reflex klystron. The repeller voltage with respect to earth is

$$V_A - V_C = (V_A - V_B) - (V_C - V_B),$$

i.e. equal to the difference of the cathode-grid voltages of both triodes and proportional to the difference between the rectified signals. If the detectors D_1 and D_2 pass no current, this voltage will be zero and the frequency will not be affected, However, as soon as one of the diodes begins to pass current, the repeller bias will change in the direction appropriate to that diode, and the frequency will be automatically corrected until the current returns to zero. The characteristics of the two output valves should of course be fairly similar. It has been found in practice that differences in valve characteristics cause frequency deviations of less than 1 in 10^6 which is considerably less than the deviation of 1 in 10^4 permissible.

The low-frequency section can be designed in another way. The multi-reflex klystron can be tuned not only electrically but also mechanically. By means of a mechanical system connected to the built-in cavity resonator, it is possible to vary the frequency by a total of 300 Mc/s, which is much more than can be achieved by adjusting the repeller potential. For this purpose a servo motor with two windings is used as the phase discriminator. The alternating voltage at the output of the circuit in fig. 7 is fed, after amplification, to one of the two windings. The other winding receives continuously a signal of the same frequency but whose phase differs from that of the first signal by 90°. As we have seen, the phase of the first signal changes by 180° if the sign of the frequency deviation reverses. In this way a rotating field is obtained which causes the rotor of the servo motor to rotate in a direction depending upon the phase of the first signal, and rotation continues until the balance is restored. With this system the frequency can be kept constant with approximately the same accuracy. The great advantage of mechanical compensation is that the repeller bias remains constant. After the latter has been adjusted to obtain the most linear possible modulation of the transmitted carrier wave (see [2]), the linearity is maintained when the compensation system comes into operation.

The stabilization ratio

The control range of the repeller, which is determined by the characteristics if the multi-reflex klystron, is limited to ± 100 V. This includes a large safety margin. The frequency change a per volt change of repeller potential determines the maximum frequency deviations which the system can correct without departing from the permissible voltage zone. In the present equipment, $a = 50$ kc/s per volt, which means that it is possible to correct frequency deviations up to 5 Mc/s.

Our object was to prevent the occurrence of frequency shifts greater than about 250 kc/s which implies an accuracy somewhat better than 1 in 10^4. The amplifier is so designed as to make full use of the repeller control range at the frequency variations concerned; the maximum shift of 250 kc/s thus gives rise to a correction voltage of 100 V. Without correction, the frequency in this case would have drifted 5 Mc/s. The ratio of the frequency variations with and without correction is called the stabilization ratio: with the present system it amounts to 20.

For a deviation of 250 kc/s, the crystal receives a voltage of about 1 mV. The voltage gain of the phase

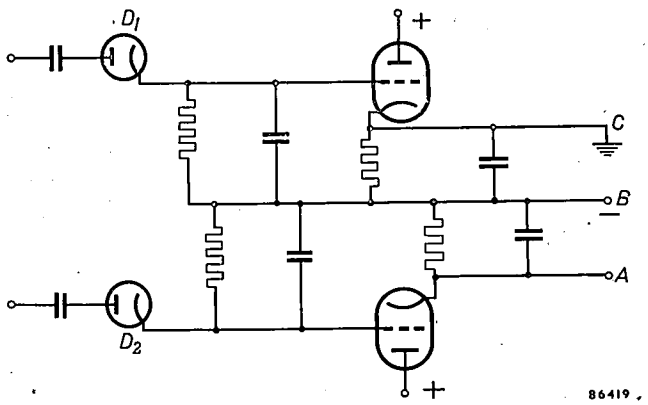


Fig. 8. The output stage of the circuit. The voltage at A is fed to the control electrode of the transmitting valve.

discriminator is 60, so that further amplification by 1700 is needed to obtain the output voltage of 100 V. At the low frequencies in question, however, this presents no difficulties. In practice we usually employ 1.5 times as much amplification in order to allow for any possible decrease in voltage gain.

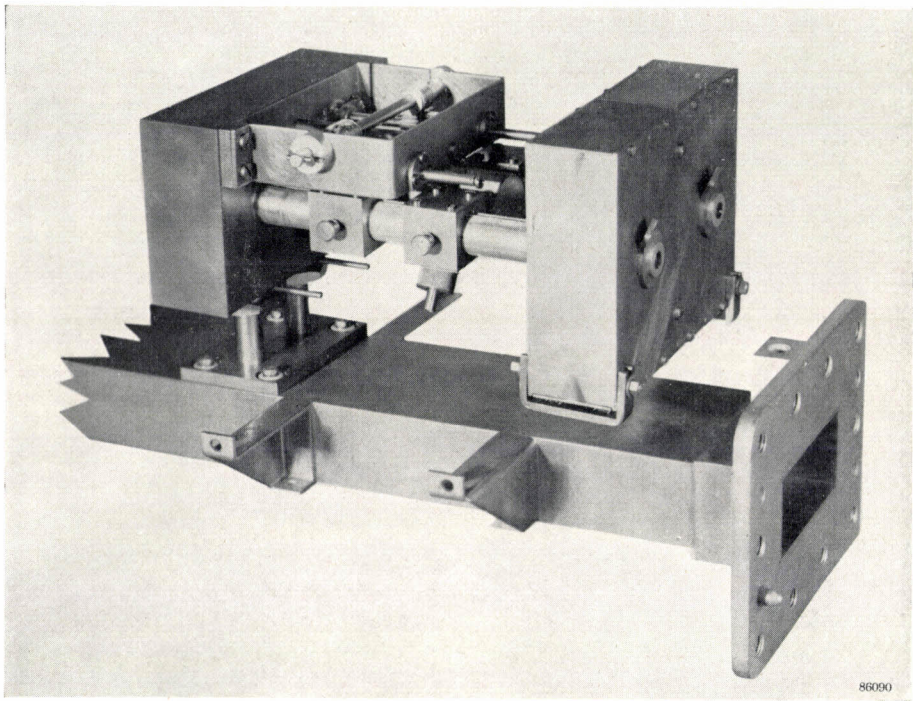


Fig. 9. The H.F. section of the stabilization circuit. The photograph shows the box with the two cavity resonators (left) and the co-axial couplers which lead to the wave guide; two coaxial lines lead to the ferroxcube modulator (right). The mechanical tuning system can partly be seen.

Adjustment to the desired frequency

In view of the fact that it is sometimes necessary to change the frequency of the transmitter quickly, the adjustment required should be as simple as possible. This means that it must be possible to change the resonant frequencies of both cavity resonators by an equal amount simultaneously without seriously influencing the Q factor or the coupling. The method adopted is to fit each cavity resonator with an adjusting screw, mechanically coupled so that they turn together. With this system a frequency band of 150 Mc/s can be covered, which is the total bandwidth within which the multi-reflex klystron can be mechanically tuned. At the limits of this band the output voltage from the crystal — and hence the stabilization ratio — is 2.5 times lower than in the centre of the band.

The difference is normally unimportant, but should it be disturbing it can be remedied by, for example, coupling the resonators a little more closely to the wave guide. This consideration apart, it is possible with the mechanism employed to change the frequency of the transmitter within a band of 150 Mc/s by a single, brief adjustment.

The H.F. section of the frequency stabilizer with the mechanical tuning system is reproduced in *fig. 9*. The thermostat enclosure is not shown in the photograph. When the apparatus is in operation, the mechanical adjustment can be made externally via a thermally insulating spindle which projects through the enclosure.

Summary. An automatic stabilizing device is described with which the frequency of a beam transmitter working with a multi-reflex klystron on a wavelength of 8.5 cm, can be kept constant within an accuracy of 1 in 10^4. Two cavity resonators, whose resonant frequencies differ by about 15 Mc/s, but which are otherwise identical, are coupled to the wave guide. If the frequency of the transmitted carrier lies midway between the two resonant frequencies, a slight change in it will produce a stronger signal in one resonator and a weaker signal in the other. The difference signal, whose amplitude and sign are a measure of the frequency shift, is amplified and fed to a control electrode of the transmitting valve, which automaticaly corrects the frequency. The signals in the two cavity resonators are measured alternately by the same crystal detector, so that the measurement is not influenced by small differences (or possible drifting) between the characteristics of two otherwise suitable detectors. The commutation is effected by "absorption modulation" with the aid of small rods of ferroxcube, whose energy absorption depends upon the strength of an alternating magnetic field. The resultant alternating voltage on the crystal is rectified and fed to the control electrode via a phase discriminator. The stabilization ratio (the ratio of the frequency shifts with and without stabilization) is 20. With the multi-reflex klystron, the transmitter frequency can also be corrected by mechanical means; the accuracy within which the frequency can thus be kept constant is comparable to that achieved by purely electronic means. The entire system can be mechanically adjusted over a band of 150 Mc/s.

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

Reprint of these papers not marked with an asterisk * can be obtained free of charge upon application to the Philips Research Laboratory, Eindhoven, Netherlands.

2241: L. Heyne: Operation and characteristics of the "Vidicon" (T. Ned. Radiogenootsch. **20**, 1-10, 1955, No. 1).

The operation of "Vidicon" is explained and compared with that of the "C.P.S. Emitron". In both tubes a charge image is generated on the light-sensitive surface by the optical image focussed on it. However the "C.P.S. Emitron" uses photo-emission for this conversion whereas in the "Vidicon", photoconductivity is used. The video signal is obtained in both tubes by scanning the charge image with a beam of low velocity electrons, which stabilize the surface of the target at the potential of the cathode of the electron gun. The photoconductor used in the "Vidicon" must have a very high specific dark resistance, good sensitivity for visible light and no retardation effects.

With lead oxide encouraging results have been obtained. Lead oxide vidicons show a very low dark current, a good sensitivity and can be made with a short decay time. The signal output is proportional to the illumination. The lead oxide layer is also sensitive to X-rays. In X-ray pictures the statistical fluctuation of the number of X-ray quanta is easily noticeable. This is the result of the fact that one X-ray quantum releases about 300 electrons. The X-ray sensitivity is too low for medical applications.

2242: F. H. J. van der Poel: A 35 mm film scanner (T. Ned. Radiogenootsch. **20**, 11-19, 1955, No. 1).

After some general remarks on the televising of motion-picture film and a brief description of a well-known film scanning method using an intermittent projector and a storage type camera tube, this paper describes a method using the flying spot scanner (see Philips tech. Rev. **15**, 221-232, 1953/1954) and continuously moving film. Television pictures of very good quality are obtained with the latter method.

2243: K. Teer: Overdrachtsystemen voor kleurentelevisie (T. Ned. Radiogenootsch. **20**, 21-34, 1955, No. 1). (Transmission systems for colour television; in Dutch).

A compatible colour television system is required to transmit the colour picture information, which can be represented by three video signals, in a normal television channel in such a way that monochrome reception on existing black and white receivers is possible. Such a transmission can be established by adding to the normal video signal (luminance signal) subcarriers modulated by two other signals (colour signals). Two subcarriers may be used for the transmission of the colour signals but the modulation of both signals on a single subcarrier is also possible. (N.T.S.C. colour system in U.S.A.) A brief description is given of an experimental transmisson system where two subcarriers are applied.

2244: A. G. van Doorn and F. W. de Vrijer: Display methods for colour television (T. Ned. Radiogenootsch. **20**, 35-48, 1955, No. 1).

This article describes some possible methods of producing colour-television images when three simultaneous colour signals representing the red, green and blue brightnesses are available. In addition to a brief review of the principles used in a direct-view colour-picture tubes, a description is given of a projection method of displaying colour images. The results obtained to date by the various systems are compared.

2245: F. A. Kröger: The physical chemistry of sulphide phosphors (Brit. J. appl. Phys., Suppl. No. 4, 558-564, 1964).

By preparing phosphors under controlled conditions, it is possible to regulate the properties in a reproducible way. From a study of the variations induced in this way, insight into the micro-structure of the phosphor can be obtained. For ZnS, various physico-chemical aspects of the preparation such as oxidation-reduction, the action of fluxes and the formation of solid solutions are discussed.

VOL. 17 No. 12, pp. 341-372 JUNE 1956

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 F.M. SECTION OF MODERN BROADCAST RECEIVERS

I. CIRCUITRY AND STRUCTURAL DESIGN

by H. de QUANT and P. ZIJP.

II. THE BUILT-IN F.M. AERIAL

by M. HUISSOON.

621.396.621:621.376.3

In the early days of broadcasting the thing which perhaps appealed most to the imagination was the idea of receiving transmissions direct from distant countries. Nowadays the attitude is less romantic: enquiries show that the vast majority of present-day listeners tune in to only a few stations (mainly in their own country) where they can be assured of good reception. The emphasis now lies on reception of good quality. Efforts in this direction are almost all concerned with the use of frequency modulation (F.M.) for broadcasting. Although this system implies short wavelengths (metre waves), whose range is limited, it has important advantages: the signal-to-noise ratio is much better, and, in comparison to most long and medium wave transmitters, the audio spectrum is much wider. Nearly all European countries have a short-wave F.M. service in operation in addition to the usual amplitude modulated transmissions on long, medium and short waves. Accordingly, most of the latest receivers in Europe (with the exception of those in the lowest price class) are adapted for both systems.

It is usually assumed that broadcasting with frequency modulation (F.M.) owes its growing popularity to the fact that with this modulation system the audio spectrum extends to about 15000 c/s, which is much higher than that of the majority of broadcasting transmitters working with amplitude modulation (A.M.). Up to a point this is true if we compare F.M. broadcasting with A.M. broadcasting on long or medium waves, where the large number of transmitters enforces a drastic limitation of the bandwidth and accordingly of the audio spectrum [1]). This limitation however, would disappear if the carrier wavelength in an A.M. transmitter were chosen in the metre waveband, as has been done for the sound channel for TV in England, France and Belgium. One might then expect there to be no difference in sound quality between A.M. and F.M. reception: nevertheless, F.M. is to be preferred for high quality sound broadcasting for the following reasons:

1) With F.M. the signal is contained in the variations in frequency of the carrier wave. All interference which causes variations in amplitude can, in principle, be eliminated and consequently only interference which affects frequency variations can give rise to audible interference after detection. This results in a signal-to-noise ratio which is much better than with A.M., provided that the frequency variation is large enough.
2) The second advantage of F.M. concerns the transmitter. With F.M. the power transmitted is constant, independent of the modulation whilst an A.M. transmitter must have a reserve such that up to 4 times the normal power can be supplied during periods of strong modulation.

As Armstrong showed in 1936, in order to achieve the favourable signal-to-noise ratio potentially available with F.M., the maximum frequency variation ("frequency sweep") must be much greater

[1]) It should be noted that some A.M. transmitters do transmit quite high notes (up to 10 000 or even 12 000 c/s), for the sake of those who can receive these transmissions without interference. With this point in mind there are receivers on the market whose bandwidth can be enlarged accordingly.

than the highest audio frequency [2]) to be transmitted, viz. 15 000 c/s. By international agreement, the highest permissible frequency sweep for broadcasting is 75 kc/s. With this sweep, a bandwidth of 160 kc/s at the very least is required, so that for F.M. we have to resort to V.H.F. Of the frequency ranges made available for F.M. broadcating by international agreement, that used in Europe is from 87.5 to 100 Mc/s. To profit fully from the wide audio spectrum, clearly the audio frequency section of the set, including the loudspeaker, must give good reproduction of frequencies up to 15 000 c/s. (Of course this is desirable not only for F.M. reception but also for "local reception" of the A.M. transmissions mentioned in note [1]) and for the reproduction of gramophone records.) The problems concerning the audio frequency section of the set will not be discussed here; this article deals with the following matters:

I) The requirements which a modern F.M. receiver should meet and the way in which these can be realized; also questions which arise in the design of a combined A.M./F.M. receiver;
II) The built-in aerial for F.M. reception.

[2]) See Philips tech. Rev. 8, 42-50 and 89-96, 1946.

I. CIRCUITRY AND STRUCTURAL DESIGN

Requirements for an F.M. receiver

In addition to the low frequency section the following points have to be considered in a modern F.M. receiver, viz: the sensitivity, the signal-to-noise ratio, the signal-to-interference ratio, the selectivity and the radiation. These points will now be considered in turn.

Sensitivity

The input signal of the low frequency amplifier, even when receiving from fairly distant transmitters, must be so strong that the volume control never needs to be in its maximum position. If this were necessary, the sound would generally be much too loud and distorted when switched over to A.M. transmitters whose field strength is generally far greater.

Signal-to-noise ratio

The quality of the reception is determined largely by the signal-to-noise ratio at the output of the set. To make this as large as possible, one should aim at:

1) As large a ratio of H.F. signal to noise as possible at the input of the set.
2) As large an H.F. amplification as possible so that the noise arising from the mixer tube and the I.F. tubes is relatively small.
3) An I.F. amplification which ensures that the signal at the control grid of the last I.F. tube is large enough to permit it to be limited (in order that the noise present as amplitude modulation may be largely eliminated).
4) Further limiting of the signal in the detector.

Item 3) above implies the use of 3-stage I.F. amplification. Combined with the large H.F. amplification this makes the total amplification so large that normal criteria for the sensitivity are no longer valid. With A.M. receivers, the usual measure of the sensitivity is the e.m.f. of the aerial which gives an L.F. output power of 50 mW at a certain depth of modulation. With an F.M. receiver, a sensitivity of, for example, 0.4 μV might be found, but at such a low e.m.f. the noise would predominate. Three cases are therefore distinguished as far as the signal-to-noise ratio is concerned according to the strength of the aerial signal. In each one, the aerial e.m.f. required to produce the signal-to-noise ratio given below should not exceed a certain value, according to present standards:

Signal-to-noise ratio dB	Aerial e.m.f. μV	Reception quality
8-10	2	This low quality is considered satisfactory for reception of very weak signals.
26	4	Generally accepted as "good" quality.
40-60	20	Very high quality.

Signal-to-interference ratio

F.M. reception can suffer interference from:

1) Transmitters working on the same central frequency as the transmitter it is desired to receive.
2) Transmitters whose central frequency differs but little from that of the one it is desired to receive.
3) Reception along paths of varying length, such as is often the case in mountainous regions.
4) Sparks from commutator motors and internal combustion engines (without any measures for

interference suppression, an engine may give rise to a field strength of, say, 500 μV/m at a distance of 10 m).

Limiting again provides an effective means of counteracting this last type of interference in F.M. In general, we require a signal-to-interference ratio of at least 30 dB (for modulation with a frequency sweep of 15 kc/s). As in the case of noise suppression, this implies three I.F. stages.

Selectivity

The carrier frequencies of the European F.M. transmitters are usually 300 kc/s apart. For this separation the selectivity must be at least 200 (i.e. a signal 300 kc/s off-tune must be amplified 200 times less than the in-tune signal). In practice, this selectivity can be achieved only by using three or more stages of I.F. amplification.

Radiation

To avoid interference with nearby F.M. and TV receivers, the oscillator of an F.M. receiver should radiate only very little both at the fundamental frequency and at the first harmonic (which comes within a frequency range used for TV in some countries). Very strict requirements are in force in Germany as regards radiation in the frequency ranges (87.5-100) + 10.7 ≈ 98-111 Mc/s (10.7 Mc/s is the intermediate frequency) and 196-222 Mc/s.

F.M. receiver circuits

We shall now examine how the requirements outlined above are met in Philips F.M. receivers.

The H.F. amplifier

The input of the H.F. section is matched to an aerial impedance of 280 ohms (see part II of this article); the characteristic impedance of the feeder between aerial and set also has this value. Theoretically, the noise factor should be 2 but for various reasons it is approximately 3.5 in practice, a value which is easy to maintain in mass-production.

In order to keep the noise from the tubes as small as possible, triodes are used in the H.F. section because, unlike pentodes, they are free of distribution noise. A tube that is frequently used is the double triode ECC 85, one half of which functions as an H.F. amplifier and the other half as an oscillator-mixer tube.

A disadavantage of triodes, however, is that the grid-anode capacitance C_{ag} is fairly large. When using a grounded-cathode circuit, it would be necessary to neutralize in order to prevent the output circuit reacting unduly on the input circuit via C_{ag}, but this is not so easy on account of the rather large frequency range (87.5-100 Mc/s). Neutralizing presents no problem in the grounded-grid circuit, for this involves the anode-cathode capacitance C_{ak}, which is much smaller than C_{ag} and moreover has less effect because the input impedance of the tube in this circuit is very low, viz. smaller than $1/S$, where S is the mutual conductance. In the ECC 85, with $S = 5$ to 6 mA/V, the input impedance in a grounded-grid circuit is roughly 150 ohms. This circuit however has the disadvantage of low amplification (ratio of anode H.F. voltage to aerial e.m.f.).

A favourable compromise is to be found in the "Zwischenbasis" circuit (*fig. 1a*) which is intermediate between the grounded-cathode and grounded-grid circuits. Neither the cathode nor the grid are common to both the input and circuits; instead, an intermediate point, for example a tapping on the coil S_2, is made common. As the tapping is moved downwards (in this figure) the circuit tends towards the grounded-cathode circuit; moved upwards, we approach the grounded-grid circuit. By a suitable choice of the position of the tapping sufficient amplification can be obtained from only one tube and the input impedance of the tube can be made considerably higher than $1/S$ (and the damping in the input circuit small in consequence); also the neutralizing is not very critical. The condition for neutralizing with a given tapping point is that there should be a specific ratio between the grid-anode capacitance and the total anode-cathode capacitance (see fig. 1*b*). In our

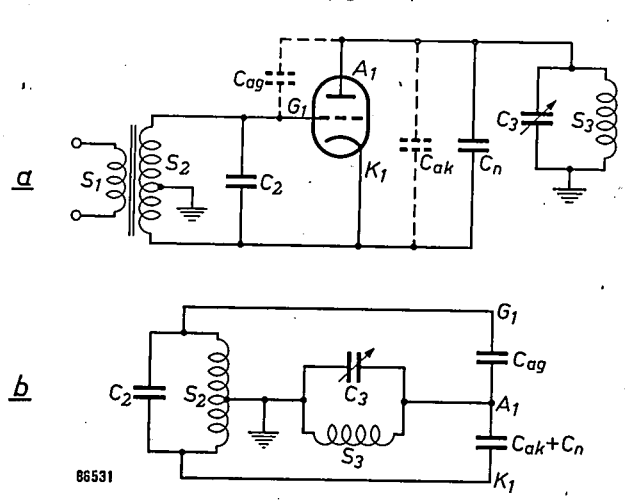


Fig. 1. *a*) Diagram of the H.F. amplifier in "Zwischenbasis" circuit. *b*) Equivalent circuit. S_1 primary coil. S_2-C_2 secondary circuit, tuned to a fixed frequency at the middle of the frequency range (94 Mc/s). A_1 anode, G_1 grid and K_1 cathode of one half of the double triode ECC 85. S_3-C_3 output circuit. C_{ag} grid-anode capacitance. C_{ak} anode-cathode capacitance. C_n neutralizing capacitance. The power supplies are not shown.

case the correct value of this ratio is obtained by adding a certain neutralizing capacitance C_n to the stray capacitance C_{ak} already present. Once the right values had been found, it was possible to use a fixed tapping and a fixed neutralizing capacitor in mass-production, with no subsequent adjustment.

The mixer stage

The H.F. amplification stage is followed by the frequency-changer stage consisting of oscillator and mixer. As already stated, 10.7 Mc/s is chosen as the intermediate frequency. As regards the mixer tube the most important properties are minimum noise and large conversion coductance. The triode is the best valve to meet these requirements and as mentioned earlier, an ECC 85 is often used. Owing to the large H.F. amplification of the previous stage the amount of noise from the mixer tube is negligible. Since a triode has only one grid, a so-called additive mixing circuit has to be used, i.e. a circuit in which the two signals to be mixed are applied to the same grid.

Fig. 2 shows the layout of the H.F. stage and the mixer stage. Two features of the circuit will now be discussed: the first concerns the damping of the first I.F. band pass filter and the second concerns the stray radiation.

The plate resistance of the mixer tube measured on D.C. amounts only to about 20 000 ohms, which implies a fairly strong damping in the band-pass filter and, as a result, low amplification. In addition, an I.F. voltage passes to the grid via the grid-anode capacitance, causing the effective plate resistance to become even lower, viz. about 5000 ohms. However, the bridge circuit shown in fig. 2*b* causes another I.F. voltage to appear at the grid such that the effective plate resistance is increased to a value at which the damping is small enough. The bridge circuit comprises four capacitances (regarding coils S_3, S_4 and S_5 as short-circuits for I.F. currents). Of these four capacitances C_{ag} is given by the tube and C_6 and C_7 are for other reasons also fixed within certain limits. Typical values in practice are: $C_{ag} = 2$ pF, $C_6 = 20$ pF and $C_7 = 200$ pF. With these values the bridge would be balanced if $C_5 + C_3 = 2000$ pF. By choosing another value for $(C_5 + C_3)$, the I.F. voltage appearing across the

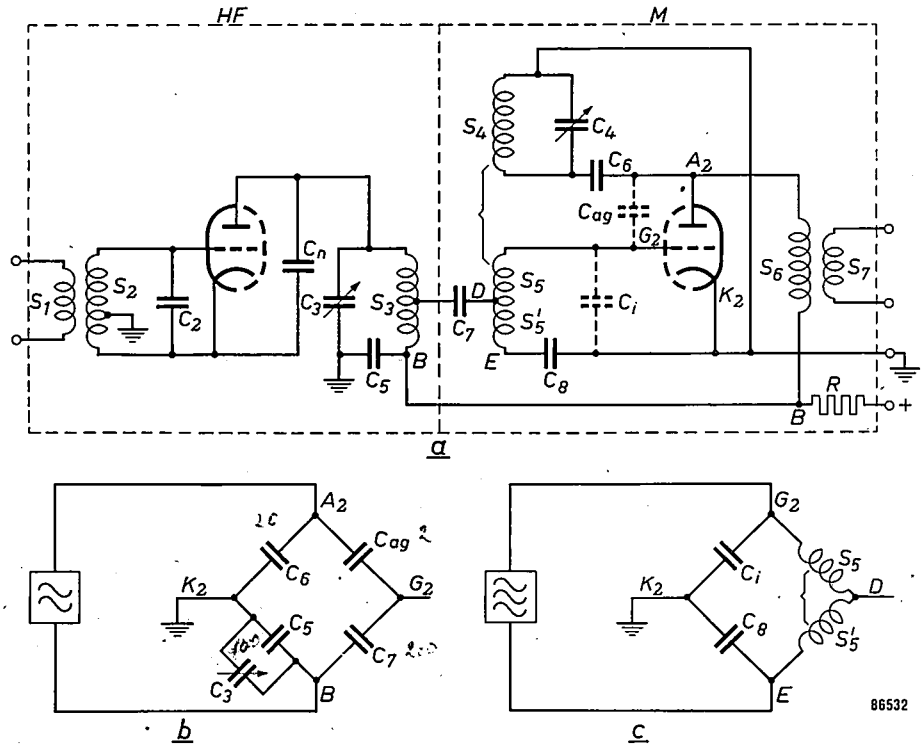


Fig. 2. *a*) Simplified diagram of high frequency amplifier *HF* and mixing stage *M*. For S_1, S_2, S_3, C_2, C_3 and C_n see figure 1. A_2, G_2, K_2 are electrodes of the other half of the double triode ECC 85. S_4-C_4 oscillator circuit. $S_5 + S'_5$ feed-back coil. S_6-C_6 primary circuit of the first I.F. band-pass filter. S_7 coupling coil. C_i input capacitance.
b) Bridge circuit in circuit (*a*) which, at the correct value of C_5, causes a reduction in the damping of the band-pass filter.
c) Second bridge circuit, incorporated in (*a*) which, at the correct value of C_8, ensures that no voltage at the oscillator frequency appears at point D, so that there is no radiation.

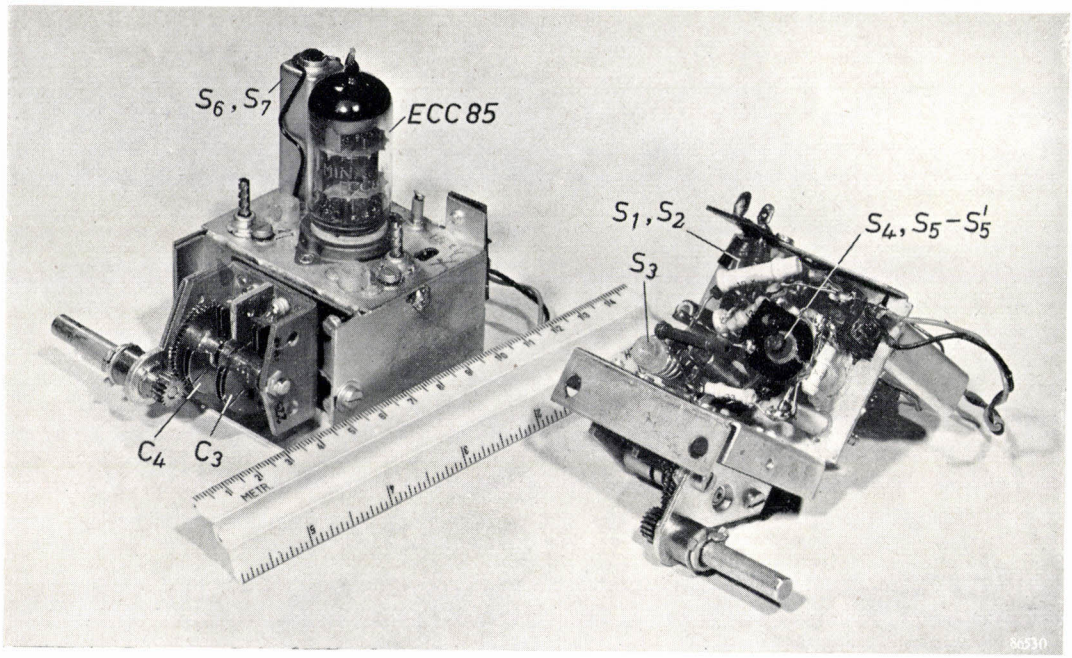


Fig. 3. One form of the circuit of figure 2*a*, seen from above and below. The letters refer to the corresponding components in fig. 2.

diagonal K_2-G_2 can be varied; with $C_5 \approx 900$ pF, this voltage has the correct magnitude and sign. The capacitance C_3 of the tuning capacitor is adjustable from 2.5 to 12.5 pF, and this range is so small compared to C_5 that the required condition is maintained over the entire tuning range.

A second bridge circuit is incorporated in the circuit of fig. 2*a*, to counteract radiation. This is shown in fig. 2*c* and is formed by the two sections of the reactance coil $S_5 + S_5'$, the input capacitance C_i of the oscillator tube and the capacitor C_8. The bridge is balanced with C_8 so that the diagonal K_2-D remains free of oscillator voltage and this voltage can therefore not reach the aerial via D and the H.F. amplifier. The self-inductances, mutual inductances and capacitances in the bridge circuit are such that the balance of the bridge is theoretically independent of the frequency. Stray self-inductances somewhat impair this independence of frequency in practice, but in the range within which the oscillator frequency varies the balance holds satisfactorily.

In this way and with very short wiring, the circuit in figure 2*a* can produce a total amplification of 200× (ratio of I.F. voltage on control grid of first I.F. tube to aerial e.m.f.). *Fig. 3* shows two views of such a H.F. section.

The I.F. amplifier

As already mentioned, there are a number of reasons for using not less than three I.F. stages in an F.M. receiver, e.g. to achieve sufficient selectivity and especially sufficient amplification in order that undesirable amplitude variations (noise and interference) can be eliminated by limiting.

In the first I.F. stage, use can be made of the heptode section of the triode-heptode (ECH 81) which does service as a mixer tube in A.M. reception. The triode section remains out of use for F.M. reception. In the second stage, the best tube to use is the EF 89 pentode since with this type, the ratio S/C_{ag}, which determines the amplification, is largest. The last stage must function as a limiter particularly for strong signals (with weak signals it is mainly the detector which functions as a limiter). The EF 85 tube has the most favourable characteristics for this. This tube is used in such a way that limiting is achieved by grid current; this prevents unwanted amplitude variations in the anode current.

Automatic volume control is arranged by taking a control voltage from between the second and third I.F. tubes and feeding this back to the control grid of the first I.F. tube. In this way, the automatic volume control has no direct influence on the limiting effect of the third I.F. tube.

The detector

The most modern F.M. detector is the "ratio detector", which has the property of not responding to amplitude variations of the signal and therefore works as a limiter [3]). The advantages of this detector

[3]) With the addition of one or two circuit elements, this detector can easily be converted into an A.M. detector; this is particularly convenient in the case of dual or multi-standard TV receivers. See Philips tech. Rev. **17**, 168, 1955/1956 (No. 6), fig. 12.

over the Foster and Seeley detector originally developed in America [4]) are as follows; 1) better limiting for weak signals, and 2) less interstation noise, i.e. the noise the receiver produces when it is not well-tuned (for example while it is being retuned to another station).

The basic circuit of the ratio detector is reproduced in *fig. 4a*.

The following discussion is intended to show how this circuit works as an F.M. detector (and not as an A.M. detector) without, however, claiming to be a complete explanation.

Consider first the property of band-pass filters that the difference in phase between the primary and the secondary voltage increases with frequency and at the resonance frequency is equal to 90°.

In fig. 4*a* L_1-C_1-L_2-C_2 form a band-pass filter whose resonance frequency is equal to the central intermediate frequency f_c. When there is no frequency modulation (frequency f equal to f_c), fig. 4*b* applies: the secondary voltages $\mathbf{V}_1$ and $\mathbf{V}_2$ are phase-shifted + 90° and − 90° respectively with respect to the primary voltage $\mathbf{V}_0$ (letters in the bold type represent vectors).

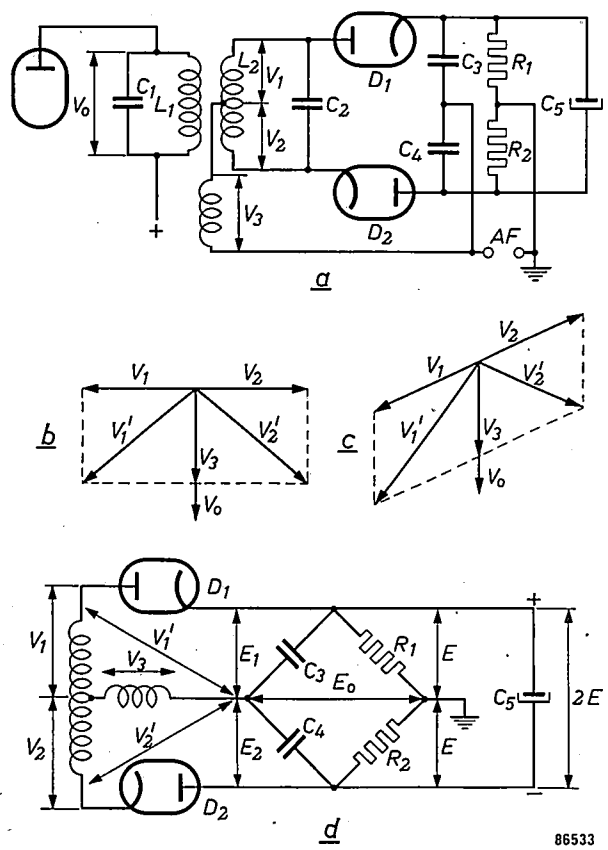


Fig. 4. *a*) Circuit of ratio detector. L_1-C_1-L_2-C_2 last I.F. band-pass filter. AF terminals across which the audio frequency voltage appears.
b) Vector representation of the I.F. voltages $\mathbf{V}_0$, $\mathbf{V}_1$, $\mathbf{V}_2$ and $\mathbf{V}_3$ (see (*a*)) in the absence of any frequency modulation. The sum voltages $\mathbf{V}_1'$ and $\mathbf{V}_2'$ which are rectified are of equal magnitude.
c) Same as (*b*), but with a certain frequency deviation. $\mathbf{V}_1'$ and $\mathbf{V}_2'$ are no longer equal.
d) Same as (*a*) but showing the direct voltages E_1 and E_2 (varying at audio frequency), the audio frequency output voltage E_0 and the direct voltage $2E$ (assumed constant) across the electrolytic capacitor C_5.

[4]) See Philips tech. Rev. 8, 48, 1946, fig. 10.

To $\mathbf{V}_1$ and $\mathbf{V}_2$ a voltage $\mathbf{V}_3$ is added which is always in phase with $\mathbf{V}_0$. The sum voltages $\mathbf{V}_1'$ and $\mathbf{V}_2'$ are each rectified separately (diodes D_1 and D_2) so that the direct voltages E_1 and E_2 appear across the smoothing capacitors C_3 and C_4 respectively. We then have:

$$\frac{E_1}{E_2} = \frac{V_1'}{V_2'} \quad \ldots\ldots\ldots\ldots \quad (1)$$

where V_1' and V_2' are the moduli of $\mathbf{V}_1'$ and $\mathbf{V}_2'$. Since $\mathbf{V}_1$ and $\mathbf{V}_2$ are equal in magnitude ($V_1 = V_2$), when there is no frequency modulation $V_1' = V_2'$, $E_1 = E_2$ and the output voltage E_0 is nil (fig. 4*d*).

If there is frequency modulation, the phase angle between $\mathbf{V}_1$ and $\mathbf{V}_0$ is alternately larger and smaller than 90°. Fig. 4*c* gives the state of affairs for a particular frequency deviation: V_1' and V_2' are now no longer equal, neither are E_1 and E_2 and there is consequently a certain output voltage E_0.

Since the voltage across the electrolytic capacitor C_5 can be regarded as constant (value, say, $2E$), for short periods, a constant voltage E appears across each of the two equal resistors R_1 and R_2. We then have:

$$\left.\begin{aligned} E_1 &= E - E_0 \\ E_2 &= E + E_0 \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots \quad (2)$$

From (1) and (2), we may write:

$$E_0 = \frac{V_2' - V_1'}{V_2' + V_1'} E \quad \ldots\ldots\ldots\ldots \quad (3)$$

With a correct choice of component values, E_0 varies almost proportionally to the frequency deviation.

If the input signal V_0 of the detector changes in amplitude, for example by a factor a, then V_1' and V_2' also change by this factor, so that the ratio $(V_2' - V_1')/(V_2' + V_1')$ remains constant (hence the name "ratio detector"). If the change in V_0 is rapid in comparison with the large time-constant $(R_1 + R_2)C_5$, then E remains practically constant and hence from (3) E_0 also remains constant. In other words, any amplitude modulation which might arise in V_0 is not manifested in E_0. Because the diodes are not ideal, the suppression of amplitude modulation is not complete.

In F.M. transmissions "pre-emphasis" of the high notes is used, i.e. the frequency sweep of the transmitter for an audio signal of constant amplitude is made to increase with frequency. At the receiver end, an equal "de-emphasis" must be applied to restore the original balance. The advantage of this method is to be found in the fact that with the de-emphasis the noise is attenuated relatively more than the signal and this results in reproduction with an improved signal-to-noise ratio. The pre-emphasis is achieved by using a high-pass LR-filter before the modulation stage; for de-emphasis in the receiver, a low pass RC-filter is incorporated after the detector stage. Pre-emphasis and de-emphasis compensate each other if both filters have the same time constant. On the continent of Europe this constant is fixed at 50 μsec by agreement.

Structural design

Circuit requirements as outlined above impose certain restrictions on the structure of an F.M.

receiver in particular with regard to the location of certain components and the control knobs.

It is not advisable to mount the A.M. and F.M. tuning condensers on a single spindle; the component locations would then preclude the use of the shortest wiring in the F.M. section. The simplest solution is for each condenser to be driven by its own knob. If the set is provided with push-buttons (on/off switch, waveband selection, etc.) the set can be left tuned in to both an A.M. transmitter and an F.M. transmitter; the set can then be switched on at either of these stations by pressing the appropriate button without needing to tune again. To save space on the front of the set, the two tuning knobs can be arranged concentrically (*fig. 5*). The condensers can be turned via the usual cord-drive.

It is also possible to use only one tuning knob which can be coupled with the A.M. or the F.M. tuning unit as desired. The mechanical coupling of the knob is done automatically when the required wave band is selected by means of buttons (*fig. 6*). This type of tuning control imposes considerable demands on the reproducibility of the adjustment since only a slight mechanical inaccuracy can adversely affect the sound quality and the freedom from interference.

Pre-tuning, with push-button selection might also be applied to more than two transmitters, but of course for each extra transmitter a separate manual tuning arrangement is necessary, and a corresponding push-button. In the more de luxe models one might use motor-driven tuning, the motor being stopped automatically at the desired point by pressing an appropriate button. Features such as this, however, are beyond the scope of this article.

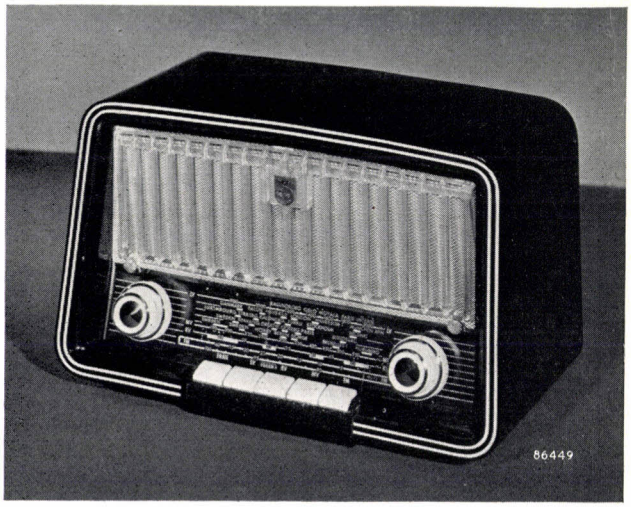


Fig. 5. Receiver with separate tuning knobs for AM and FM (type BX 253 U). In the right is a double knob for tuning; the front knob is used for AM and the rear for FM. The left-hand knob is also double; the front knob is used for the volume and the rear one for the tone control. The buttons in the centre are (from left to right); on/off switch, long wave, short wave, medium wave, and F.M. reception. For gramophone reproduction, the long and the short wave buttons are pressed in together.

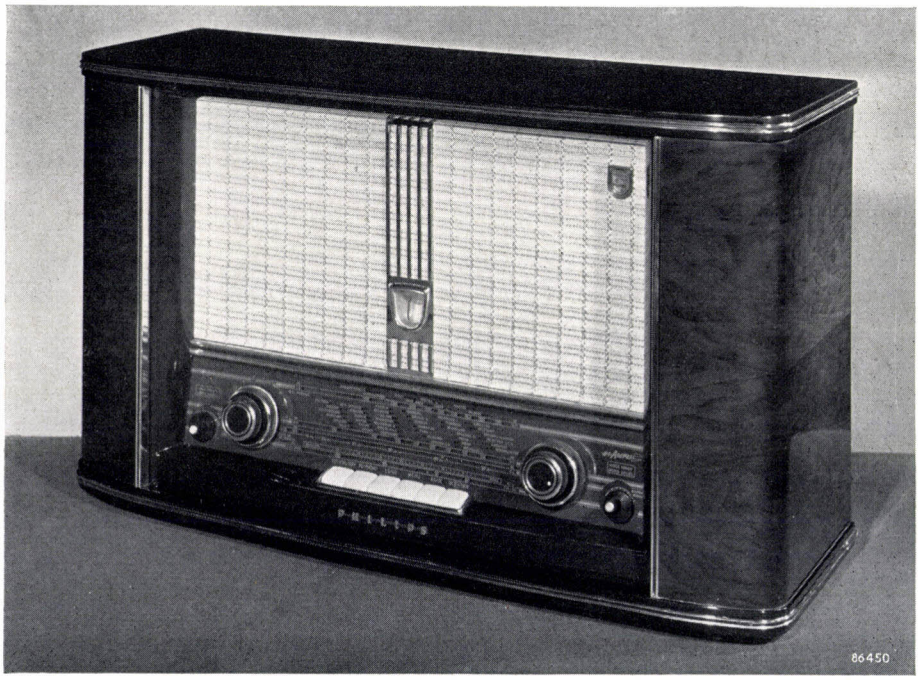


Fig. 6. Receiver (type BX 653 A) with a common tuning knob (the large knob on the right which appears to be double but is, in fact, single). This controls the A.M. tuning when an A.M. button is pressed in and the F.M. tuning when the F.M. button is pressed in. The knob on the left is double; the front knob is the volume control and the rear one adjusts the orientation of the "Ferroceptor" aerials. The knobs on the extreme left and right are independent tone controls for low and high notes.

II. THE BUILT-IN F.M. AERIAL

Many listeners are unable for various reasons to instal a good outside aerial. For short-wave reception (10-100 m, A.M.) this is not necessary since the small plate aerial with which the majority of sets are now equipped ensures good and practically interference-free reception on these wavelengths nearly everywhere. In the case of medium and long waves, however, the situation is different. In the absence of an outside aerial, use is often made of an indoor aerial. In many cases this gives sufficiently strong reception but especially in towns, interference may be troublesome. Improvements in long and medium wave reception have been effected by equipping sets with inductive aerials for these wavebands (either frame aerials with one or two turns or "Ferroceptor" aerials [5]).

For receivers incorporating a frequency modulation unit it is convenient to equip them with a built-in aerial for F.M. reception as well, especially as with ultra short waves the circuit parameters and matching require some care to get the best results.

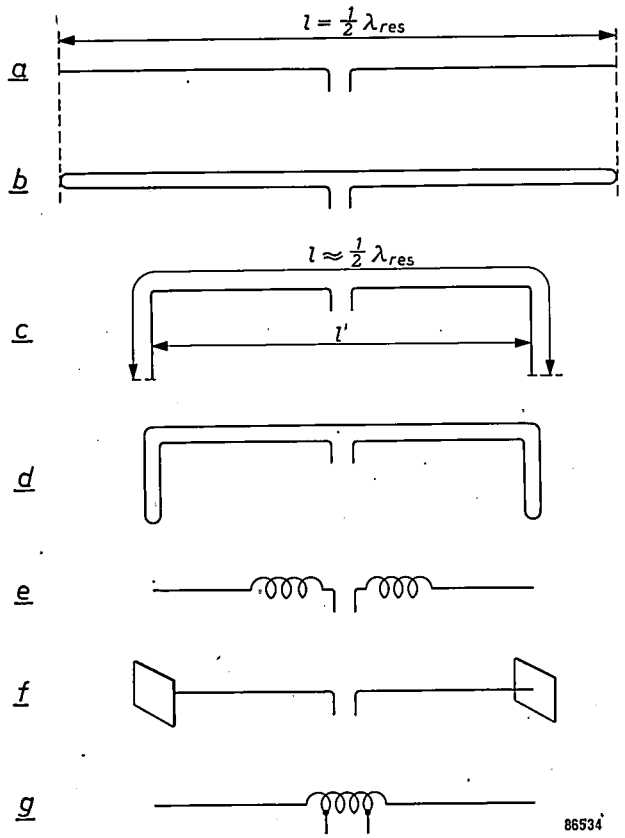


Fig. 7. Forms of dipole aerials. *a*) Normal dipole; length l equal to half the resonance wave length λ_{res}. *b*) Folded dipole.. In order to fit the dipole into a cabinet of length $l' < l$, it can either be bent (*c*, *d*), or provided with two coils (*e*), two end plates (*f*) or one coil (*g*).

[5]) H. Blok and J. J. Rietveld, Philips tech. Rev. **16**, 181-194, 1954/1955 (No. 7).

This F.M. aerial — a dipole — can also be used for reception on amplitude modulated short waves. Many modern radio sets therefore incorporate three built-in aerials, viz. two inductive aerials and a dipole.

A dipole in its simplest form is shown in *fiig. 7a* Electrically speaking it may be replaced by an equivalent generator with a certain internal impedance. The magnitude and the nature of this impedance are important for the matching, which will be dealt with presently. In the case of fig. 7*a*, the impedance has the same character as a resonant circuit of self-inductance, capacitance and resistance in series. At resonance (wavelength $\lambda_{res} = 2 \times$ length l) the impedance is a resistance of roughly 70 ohms; at higher frequencies it is inductive in character and at lower frequencies capacitative. The folded dipole (fig. 7*b*) behaves similarly, only here (the leads having the same diameter) the resonance resistance is 4 × as large, i.e. roughly 280 ohms.

Since the wavelength in the F.M. band is three metres or more and a dipole such as in fig. 7*a* or *b* must accordingly be about 1.5 metres long, it cannot be housed in the cabinet of a radio set without its shape being modified as shown in figures 7*c* and *d*. It appears that this bending of the aerial hardly alters the resonance frequency but that the radiation resistance (which forms the greater part of the resonance resistance) decreases considerably. In the case of a cabinet whose largest dimension is between 20″ and 30″, the resonance resistance is 20 to 30 ohms with the normal dipole (fig. 7*c*) and 100 to 120 ohms with the folded dipole (fig. 7*d*).

It is possible to make do with a dipole shorter than $\frac{1}{2}\lambda_{res}$ by adding self-inductance or capacitance. In fig. 7*e* the self-inductance is increased by two coils and in fig. 7*f* the capacitance is increased by two end plates. These kinds of aerial have almost the same characteristics as a bent dipole (fig. 7*c*). If in fig. 7*e* the two coils are replaced by one coil having a self-inductance twice as large (fig. 7*g*), the required impedance may be achieved by means of tappings on the coil.

In Philips sets a bent folded dipole is used (fig. 7*d*). The length is so chosen that the resonance occurs in the middle of the frequency range, i.e. at 94 Mc/s.

Matching of aerial and receiver

The signal at the input terminals *c-d* of the receiver (*fig. 8*) is strongest if the aerial impedance $Z'_a = R_a + jX_a$ is matched to the impedance $Z_i' = R_i' + jXi'$ which the aerial "sees" at the input *a-b* of the lead connecting it to the set, i.e. if

$R_a = R_i'$ and $X_a = -X_i'$. The impedance Z_i' generally differs from the input impedance $Z_i = R_i + jX_i$ between the terminals *c-d* of the receiver itself in that the feeder cable functions as an impedance transformer. Only if Z_i is equal to the characteristic impedance ζ of the feeder over the whole frequency range, is $Z_i' = Z_i$.

A folded dipole used as an outside aerial (with $R_a = 280$ ohms) can be properly matched by connecting it via a suitable lead ($\zeta = 280$ ohms) to a receiver with an input impedance at resonance likewise of 280 ohms.

With a bent folded dipole inside the set (fig. 7*d*), we have $R_a \approx$ 100-120 ohms whilst $R_i = 280$ ohms. The short connecting line has to make as good matching as possible. In order to check the matching we can profitably use a

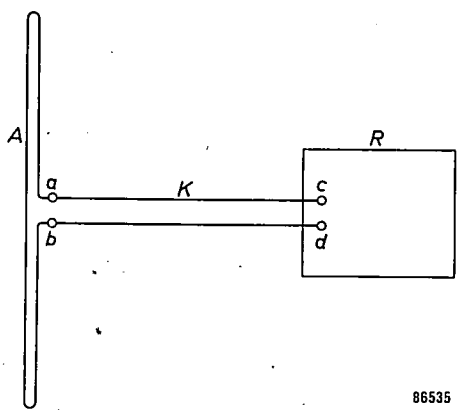


Fig. 8. Folded dipole *A* connected by a feeder *K* to the receiver *R*.

measuring technique [6]) by means of we obtain values of the components of Z_a and Z_i'. It is then possible to see whether the conditions for matching: $R_a = R_i'$ and $X_a = -X_i'$, are satisfactorily met. If this is not the case, the measurement may be repeated with a connecting line of other dimensions: by such a process of trial and error, satisfactory matching can be achieved.

[6]) J. C. van den Hoogenband and J. Stolk, Philips tech. Rev. **16**, 309-320, 1954/1955 (No. 11).

On the last page of the article quoted in [6]) a number of variants of the measuring technique described are given. The last of these has been used in the present work: a short description of the measuring arrangement will now be given.

In *fig. 9*, *G* is a signal generator with a known frequency

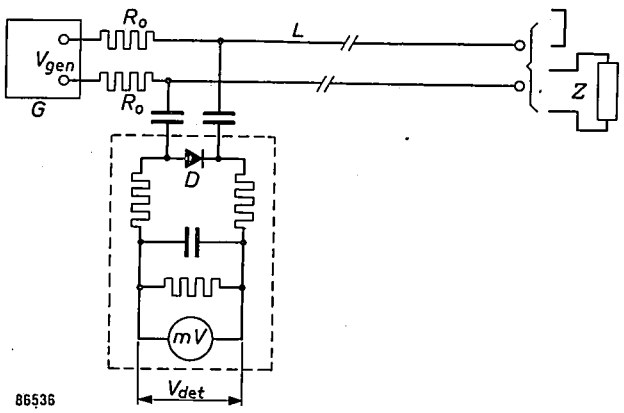


Fig. 9. Circuit for measuring impedances. *G* is a signal generator with adjustable frequency and voltage, amplitude-modulated. *L* is a long transmission line. *Z* is the impedance to be measured, *D* the detector and *mV* the electronic millivoltmeter.

f which is adjustable from 87.5 to 100 Mc/s. The output voltage V_{gen} is continuously adjustable and its relative value is known. The generator is connected in series with resistances R_0 to a transmission line roughly 60 m long. The value R_0 is so chosen that $2R_0$ plus the internal resistance of the generator is equal to the characteristic impedance of the transmission line. The other end of the latter is either short circuited or connected to the impedance Z which is to be measured (i.e. the impedance Z_a of the dipole or the input impedance Z_i' of the feeder *K* in fig. 8). At the generator end of the transmission line there is a detector *D* and an electronic millivoltmeter *mV*. Since V_{gen} is modulated up to a particular depth in amplitude, the detector voltage V_{det} read off from the meter *mV* is a measure for the H.F. voltage at the beginning of the transmission line.

Measuring the sensitivity

For measuring the sensitivity of a built-in dipole a millivoltmeter is connected in the limiting circuit or in the detector circuit of the receiver. This meter is calibrated in H.F. voltage at the input of the receiver.

The receiver is set up as far away as possible from objects which might reflect the waves. A transmitting dipole is placed not less than 30 metres away so that its plane of symmetry coincides with that of

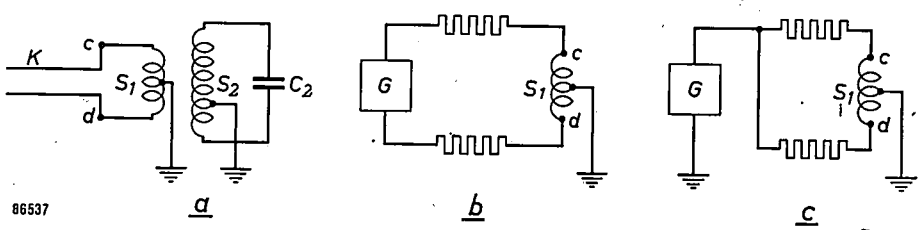


Fig. 10. *a*) Input of an F.M. receiver showing the feeder *K* of the built-in dipole. S_1, S_2 and C_2 as in fig. 1. *b*) and *c*). Checking for symmetry. *G* signal generator. The millivoltmeter connected in the set should show a deflection no greater in case (*c*) then it was in (*b*), when the generator voltage is made 10 times smaller than in the latter case.

the receiving aerial. The transmitting dipole produces a field whose strength at the receiver is known. The magnitude of the input signal resulting from this field strength is then read from the meter in the receiver.

The possibility of asymmetry at the input of the receiver has to be considered. The two wires of the feeder K (*fig. 10a*; the dipole is assumed to be not yet connected) function themselves rather like an aerial and supply in-phase signals to points c and d. When the input it absolutely symmetrical these signals cancel each other out, but when asymmetrical, as they are in general, a certain residual signal remains, say V_1. Let the desired signal provided by the dipole be V_2. According to the way in which the dipole is connected (connection is done by means of interchangeable plugs), the total signal becomes the sum of the vectors of V_1 and V_2 or V_1 and $-V_2$. The difference between the two sums and therefore the difference in sensitivity can be considerable.

For this reason it is desirable to check the symmetry beforehand. To do this a signal generator is connected in turn as shown in figs. 10*b* and *c*. In the latter case the meter would show zero deflection if there was perfect symmetry. The symmetry attained may be regarded as satisfactory if the deflection obtained in case *c* is less than that which would be obtained in case *b* with a 10× smaller input signal.

Summary. The principal electrical properties required of an F.M. receiver concern the sensitivity, the signal-to-noise ratio, the signal-interference ratio, the selectivity, the radiation and the quality of the audio-frequency section. These requirements have led to the design of sets incorporating the following features: single stage H.F. amplification with one half of a double triode ECC 85 in "Zwischenbasis circuit", a frequency changer stage using the other half of the ECC 85, three stage I.F. amplification (with the heptode section of an ECH 81, a pentode EF 89 and a pentode EF 85), a ratio detector and an audio frequency section (not described in this article) with good response up to 15 000 c/s. Two bridge circuits are incorporated in the mixing stage, the one ensures that the signal attenuation in the first I.F. band pass filter is small and the other prevents radiation. Sets designed for both A.M. and F.M. reception are equipped either with two independent tuning knobs or with one tuning knob which is coupled to the appropriate tuning unit by means of press-buttons.

The second part of this article deals with the built-in F.M. aerial: a folded dipole, the ends of which are bent to enable it to be housed in a radio cabinet. The bent ends change the aerial impedance from 280 ohms to 100-120 ohms. A short feeder of suitable length ensures matching with the input impedance of the set, which, with an eye on the use of an external dipole, is kept at 280 ohms. A method for checking the matching is briefly described.

A RECEIVER FOR THE RADIO WAVES FROM INTERSTELLAR HYDROGEN

II. DESIGN OF THE RECEIVER

by C. A. MULLER *).

522.6:621.396.9

Under the auspices of the Netherlands Foundation for Radio Astronomy, extensive observations are being made of the 21 cm radiation emitted by interstellar hydrogen. This radiation, which has the character of noise, is extremely weak and special measures are required to be able to distinguish if from the very much stronger noise inherent in the receiver. The first part of this article outlined the principles on which these investigations are based. The second part, printed below, describes the receiver circuits employed, with special reference to the function and design of certain components.

Principle of the receiver

The receiver about to be described was designed for investigations on the 21 cm wavelength (1,420 Mc/s) region of the radio spectrum from the Galatic System. As discussed in part I of this article [1]), there is a spectral line at this wavelength, which is radiated by clouds of hydrogen atoms in interstellar space.

The radiation from interstellar hydrogen has the character of noise, that is to say its intensity varies randomly with time. Considered as a function of frequency, the average intensity shows a marked maximum around 1,420 Mc/s. The form of the spectral line, i.e. the line profile to be measured, depends upon the bearing from which the radition is received A typical profile is shown schematically in *fig. 1*. The origin of these profiles and their significance were outlined in part I.

The principle of the measurement is as follows. The receiver, which is a superhet with three mixing stages, is switched to and fro at a rate of 400 times per second between two tuned positions between which a constant difference is maintained of $\Delta f =$ 1080 kc/s. These tuned positions are together displaced in such a way that one of them always falls outside the frequency band occupied by the spectral line (the width of the line is generally less than 1 Mc/s) and the difference in output voltage at both frequencies is recorded. In this way the complete profile of the spectral line is traversed. As the signal at the one frequency is always somewhat stronger than at the other, the output voltage after rectification displays a square waveform with a frequency of 400 c/s. Its amplitude is a measure of the difference in the noise power received at both frequencies.

Since the receiver is tuned to a frequency outside the spectral line (the comparison frequency) during each half period of the measurements, the available information over the line profile is received for only half the time. The accuracy of the measurement is thus reduced by a factor $\sqrt{2}$. (We shall return to this point later.)

In the actual receiver this drawback has been overcome by using a parellel arrangement of two I.F. amplifiers, tuned to frequencies which differ from each other by the above-mentioned $\Delta f =$ 1080 kc/s. Two measurements can now be carried out simultaneously; when the receiver is in the one tuned position, the first I.F. channel amplifies the comparison frequency f_2 while the second (owing to the frequency difference) amplifies the spectral line

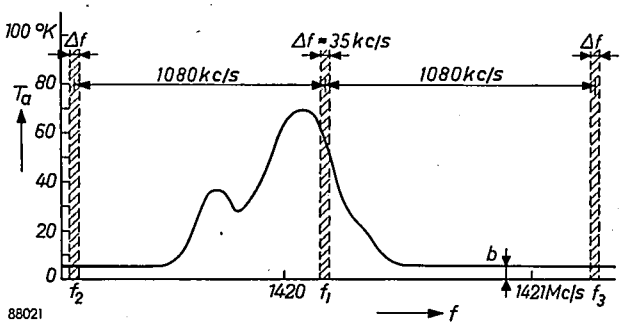


Fig. 1. Schematic diagram of a typical profile of the 21 cm line radiated by interstellar hydrogen. The noise intensity received is expressed in "antenna temperature", T_a. Continuous background radiation is present (b), the temperature of which is normally less than 10 °K. The receiver is switched to and fro between the frequencies f_1 (on the spectral line) and f_2 and f_3 (adjacent to it, on opposite sides) The bandwidth in each case is about 35 kc/s. All three frequencies are simultaneously displaced through the spectrum, a constant frequency interval of 1080 kc/s being maintained between them.

*) Director of the observations group of the Netherlands Foundation for Radio Astronomy.

1) C. A. Muller, A receiver for the radio waves from interstellar hydrogen, Part I, Philips tech. Rev. **17**, 305-315, 1955/56.

frequency f_1. In the other tuned position, however, the first I.F. channel amplifies frequency f_1, and channel 2 amplifies a second comparison frequency f_3, which, like f_2, differs from f_1 by $\Delta f = 1080$ kc/s, but in the opposite direction (see fig. 1). All three frequencies are together slowly displaced, so that f_1 again traverses the spectral line, while both f_2 and f_3 remain always outside it. After rectification, the output voltages of both channels each display a square waveform, and these two squarewave voltages, which are in anti-phase, are then added together in a push-pull amplifier and their combined amplitude measured. In this way, therefore, we are able to observe the spectral line uninteruptedly and to compare the received power with the average power at the two frequencies, f_2 and f_3.

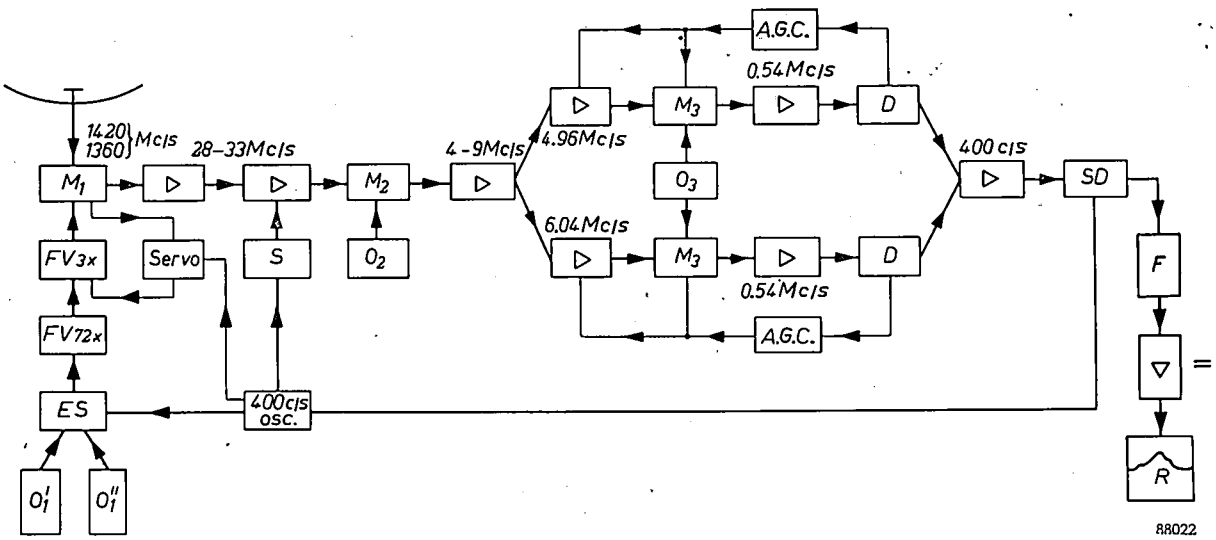


Fig. 2. Block diagram of the receiver. The units joined by thick lines constitute the receiver proper; those joined by thin lines represent control circuits. O'_1 and O''_1 are the crystal oscillators of the first mixer stage M_1; O_2 and O_3 are the oscillators of the second and third mixers, M_2 and M_3 respectively. *FV* frequency multipliers, *ES* electronic switch, *D* detectors, *AGC* automatic gain control, *SD* synchronous detector, *F* low pass filter, *R* strip-chart recorder.

A receiver designed on these principles must satisfy two stringent requirements. In the first place, it must be completely identical in both switching positions, as otherwise a 400 c/s square-wave voltage might appear even in the absence of a signal from the Galaxy. Secondly the switching process must have no effect on the measurement, that is to say, it must not cause the appearance in the output voltage of any component with a frequency of 400 c/s. These requirements principally determine the design of the receiver.

Fig. 2 shows the block diagram of the receiver. The incoming signal, which contains a wide-range of frequencies, is fed to the first mixer stage M_1, where it is combined with an oscillator signal with a frequency of about 1390 Mc/s. The difference frequency is fed to a wide-band amplifier (bandwidth approximately 5 Mc/s) operating in the region of 30 Mc/s. Thus, the amplified signal originates from those components of the incoming signal which contain frequencies in bands of 5 Mc/s around the spectral line frequency of 1420 Mc/s and the image frequency of 1360 Mc/s. The 30 Mc/s signal is combined in a second mixer stage (M_2) with an oscillator signal slowly changing in frequency between 34 and 39 Mc/s. The signal emerging from this stage, which consists of frequencies of a few Mc/s, is amplified and fed to two selective amplifiers in parallel operating at frequencies of 6.04 and 4.96 Mc/s. This produces the two I.F. channels mentioned above differing from each other by 1·080 Mc/s. The output signals of both amplifiers are combined in two third mixer stages (M_3) with the 5·50 Mc/s signal from a common oscillator. The resultant difference-frequency in both channels is therefore 0·54 Mc/s. The two amplifiers used for subsequently amplifying the signals in both channels have the same band-pass response, characterized by the narow bandwidth of 35 kc/s. These are the amplifiers which restrict the signals to two narrow bands within the total frequency spectrum received, that is to say, two bands 1·080 Mc/s apart in the region of 1420 Mc/s, and two similar bands around the image frequency. (The latter are of no further interest and, in any case, the signal received in these bands is extremely weak).

The exact frequencies in the centre of the bands (the tuning frequencies) are determined by the os-

cillator frequencies of the first and second stages. As the frequency of the second oscillator is slowly varied, the tuning frequencies vary likewise; they move slowly over several Mc/s through the spectrum thereby traversing the spectral line.

The oscillator signal for the first mixer stage is taken alternately from two oscillators, O_1 and O_1'', which differ in frequency by 1·050 Mc/s, and which are alternately gated by a separate 400 c/s oscillator. The result is that both tuned bands of the receiver are switched to and fro at a rate of 400 times per second between two positions lying 1·080 Mc/s apart so that, in turn, one of these frequencies lies within the spectral line while the other is adjacent to it. Since the amplitude of the ultimate square-wave voltage depends upon the gain in both 0·54 Mc/s amplifiers, automatic gain control is applied in both channels and the average output held constant.

Finally, after detection in both channels, the absolute values of the 400 c/s signals obtained are added in a push-pull amplifier, and the summed voltage is fed to a synchronous voltmeter. This measures and records the fundamental of the 400 c/s squarewave voltage which appears at the output of the push-pull amplifier.

Apart from the radio receiver itself, the equipment also contains many semi-automatic calibrating and control circuits.

We shall now consider the various parts of the receiver individually.

H.F. section and crystal mixer

First mixer

The antenna consists of a half-wave dipole with a flat, circular reflector mounted at the focus of a paraboloid of 7½ m diameter (see I). It is connected to the first mixer by a coaxial feeder approximately two metres long; the outer conductor is of rigid copper tubing which also serves to support the dipole. The line is air-insulated apart from spacers of foamed-plastic ($\lambda/2$ in length to avoid reflections). The antenna is carefuly matched to the feeder (which has a characteristic impedance of 50 Ω) for two frequency bands, viz. for a bandwidth of several Mc/s around the spectral line of frequency 1420 Mc/s and for a similar band around 1360 Mc/s. The latter is the image frequency with respect to the oscillator frequency of about 1390 Mc/s, which is used in the first mixer. The degree of matching is dependent on the standing wave ratio [2]) which should approximate as closely as possible to unity. An illustration of the matching is given in *fig. 3*. The matching, and hence the output impedance of the antenna - coaxial line system changes only slightly with frequency in the bands concerned.

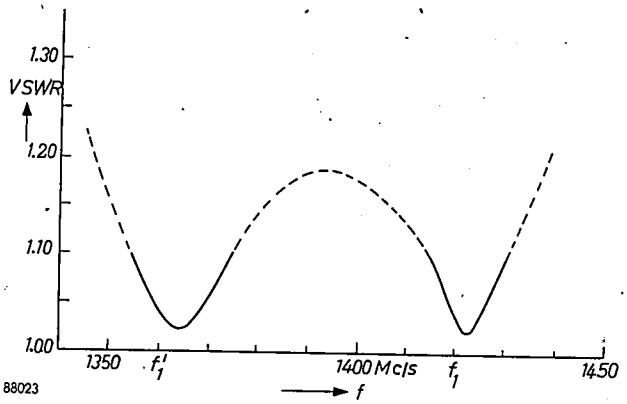


Fig. 3. Matching of antenna to 50 Ω coaxial feeder at frequencies of $f_1 = 1420$ and $f_1' = 1360$ Mc/s (wavelengths 21·12 and 22·08 cm). The degree of matching achieved is expressed by the measured $VSWR$ (voltage standing wave ratio). Perfect matching occurs when $VSWR = 1$.

The weak antenna signal is fed, via a high-frequency matching transformer, to a germanium diode, which is used as the mixer valve. Thus, with no H.F. amplification and no image suppression, the H.F. section is kept relatively simple. Apart from the fact that it is not yet technically possible to achieve any appreciable gain in sensitivity by means of high-frequency amplification, its use in this receiver would in any case be undesirable. The tuned circuits involved would mean that the characteristics of the system would change too rapidly with frequency. Since the operation of the crystal mixer is closely dependent upon the impedance of its external circuit, the result would be to make it much more difficult to fulfil the requirement of equivalence between the two switching positions. This applies not only to the impedance at the signal, image and intermediate frequencies, but also to the impedance at the oscillator frequency and its multiples, and to that at frequencies which differ from these multiples by an amount equal to the intermediate frequency. The latter frequencies are also mixed, resulting in the intermediate frequency.

If the oscillator signal is very strong with respect to the input signal (as it is in this case) the current through the diode is determined exclusively by the oscillator signal. For a sinusoidal voltage across the diode, the current through it consists of unidirectional pulses. The diode current thus contains higher harmonics at multiples of the oscillator frequency. Only if the external circuit constitutes a short-circuit to all these frequencies can the voltage on the diode be purely sinusoidal. If, however, the impedance of the external circuit is infinite to one of the harmonics, the current component concerned

[2]) See J. C. van den Hoogenband and J. Stolk, Reflection and impedance measurements by means of a long transmission line, Philips tech. Rev. **16**, 309-320, 1954/55 (No.11).

must be zero. This means that the operation of the diode depends upon the behaviour of the external circuit at multiples of the oscillator frequency. In our receiver, the two oscillator frequencies remain constant when the entire tuning is changed. Consequently, the matching to the oscillator frequency is less important than to the signal and image frequencies, since a deviation from the requirement of equilvalent switching positions does not depend upon the tuning of the receiver.

The operation of the diode as a mixer stage may be likened to that of a linear network, except that in this case the input and output frequencies are different. The parameters of this network are determined by the form of the current pulse through the diode; they thus depend upon the amplitude of the oscillator signal and upon the impedance of the external circuit to multiples of the oscillator frequency. Regarding the mixer stage as a linear network, it is evident that its response, particularly at signal and image frequencies but also at the other higher frequencies mentioned above depends on the impedances connected to it.

During the switching process the high frequency section itself remains unchanged, but the frequencies in which we are interested do change. Thus, the output impedance of the H.F. section must be as independent of frequency as possible, in order to prevent the switching from affecting the circuit. Selective amplification and image suppression demand tuned circuits, which cannot easily be given a flat band-pass response. That is why neither is employed in this stage of the receiver.

A constant output impedance of the antenna - coaxial feeder system is also required for the region around the image frequency. For this reason the antenna feeder must be well matched to both the signal and the image frequency. An incidental advantage of passing the image frequency is that it facilitates the calibration of the receiver. We shall return to this point later.

It will now be clear why the switching must be effected in the first mixer stage. The equivalence requirement could certainly be fulfilled by the high-frequency section alone, but if the switching were to take place at a later stage, the preceding section of the receiver would have to have a flat band-pass response which is very difficult to achieve in practice. What is more, the effect of the inherent noise in the first amplifying stages would be much greater.

First oscillator

The antenna signal is combined in the first mixer with each of the oscillator signals in turn which are fed to the germanium diode via a capacitative coupling. The oscillator signals of about 1390 Mc/s are derived by frequency multiplication from two crystal oscillators working at 6431 and 6436 kc/s respectively. These frequencies can be varied a few kc/s by means of a variable capacitor connected in series to each crystal. The nominal values are adjusted as accurately as possible with the aid of a frequency calibration circuit.

The output voltages of both oscillators are fed to two separate two-stage amplifiers, which form part of an electronic switch. The amplifiers are alternately cut off by a negative voltage of about 50 V applied to their control grids from a multivibrator which is synchronized by a switching voltage of 400 c/s generated in a special oscillator. Two stages of amplification were found to be necessary in order to avoid capacitative coupling.

The electronic switch is followed by a frequency multiplier consisting of several stages including, inter alia, a type E 03/20 double tetrode. The total multiplication is $\times 216$. The last multiplier stage is a trebler, which works with a disc-seal triode, type EC 55, tuned by a coaxial quarter-wave filter. This stage, together with the first mixer and a part of the first I.F. stage which follows it, is hermetically sealed in a container immediately behind the parabolic reflector. In order to lose as little as possible of the weak antenna signal to the oscillator, the coupling between oscillator and diode is very loose. For this reason it is necessary to raise the available power of the oscillator to as much as 50 or even 100 mW, although only 1mW is needed for the crystal mixer.

The signals finally delivered by the oscillators to the crystal have frequencies, f_0' and f_0'', of 1389.196 and 1390.276 Mc/s respectively, which differ by $\Delta f = 1080$ kc/s; these signals are switched to and fro at a frequency of 400 c/s.

Characteristics of the mixer circuit

The properties of the reciever may be characterized by two important quantities, viz. the conversion loss L and the noise temperature t of the mixer stage. The conversion loss is the ratio of the available antenna signal power at the signal frequency to the corresponding I.F. signal power at the output of the mixer. The value of L with a good crystal should be 5 dB. For reasons already mentioned, the conversion loss is considerably dependent on the frequency characteristics of the external circuit.

The noise temperature, which is really only a temperature ratio, expresses the relationship between the noise power originating from the mixer at an antenna temperature of 300 °K and the noise power of a resistance having the same value as the impedance of the mixer output at the intermediate frequency concerned, both being measured at the I.F. input, i.e. behind the mixer stage. The noise

inherent in the mixer will always be greater than in an ordinary resistance, but with a good mixer crystal the value of t is found to be less than 1·5.

Like L, t depends upon the impedances in the equivalent circuit of the mixer stage, i.e. upon the external circuit, and also upon the output voltage of the oscillator circuit. In principle, therefore, there might be slight differences in L and t (and thus in the receiver characteristics) between the two frequencies which are switched to and fro, resulting in the appearance at the receiver output of a 400 c/s voltage component that is not due to interstellar radiation. It is therefore important to make the oscillator output signal exactly equal in both switching positions. (The use of two I.F. channels reduces somewhat the adverse effect of such a difference, as the extra signals are generally in phase in both channels and are accordingly eliminated. Nevertheless, this in itself is not enough.)

To achieve equal oscillator output a servo system is employed. The appearance of a 400 c/s A.C. component in the diode current is used as an indicator of any difference in the amplitude of the oscillator output voltage. This component is amplified, rectified in a synchronous detector circuit, and finally converted into a 50 c/s alternating voltage which drives a 50 c/s servo motor. (In principle, a 400 c/s servo motor might also have been used; present-day servo-amplifiers are available for this and other frequencies). The servo motor regulates the position of a stub in the last (coaxial) circuit of the frequency multiplier relating to the first oscillator. The time constant of this regulation is fairly large, which reduces circuit sensitivity to interfering voltages of short duration. The servo system keeps the oscillator output power in both positions equal within approximately 0.1%, which is amply sufficient at the present sensitivity of the receiver.

A further characteristic of receiver noise is the noise factor. With reference to the concept of "antenna temperature", discussed in I, the noise factor can be simply defined as follows. With an ideal receiver, i.e. one which itself produces no noise, a change in antenna temperature from room temperature T_0 to twice the value, $2T_0$, would result in twice the noise power at the output. With a non-ideal receiver the noise power at the output increases by a factor of only $(1 + 1/N)$, in which N (>1) represents the noise factor of the receiver. The receiver behaves as if there appeared on the input terminals not merely the signal from the antenna but also a noise power equivalent to a temperature of $(N-1)\ T_0$. The cause is to be found in the receiver itself. The extra noise is called the inherent noise of the receiver, and it arises chiefly in the first mixer and in the first I.F. stage which follows it. Of course, noise is produced in the other stages too, but it is neglible compared with the greatly amplified noise of the first stages.

If N_{IF} is the noise factor of the I.F. amplifier alone, the noise factor of the whole receiver is

$$N = L(t + N_{IF} - 1).$$

Since the noise temperature t for a good mixer crystal is at the most 1.5, it follows that N_{IF} is an important factor. The I.F. input must accordingly produce a minimum of noise. In the receiver under discussion, an N_{IF} of about 1.5 was obtained at an intermediate frequency of 30 Mc/s, which gives a noise factor of $N \approx 6$ for reception at 1420 Mc/s alone. However, the noise factor is also affected by the use of image suppression. In our case it is not used, and the total noise factor, as defined above is only half as large, i.e. $N \approx 3$, noise being received at both the signal and image frequencies instead of at the signal frequency alone. The manner in which the noise factor is measured will be discussed later.

The first I.F. section

As shown by the block diagram in fig. 2, the signal of about 30 Mc/s from the first mixer is fed to the first I.F. amplifier, which has a bandwidth of approximately 5 Mc/s. The reason for choosing such a large bandwidth is that the frequencies at which the measurement is made vary by several Mc/s in the course of traversing the line profile. As stated above, the input of this amplifier is designed to produce the minimum of inherent noise, with a view to obtaining the lowest possible noise factor. For this purpose a "cascode" circuit is used, the principles of which will not be dealt with here [3]). The amplifier consists of two sections: the cascode circuit, together with one of the amplifier stages, is mounted directly behind the parabolic reflector in the hermetically sealed container already mentioned, and the next two conventional amplifier stages are situated in the observation cabin with the rest of the receiver. The sections are joined by a screened coaxial cable.

To the second section of the amplifier a suppressor circuit is connected, the purpose of which is to ensure that the switching process itself has no effect on the measurements. In spite of all the measures mentioned, it was impossible to satisfy this condition

---

[3]) H. Wallman, A. B. MacNee and C. P. Gadsden, Proc. Inst. Radio Engrs. **36**, 700-708, 1948. See also H. H. van Abbe, B. G. Dammers and A. W. G. Uitjens, Noise of the cascode amplifier, Electronic Appl. Bull. **14**, 141-150, 1953.

without using a suppressor, for the following reason. The process of switching produces voltage pulses which manifest themselves throughout a wide band of frequencies and which, owing to the first mixer, also appear with appreciable amplitude in the I.F. band at 30 Mc/s. This would present no problem if switching from the second to the first position produced exactly the same waveform as switching from the first to the second. In that case the frequency of the switching pulses would be 800 c/s, to which the synchronous detector at the end of the receiver is insensitive. In actual fact, however, the pulses are dissimilar, with the result that a 400 c/s component appears at the output of the I.F. amplifier, causing spurious deflection of the recorder. As the amount of this spurious deflection depends on the tuning frequency and on the temperature, it is not constant and is therefore very troublesome during measurements.

The suppressor cuts out the I.F. amplifier just before the moment at which the oscillators are switched over, and switches it on again immediately after. This, ofcourse, introducess witching pulses from the suppressor, but the form of each pulse is now identical. The suppressor consists of a mono-stable "puller" circuit[4]), driven by pulses with a frequency of 800 c/s derived from the 400 c/s switching voltage appropriately shifted in phase. The second section of the first I.F. amplifier remains switched off on each occasion for approximately 50 μsec.

Nevertheless, a 400 c/s component not originating from the antenna might still appear if the spacing of the cut-out pulses is unequal, owing to incorrect adjustment of the 400 c/s driving voltage in the suppressor. However, this fault can easily be corrected, and as it is independent of frequency and thus identical and in phase in both I.F. channels, it does not in the end affect the measurement.

The remaining I.F. stages

The 30 Mc/s signal is fed to a second mixer stage, containing a standard ECH81 heptode, in which it is mixed with the signal from a second oscillator. The frequency of this oscillator is varied continuously between 34 and 39 Mc/s by means of a small high stability tuning capacitor driven via a worm gear by a synchronous motor. In this way the whole profile of the spectral line is traversed.

The frequency of the difference signal leaving the mixer varies between 4 and 9 Mc/s. After further amplification, the signal divides into the two separate I.F. channels via two amplifiers in parallel operating at 4·96 and 6·04 Mc/s respectively. The output signals from these amplifiers are each combined in third mixers with the 5·50 Mc/s signal from a common crystal oscillator. Thus, the resultant difference signals in both channels have the same frequency, viz. 0·54 Mc/s. Each amplifier attenuates the image frequencies by more than 30 dB.

The two 0·54 Mc/s signals are then each fed to 3-stage amplifiers, each having a frequency characteristic determined by three identical resonant elements. The bandwidth of both is 35 kc/s. As already mentioned, these amplifiers determine the bandwidth of the channels; all other amplifiers have much larger bandwidths.

It is not necessary, nor is it desirable, to have flat band-pass characteristics. The present characteristics are approximately Gaussian in form which facilitates frequency adjustment. This is an important point, because the tuning to 0·54 Mc/s must be exact. Actually, the tuning is so stable that it scarcely needs any readjustment. Diode detection follows the last stage in both channels. The D.C. output is taken from the diode load resistor and fed via a low-pass filter to the amplifier for automatic gain control. The latter is necessary because the amplitude of the 400 c/s square wave voltage naturally depends upon the gain of the I.F. amplifier For reasons of simplicity, it is not actually the gain that is kept constant but the average output voltage, which can be tapped off directly from the detector output. Strictly speaking the noise output voltage (i.e. in absence of signal) rather than the average output voltage should be kept constant in order to achieve the desired constant gain. However the errors due to use of the average output voltage are negligible and corrections are unnecessary

In our reciever the A.G.C. works on the last mixer valve and on the amplifier valves preceding it. Its effective time constant is approximately 0·2 sec, so that the 400 c/s component is not affected.

The synchronous detector circuit

The low-frequency alternating voltage across the two load resistors consists of different components. Apart from the 400 c/s square-wave voltage to be measured, it includes the suppressor pulses, which have a repetition frequency of 800 c/s, and the inherent noise of the receiver, in which frequencies from 0 to 35 kc/s are represented. The two 400 c/s components are equal and in antiphase, the 800 c/s pulses are also equal but in phase, while the noise voltages are mutually uncorrelated.

The synchronous voltmeter measures the fundamental of the 400 c/s square-wave voltage and the

[4]) B. Chance, Wave forms, Radiation Laboratory Series No. 19, McGraw Hill, New York 1949, page 181.

amplitude of that part of the noise spectrum which is immediately adjacent to this frequency. Of the entire spectrum only a very narrow band, about $^1/_{200}$ c/s around the 400 c/s switching frequency, passes beyond the diode detectors and is recorded. The detector circuit which achieves this high selectivity functions as follows.

The two 400 c/s components are added in a push-pull input circuit and the two in-phase 800 c/s components thus eliminated. This circuit is followed by a selective amplifier with a bandwidth of about 100 c/s. Its function is to attenuate the odd harmonics of 400 c/s (to which the synchronous detector is also sensitive) sufficiently to make the total sensitivity to these frequencies neglible.

The synchronous detector itself (*fig. 4*) is really a mixer stage, to which is applied not only the 400 c/s signal voltage, but also a relatively large oscillator-voltage with exactly the same frequency.

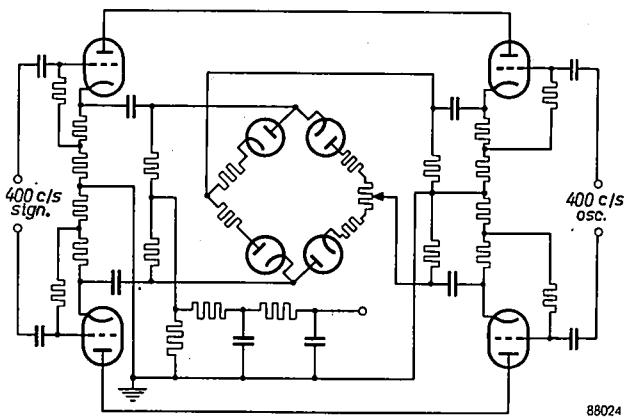


Fig. 4. Simplified circuit of the synchronous detector.

The difference frequency is therefore zero, that is to say the result is a direct voltage proportional to the amplitude of the signal — and to the cosine of the phase angle between the two 400 c/s voltages. The detector proper is a bridge circuit, each arm consisting of a vacuum diode and a 100 kΩ wire resistor in series [5]). The signal and the oscillator voltages are symmetrically applied to opposite diagonals and the D.C. output voltage appears between the centres of two resistance arms across each push-pull cathode follower output (see fig. 4). As the oscillator voltage is large compared to the signal voltage one half of the diode bridge periodically becomes conducting while the other half presents an infinite resistance. The resistance of the conducting branch is determined mainly by the wire resistors, so that differences in the diode characteristics have very little effect. An examination of the current flow in the bridge shows that, if the two voltages are in phase, a D.C. voltage appears across the output.

The fundamental of the 400 c/s square-wave voltage having thus been filtered out, the output D.C. voltage is fed to a low-pass filter with a bandwidth of about 0·005 c/s, which determines the bandwidth of the voltmeter. The filter consists of a double RC network, which acts as an almost critically damped filter. The time constant of the filter is inversely proportional to the bandwidth, and may therefore be used instead of the bandwidth to describe the characteristics of the filter. The time constant, as defined by Nechleba [6]), is 54 secs in the present case.

The filter is followed by a D.C. amplifier which is used purely and simply as an impedance transformer: the amplifier input does not load the low-pass filter. The output impedance of the amplifier [7]) is about 1 Ω, which is lower than the input impedance of the recording meter.

The narrower the bandwidth of the circuit the less will be passed of the interfering noise spectrum. However, this also necessitates slower variation of the second oscillator, because the tuning rate, expressed in I.F. bandwidth per second, determines the smallest admissible bandwidth of the voltmeter. It can be shown that at a tuning rate of n bandwidths per second the cut-off frequency of the low-pass filter must lie at about n c/s.

If the intensity of the spectral line is measured as a function of the frequency with a band-pass filter of bandwidth Δf (in our case 35 kc/s) the measured intensity will always be an average of the "true" intensity distribution over this bandwidth. This implies that details in the "true" intensity distribution, whose width is narrower than the bandwidth, are not resolved, or are observed only in an attenuated and distorted form.

If we analyse this further, assuming the bandpass curve to be sinusoidal (*fig. 5*), we find that, at a tuning rate of n bandwidths per second, hardly any frequencies equal to or greater than n can occur in the output voltage behind the synchronous detector. The low-pass filter does not, therefore, require a bandwidth larger than n. However, if this bandwidth were much narrower, the highest frequencies in the output voltage would not be passed, so that the profile observed behind the filter would look as if the bandwidth of the I.F. band-pass filter were larger. In our receiver, the bandwidth of the low-pass filter has been chosen such that its 3dB cut-off frequency (i.e. the frequency at which the power passed has fallen to one half of its value) just coincides with the above-mentioned limit of n c/s. In this case the low-pass filter has hardly any influence upon the form of the observed line profile. The only noteworthy effect is that the profile is subject to a delay during its passage through the low-pass filter. This delay is equal to the time constant of the filter (54 sec) so that in the

[5]) Cf. a discussion of the ring modulator by F. A. de Groot and P. J. den Haan in Modulators for carrier telephony, Philips tech. Rev. **7**, 83-91, 1942.

[6]) F. Nechleba, Elektrotechn. Z. **74**, 98-101, 1953.

[7]) C. Morton, Electronic Eng. **25**, 312, 1953.

receiver under consideration the intensity recorded on the meter at any given moment refers to the frequency to which the receiver was tuned 54 seconds earlier.

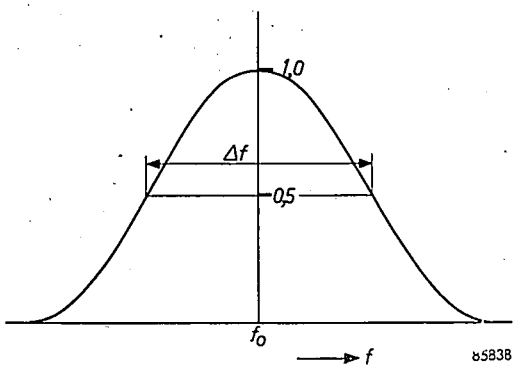


Fig. 5. Sinusoidal band-pass response of the I.F. filter.

The bandwidth of a filter with a rectangular band-pass characteristic must be considerably larger, since the intensity of the voltages with frequencies above n c/s is then no longer neglible. However, these frequencies are not very important in the measurement of the profile as they are considerably distorted with respect to the true profile. (Between n and $2n$ there is a phase shift of exactly 180° with respect to the true profile). Quite apart from these considerations, a rectangular band-pass characteristic is more difficult to achieve in practice than a sinusoidal one; for this reason the sinusoidal characteristic adopted in the present receiver is to be preferred.

The theoretical sensitivity of the receiver

By theoretical sensitivity is meant the smallest difference in antenna temperature between signal and comparison channels that is still theoretically measurable. It depends upon the method of measurement and upon the noise factor, the I.F. bandwidth and the time constant of the voltmeter. The determining factor for the sensitivity is the magnitude of the noise fluctuations during the recording of the line profile; their amplitude can be directly expressed in units of antenna temperature i.e. in degrees Kelvin. If we assume, according to accepted usage, that the smallest measurable change in antenna temperature is equal to the effective value of the noise fluctuations, then the theoretical sensitivity is equal to the theoretically calculated effective value of the noise fluctuations, expressed in degrees K.

It should be noted that the smallest measurable difference in antenna temperature between the two channels is, in reality, dependent not only upon the magnitude of the noise fluctuations but also upon the time duration of such a difference. If the antenna temperature first has a substantially constant value for a period of time that is long with respect to the time constant, and if it then quickly moves to another constant value, the smallest measurable difference will indeed be approximately equal to the effective value of the noise fluctuations. For a difference of only short duration, however, the sensitivity will be lower, because there is a greater chance that the deflection on the meter was not caused by a difference in antenna temperature but by an incidental noise-fluctuation, which may well be greater than the average fluctuation.

It will be sufficient for our purpose to consider only the effective value of the noise fluctuations. Expressed in degrees K it is given by:

$$\Delta T = 1{\cdot}03 \frac{\pi}{\sqrt{2}} (N-1)\, T_0\, (F \Delta f \cdot \tau)^{-\frac{1}{2}},$$

where N is the noise factor for simultaneous reception at signal and image frequencies, $T_0 = 293$ °K, Δf is the bandwidth of the I.F. amplifier, τ is the time constant of the low-pass filter, and F is a factor determined by the shape of the I.F. band-pass characteristic.

A similar formula was first derived by Dicke [8]) for a receiver with one I.F. channel and a square law detector. The theoretical sensitivity of our receiver with two I.F. channels is superior by a factor $\sqrt{2}$, which may be simply explained as follows. The addition of the voltages behind the two diode detectors doubles the magnitude of the 400 c/s signal voltage and increases the noise voltage by a factor of only $\sqrt{2}$, there being no correlation between the two noise voltages. The signal-to-noise ratio is therefore $\sqrt{2}$ times larger. Taking into account that the spectral line is received only at 1420 Mc/s and not at the image frequency, 1360 Mc/s, and that a linear diode detector is used instead of a square law detector, which reduces the sensitivity by 3%, the above formula can be derived directly from that of Dicke.

In the network used, which consists of three identical resonant circuits, $F = 1{\cdot}53$, $N \approx 3$, $\Delta f \approx 40$ kc/s (as regards the fluctuations of inherent noise, it is better to use this value of the 3dB cut-off frequency than the earlier-mentioned bandwidth of 35 kc/s) and $\tau = 54$ secs. The theoretical sensitivity is thus:

$$\Delta T = 0{\cdot}73 \text{ °K}.$$

If the theoretical sensitivity of this receiver is compared with that of a receiver in which no switching takes place, the present receiver is found to be inferior by a factor of 1·57. This factor is mainly attributable to the fact that the present receiver records the difference in antenna temperature in signal and comparison channels (with the same bandwidth), the uncorrelated noise voltages being quadratically added, whereas the ideal receiver records only the increase of antenna temperature in one channel. The sensitivity thereby deteriorates by a factor of $\sqrt{2}$; the remainder, a factor of 1·11,

[8]) R. H. Dicke, Rev. sci. Instr. 17, 268-275. 1946.

is concerned with the synchronous detector. The loss in sensitivity resulting from the method of frequency comparison is amply offset by the much greater stability obtained. In practice, the zero stability during recording decides the maximum sensitivity obtainable with a receiver. From the formula given for ΔT it would appear that, in principle, it is possible to achieve any sensitivity required by increasing the time constant (or the I.F. bandwidth, which is not desirable in this case, however, as it entails the loss of details in the spectrum). However, a corresponding reduction in the tuning rate is then necessary and this prolongs the measurements. The real limit is therefore set by the zero stability, which is subject to a variety of factors such as fluctuations in the supply voltages and changes in ambient temperature. The sensitivity of the receiver while it was located at Kootwijk was restricted by interference from neighbouring transmitters to a limit of about 1 °K. This means that variations in signal intensity amounting to less than 0.1% of the noise inherent in the receiver could still be recorded. The performance of the receiver in its new and quieter location at Dwingelo is not yet known. A substantial improvement in sensitivity is to be expected.

Frequency and intensity calibration

The various frequencies generated in the receiver itself must be kept strictly constant if the measurement of the line profile is to be reliable. This applies particularly to the oscillator frequencies. The switching frequency need not be exactly constant, since the electronic switch and the tuned detector operate at the same frequency, derived from the same oscillator.

The accuracy required is 1 in 10^6, and the frequency scale on the recorder is therefore accurate to within 1·4 kc/s. The frequency calibration is carried out in three stages (*fig. 6*):

1) Adjustment of the two crystal oscillators in the first mixer circuit.
2) Calibration of the slowly varying frequency of the second oscillator.
3) Calibration of the pass bands of both I.F. amplifiers behind the third mixer to a width of 35 kc/s.

Kootwijk Radio Station generates standard frequencies of 1 to 10 kc/s with an accuracy of 1 in 10^7; they are derived from the frequency standard adopted by the Netherlands Post Office in the Hague the accuracy of which is better than 1 in 10^9. These frequencies come within the range of the present receiver and are therefore used as a primary standard for calibrating the secondary frequency standard in the receiver itself. The secondary standard is a 100 kc/s oscillator, whose driving crystal is contained in a thermostatted enclosure and the frequency it generates is compared on an oscilloscope with the primary calibration frequency. The accuracy of the calibration is about 1 in 10^7.

The frequencies of the oscillators in the first mixer stage should, after being trebled, be 19.308 and 19.293 Mc/s respectively. A frequency of 19.300 Mc/s is now derived from the secondary standard (by multiplying it by 193) and a voltage of this frequency is mixed with the two signals mentioned. The difference frequencies of 7 and 8 kc/s can be directly compared on an oscilloscope with the primary standard of 1 kc/s.

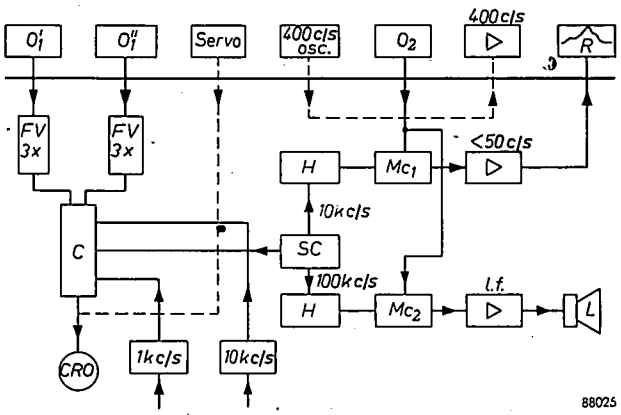


Fig. 6. Simplified block diagram of the frequency calibration circuits (full lines) and control circuits (broken lines). The units above the horizontal line are parts of the receiver proper (see fig. 2). *FV* frequency multipliers, *C* calibration panel, *KO* cathode ray oscilloscope, *SC* secondary frequency standard, *H* harmonic generators, Mc_1 and Mc_2 mixer circuits, *L* loudspeaker.

The second oscillator varies slowly in frequency, and a different method of calibration has to be adopted. The secondary standard is divided in order to obtain a frequency of 10 kc/s, and multiples are then derived from this with frequencies between 33 and 39 Mc/s. After amplification, this entire group of harmonics is mixed with a voltage having the varying oscillator-frequency, one result of which is the appearance of a difference frequency somewhere between 0 and 5 kc/s. This signal is fed to the recording meter via an L.F. amplifier with a pass band of only 0·5 to 50 c/s. Only if the difference between the oscillator frequency and one of the multiples (all are multiples of 10 kc/s) is smaller than 50 c/s will a signal be perceptible at the output of the receiver. After rectification, this signal actuates a relay which short-circuits for a moment the recording meter and causes the pen to fall towards zero. Thus, every time the oscillator frequency

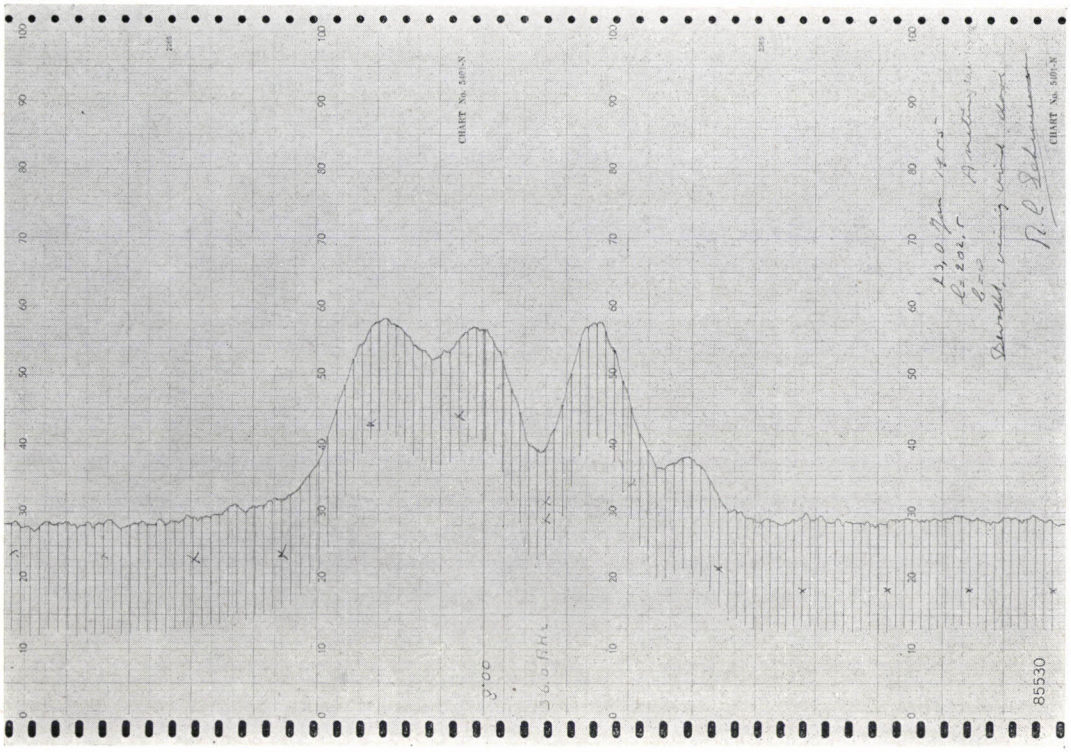


Fig. 7. Example of a recording. The vertical marks are at frequency intervals of 10 kc/s. Multiples of 100 kc/s are marked with a cross. This recording was made on the night of 23rd January 1955, for radiation from a point in the sky given by $l = 202{\cdot}5°$, $b = 0$. At this bearing several maxima are observed (implying spiral distribution of the interstellar hydrogen).

passes a multiple of 10 kc/s, a mark is automatically made on the strip chart (*fig. 7*).

A similar mark is made at multiples of 100 kc/s or 1 Mc/s is passed (with the aid of specially generated multiples of these frequencies), but they are also made audible via a loudspeaker. The observer can thus indicate on the recording what marks correspond to what multiples.

The band-pass characteristics of the two I.F. amplifiers following the third mixer are automatically recorded in the receiver with the aid of a multiple of the 100 kc/s secondary standard which lies in the proximity of 30 Mc/s and which can be fed to the first I.F. amplifier. The pass-band has proved to be very constant and need be checked only once a week at the most.

Intensity calibrations are only relative measurements carried out during recordings over several months. A daily measurement is made of one part of a line profile received from a fixed point in the sky. This profile is taken as the standard with which the intensity of all other profiles measured is compared. To determine the absolute intensity scale it is therefore only necessary to calibrate now and then the intensity of the standard profile.

Direct absolute calibration with an artificially obtained spectral line is not possible, as no noise source is available that emits radiation in a spectral line and whose intensity can be determined independently of a radio wave measurement by using a well-known physical relation (in the way, for example, the radiation from a black body can be ascertained by measuring the temperature of the body). Absolute calibration therefore has to be done indirectly by calibrating the various receiver-constants concerned, such as the noise factor, the rectification characteristic of the two diode detectors and the sensitivity of the 400 c/s voltmeter. The diode characteristic is easily determined by connecting a noise diode instead of the first mixer to the 30 Mc/s I.F. amplifier. The 400 c/s voltmeter can be calibrated quite simply by means of a known 400 c/s voltage.

Several factors must be taken into account, however, when measuring the noise factor, Since no image-suppression is employed, the noise factor ascertained with a given noise source always holds for simultaneous reception on signal and image frequencies. Now for measuring the spectral line, we are concerned only with reception at 1420 Mc/s, so that the sensitivity ratio at the two frequencies must be found in another way. The method used here is to make the sensitivity at both frequencies approximately equal by increasing the bandwidth

of the mixer stage. It remains to be seen how sensitive the receiver will be to the odd multiples that arise during mixing. (A short-circuit is used for the even multiples in the mixer stage and in the antenna). It is to be expected that the noise power received at these frequencies will amount to less than 10% of that received at the signal frequency. Since these properties have not yet been determined in detail, the antenna temperature scale in use at present is only temporary and may conceivably involve an error of 20%.

For determining the noise factor earth radiation is used as the noise source. For this purpose the antenna was first directed towards the zenith and then upon a pine-wood situated on a slope to the north-east of the antenna site. The continuous radiation received was found to increase by 50%. If we assume that the earth emits radiation like a black body with a temperature of 300 °K, while the antenna temperature in the other position (zenith) amounts to only a few °K, say 5 °K, then the difference in antenna temperature between the two positions will be 295 °K, provided that the pine-wood radiation fills the whole antenna pattern. If it does not, the temperature difference will be smaller, which means that the real noise-factor will be lower than the value found by measurement. This method, therefore, allows only the maximum value of the noise factor to be determined, which in our case will not differ much from the real value. Thus, for simultaneous reception at 1420 and 1360 Mc/s, we find $N_{\max} = 3.0$; for reception at 1420 Mc/s alone, assuming equal sensitivity, the noise factor is then $N'_{\max} = 6.0$.

From the temporary intensity scale calibrated in this way the highest value of the spectral line temperature measured was 118 °K.

Power supply

The entire receiver is fed from the electric light mains. The mains voltage is kept constant to 0·5% by a voltage stabilizer of 2 kVA. The high tension is taken from a number of electronically stabilized supply units of the same design, incorporating a neon tube type 85A1 or 85A2. Further stabilization is unnecessary in this receiver since, with the method of switching used and the automatic gain ocntrol, the operation of the measuring circuit is largely independent of the supply voltages.

The receiver discussed in this article is the result of several years of gradual development, periods of improvement having alternated with prolonged periods of continuous measurement. In the form described, the receiver has been in almost uninterrupted operation from August 1954 to August 1955. As regards sensitivity it has come up to all expectations. Zero stability, too, is excellent, so that reduction of the observations to give the line profile presents no special problems. Although the receiver contains more than 200 valves, scarcely any trouble has occurred during operation. In this respect it may be noted that the receiver is left switched on during intervals between recordings and that the valves in general operate far below their maximum anode dissipation. Another favourable factor is the stabilization of the supply voltage. All valves are inspected once every six months, when 10 to 20 valves usually have to be replaced.

The equipment is operated mainly by students of Leiden University, most of whom are not specialists in electronics. After a day or two of instruction, however, they can safely be left to carry out the measurements without supervision.

Summary. This article deals with the design of a receiver for radio waves with a frequency of 1420 Mc/s which are emitted by interstellar hydrogen. In order to be able to distinguish the very weak signal from the noise inherent in the receiver, a method of frequency comparison is employed. The receiver is switched to and fro at a rate of 400 times per second between two frequencies, one lying on the spectral line and the other adjacent to it. The spectral line is traversed by slowly varying both frequencies simultaneously while maintaining a constant difference of 1080 kc/s between them. Two separate I.F. channels are used in order to keep the receiver tuned to the spectral line all the time (instead of only half the time, as would be the case if only one I.F. chamnel were used). Each channel has its own comparison frequency on opposite sides of the measuring frequency, and each is tuned to the line in turn. The receiver must satisfy two primary requirements: its characteristics must be identical in both switching positions, and the process of switching must in no way affect the measurement. To fulfil the first condition, the high-frequency section is of simple design and is carefully matched to the first mixer stage at the signal and image frequencies. The oscillator frequency of this stage springs to and fro at a rate of 400 c/s between two settings 1080 kc/s apart. The output voltage of the oscillator circuit is made identical in both positions by means of a servo system. The second condition is fulfilled with the aid of a special suppressor circuit, which cuts off the receiver for a moment during the switchover. Some stages of amplification are then followed by the second mixer, the oscillator frequency of which is slowly varied for the purpose of traversing the spectral line. Division now takes place into the two I.F. channels mentioned, which are each provided with automatic gain control. These two channels determine the ultimate bandwidth of the receiver. After further mixing and subsequent detection the two signals are fed to a push-pull amplifier, where they are added together and where unwanted voltages are, as far as possible, eliminated. The fundamental frequency of the 400 c/s square-wave voltage, which is a measure of the intensity of the interstellar radiation, is rectified in a tuned detector circuit and, after emerging from a low-pass filter of only $^1/_{200}$ c/s bandwidth, it is fed to a strip-chart recorder. The frequencies generated in the receiver and the bandpass characteristics of the I.F. filters, which must be identical and constant in both channels, can be verified and adjusted by means of various semi-automatic calibration circuits.

The article finally deals briefly with the noise characteristics of the high-frequency section and of the first mixer, and discusses the factors that determine the sensitivity of the receiver. The sensitivity achieved is better than 0.1% of the inherent noise. In terms of antenna temperature, this means that it is possible to detect temperature differences down to about 1 °K.

SCIENCE GOES TO THE FAIR

by G. REMEDI.

53:79

The post-war growth of Philips has been accompanied by a continual increase in the personnel of the Research Laboratories. From 1946 to 1956 the staff increased from about 700 to about 1300. To provide accomodation for these growing numbers the laboratory has been repeatedly extended. The fortieth anniversary of the laboratory in June 1955 saw the completion of a new wing. These events were celebrated at a specially organized fair which, on account of its scientific nature, deserves some description in this review. The author of this article is one of the leading figures among the organizers, writing under a nom-de-plume.

Scientific research has been compared to a game and even in an industrial laboratory daily activities often have something of this character. One of the most interesting aspects of this game with Nature is that as it progresses both the strategy and the equipment must be continually revised and adapted to the changing situation.

The impact of this element of play on the staff of the Research Laboratory found another outlet on the 2nd of July 1955, when the entire laboratory personnel and many former members, their wives and fiancées, celebrated the fortieth anniversary of the laboratory and the official opening of a new wing (*fig. 1*).

For this monster celebration (2800 people were present) a large hall in Eindhoven was converted into a Brabant village square on which a "scientificfair" was set up. The title photograph, taken during its construction, gives an impression of the fair. The side-shows were prepared by the various groups of the laboratory, each contribution bearing some reference to the daily work and equipment of the relevant group. Relays of laboratory personnel took it in turns to run the fair so that everyone had an opportunity to enjoy it as visitors. In addition to the sideshows there was a general programme with dancing, competitions, a ballet and an opera with well-known arias but a new libretto appropriate to the occasion.

A "glass" dance-floor was made by a method commonly used for the construction of film scenery. Strong paper is stuck on a flat foundation and painted with coloured motifs. A few layers of water-glass, painted-on, provide both an ideal surface and excellent durability (see title photograph). The durability of the paper and other materials was first tested on a small scale in the laboratory with a thoroughness typical of the enthusiasm with which all the preparations were tackled. Advantages of a dance-floor of this type are its glossy and colourful

appearance which breaks up the monotony of a large expanse of floor, its adaptability as regards shape and extent and the low price. The life of such a floor can be easily extended, if need be, simply by applying a new layer of water-glass.

The musical background for the event was provided by a professional dance-orchestra, a barrel organ and the laboratory's own band and, in addition, an "electronic orchestra". The instruments of the latter, designed and played by members of the acous tical group, each have their own amplifier and loudspeaker, together covering the whole audible frequency range and giving an output loud enough to be heard everywhere above the din of the fair. The melody section was made up of four non-polychromatic instruments. The soprano and tenor parts were played on two multivibrators, the pitch being controlled by variable resistors. The alto part was supplied by an A.F. generator, played by means of the frequency control knob whilst a low-frequency *RC*-oscillator was used for the bass. The rhythm section included kettle drums in the form of two large capacitor-microphones, the diaphragms (8″ and 6″) being played with the fingers like drum-skins. In addition, there was a pulsed noise generator that acted as cymbals, and a device imitating a bass drum. In the circuit designed for this drum, a capacitor was discharged across an *LC* circuit; the amplifier was provided with positive feedback to give the required resonance (*fig. 2*). An electrical clavichord completed the ensemble. Its strings were stretched tightly across a metal plate, forming a

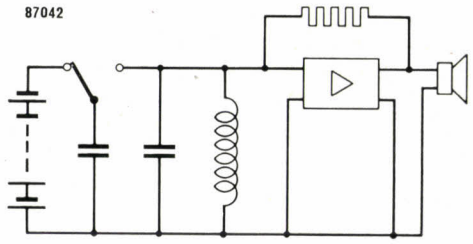


Fig. 2. Circuit diagram of the electronic bass drum.

series of capacitors, so that, when a string vibrates, an electrical signal is produced. Just in the same way as on a normal clavichord it was possible to play polyphonically and even to reproduce the characteristic vibrato of a clavichord. The volume could be controlled electrically by a pedal as well as by the touch on the keys. Visitors were particularly fascinated by the apparent ease with which the members of the orchestra produced so much volume and richness of tone.

Points were awarded in both the general competitions and in the competitive side-shows. Prizes were awarded according to the total number of points which each participant had collected. With the help

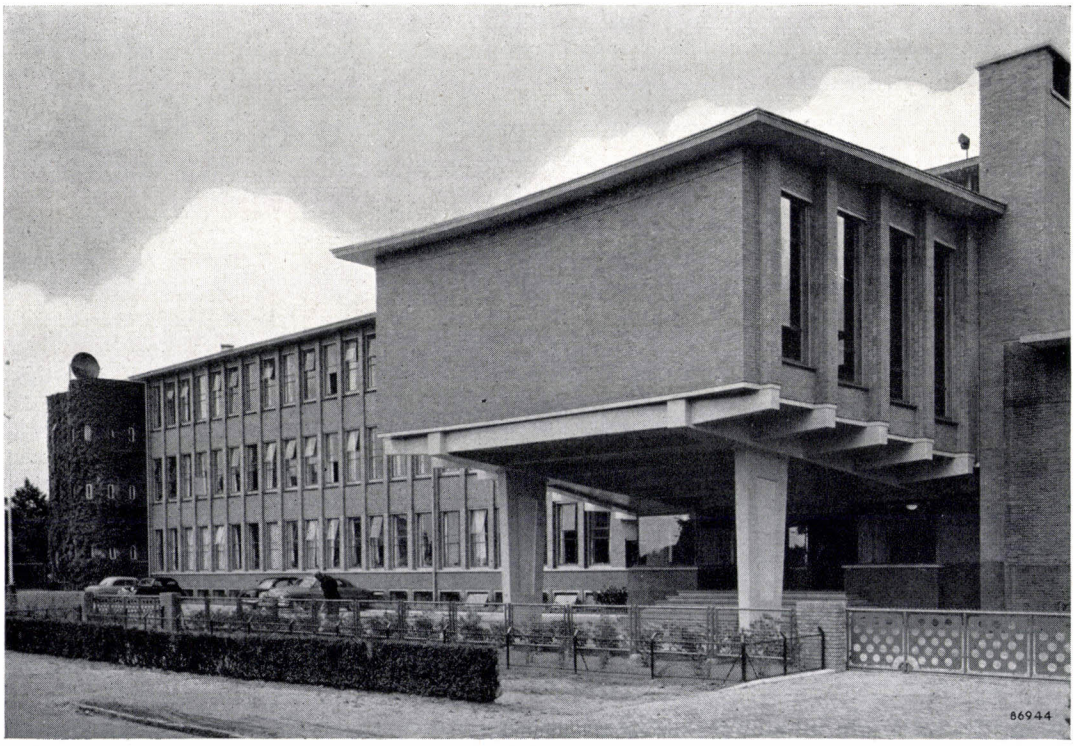


Fig. 1. The new wing of the Research Laboratories at Eindhoven, the official opening of which on 2 July 1955 was celebrated by the scientific fair described in these pages.

of the statistical department the system of allocating points was so arranged that, at the higher end of the scale, the chance of two or more persons ending up with the same number of points was very small.

A number of original contrivances were prepared for the main competitions. Among them were the "Hariguars", 3-wheeled racing cars equipped with independent front wheel hand drive with a different transmission ratio for left wheel and right wheel, and with a swivelling wheel at the back. To drive in a straight line in this vehicle requires considerable skill.

During the rehearsal of the opera, it was found that a well-nigh perfect performance could be ensured by making use of the methods of modern electronics. Because of the poor acoustics and the size of the hall it was necessary to reproduce the dialogue and the arias through amplifiers to make them intellegible. Direct reproduction has the disadvantage that the players are more less or bound to the microphone. This was avoided by recording the whole opera (with the exception of the overture which was played *au naturel* by the wind-orchestra) on magnetic tape, the players (without microphones) merely singing and acting synchronously with the tape during the actual performance. It is remarkable how word-perfect the performers can be by this technique!

The stalls at the fair were of two types: those with a competitive element where points could be won towards the prizes and those sideshows which were merely for entertainment. As already mentioned, the stalls were run by members of the laboratory staff in rotation; each person spent about one hour on duty and in this way more than 1000 fair-goers took an active part in the success of the fair. In all there were about 60 stalls, many of which were completely novel in character. To give some idea of the nature of these scientific playthings, technical details of some of them are given below.

Rubber membrane races (fig. 3). Races between miniature motorcycles on the taut rubber membrane used in the laboratory as an analogue for problems of electronic tubes and wave guides [1]). One end of the rubber membrane is made to vibrate vertically by means of a motor-driven eccentric so that a standing wave is set up on the membrane. The positions of nodes and antinodes depend upon the frequency, which is controlled by a "frequency twist-grip" on a pair of motorcycle handlebars. A miniature motorcycle is placed on the membrane and is so arranged that it moves forward when it is in a place where the membrane is moving. The forward speed is greatest at the position of the antinodes whilst the motorcycle remains still at the nodes. In order to reach the finishing line as quickly as possible it is necessary to vary the standing wave pattern by altering the frequency whenever the motorcycle reaches a node.

Child's play, or blowing up balloons with liquid air. Air taken from the room is liquified by means of a gas refrigerating machine [2]) and flows continuously into a Dewar-flask. The liquid is pumped out of the flask and up to the desired level by means of a vapour bubble pump. A deflated balloon fitted with a plastic funnel is then held under the stream of liquid air so that the liquid air runs into the balloon,

Fig. 3. The "Rubber membrane races". On the extreme left and near the centre of the photograph can be seen two miniature motorcyclists on the taut rubber membrane. The two people seated behind the membrane each grasp the frequency twist-grip on a pair of motor-cycle handle-bars. By controlling the frequency, the motorcyclist can be made to move forwards.

after which the funnel tube is closed tightly with a stopper. The liquid air in the balloon vaporizes and inflates the balloon; however, the rubber of the balloon wetted by the liquid air is hard and cannot stretch. Contestants must then try to warm up these cold hard regions with a jet of air quickly enough to prevent the balloon bursting. The larger the balloon, the more points gained.

Philirose. This is a shop window dummy with a glass head, the glass being made locally conducting by a coating of tin oxide (*fig. 4*). Current is supplied to these regions through silver strips which are

[1]) K. S. Knol and G. Diemer, A model for studying electromagnetic waves in rectangular wave guides. Philips tech. Rev. **11**, 156-163, 1949/50. Further articles on the application of the rubber membrane can be found in Philips tech. Rev. **2**, 338-345, 1937 and **14**, 336-344, 1952/1953.

[2]) J. W. L. Köhler and C. O. Jonkers, Philips tech. Rev. **16**, 69-78 and 105-155, 1954.

connected to a variable transformer via connecting terminals concealed in the hair. The head is coated with a paint containing Ag_2HgI_4. A handle is provided which via a train of gears rotates a governor; the latter operates the variable transformer and thus the current through the conducting

Fig. 4. Mounting the head of "Philirose". As the hair is still missing, the connecting terminals for the current supply wires can be seen on the head.

glass. The heat produced makes the dummy's cheeks blush by changing the colour of Ag_2HgI_4 in the paint, which is bright yellow below 50° C and orange-red above. In addition, the brown-painted pupils dilate into a red iris by a change in the colour of Cu_2HgI_4 from bright red to dark brown at 70° C. If the handle is turned too rapidly the governor switches the current off; the handle has therefore to be turned at a particular rate to get the quickest results.

Potentiometer. Quite unlike the instrument normally associated with this word, this contrivance pretends to be the substantiation of a literal interpretation of the word — a test of strength. It consists of a bicycle on a stand, a copper disc acting as the rear wheel, which can be braked by an electromagnet (eddy current brake). This bicycle is actually used in the laboratory for physiological tests. Brute force, however, is of little use on this "potentiometer". Instead, patience and attention are required. Against the copper disc runs a bicycle dynamo which generates an A.C. voltage having a frequency proportional to the speed of the disc. This A.C. voltage is supplied via an amplifier to one of the windings of an induction motor. The other winding, perpendicular to the first, is supplied from the mains. The motor can turn only if the frequency of the voltage supplied by the bicycle dynamo is 50 c/s; and the direction in which it then turns depends on whether this voltage is leading or lagging in phase with respect to the mains voltage. By means of a reduction gear, the motor drives an endless cord on which is fastened a miniature racing cyclist who cycles over a course in a painted mountain landscape. It is quite feasible to keep the frequency practically constant at 50 c/s with pedalling but it is impossible to control the phase so that the cyclist moves mainly forwards. For this reason a slip coupling is incorporated in the transmission from motor to cord which ensures that cycling forwards is easier than cycling backwards; with this modification there is a fair chance of making the cyclist reach the finishing line within $1\frac{1}{2}$ minutes. When the cyclist comes to the foot of a mountain the brake magnet is energized and the eddy current brake comes into action. As the slope becomes steeper the excitation becomes stronger and with a little imagination the competitor has the feeling of struggling up a mountain.

Knijp van Jut, or Test your Clench. Two players sit at opposite sides of a table each operating a hand dynamo [3]). The hand dynamo supplies an A.C. voltage whose frequency is proportional to its speed of rotation. This A.C. voltage is fed to a low-pass filter and is then rectified (*fig. 5*). Since the amplitude of the A.C. voltage also increases with the frequency, a maximum D.C. voltage is obtained if the hand dynamo is operated at such a rhythm that the frequency lies just below the cut-off frequency of the filter. The difference ΔV of the D.C. voltages produced by the two

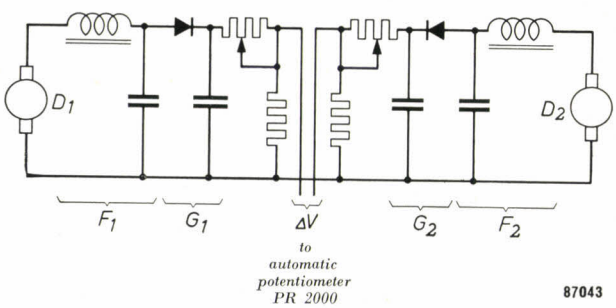


Fig. 5. Circuit diagram of the "Knijp van Jut" (Test your Clench). D_1, D_2 are hand-dynamos; F_1, F_2 low-pass filters and G_1, G_2 rectifiers.

opponents is fed to the input of an automatic recording potentiometer (type PR 2000) [4]). The servo drive of the latter turns a horizontal arm about a fixed vertical spindle, the free end of the arm moving towards the competitor producing the high-

[3]) E. A. van IJzeren, The dynamo pocket torch, Philips tech. Rev. **8**, 225-230, 1946.

[4]) H. J. Roosdorp, An automatic recording potentiometer for industrial use. Philips tech. Rev. **15**, 189-220, 1953/1954.

Fig. 6. "Knijp van Jut" (Test your Clench). This photograph was taken not at the scientific fair but during a TV transmission of the Philips "Electronische Tienkamp" ("Electronic Decathlon") at the E 55 exhibition held in Rotterdam in 1955.

Fig. 7. The "Senilograph" (photograph taken at the E 55).

est voltage. If one of them succeeds in turning the arm completely round to his side a lollipop fastened on the arm lands in his mouth and he is the winner (*fig. 6*).

Senilograph. This competition involves beating with the fists on a table, under which is fitted an electronic counting meter which records the number of sufficiently strong blows within a stipulated length of time (*fig. 7*).

Phililoei — a contest in screaming power. Two amplifiers (see *figs. 8* and *9*) amplify the signals from two microphones each of which is held by one of the contestants. The A.C. voltages supplied by the amplifiers are rectified and each charges a condenser via a resistor whereby the voltage of the condenser gradually increases. When the condenser, for example C_1, reaches a potential of 125 V, a neon lamp N_1 connected across it ignites and the condenser begins to discharge via a relay in series with N_1. This closes contact K_1 which causes C_1 to discharge rapidly through the coil Rel_1 of a multi-contact switch W which thereby changes one position. When the voltage of C_1 drops below the working voltage of the neon lamp, K_1 re-opens and C_1 begins to re-charge. The multi-contact switch has two opposed coils corresponding to the two micro-

phones, which cause the switch to rotate in opposite directions. A row of lamps L signals the position of the switch, giving visible indication of the louder and more persistent scream. It turned out that the greatest difficulty in this competition was that competitors had to retain sufficient self-control not to burst out laughing and so become short of breath.

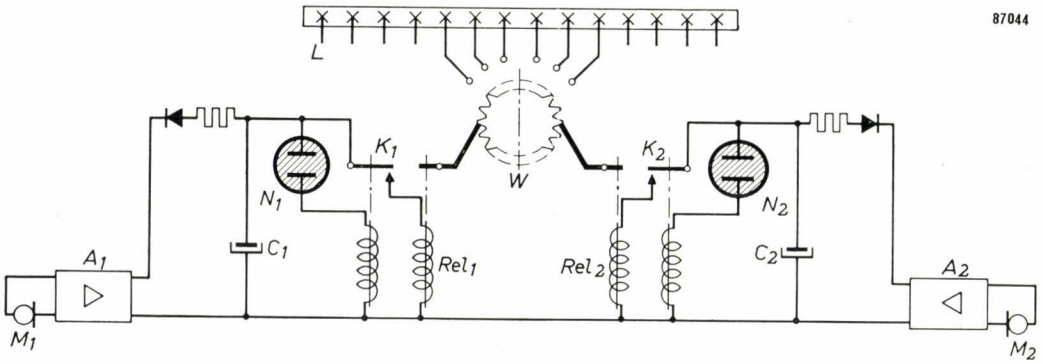


Fig. 8. Circuit of the "Phililoei". M_1, M_2 are microphones; A_1, A_2 amplifiers; C_1, C_2 condensers; N_1, N_2 neon lamps 85 A 2; K_1, K_2 relay contacts; Rel_1, Rel_2 exciting coils of the multi-contact switch W and L indicator lamps.

Magnoball — magnetic table football. The "ball", in the form of a vertically magnetized disk of ferroxdure [5]) (diameter 1″, height $^1/_2$″), lies between two horizontal sheets of glass (about 30″×15″) with just sufficient clearance to slide between them.

Fig. 9. "Phililoei" (photograph taken at the E 55). The two opponents each have a microphone in their hand. Above their heads is a row of indicator lamps.

The two players have each a disk in their hand identical to the "ball". One player holds his disk underneath the lower sheet whilst the other holds his on the upper sheet, each in such a way that both disks exert a repelling force on the "ball" (*fig. 10*). The field of play is bounded by a frame of ferroxdure, magnetized vertically in the same sense as the ball: consequently the ball never rests against the wall but is always repelled back into the field. The ferroxdure frame is absent at the two goals and each player tries to steer the ball into his opponent's goal. This game is surprisingly fast and requires great agility because of the somewhat unpredictable movements of the ball due to the action at a distance. Goals are shown electrically.

All the above mentioned attractions are of a competitive nature. We may now mention a few attractions which were organized just for entertainment.

Nilando the calculating genius. The "genius" was in

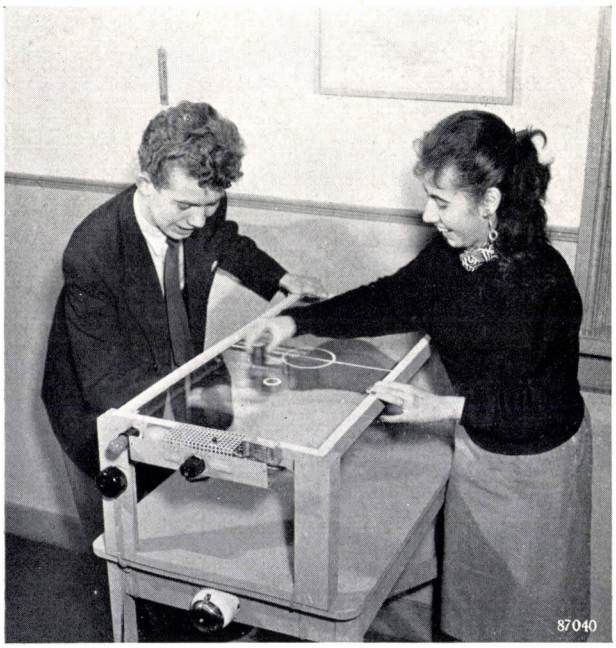


Fig. 10. "Magnoball", a magnetic table football game for two persons.

[5]) Also known as Magnadur. See Philips tech. Rev. **13**, 194-208, 1951/1952.

touch with a computing room by means of an induction loop [6]) and a concealed hearing aid. Consequently he was able to solve the most difficult problems in the twinkling of an eye.

Boffin mirror. An electronic distorting mirror in which people saw their own image, picked up by a TV camera en reproduced with adjustable distortion on a TV screen.

Lotty — the right type. Lotty was a tele-typewriter used in the laboratory for printing the output from an electronic computer. At the fair the orders were fed into it in the form of punched teleprinter tapes. The tapes were prepared beforehand and contained answers to various questions which could be put to Lotty. The tape containing the answer was fed into the instrument by a concealed accomplice, so that Lotty typed out the correct answer. By suitable arrangements of characters and spaces she was able to produce caricatures of well-known personalities (*fig. 11*).

As may be gathered from the photographs shown here, the attractions of the fair also did service on a subsequent occasion, namely at the Dutch exhibition "Energy, 1955" held in Rotterdam last year. A number of the competitive sideshows were made available to the Publicity Department of Philips for an experimental commercial TV programme which ran successfully for several months at the "E 55". The Philips programme, entitled the "Electronic Decathlon", gave members of the audience the opportunity to test their skill and strength on these games. The programme ran daily

Fig. 11. A well-known figure caricatured by Lotty.

from 15 July to 3 September 1955, and was an unqualified success. The popularity of these electronic playthings show not only their fascination for the contestants but also their considerable entertainment value for audiences.

[6]) See p. 42 of P. Blom, An electronic hearing-aid, Philips tech. Rev. **15**, 37-48, 1953/1954.

ABSTRACTS OF RECENT SCIENTIFIC PUPLICATIONS BY THE STAFF OF N.V. PHILIPS' GLOEILAMPENFABRIEKEN

Reprints of these papers not marked with an asterisk * can be obtained free of charge upon application to the Philips Research laboratory, Eindhoven, Netherlands.

2246: J. F. Klinkhamer: Lineaire vierpolen met voorgeschreven overdrachtsimpedantie (De Ingenieur **67**, E1-E8, 1955). (Linear four-poles with specified transfer impedance functions; in Dutch).

After a discussion of the circumstances in which the synthesis of a four-pole with a prescribed transfer impedance is of practical importance, the synthesis is treated for the case that a) the prescribed transfer impedance is given as a rational function of frequency; b) the frequencies of infinite attenuation are real frequencies; c) the four-pole is allowed to consist of a reactive network together with one terminating resistance. After this a short discussion is given on the problem of deriving a feasible rational transfer impedance function with a prescribed attenuation curve, within given tolerances.

2247: G. W. van Santen: Het automatisch regelen van de temperatuur in een melkpasteur met behulp van electronische apparatuur (De Ingenieur **67**, 051-055, 1955, No. 14) (Automatic control of temperature in a milk pasteurizing plant by means of electronic equipment; in Dutch).

Several systems exist for pasteurizing milk; the best one, using a temperature of 72 °C, requires a

very close temperature control. An installation with electronic controlling instruments is described; the practical difficulties arising from various disturbances are mentioned. Special attention is paid to disturbances caused by the action of the flow diversion valve. The electronics behind the Philips controlling and recording equipment are outlined.

2248: J. M. Stevels, C. van Amerongen and J. Volger: Dielectric losses in quartz at low temperatures (Z. physik. Chem. **3**, 382-385, 1955, No. 5-6).

Investigations of the deformation losses in quartz at temperatures of about 50 °K and at frequencies of a few kc/s. Clear crystals show several relaxation effects, obviously correlated with lattice defects of different types, the relaxation time being determined by an activation energy of the order of 0.1 eV. Quartz coloured by irradiation in the laboratory, on the other hand, and also natural smoky quartz, shows relatively high dielectric losses at very low temperatures particularly below 20 °K. Amethyst, however, differs essentially from the other crystals investigated.

2249: H. G. van Bueren and P. Jongenburger: Resistivity changes by plastic deformation of polycrystalline metals (Nature **175**, 544-545, 1955, No. 4456).

Annealed polycrystalline wires of copper and silver were plastically deformed at 20 °K and 77 °K by first extending them to about 10%, then applying a twist corresponding to a surface shear of about 0,1, and subsequently extending them further. The increase in resistivity caused by this deformation was measured. It is caused by the formation of point-defects (vacancies, interstitial atoms) and line-defects (dislocations). The observations, combined with recent theories, lead to the conclusion that the major part of the resistivity increase caused by plastic deformation is contributed by the point-defects.

2250: P. Jongenburger: Some remarks on the extra-resistivity due to interstitial atoms in copper (Nature **175**, 545-546, 1955, No. 4456).

Previous calculations of the extra-resistivity due to the presence of vacancies in copper are extended to the case of interstitial atoms. The displacements of the neighbouring atoms are very large in this case, and a rough estimation shows them to be responsible for a large part of the extra-resistivity. The contribution of the interstitial atom itself is also calculated. Supposing both parts of the extra-resistivity to be additive the conclusion is drawn that interstitial atoms cause an increase in resistivity of about 5 $\mu\Omega$ cm per atom percent.

2251: R. Vermeulen: Validity of hypotheses (Synthese **9**, 385-394, 1955).

Remarks on the evaluation of the validity of hypotheses by means of Bayes' theorem. The author develops an idea of H. Jeffreys and gives an example to show that the results agree with common sense.

2252: D. Kleis: Experimente zur Verbesserung des Raumwirkung von Schall (Elektronische Rundsch. **9**, 64-68, 1955, No. 2). (Experiments to improve the acoustical qualities of halls; in German).

An account of the work described in Philips tech. Rev. **17**, 258-266, 1955/56, No. 9.

2253: M. C. Teves: Image intensification: instrumental aspects (Brit. J. Radiology **28**, 216-217, April 1955).

Short description of the image intensifier tube as now produced by Philips, a development based on requirements laid down by Chamberlain in Brit. J. Radiology, 1942. The factors limiting the contrast detail rendering and a fundamental factor limiting the minimum X-ray intensity are discussed briefly. Finally the use and advantages of the tube for photographic and cinematographic recording are discussed. See also Philips tech. Rev. **17**, 69-77, 1955/56 (No. 3).

2254: J. Feddema: Image intensification: some possible diagnostic applications in cineradiography (Brit. J. Radiography **28**, 217-220, 1955/56, No. 3).

Clinical experiences in the use of the X-ray image intensifier for diagnostic purposes and comparison with other X-ray diagnostic techniques. See also Philips tech. Rev. **17**, 88-93, 1955 (No. 3).

2255: A. van Weel: A direct-indicating phase meter (J. Brit. Instn. Rad. Engrs. **15**, 143-152, 1955, No. 3).

A phase-measuring method is described, in which the unknown phase angle is introduced in an oscillating circuit, thus influencing the oscillator frequency. The phase angle can be determined by measuring the frequency shift of the oscillator. Both very small (10^{-4} radian) and larger phase angles (up to 2π radians) can be measured in this way. A descrip-

tion of a complete phase meter is given. See also Philips tech. Rev. **15**, 307-316, 1953/54.

2256: H. J. G. Meyer: On the theory of radiative transitions of trapped electrons in polar crystals (Physica **21**, 253-268, 1955, No. 4).

Further consideration of the theory of the properties of F-centre absorption bands in ionic crystals (see previous article by the same author, these abstracts No. **2223**).

2257: H. O. Huisman and A. Smit: A new synthesis of 2,4,5-trichlorophenoxyacetic acid (2,4,5-T) (Rec. Trav. chim. Pays-Bas **74**, 155-160, 1955, No. 2).

Description of a new synthesis for the above-named compound. A mixture of the insecticidally inactive α and β isomers of hexachlorocyclohexane, obtained as a by-product in the manufacture of the γ isomer is converted into 1,2,4-trichlorobenzene. After converting the 1,2,4-trichlorobenzene into 2,5-dichlorophenol, the latter is reacted with monochloroacetic acid to 2,5-dichlorophenoxyacetic acid and chlorinated into 2,4,5-T. The overall yield of 2,4,5-T calculated on hexachlorocyclohexane is about 40-50 mole %.

2258: N. W. H. Addink: An improved method for the spectrochemical determination of zinc in blood (Rec. Trav. chim. Pays-Bas **74**, 197-205, 1955, No. 3).

A method for determining the concentration of zinc in blood with greater precision than the earlier methods described by the same author (see these abstracts, Nos. **1947** and **1972**). The main points of the method are: a) complete evaporation of the metal out of a weighed amount (10 mg) of blood ash; b) a constant arc temperature independent of a varying sodium content of the blood ash; c) a rotating anode of a definite shape. By using this method, concentration values have been determined with a standard deviation of 5%. As for the absolute concentration of zinc in various blood samples, values found by applying the method of successive additions, agreed satisfactorily with those determined chemically.

2259: P. B. Braun and J. H. N. van Vucht: Al-Th intermetallic compounds, I (Acta Cryst. **8**, 117,1955, No. 2).

The system thorium-aluminium has been investigated. Six intermetallic compounds have been found. In this paper the structure and lattice constants are given of Al_3Th (hexagonal), Al_2Th (hexagonal) and Al_2Th_3 (tetragonal).

2260: P. B. Braun and J. H. N. van Vucht: Al-Th intermetallic compounds, I (Acta Cryst. **8**, 246, 1955, No. 4).

The structure and lattice constants are given of the compounds Al-Th (orthorhombic) and $AlTh_2$ (tetragonal) and a probable unit cell for a compound with the probable composition Th_4Al_7 which is stable only in a narrow temperature range at about 1300 °C.

2261: J. B. de Boer: A "Duplo" headlight with asymmetric passing beam (Light and Lighting **48**, 137-141, 1955, No. 4).

Description of a car headlamp bulb (briefly described in Philips tech. Rev. **16**, 351-352, 1954/1955, No. 12) in which the cap under the "dip" filament is partially cut away. In conjunction with a suitable headlight glass, this bulb largely combines the advantages of the European and the American dipping systems. Tests with the lamps are described.

2262: J. P. Spruyt: Accelerated stability test for vitamin A in oils and fats by means of surface enlarging at room temperature (J. Amer. Oil Soc. **32**, 197-200, 1955, No. 5).

Abbreviated version of the article abstracted below: see No. **2263**.

2263: J. P. Spruyt: Versnelde houdbaarheidsproef door middel van oppervlaktevergroting bij kamertemperatuur voor vitamine A in oliën en tranen (Voeding **16**, 416-429, 1955, No. 5). (Title as in **2262**; in Dutch).

A new, simple and rapid method for determining the stability of vitamin A in oils and fats. It is performed at a constant temperature of 20 °C; the acceleration is obtained by enlarging the surface exposed to the air by spreading the oil over a neutral carrier.

2264: J. L. Meijering: Het systeem koper-nikkel-chroom (Chem. Weekblad **51**, 438-441, 1955, No. 23) (The system copper-nickel-chromium; in Dutch).

In the ternary system Cu-Ni-Cr, two face-centered cubic phases appear, besides the body-centered phase rich in Cr (see also J. Inst. Metals **84**, 118-120, 1955/1956). This looks at first sight rather unexpected, but is shown to be in line with the result of thermodynamical calculations based on the regular solution model.

2265: H. G. van Bueren: Elektrischer Widerstand und plastische Deformation von Metallen (Z. Metallk. **46**, 272-282, 1955, No. 4). (Electrical resistance and plastic deformation of metals; in German).

A review is given of the observations concerning the relation between electrical resistivity and plastic flow at low temperatures of pure metals, especially copper. A theoretical foundation of the observed relationship is proposed. There is strong evidence that the resistivity increase is caused in the first instance by the vacancies and interstitial atoms formed in the lattice during plastic flow. The recovery phenomena of the electrical conductivity on heating are described and compared to similar phenomena in irradiated and quenched metals. Five recovery steps at different temperatures are indicated. A tentative explanation is proposed in terms of various diffusion phenomena of the above-mentioned lattice defects.

2266: A. Claassen and A. Daamen: The photometric determination of cobalt by extraction with β-nitroso-β-naphthol (Anal. chim. Acta **12**, 547-553, 1955, No. 6).

An improved procedure is given for the photometric determination of cobalt by extraction of the α-nitroso-α-naphthol complex with chloroform. In contrast to former procedures it can be applied in the presence of large amounts of foreign elements, in particular iron and nickel. Interfering elements are gold, palladium and more than 25 mg of copper. The method has been applied with good results to the determination of cobalt in steels, non-ferrous alloys and in nickel.

2267: J. A. Lely: Darstellung von Einkristallen von Siliciumkarbid und Beherrschung von Art und Menge der im Gitter eingebauten Verunreinigungen (IUPAC, Colloq. Münster, Westf. 2-6 Sept. 1954, published by Chemie, Weinheim 1955) (Preparation of single crystals of silicon carbide and control of type and quantity of lattice impurities; in German).

Single crystals of the hexagonal form of silicon carbide (SiC) have been prepared by sublimation of commercial SiC in a protective gas at atmospheric pressure in a graphite furnace at a temperature of 2500-2600 °C (4500-4700 °F).

The sublimation of SiC is complicated by the fact that together with the molecular sublimation there occurs also dissociation into Si-vapour and solid graphite. By lining a graphite crucible on the inside with blocks of SiC one obtains a chamber in which clear, graphite-free SiC crystals are deposited. In 6-8 hours crystals are formed up to 10 mm wide and 2-3 mm thick. Very pure crystals ($\leq 10^{-5}\% = 10^{10}$ impurity atoms/cm^3) are practically colourless; when Al or B is built into the lattice they are blue and show hole (p-type) electrical conductivity. By introducing nitrogen or phosphorus the crystals acquire a green colour and show electron (n-type) conductivity.

The impurity to be introduced is added preferably to the protective gas (in most cases Argon) in the form of the element (N_2) or a volatile compound ($AlCl_3$, BCl_3, etc.). In this manner the crystals show a homogeneous impurity content. By changing over during a run from N_2 to $AlCl_3$, it is possible to obtain crystals with differently coloured zones, so that the blue zone shows p-type conductivity and the green zone n-type conductivity.

2268: Y. Haven: Concentration and association of lattice defects in NaCl (Rep. Conf. Defects in Crystalline Solids, pp. 261-272, Bristol 1954).

In NaCl containing small amounts of divalent impurities like Ca^{2+}, Cd^{2+}, Mn^{2+} etc., the (—) Na vacancies are captured by the (+) divalent ions. The concentration of associates (Ca^{2+}-Na vacancy) can be deduced rather accurately from dielectric measurements. The concentration of free ions can be deduced much less accurately from conductivity measurements. It appears that not only the equilibrium Ca^{2+} + Na vacancy $\rightleftarrows$ (Ca^{2+}-Na vacancy) plays a role, but also the associates consisting of more individuals. The different behaviour of Cd^{2+} and Ca^{2+} is largely determined by the segregation of mixed crystals for Cd^{2+}, rather than by the pair association. At room temperature the vacancies have still a relatively large mobility. The mobility of the associated vacancies seems to be higher than of free vacancies. The Cl^- vacancies seem to have a mobility about .6 times smaller than that of Na vacancies. Isotypic mixed crystals like NaCl-NaBr do not exhibit an excessive concentration of lattice defects.

R 274: A. Bril and H. A. Klasens: Phosphors for tricolour television tubes (Philips Res. Rep. **10**, 305-318, 1955, No. 5).

The efficiencies and spectral distributions were measured of various blue, green and red phosphors which might be considered for use as primaries in tricolour tubes. It is shown that it is mainly the red phosphor that determines the brightness of the screen and that a compromise has to

be found between colour reproduction and light output. Of the known red phosphors, a silver-activated zinc-cadmium sulphide and a copper-activated zinc-cadmium selenide can be expected to produce higher efficiencies for near white colours than is possible with the commonly used manganese-activated zinc phosphate.

R 275: J. van den Boomgaard: Zone-melting processes under influence of the atmosphere (Philips Res. Rep. **10**, 319-336, 1955, No. 5).

The mathematical description of the zone-melting process, as given by Pfann and by Reiss concerns only the cases of ingots consisting of an element or a non-decomposing compound with non-volatile solutes. In this paper the theory is extended to elements containing volatile solutes. It is shown that it is possible to bring a volatile impurity element homogeneously into an ingot by means of zone melting under a constant vapour pressure of that element, providel that diffusion in the liquid phase is rapid.

R 276: B. H. Schulz: Analysis of the decay of photoconductance in germanium (Philips Res. Rep. **10**, 337-348, 1955, No. 5).

A method of analysing surface and volume effects in the recombination of injected charge carriers is described. The influence of the capacity of a surface double layer is discussed and some of the results of measurements are given.

R 277: A. G. Th. Becking, H. Groendijk and K. S. Knol: The noise factor of four-terminal networks (Philips Res. Rep. **10**, 349-357, 1955, No. 5).

In this paper a simple formula is given for calculating the noise factor of a linear four-terminal network with internal noise sources for any input circuit and any value of signal source impedance. This formula is derived from the theorem that the noise behaviour of the network can be characterized by two noise sources introduced at the input of the network. A proof of this theorem is given. The noise sources and their mutual correlation can be denoted by four quantities, which may be determined by four simple measurements.

R278: H. K. Hardy and J. L. Meijering: Phase separation in quaternary regular solutions (Philips Res. Rep. **10**, 358-400, 1955, No. 5).

Calculation of the phase separation of quaternary solutions based on the "regular" approximation for the free energy (see earlier calculations on ternary solutions: these abstracts **R149** and **R166**). The regular approximation includes the Gibbs expression for the entropy of mixing and six terms representing the heat of mixing. Each of these six terms consists of the product of one interaction parameter (corresponding to one of the six binary boundary systems) and the concentrations of the two components concerned in that binary system. In particular, the existence of quaternary critical points and the shift of the critical curves with temperature are investigated.

R 279: G. Diemer, H. A. Klasens and J. G. van Santen: Solid state image intensifiers (Philips Res. Rep. **10**, 401-424, 1955, No. 6).

A theory is given of the characteristics of solid-state image-intensifying screens consisting of a photoconducting and an electroluminescent layer, including the influence of positive internal feedback and negative electrical feedback. The possibilities of brightness as well as contrast amplification are discussed. With intermittent irradiation it is possible to increase the amplification factor by storage, due to the decay, or by triggering of a feedback amplifying-screen. In the latter case the gradation need not be lost. The dimensioning of the parameters with regard to specific applications (e.g. radar, X-ray images, television) is discussed. Preliminary experimental results are in agreement with theory. With X-rays a brightness-amplification factor of 30 was obtained with respect to a normal fluorescent X-ray screen.

R 280: Th. A. J. Payens: Ionized monolayers (Philips Res. Rep. **10**, 425-481, 1955, No. 6).

This thesis deals with monomolecular layers of long-chain weak electrolytes at oil-water and air-water interfaces. The ionization of such films sets up an electronic double layer at the interface. The contribution of this double layer to the interfacial tension is derived from general thermodynamical considerations. The theory is put to the test with monolayers of various fatty acids, stearyl phosphonic acid and octylamine on sub-phases of various *p*H and salt concentration. Deviations from the theory at salt concentrations exceeding 0.1 N NaCl can be explained by the penetration of the counter-ions between the film charges.