Philips Research Reports

Volume 2

1947

CHRONOLOGICAL INDEX

PHILIPS RESEARCH REPORTS, VOL. 2

VOL. 2 No. 1 FEBRUARY 1947

Philips Research Reports

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

R 32 *Philips Res. Rep.* 2, 1-19, 1947

COUPLED CIRCUITS *)

by B. D. H. TELLEGEN 538.565.5:621.396.611.3

Summary

The theory of coupled circuits leads, in the determination of the frequencies and dampings of the free oscillations, to an equation of the fourth degree. If we confine ourselves to cases in which the resonance curve has a small relative width, that equation can be reduced to one of the second degree which can be resolved into factors in a simple manner. Each of the factors determines one of the free oscillations. Inductive or capacitive coupling leads to a real coupling factor k, resistance coupling to an imaginary k, mixed coupling to a complex k. If the circuits are equally damped and equally tuned, then, when k is real, the factors have equal dampings but different characteristic frequencies; when k is imaginary, they have equal characteristic frequencies but different dampings; when k is complex, they have both different characteristic frequencies and different dampings. If the circuits are equally damped but differently tuned, the effect on the shape of the resonance curve is the same as that of an increase of k^2. If the circuits are equally tuned but unequally damped, k^2 is apparently diminished. If the circuits are differently tuned and unequally damped, k^2 is apparently altered by a complex amount. All symmetrical resonance curves can be drawn in a single family. The height of the resonance curve is also investigated. Finally, systems are considered in which the circuits are coupled over an arbitrary four-terminal network which may also contain an amplifying valve.

1. *Introduction*

Much has been written about coupled circuits, and they have been studied in all kinds of manners. Many of the results obtained in this article have thus been obtained previously by others, although they were often formulated differently, and they may be found in the surveys of Wagner [1]) and Aiken [2]). After we have set up the equations for the coupled circuits, our method of treatment deviates, however, from that followed until now in various respects, and in many cases it makes it possible to emphasize certain characteristics with a minimum of calculation, which is an advantage as far as physical insight is concerned. We shall moreover generalize the concept of coupled circuits by considering circuits that are coupled over an arbitrary four-terminal network which may also contain an amplifying valve.

*) Published in Dutch in Tijdschr. Ned. Radiogenoot. **10**, 147, 1943.

Coupled circuits are often used in radio engineering as the coupling element between two valves and are then often called *bandpass filters*. A certain alternating current I is passed by the first valve to the first circuit, while the second circuit furnishes an alternating voltage V to the next valve. It is the ratio of this voltage to that current as a function of the frequency, and especially its absolute value, which determines the *resonance curve* of the bandpass filter, which must be known for the applications.

For any given linear system of constant elements this ratio has the form

$$\frac{V}{I} = \frac{a_0\lambda^n + a_1\lambda^{n-1} + \ldots + a_n}{b_0\lambda^n + b_1\lambda^{n-1} + \ldots + b_n},$$

where $\lambda = j\omega$, when ω is the angular frequency, and the a's and b's are all real and determined by the value of the elements of which the system is composed. Setting the denominator equal to zero gives an equation of the n^{th} degree in λ, whose roots determine the *free oscillations* of the system. An imaginary root leads to an undamped oscillation, a complex root to a damped or augmenting oscillation, a real root to an aperiodic phenomenon. If the system consists of one oscillation circuit, then $n = 2$ and the expression is easily discussed. If the system consists of two circuits, $n = 4$; for investigation it is then advantageous to resolve numerator and denominator into two factors of the second degree. The resonance curve of a system with two circuits can thus always be conceived of as the product of the resonance curves of two systems with one circuit.

The factorization of a fourth-degree polynomial into two second-degree polynomials comes down to the solution of an equation of the fourth degree. In the most general form this solution is by way of that of a third-degree equation. This method was followed for two coupled circuits by Wien [3]), Kiebitz [4]), and recently by Schneider [5]). The results are, however, not very instructive. By assuming that the circuits are little damped, little detuned with respect to each other, and loosely coupled, it is however possible to find approximate solutions which are simpler. These have been calculated by Wien via the third-degree equation to the third approximation, and later by Van der Pol [6]) directly in a simpler way to the fourth approximation. Van der Pol points out that the quintessence of the phenomena occurring with coupled circuits under the conditions mentioned can already be satisfactorily described with a solution to the second approximation. If we confine ourselves to this approximation, we may introduce various simplifications from the very beginning, for instance instead of the frequency itself we may use as variable the difference between the frequency and the resonance frequency of the circuits. In that way we arrive for a single circuit at expressions of the first

degree, and for coupled circuits at expressions of the second degree which are easily resolved into two factors of the first degree.

2. *The single circuit*

Before discussing coupled circuits we shall review several properties of the single circuit. For the circuit shown in *fig. 1*, which consists of a self-induction L, resistance R, capacity C and source of voltage V in series, the current I can be calculated from,

$$\left(j\omega L + R + \frac{1}{j\omega C}\right) I = V. \tag{1}$$

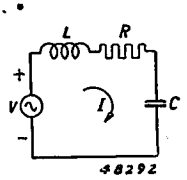


Fig. 1. Single oscillation circuit.

When we put

$$\frac{1}{\sqrt{LC}} = \omega_0,$$

the *resonance frequency* of the circuit, and divide both members by $\omega_0 L$, we obtain

$$\left(j\frac{\omega}{\omega_0} + \frac{R}{\omega_0 L} - j\frac{\omega_0}{\omega}\right) I = \frac{V}{\omega_0 L}. \tag{2}$$

When we put

$$\frac{R}{\omega_0 L} = \delta,$$

the *damping* of the circuit, and

$$\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega} = \beta,$$

the *detuning*, we may write

$$(\delta + j\beta)\, I = \frac{V}{\omega_0 L}. \tag{3}$$

The current is thus determined by

$$\frac{I}{V} = \frac{1}{\omega_0 L} \cdot \frac{1}{\delta + j\beta}. \tag{4}$$

In order better to understand the significance of β we introduce the

difference $\Delta\omega$ between the frequency of the source of voltage and the resonance frequency of the circuit and thus put

$$\omega = \omega_0 + \Delta\omega .$$

Therefore, when $\frac{\Delta\omega}{\omega_0} \ll 1$,

$$\beta = 1 + \frac{\Delta\omega}{\omega_0} - \frac{1}{1 + \frac{\Delta\omega}{\omega_0}} = \frac{2\Delta\omega}{\omega_0} - \left(\frac{\Delta\omega}{\omega_0}\right)^2 + \ldots \approx \frac{2\Delta\omega}{\omega_0}.$$

In that case therefore β is directly proportional to $\Delta\omega$ and at the same time $\beta \ll 1$. In order that the main part of the resonance curve shall lie at these small values of β, it is necessary also that $\delta \ll 1$, which we shall for the rest assume. In general R will depend upon the frequency, since the losses of a circuit, which may for instance originate in the coil resistance, are not constant. If we consider only a small frequency region around the resonance frequency, we may disregard this and consider R and thus also δ as constant. The only variable quantity in (4) is then β, which occurs only in the denominator, and to the first degree. The absolute value of (4) as a function of β gives the familiar symmetrical resonance curve whose width is determined by δ. Due to the fact that we confine ourselves to small values of δ, this width in cycles per second compared with the resonance frequency in cycles per second, the *relative band width*, will be small.

Upon the application of a single circuit as coupling element between two valves the situation deviates somewhat from fig. 1. The system then consists of a self-induction L, a capacity C and a source of current I in parallel, while the voltage V on the circuit is passed on to the next valve. It is advisable here to represent the circuit losses by a conductance G in parallel with the circuit (*fig. 2*). By the restriction to a small relative band

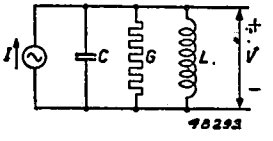


Fig. 2. Another form of a single oscillation circuit.

width we may represent the losses of the circuit not only by a constant resistance in series, but also by a constant conductance in parallel. The internal resistance of the preceding valve may be considered to be included in G. The calculation now corresponds entirely to the above. We may write

$$\left(j\omega C + G + \frac{1}{j\omega L}\right) V = I, \qquad (5)$$

which, upon division of both members by $\omega_0 C$ and introduction of the damping by now putting

$$\frac{G}{\omega_0 C} = \delta ,$$

while ω_0 and β have the same significance as above, can be converted into

$$\frac{V}{I} = \frac{1}{\omega_0 C} \cdot \frac{1}{\delta + j\beta}, \tag{6}$$

which corresponds entirely to (4).

3. *Two coupled circuits*

3a. *Inductive or capacitive coupling*

We shall begin by considering two inductively coupled circuits, the first of which contains a source of voltage. We ask for the current in the second circuit (*fig. 3*). The equations for the two circuits may be written as follows:

$$\left.\begin{aligned} \left(j\omega L_1 + R_1 + \frac{1}{j\omega C_1}\right) I_1 - j\omega M I_2 &= V, \\ -j\omega M I_1 + \left(j\omega L_2 + R_2 + \frac{1}{j\omega C_2}\right) I_2 &= 0 . \end{aligned}\right\} \tag{7}$$

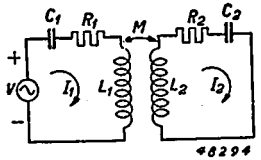


Fig. 3. Two inductively coupled circuits; coupling factor real.

If we put

$$\frac{1}{\sqrt{L_1 C_1}} = \omega_1 \quad \text{and} \quad \frac{1}{\sqrt{L_2 C_2}} = \omega_2 ,$$

and divide the first equation by $\sqrt{\omega_1 L_1}$ and the second by $\sqrt{\omega_2 L_2}$, we obtain

$$\left.\begin{aligned} \left(j\frac{\omega}{\omega_1} + \frac{R_1}{\omega_1 L_1} - j\frac{\omega_1}{\omega}\right) \cdot I_1 \sqrt{\omega_1 L_1} - j\frac{\omega}{\sqrt{\omega_1 \omega_2}} \cdot \frac{M}{\sqrt{L_1 L_2}} \cdot I_2 \sqrt{\omega_2 L_2} &= \frac{V}{\sqrt{\omega_1 L_1}}, \\ -j\frac{\omega}{\sqrt{\omega_1 \omega_2}} \cdot \frac{M}{\sqrt{L_1 L_2}} \cdot I_1 \sqrt{\omega_1 L_1} + \left(j\frac{\omega}{\omega_2} + \frac{R_2}{\omega_2 L_2} - j\frac{\omega_2}{\omega}\right) \cdot I_2 \sqrt{\omega_2 L_2} &= 0 . \end{aligned}\right\} \tag{8}$$

If we put

$$\frac{R_1}{\omega_1 L_1} = \delta_1, \frac{R_2}{\omega_2 L_2} = \delta_2, \frac{\omega}{\omega_1} - \frac{\omega_1}{\omega} = \beta_1, \frac{\omega}{\omega_2} - \frac{\omega_2}{\omega} = \beta_2, \sqrt{\omega_1 \omega_2} = \omega_0, \frac{\omega}{\omega_0} \cdot \frac{M}{\sqrt{L_1 L_2}} = K ,$$

then we may write

$$\left.\begin{aligned}(\delta_1 + j\beta_1) \,.\, I_1 \sqrt{\omega_1 L_1} - jK \,.\, I_2 \sqrt{\omega_2 L_2} &= \frac{V}{\sqrt{\omega_1 L_1}},\\ -jK \,.\, I_1 \sqrt{\omega_1 L_1} + (\delta_2 + j\beta_2) \,.\, I_2 \sqrt{\omega_2 L_2} &= 0\,.\end{aligned}\right\} \tag{9}$$

From this it follows for the current in the second circuit that

$$\frac{I_2}{V} = \frac{j}{\omega_0 \sqrt{L_1 L_2}} \cdot \frac{K}{(\delta_1 + j\beta_1)(\delta_2 + j\beta_2) + K^2}. \tag{10}$$

When we confine ourselves to bandpass filters with a small relative band width of the resonance curve, δ_1, δ_2, β_1, β_2 and also K are all small compared with unity. We then need to distinguish between ω, ω_0, ω_1 and ω_2 only in β_1 and β_2. When we introduce the differences $\Delta\omega_1$ and $\Delta\omega_2$ between the frequency of the source of voltage and the resonance frequencies of the circuits by putting

$$\omega = \omega_1 + \Delta\omega_1 = \omega_2 + \Delta\omega_2,$$

then

$$\beta_1 \approx \frac{2\,\Delta\omega_1}{\omega_0},\ \beta_2 \approx \frac{2\,\Delta\omega_2}{\omega_0},\ \beta_1 - \beta_2 \approx \frac{2\,(\omega_2 - \omega_1)}{\omega_0} \text{ and } K \approx \frac{M}{\sqrt{L_1 L_2}} = k,$$

the *coupling factor*. In the small region around the resonance frequencies to which our attention is confined, we may consider R_1 and R_2, and thus also δ_1 and δ_2, as well as K, to be constant. The only variable quantities in (10) are then β_1 and β_2 and they occur only in the denominator, while their difference may be considered as constant.

We have derived (10) here for inductively coupled circuits with series resistances. For any given coupled circuits we can, however, always derive an expression of the form of (10), which is thus characteristic of two coupled circuits. If for example they form the coupling element between two valves, a source of current I is in parallel with the first circuit and the voltage on the second circuit is passed on to the next valve. It is then again advisable to represent the circuit losses by conductances in parallel with the circuits. If we imagine the circuits coupled by a small capacity between the tops of the circuits, the system of *fig. 4* is obtained. Its calculation corresponds entirely to the above, when we consider the voltages on the circuits

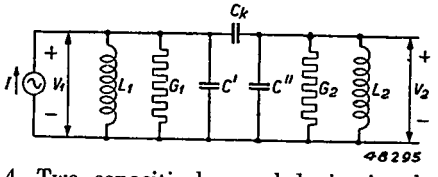


Fig. 4. Two capacitively coupled circuits; k real.

as the unknowns and set the sum of the currents of the first circuit equal to I and the sum of the currents of the second circuit equal to zero. We then obtain

$$\left.\begin{aligned} \left\{j\omega(C' + C_k) + G_1 + \frac{1}{j\omega L_1}\right\} V_1 - j\omega C_k V_2 &= I, \\ -j\omega C_k V_1 + \left\{j\omega(C'' + C_k) + G_2 + \frac{1}{j\omega L_2}\right\} V_2 &= 0. \end{aligned}\right\} \quad (11)$$

When this is compared with (7) it is found that $C' + C_k$ and $C'' + C_k$ must be considered as the capacities of the circuits. We therefore put

$$C' + C_k = C_1 \quad \text{and} \quad C'' + C_k = C_2.$$

Thus we must consider as first circuit the system that is obtained from fig. 4 by short-circuiting the second circuit, and *vice versa*. If we divide the first equation by $\sqrt{\omega_1 C_1}$, the second by $\sqrt{\omega_2 C_2}$ and put

$$\frac{G_1}{\omega_1 C_1} = \delta_1, \frac{G_2}{\omega_2 C_2} = \delta_2, \frac{C_k}{\sqrt{C_1 C_2}} = k,$$

while ω_0, ω_1, ω_2, β_1 and β_2 have the same significance as above, and we again confine ourselves to bandpass filters with a small relative band width, as will also be done in the following, we arrive at

$$\frac{V_2}{I} = \frac{j}{\omega_0 \sqrt{C_2 C_1}} \cdot \frac{k}{(\delta_1 + j\beta_1)(\delta_2 + j\beta_2) + k^2}, \quad (12)$$

which corresponds entirely to (10).

If in fig. 4 it were a question of inductive coupling instead of capacitive coupling, that would correspond to a large self-induction L_k instead of the small capacity C_k. For that also (12) can again be derived.

3b. *Resistance coupling*

The circuits can also be coupled by a conductance G_k as indicated in *fig.* 5. If we follow the method indicated above, we put

$$G' + G_k = G_1, \; G'' + G_k = G_2, \; \frac{G_k}{j\omega_0 \sqrt{C_1 C_2}} = k \; *),$$

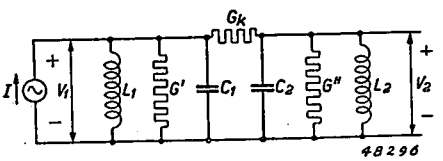


Fig. 5. Two circuits with resistance coupling; k imaginary.

*) In the literature a definition is often found for this case that corresponds to $G_k/\sqrt{G_1 G_2} = k$. G. W. O. Howe, Wireless Eng. 9, 485, 1932, has already pointed out that this cannot be a correct measure of the coupling.

while the other quantities retain the same significance. We then again obtain (12). *The coupling factor is now imaginary*, however.

3c. *Mixed coupling*

When we have capacitive as well as resistance coupling, as indicated in *fig. 6*, we must put

$$\frac{G_k + j\omega_0 C_k}{j\omega_0 \sqrt{C_1 C_2}} = k$$

and we then again obtain (12). *The coupling factor is now complex.*

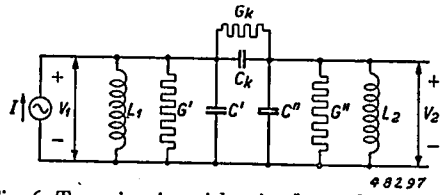


Fig. 6. Two circuits with mixed coupling; k complex.

4. *Equally damped and equally tuned circuits*

The frequency dependence of a bandpass filter is determined entirely by the denominator of (12)

$$N = (\delta_1 + j\beta_1)(\delta_2 + j\beta_2) + k^2 . \qquad (13)$$

If k^2 is so small that it may be neglected in N, N consists of the product of two factors each of which corresponds to one of the circuits. In that case the form of the resonance curve is that of the product of the resonance curves of the separate circuits. If k^2 may not be neglected, N may again be resolved into two factors which now no longer correspond to the separate circuits, but are related to the free oscillations which the system can execute.

We shall begin with the simplest case, namely that in which the two circuits have equal damping

$$\delta_1 = \delta_2 = \delta,$$

and in which they are equally tuned, so that we may put

$$\beta_1 = \beta_2 = \beta .$$

Then

$$N = (\delta + j\beta)^2 + k^2 . \qquad (14)$$

The coupling factor will in general be complex, so that we put

$$jk = m + jl \text{ *}). \qquad (15)$$

*) If it were still possible to define freely the concept coupling factor, we should prefer for it the quantity jk, with which resistance coupling would have a real, and inductive or capacitive coupling an imaginary coupling factor. For this reason we have written jk and not k as the sum of a real and an imaginary part. This also fits best with the expression for k given in section 3c for the case of mixed coupling.

4a. *k real*

We shall begin with the case where k is real, thus $k = l$. With this we obtain

$$N = (\delta + j\beta)^2 + l^2 = \{\delta + j(\beta + l)\}\{\delta + j(\beta - l)\}. \qquad (16)$$

These factors correspond to two circuits with equal dampings δ whose characteristic frequencies are shifted in opposite directions with respect to that of the circuits of the bandpass filter, each by an amount l (*fig.* 7). If l is varied

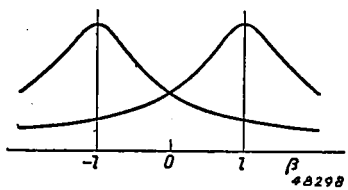


Fig. 7. The two factors have equal dampings but different characteristic frequencies.

continuously from zero, these characteristic frequencies are shifted equally to both sides. The resulting resonance curve is always *symmetrical* and has initially one maximum in the middle. Upon increasing the coupling two maxima are formed, one on either side of a minimum in the middle. In order to ascertain at what coupling the transition between the two cases lies, we must write the absolute value of N^2. That is

$$|N^2| = (\delta^2 + l^2)^2 + 2(\delta^2 - l^2)\beta^2 + \beta^4. \qquad (17)$$

The term with β^2 disappears for $l^2 = \delta^2$ and thus the transition lies at that point. The resonance curve here has a flat top. The position of the maxima occurring with stronger coupling is found by determining the minima of (17). They lie at $\beta^2 = l^2 - \delta^2$.

Since the characteristic frequencies of the factors of (16) correspond to the frequencies of the free oscillations of the system, we can indicate them immediately in fig. 4, when we assume here also that the two circuits are exactly alike, thus, in addition to equal tuning and equal damping, that they also have equal capacities $C' = C'' = C$. In the case of the one free oscillation V_1 and V_2 are equal and in phase and thus C_k carries no current. In the case of the other oscillation V_1 and V_2 are equal and in opposite phase, so that C_k does carry current. If we imagine C_k replaced by two capacities $2C_k$ in series, there is in the latter case no alternating voltage between the point of connection of these capacities and the lower side of the circuits, so that we may connect these points without any alteration of the state of oscillation. We then see that the capacity determining the characteristic frequency in this case is $C + 2C_k$, while for the case

when the voltages on the circuits are in phase, this is C. If we change the coupling between the circuits by changing only C_k, the one characteristic frequency remains unchanged while the other is shifted. This is in contrast to our statement above where we said that upon change in l the two characteristic frequencies are shifted equally but in opposite directions. The reason for this lies in the fact that when we vary only C_k, we vary not only l but also the tuning of the circuits. In order to prevent this we must keep $C + C_k$ constant and thus upon increasing C_k we must diminish the value of C by the same amount. If we have two circuits that are coupled by a mutual induction and we change the coupling without altering the self-inductions of the coils, then indeed we obtain an equal and opposite shift of the characteristic frequencies. From this it also follows that in order to obtain a variable width of the resonance curve such as is employed for example in intermediate-frequency amplifiers of receiving sets, a coupling by a variable mutual induction is better than a variable capacitive coupling.

4b. *k imaginary*

If k is imaginary, thus when $jk = m$, we have

$$N = (\delta + j\beta)^2 - m^2 = \{(\delta + m) + j\beta\}\{(\delta - m) + j\beta\}. \qquad (18)$$

These factors correspond to two circuits with equal characteristic frequencies but with dampings that are altered in opposite directions with respect to that of the circuits of the bandpass filter, each by an amount m (*fig. 8*).

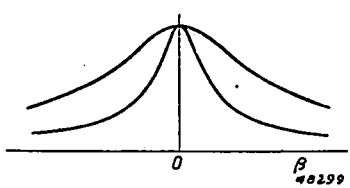


Fig. 8. The two factors have equal characteristic frequencies but different dampings.

The resulting resonance curve is now also *symmetrical* and always has a maximum in the middle. With increasing value of m the width of the resonance curve decreases, at $m = \delta$ it even becomes infinitesimally narrow, so that self-oscillations may occur. With the system of fig. 5 this cannot be fulfilled. Systems in which m can be equal to δ, will be discussed in section 10.

Since the factors of (18) correspond to the free oscillations of the system, we can immediately indicate them in fig. 5 if we assume here that in addition to equal tuning and equal damping the circuits also have equal conduc-

tances $G' = G'' = G$. In the case of the one free oscillation V_1 and V_2 are again equal and in phase so that G_k carries no current. In the case of the other oscillation V_1 and V_2 are equal and in opposite phase so that G_k then does carry current. If we imagine G_k replaced by two conductances $2G_k$ in series, there is then in the latter case no voltage between the point of connection of these conductances and the lower side of the circuits, so that these points may be connected without altering the state of oscillation. We then see that the conductance determining the damping in this case is $G + 2G_k$, while for the case where the voltages on the circuits are in phase this amounts to G. If we change the coupling between the circuits by changing only G_k, one of the dampings of the free oscillations remains unchanged while the other is increased. In order to attain the above described state, at which by the increase of the coupling the dampings of the free oscillations are changed equally but in opposite directions, upon increasing G_k we must diminish the value of G by an equal amount so that $G + G_k$ remains constant.

4c. *k complex*

If k is complex so that $jk = m + jl$, we obtain

$$N = (\delta + j\beta)^2 - (m + jl)^2 = \{(\delta + m) + j(\beta + l)\}\{(\delta - m) + j(\beta - l)\}. \quad (19)$$

Here the factors have different characteristic frequencies as well as different dampings (*fig. 9*). The result is that their product gives an *unsymmetrical* resonance curve. The free oscillations corresponding to these factors can be indicated in fig. 6 in a manner analogous to that above, when the two circuits are exactly alike.

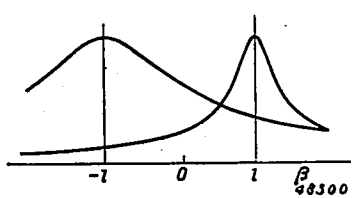


Fig. 9. The two factors have both different characteristic frequencies and different dampings.

The system of fig. 6 may for example occur when the coupling capacity C_k is subject to losses. In order in that case to prevent the occurrence of an unsymmetrical resonance curve, we may give an extra damping to the least damped free oscillation. For that purpose we split C_k and G_k into two capacities and two conductances $2C_k$ and $2G_k$ in series and introduce between the connecting point of these and the lower side of the circuits a conductance $4G_k$ (*fig. 10*). A similar case is obtained when the circuits are coupled

over a common series condenser with losses (*fig. 11a*). If I_1 and I_2 oscillate in phase, no current flows through the coupling capacity and thus R does not damp; if I_1 and I_2 oscillate in opposite phase, current does flow through

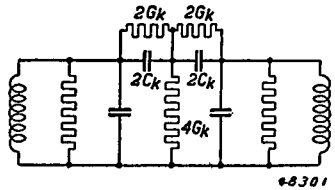


Fig. 10. By splitting the coupling condenser and adding a resistance the resonance curve can be prevented from becoming unsymmetrical by the losses of the coupling condenser.

it and there is therefore extra damping. By changing the system into that of *fig. 11b* the two free oscillations are again equally damped and the occurrence of an unsymmetrical resonance curve is prevented.

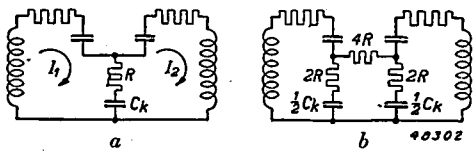


Fig. 11. In *a* the resonance curve is made unsymmetrical by the losses of the coupling condenser, while in *b* it has been made symmetrical again.

5. *Equally damped but differently tuned circuits*

If the circuits have equal dampings $\delta_1 = \delta_2 = \delta$, but different tunings, the detuning can best be measured from the point midway between the tunings of the two circuits, and thus we put

$$\beta_1 = \beta + \gamma, \quad \beta_2 = \beta - \gamma, \tag{20}$$

where 2γ is thus a measure of the detuning of the circuits with respect to each other. Substituting this in (13) we obtain

$$N = \{\delta + j(\beta + \gamma)\}\{\delta + j(\beta - \gamma)\} + k^2 = (\delta + j\beta)^2 + k^2 + \gamma^2. \tag{21}$$

Comparing this with (14) it is evident that *the detuning γ has the same effect on N as an increase of k^2 to $k^2 + \gamma^2$*. Thus if k^2 is zero or positive, the system behaves as the product of two equally damped but differently tuned circuits. When k^2 decreases the dampings of the factors at first remain unchanged but the characteristic frequencies approach each other until $k^2 = -\gamma^2$, where the characteristic frequencies of the factors are equal. If k^2 then decreases still further, the characteristic frequencies of the factors remain further unchanged but the dampings are changed in opposite directions. [7])

The resonance curve has a flat top at $k^2 + \gamma^2 = \delta^2$.

6. *Equally tuned but unequally damped circuits*

If the circuits are equally tuned so that we may write $\beta_1 = \beta_2 = \beta$, but their dampings are different, we put

$$\delta_1 = \delta + \varepsilon, \quad \delta_2 = \delta - \varepsilon. \tag{22}$$

The quantity δ thus represents the average damping $\frac{1}{2}(\delta_1 + \delta_2)$, the quantity 2ε represents the difference between the dampings of the circuits. With this (13) becomes

$$N = (\delta + \varepsilon + j\beta)(\delta - \varepsilon + j\beta) + k^2 = (\delta + j\beta)^2 + k^2 - \varepsilon^2. \tag{23}$$

When we compare this with (14) we see that the *inequality of damping* ε *has the same effect on* N *as a decrease of* k^2 *to* $k^2 - \varepsilon^2$. Thus if k^2 is zero or negative, the system behaves as the product of two equally tuned but unequally damped circuits. If k^2 increases the characteristic frequencies of the factors at first remain unchanged, but the dampings approach each other until $k^2 = \varepsilon^2$, where the dampings of the factors are equal. If k^2 then increases still further, the dampings of the factors remain unchanged but the characteristic frequencies are shifted in opposite directions.[8]) The latter are then determined by $\beta = \pm \sqrt{k^2 - \varepsilon^2}$, which shows, as was also emphasized by Van der Pol[9]), that with coupled circuits the characteristic frequencies may depend considerably upon the dampings of the circuits, in particular upon the difference between their dampings. This is therefore contrary to the behaviour of a single circuit where the influence of a small damping on the frequency of the free oscillation is only slight.

The resonance curve has a flat top at

$$k^2 - \varepsilon^2 = \delta^2, \; k^2 = \tfrac{1}{2}(\delta_1^2 + \delta_2^2).$$

In the literature[10]) it is sometimes stated that in the case of coupled circuits also the dampings have, in a first approximation, no effect on the characteristic frequencies, which statement appears to conflict with the above. The reason for this discrepancy is the following. For loss-free circuits the characteristic frequencies can simply be calculated exactly, thus for any given tunings and for any given coupling, because the odd-power terms in the fourth-degree equation disappear in this case. If, starting with that solution, the losses are taken into account in a first approximation, it is indeed found that the characteristic frequencies do not change. The assumption has then been made, however, that the dampings are small compared with the coupling, thus $\varepsilon \ll k$ also. In that case in $\sqrt{k^2 - \varepsilon^2}$ the ε has indeed no effect in a first approximation. The important cases in which the dampings are small, but the coupling factor and the detuning of the circuits with respect to each other are likewise small, and all are of the same order of magnitude, cannot be approached in this way.

7. *Unequally damped and differently tuned circuits*

If the circuits are not only unequally damped but also differently tuned we put

$$\beta_1 = \beta + \gamma, \; \beta_2 = \beta - \gamma, \; \delta_1 = \delta + \varepsilon, \; \delta_2 = \delta - \varepsilon.$$

With this (13) becomes

$$N = \{\delta+\varepsilon+j(\beta+\gamma)\}\{\delta-\varepsilon+j(\beta-\gamma)\} + k^2 = (\delta+j\beta)^2 + k^2 - (\varepsilon+j\gamma)^2, \quad (24)$$

which corresponds to a *change of* k^2 *into* $k^2 - (\varepsilon + j\gamma)^2$. In general this will lead to an unsymmetrical resonance curve, unless $k^2 - (\varepsilon + j\gamma)^2$ happens to be real, thus when $(m + jl)^2 + (\varepsilon + j\gamma)^2$ is real. This is the case when

$$ml + \varepsilon\gamma = 0.$$

The above discussed special cases of this are those where both ml and $\varepsilon\gamma$ are zero. In order in the most general case to resolve (24) into factors we must calculate

$$\sqrt{(m + jl)^2 + (\varepsilon + j\gamma)^2} = q + jp.$$

With this the result is

$$N = \{(\delta + q) + j(\beta + p)\}\{(\delta - q) + j(\beta - p)\}.$$

8. *Universal resonance curves*

From the foregoing it follows that in the investigation of the possible forms of the resonance curves we may confine ourselves to the case of equally damped and equally tuned circuits, because the other cases can be reduced to that. Since the symmetrical resonance curves are the most important ones, we shall confine ourselves to them and we may therefore in addition assume that k^2 is real; it may however be either positive or negative. If we assign the height unity to the middle point of all the resonance curves, they are determined by

$$\left|\frac{N_0}{N}\right| = \left|\frac{\delta^2 + k^2}{(\delta + j\beta)^2 + k^2}\right| = \frac{\delta^2 + k^2}{\sqrt{(\delta^2 + k^2)^2 + 2(\delta^2 - k^2)\beta^2 + \beta^4}}.$$

β is here the variable quantity and δ^2 and k^2 are to be considered as parameters. By dividing numerator and denominator by δ^2 we may write

$$\left|\frac{N_0}{N}\right| = \frac{1 + k^2/\delta^2}{\sqrt{(1 + k^2/\delta^2)^2 + 2(1 - k^2/\delta^2)\cdot\beta^2/\delta^2 + \beta^4/\delta^4}}, \quad (25)$$

whereby β/δ has become the variable and only one parameter, k^2/δ^2, remains. *The curves can thus be drawn as a single family*, as has been done in *fig. 12*. If the circuits are equally damped but differently tuned, the curves remain correct when k^2 is replaced by $k^2 + \gamma^2$. If the circuits are equally tuned but unequally damped k^2 must be replaced by $k^2 - \varepsilon^2$, and δ represents the average damping $\frac{1}{2}(\delta_1 + \delta_2)$. *)

*) This possibility of drawing all symmetrical resonance curves in a single family was, for inductive or capacitive coupling, in 1931 pointed out and carried out by Ir F. H. Stieltjes, who was at that time connected with this laboratory.

9. *The height of the resonance curve*

Until now we have dealt only with the denominator N of (12), which determines the shape of the resonance curve. In addition to that the height of the resonance curve is also important, because it determines the ampli-

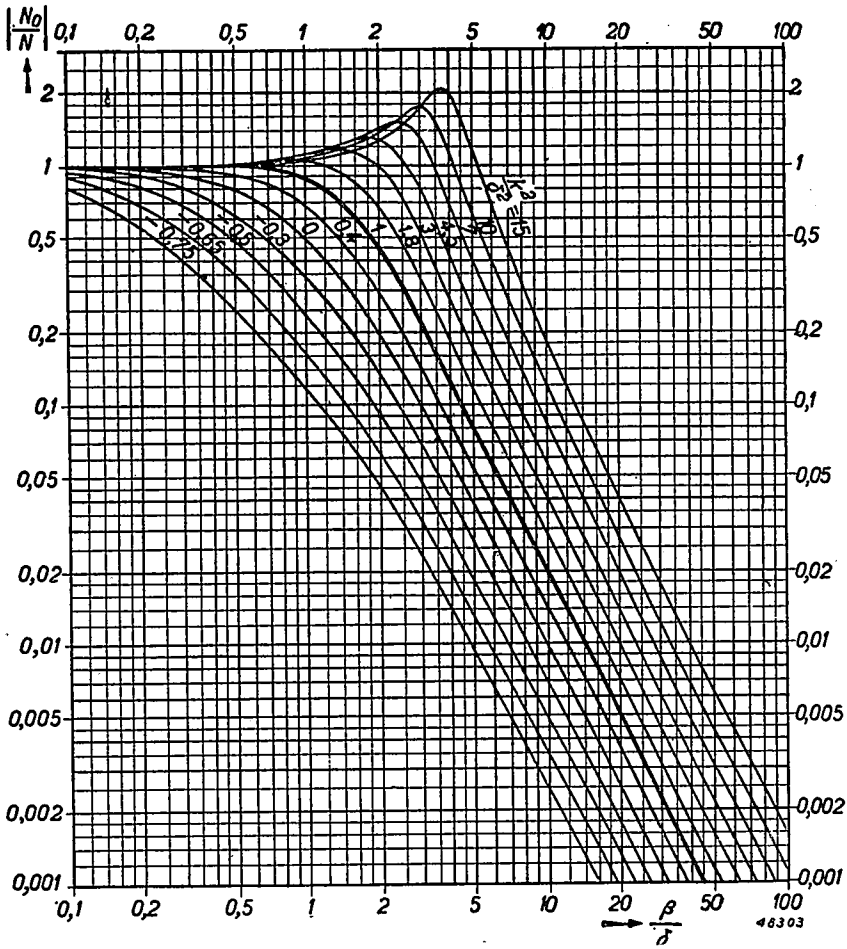


Fig. 12. Universal symmetrical resonance curves of bandpass filters with a small relative band width.

$\omega_1 = \frac{1}{\sqrt{L_1 C_1}}$, $\omega_2 = \frac{1}{\sqrt{L_2 C_2}}$, $\omega_0 = {}^1/_2\,(\omega_1 + \omega_2)$, $\Delta\omega = \omega - \omega_0$, $\beta = \frac{2\Delta\omega}{\omega_0}$, $\gamma = \frac{\omega_2 - \omega_1}{\omega_0}$.

$\delta_1 = \frac{R_1}{\omega_1 L_1}$ or $\frac{G_1}{\omega_1 C_1}$, $\delta_2 = \frac{R_2}{\omega_2 L_2}$ or $\frac{G_2}{\omega_2 C_2}$, $\delta = {}^1/_2\,(\delta_1 + \delta_2)$, $\varepsilon = {}^1/_2\,(\delta_1 - \delta_2)$.

$k^2 = \frac{M^2}{L_1 L_2}$ in inductive, $\frac{C_k^2}{C_1 C_2}$ in capacitive, $\frac{G_k^2}{-\omega_0^2 C_1 C_2}$ in resistance coupling. The curves can be applied directly for equally damped and equally tuned circuits.
With equally damped but differently tuned circuits k^2 must be replaced by $k^2 + \gamma^2$.
With equally tuned but differently damped circuits k^2 must be replaced by $k^2 - \varepsilon^2$.

fication when the system is employed as coupling element between two valves. If we desire to study the influence of the coupling factor on the height of the resonance curve we cannot confine ourselves to N since k also occurs in the numerator, and we must therefore investigate k/N.

In the case of *equally damped and equally tuned circuits* at the middle point $\beta = 0$

$$\frac{k}{N} = \frac{k}{\delta^2 + k^2}.$$

When k is real this has a maximum at $k^2 = \delta^2$, thus at *the same coupling* as that for obtaining a resonance curve with a flat top.

In the case of *equally damped but differently tuned circuits* at $\beta = 0$

$$\frac{k}{N} = \frac{k}{\delta^2 + \gamma^2 + k^2}.$$

When k is real this has a maximum at $k^2 = \delta^2 + \gamma^2$, thus at a *coupling tighter* than that for obtaining a flat top.

In the case of *equally tuned but unequally damped circuits* at $\beta = 0$

$$\frac{k}{N} = \frac{k}{\delta^2 - \varepsilon^2 + k^2}.$$

When k is real this has a maximum at $k^2 = \delta^2 - \varepsilon^2 = \delta_1\,\delta_2$, thus at a *coupling looser* than that for obtaining a flat top.

From (12) it also follows that in order to attain a large height it is important to make C_1 and C_2 small. In the amplification of very wide frequency bands, as for instance in television, C_1 and C_2 are kept equal to the valve capacities increased by the unavoidable wiring and coil capacities which are limited to a minimum. In addition, with a given shape of the resonance curve, thus with a given N, it remains desirable to be able to use a large coupling factor. Now in general it is desirable that the resonance curve should be as rectangular as possible, and this can be approximated by a symmetrical resonance curve with a flat top. If the circuits are equally tuned and equally damped we must therefore choose $k = \delta$. If the circuits have equal dampings but are differently tuned, we saw above that the detuning has the same effect on the shape of the resonance curve as an increase in the coupling factor. If we do not wish to alter the shape of the resonance curve by the detuning the coupling factor will therefore have to be diminished, which is disadvantageous for the height. If we tune the circuits alike but give them different dampings, leaving the mean damping constant, we saw above that this has the same effect on the resonance curve as a decrease in the coupling factor. If we do not wish to alter the shape of the resonance curve by unequal dampings, we shall therefore have to

increase the coupling factor, which is advantageous for the height. We may proceed with making the dampings unequal until one circuit has a zero damping and the other a damping of 2δ [11]). k^2 is thereby apparently decreased by an amount δ^2. Originally k^2 was equal to δ^2, so that we must now make $k^2 = 2\delta^2$ in order to retain the shape of the resonance curve. The coupling factor and thereby the height of the resonance curve thus become $\sqrt{2}$ times as large.

10. *Coupling over a four-terminal network*

The most general possibility of coupling between two circuits is found likewise to lead to the expressions studied, when we investigate two circuits coupled over any given four-terminal network (*fig. 13*).

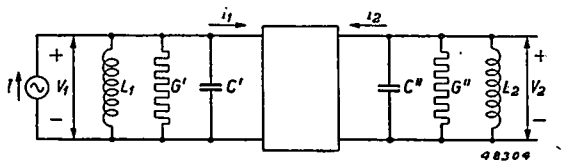


Fig. 13. Two circuits coupled over a four-terminal network.

Since we wish to investigate a system with only two circuits, the four-terminal network contains no circuit tuned approximately to the resonance frequencies of those circuits. We need not therefore pay attention to the frequency dependence of the four-terminal network, and its equations may be given in the form

$$\left.\begin{aligned} I_1 &= \quad Y_{11}\, V_1 - Y_{21}\, V_2\,, \\ I_2 &= -Y_{12}\, V_1 + Y_{22}\, V_2\,, \end{aligned}\right\} \tag{26}$$

where the Y's may be considered as constant. Since we also wish to investigate the cases where the four-terminal network contains an amplifying valve, in general Y_{12} will not be equal to Y_{21}. If we use this to write the equations of the system, we obtain

$$\left.\begin{aligned} \left(j\omega\, C' + G' + \frac{1}{j\omega L_1} + Y_{11}\right) V_1 - Y_{21}\, V_2 &= I\,, \\ -Y_{12}\, V_1 + \left(j\omega\, C'' + G'' + \frac{1}{j\omega L_2} + Y_{22}\right) V_2 &= 0\,. \end{aligned}\right\} \tag{27}$$

We see that Y_{11} and Y_{22} only affect the damping and the tuning of the circuits, so that we can include the latter in it by changing G', C', G'' and C'' into G_1, C_1, G_2 and C_2. The quantities Y_{12} and Y_{21} cause the coupling between the circuits and we now obtain

$$k^2 = \frac{Y_{12}\, Y_{21}}{-\omega_0^2 C_1 C_2}\,,$$

thus in general a complex coupling factor. If we calculate V_2/I, we arrive at an expression which differs from (12) only by a factor $\sqrt{Y_{12}/Y_{21}}$ in the numerator. The quantity N which determines the shape of the resonance curve remains unchanged and thus all the considerations given for N are also valid now.

Examples of couplings where $Y_{12} \neq Y_{21}$ are obtained when two circuits are coupled over an amplifying valve. If there is some feedback, caused for example by the anode-grid capacity C_{ag} of the valve (*fig. 14*), then

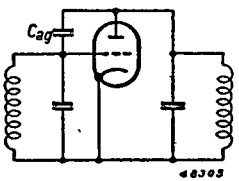


Fig. 14. Two circuits coupled over a valve with anode-grid capacity; k^2 is purely imaginary.

$Y_{12} = -S + j\omega_0 C_{ag}$ and $Y_{21} = j\omega_0 C_{ag}$, where S is the slope of the valve. Since in general $\omega_0 C_{ag} << S$, this leads to a purely imaginary k^2, thus to a complex k, so that the resonance curve in general becomes unsymmetrical. If we feed back over a conductance G_{ag} between anode and grid instead of over a capacity (*fig. 15a*), then $Y_{12} = -S + G_{ag}$ and $Y_{21} = G_{ag}$. Since $G_{ag} << S$ can already give a firm coupling, this leads to a positive k^2, thus to a real k, so that in this way we can also obtain resonance curves with a flat top. In order to carry this out in practice, a resistance cannot be introduced between anode and grid, since that would lead to very large values of the resistance (with circuits with an impedance at tuning of 100 000 Ω and a valve with a slope of 1 mA/V we obtain a flat top with an anode-grid resistance of 10 MΩ), so that the unavoidable self-capacity of this resistance would have a dominating influence. However, by tapping the resistance on one or both circuits (*fig. 15b*), which makes it necessary to give it a lower

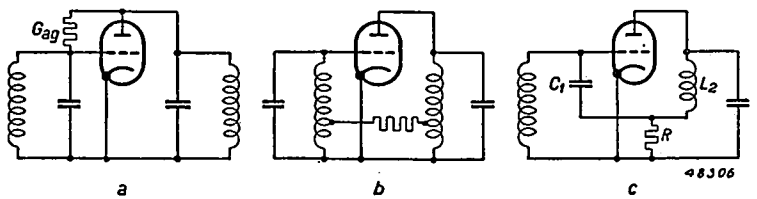


Fig. 15. Two circuits coupled over a valve with resistance between anode and grid (*a*) and two equivalent systems (*b* and *c*); k real.

value, this objection can be met. The same result can be achieved by coupling the two circuits over a small common series resistance (*fig. 15c*), which can most easily be proved by converting the star C_1, L_2, R into a triangle, whereupon a conductance C_1R/L_2 is formed between anode and grid. By reversing the phase of the feedback (*fig. 16a* and *b*) a system is formed with negative k^2, thus imaginary k. If the coupling is made strong enough self-oscillations will occur.

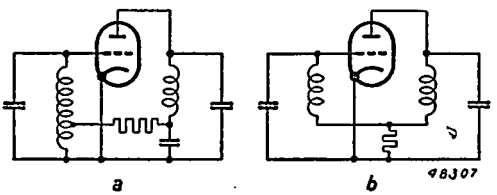


Fig. 16. Two circuits coupled over a valve and a resistance in such a way that self-oscillations may occur; k imaginary.

In all these systems with amplifying valves the coupling elements cause a much smaller damping or detuning of the circuits than in the case of the systems previously discussed without amplifying valves.

Eindhoven, February 1943

REFERENCES

1) K. W. Wagner, Telegr. u. Fernspr. Techn. **24**, 191, 1935, Ann. d. Phys. **32**, 301, 1938.
2) C. B. Aiken, Proc. Inst. Radio Eng. **25**, 230, 1937.
3) M. Wien, Ann. d. Phys. **61**, 151, 1897.
4) F. Kiebitz, Ann. d. Phys. **40**, 138, 1913.
5) P. Schneider, Ann. d. Phys. **41**, 211, 1942.
6) Balth. van der Pol, Physica **6**, 56, 1926, Jahrb. d. drahtl. Tel. **28**, 12, 1926.
7) K. W. Wagner, *loc. cit.* p. 307.
8) M. Wien, *loc. cit.* p. 167.
9) Balth. van der Pol, *loc. cit.* p. 61.
10) Rayleigh, Theory of sound, Vol. **1**, § 102, Helmholtz, Vorlesungen über theoretische Physik, Vol. **3**, § 21, F. Kiebitz, Jahrb. d. drahtl. Tel. **8**, 45, 1914, W. Rogowski, Arch. f. Elektr. **9**, 427, 1921, W. Grösser, Arch. f. Elektr. **10**, 257, 1921.
11) *Cf.* H. A. Wheeler, Proc. Inst. Radio Eng. **27**, 429, 1939, fig. 6.

R 33 *Philips Res. Rep.* **2**, 20-41, 1947

ON THE EXCITATION TEMPERATURE, THE GAS TEMPERATURE, AND THE ELECTRON TEMPERATURE IN THE HIGH-PRESSURE MERCURY DISCHARGE

by W. ELENBAAS

537.527:536.5

Summary

The excitation temperature of a high-pressure discharge, the temperature of which was formerly determined from the intensity decrease perpendicularly to the axis of spectral lines having different initial levels, is determined from the absolute value of the intensity of the yellow lines 5770/5791 Å, combined with the transition probability of these lines, as measured by Schouten and Smit. The temperature found by this last method is considerably lower than that previously determined. The cross-section of the *Hg* atom which must be used in the Langevin equation in order to arrive at the right current (the density of the electrons being determined by the Saha equation), as well as the vapour pressure, are computed by assuming the excitation temperature to coincide *a*) with the gas temperature and *b*) with the electron temperature. In case *b*) we determine the gas temperature by assuming that the electrons deliver energy to the atoms through elastic collisions just sufficiently to compensate the loss by conduction of heat. In both cases we find for the cross-section of the *Hg* atom a reasonable value, *viz.* 4.10^{-15} cm^2. The pressure as computed from the gas temperature and the mean vapour density agrees better with the measured value in the case *b*) than in the case *a*) (differences 2·5 and 8 per cent, respectively). In case *b*) we have to apply a correction to the temperature as formerly determined from the intensity decrease perpendicularly to the axis, resulting in a satisfactory agreement between the two methods. Finally, the excitation probability of the 3*d* and 3*D* levels from the ground level was estimated. The excitation by electron impact is much more frequent than the excitation by mutual collisions of the atoms, but yet it is far too small to explain temperature equilibrium.

1. *Introduction*

The temperature T_g of the gas of a discharge may be determined by measuring the density and the pressure [1) 2) 3)]. The excitation temperature T_{ex} may be found from intensity measurements of molecular rotation or vibration lines [4)]. No reliable methods are known to determine the electron temperature T_e in high-pressure discharges, the probe method of Langmuir [5)] not being applicable at high pressures [27)]. The method of determining the excitation temperature from the radial decrease of the intensity of spectral lines with different initial levels [6)], is not accurate. However, this method gives $T_{ex}(r)$ accurately if the excitation temperature in the axis T_{ex} (0) is known.

Further, there are more indirect methods of determining the excitation

temperature. If the ratio of the transition probabilities of two lines with different initial levels is known, the temperature of the discharge follows from the ratio of the intensities of these two lines [7]). Schulz [8]) used the width of spectral lines to determine the electron concentration n_e with the aid of a formula given by Unsöld [9]). The excitation temperature then follows from the Saha equation. An advantage of these latter two methods is that T_{ex} occurs in the exponential. The disadvantage of the first method is that the distance between the initial levels is often small, and of the second method that the error in n_e may be rather large as a result of the uncertain relation between the width of the lines and n_e. A further disadvantage of the last method is that T_{ex} is derived from V_i/T_{ex} (V_i = ionization potential) where V_i is uncertain at the high pressures necessary to achieve a large width of the spectral line. For instance Schulz used a value of 9·5 V for V_i whereas the ionization potential of the isolated mercury atom is 10·4 volts.

As the electrons obtain their energy directly from the field, and transfer it to the atoms and the ions, the electron temperature is higher than the gas temperature. The difference between T_e, T_{ex} and T_g, however, is small in a high-pressure discharge [10]).

In *figs 1 a, b* we have drawn these three temperatures schematically for two extreme cases:

a) The excitation temperature and the gas temperature coincide. The excitation and ionization occur mainly through mutual collisions between the atoms. The electron temperature lies so much higher that the difference is sufficient to allow the electrons to transfer to the atoms by means of elastic collisions the total energy supplied to the discharge.

b) The excitation temperature and the electron temperature are equal. The excitation and ionization occur mainly by impacts between electrons and atoms. The gas temperature lies so much lower that the gas receives just enough energy through elastic collisions with the electrons to compensate the losses by conduction of heat.

In this paper we shall try to find some evidence as to whether in the high-pressure mercury discharge of about one atmosphere the situation of fig. 1*a* or that of fig. 1*b* is the more probable. For this purpose we have studied the same discharge for which the behaviour of $T_{ex}(r)$ has already been determined [6]). Moreover we had seen that in this discharge the absorption of the yellow lines (5770/91 Å) is negligible and that the population of the different levels of the mercury atom may be described by a Boltzmann factor, the temperature figuring therein being called the excitation temperature.

Recently the absolute values of the transition probabilities of a number of mercury lines have been determined [11]). Since the absorption of the

yellow lines is negligible in our discharge, we may determine T_{ex} from a measurement of the intensity of these lines.

The discharge used in [6]) and in these measurements is characterized by an inner tube diameter of 41 mm, an input of 35 watts per cm of arc length (direct current) and an amount of vaporized mercury of 12 mg per cm of tube length. The tube is operated in the vertical position.

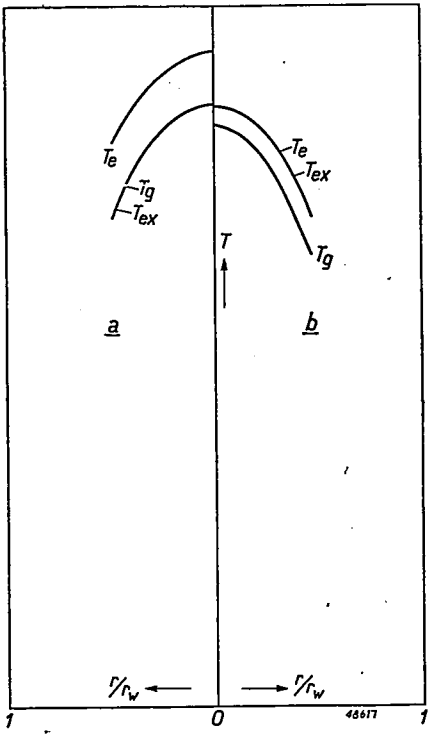


Fig. 1. Excitation temperature T_{ex}, electron temperature T_e, and gas temperature T_g as a function of the distance to the axis for two extreme cases. Tube diameter $= 2\ r_w$ (schematically). Two possibilities are considered:
a) Excitation temperature and gas temperature are the same. The electron temperature is higher.
b) Excitation temperature and electron temperature are the same. The gas temperature is lower.

2. *The measurements*

As we want to know a number of properties of this discharge we shall describe their measurements successively.

2A. *The pressure*

The pressure was determined in two ways:

a) Tubes with inner diameters of about 40 mm (diameters were exactly measured by filling with water and weighing) were filled with quantities of mercury between 12 and 13 mg per cm of length. The tubes were operated vertically and when the mercury was totally vaporized and the equilibrium state reached, a small spot on the tube wall was heated with a small flame to such an extent that the glass softened there. It was then possible to ascertain whether the pressure in the tube was below or above the atmospheric pressure. In the first case the input was increased, whilst in the second case the input was decreased and a second spot heated. In this way we could determine the input at which the pressure equalled the atmospheric pressure. That this can be done accurately may be proved by table I.

TABLE I

(inner tube diameter $d = 39{\cdot}5$ mm; total length 53·5 cm; arc length 50·0 cm; 12·05 mg *Hg* per cm; no rare gas; atmospheric pressure 768 mm)

current (amps)	6·05	6·96	7·8	7·45	7·65	7·54	7·50
behaviour of heated spot	—	—	+	—	+	+	—

(— signifies that the tube sucks in and + that the tube blows out at the overheated spot).

For the tube of table I a pressure of 768 mm at 7·52 A was thus obtained corresponding to an input (L) per cm of length of 47 watts. Writing:

$$p = C\, m/d^2, \qquad (1)$$

where p is measured in atmospheres, m in mg/cm and d in mm, we find for C a value of 130·7. If we carry out the same experiment with a larger value of m we need a smaller input to obtain a pressure of one atmosphere. C is thus a function of the input and increases with it. *Fig. 2* shows the experimental value of C as a function of L. This figure is only valid for diameters of about 40 mm and pressures of about 1 atm, as equation (1) will only be valid in a limited range of m's and d's.

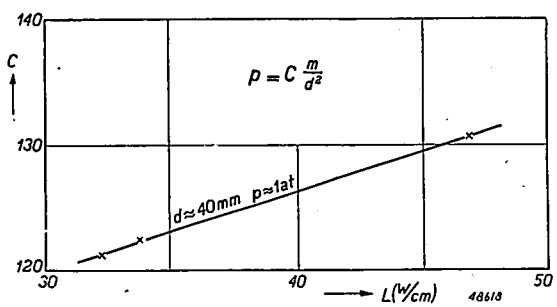


Fig. 2. The value of C in the equation $p = C\, m/d^2$ as a function of the input per cm L, for $d \approx 40$ mm and $p \approx 1$ atm (p in atm, m in mg/cm, d in mm).

For our tube of 41 mm diameter and $m = 12$ we find, as $C = 123$ for $L = 35$ W/cm (fig. 2):

$$p = \frac{123 \times 12}{41^2} = 0{\cdot}88 \text{ atm.}$$

b) A tube of 41 mm inner diameter (actually the same tube that was used in [6])) was provided with a small side tube (4 mm inner diameter and 20 mm length) and filled with 17·3 mg *Hg* per cm of length. The side tube was enclosed in a strong silver tube, to which, insulated by a sheet of mica, a heating wire was connected. A chromel-alumel thermocouple of 100 μ was placed between the side tube and the silver tube. The discharge was then caused to take place and the temperature of the side tube kept lower than that of the rest of the discharge tube, as could be ascertained from the fact that no mercury condensed except in the side tube. In this case the mercury pressure is determined by the temperature of the side tube. The discharge was made at 6 A and the temperature T of the side tube and the voltage V_{ss} between the exploring electrodes were measured at different equilibrium states (dependent on the heating of the silver tube). Plotting V_{ss} as a function of T we find 346·5 °C at $V_{ss} = 151$ V, this being the voltage occurring in [6]) at $m = 12$. The mercury pressure at this temperature (using the critical tables) is 630 mm or 0·83 atm.

As we consider this second method less accurate than the first, we will assume for the pressure in a tube of 41 mm inner diameter with 12 mg *Hg*/cm and 35 W/cm a value of 0·87 atm. This pressure will depend on the absorption of the mercury radiation by the tube wall, as the temperature of the tube wall depends thereon. As we used the same hard glass in all our experiments this dependency does not enter into our considerations.

2B. *The temperature of the tube wall*

The temperature of the outside of the tube wall was measured with a *Pt* resistance thermometer. By means of two springs a *Pt* wire of 100 μ was arranged tightly around the tube halfway between the electrodes. As the wire is thin, its temperature T_u will practically be that of the outside of the tube wall. The resistance of the wire was calibrated as a function of temperature by placing the wire in a thick-walled copper box, which was heated from the outside, so that the temperature in it was the the same everywhere. This temperature was determined by means of a chromel-alumel thermocouple.

T_u was determined for the tube with the side tube (the same as used under 2A, b) for 6 A and $V_{ss} = 151$ V. A temperature of 755 °K was found. For another tube of 41·4 mm outer diameter and $m = 12$, T_u was measured as a function of L. The results of these measurements are given in *fig. 3.*

To derive the temperature of a tube of 43 mm outer diameter at 35 W/cm we read in fig. 3 the temperature for $L = 35 \times 41{\cdot}4/43 = 33{\cdot}7$ W/cm, which gives $T_u = 757$ °K.

We may therefore assume a mean value of 756 °K. As in 2A, this value holds for the hard glass used in our experiments.

As we need the temperature T_w of the inside of the tube wall, we have to estimate the temperature difference between the inner and outer sides of the glass wall. This difference depends on the energy passing through the wall by conduction of heat and on the heat conductivity of the glass used.

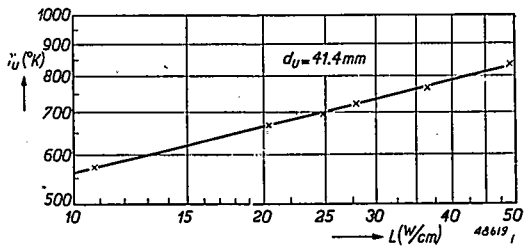


Fig. 3. Temperature T_u of the outside of the glass tube of a high-pressure mercury discharge as a function of the input per cm, for a tube of an outer diameter of 41·4 mm, filled with 12 mg *Hg* per cm of length.

To determine the first quantity we proceed as follows: The total radiation S per cm of length of a high-pressure mercury discharge in a quartz tube of large diameter amounts to [12]):

$$S = 0{\cdot}72\ (L\text{—}10)\,. \tag{2}$$

With $L = 35$ W/cm we find a total radiation of 18 W/cm. Taking into account the absorption in the quartz wall and in the air between the tube and the thermopile (formation of ozone), we may assume that say 19 watts arrives at the inner surface of the tube wall in the form of radiation. The remainder, thus 16 watts, arrives as conducted heat. The same holds for our hard glass tube. In the glass the radiation will be partly absorbed. With 10% absorption we obtain a heat flow of 16 to 18 watts through the glass wall and, because most of the absorption will occur near the inner side of the wall, we assume a constant heat flow of 17·5 W = 4·2 cal/sec through the glass wall per cm of tube length.

As the thickness of the glass wall is small compared with the mean diameter of 4·2 cm, we may write:

$$\pi \times 4{\cdot}2\ \lambda\ \mathrm{d}T/\mathrm{d}r = 4{\cdot}2\,. \tag{3}$$

The value of λ at a temperature of 760 °K is estimated from measurements of Seemann [13]) as $4{\cdot}2.10^{-3}$. With (3) we find $dT/dr = 76$. As the wall is 1 mm thick, we have $T_w - T_u = 7{\cdot}6°$, so that with $T_u = 756$ °K we obtain: $T_w = 764$ °K.

2C. *The radiation of the lines 5770/91 Å*

Schouten and Smit [11]) measured $A.g$ (A = transition probability, g = statistical weight of the initial level) for the 5770 line and $A.g$ for the two 5790/91 lines. The heights of the initial levels of these three lines being practically the same (8·845 eV) it is sufficient for our purpose to measure the absolute intensity of these three lines together. This was performed in three ways:

a) The lumen output and the candle power of the discharge were measured through a filter containing the Schott filters OG2 (2 mm) + VG3 (1 mm) + BG18 (2 mm). This filter absorbs practically all the light except that of the yellow lines. The transmission of the filter for 5780 Å radiation was determined to be 11·2%. The ratio of the lumen output to the candle power was found to be 13·1. As self-absorption causes a value smaller than 4π, the fact that a larger value was found must be attributed to errors in the measurements. The value of 13·1 makes it improbable that self-absorption is of importance for the yellow lines of our discharge, as was already concluded from the measurements of [6]). As an average from the lumen output and the candle-power measurement a radiation of 2·46 W/cm was found for the lines 5770/91 Å.

b) With a thermopile. In front of the thermopile we placed successively a cuvette with 1 cm of water and a filter comprising the filters as used under a), to which was added the Schott filter BG19 (2 mm), and again a cuvette with 3 cm water. The only radiation of the lamp reaching the thermopile originates from the yellow lines. The transmission of the filter combination for 5780 Å was 9·3%, whereas the two cuvettes with water transmitted 80% of the 5780 Å radiation. The thermopile was calibrated with a vacuum incandescent lamp, the radiated energy of which was found by substracting from the input the loss by conduction and convection of the bulb. A radiation of 2·54 W/cm was found for the three yellow lines.

c) Comparision with the U.V. radiation standard. Krefft, Rössler and Rüttenauer [14]) described a special high-pressure mercury discharge and measured the radiation of several lines. For the three yellow lines a mean value of 54 μW/cm^2 is found at a distance of 1 metre. We constructed some of these lamps and compared the 5770/91 Å radiation with that of the discharge under consideration, with the aid of the filter combination used under a) and a photo cell. A radiation of 2·10 W/cm was found for the three lines.

As an average of these three methods we thus obtain a radiation of 2·4 W/cm in the lines 5770/91 Å.

3. *The temperature trend*

The excitation temperature of the discharge path is known from fig. 6 of [6]) if T_{ex} (0) is known. In *fig. 4* we reproduce this figure. As beyond $r = 1$ cm the light is too weak, the trend of $T_{ex}(r)$ was only measured between $r = 0$ and $r = 1$ cm.

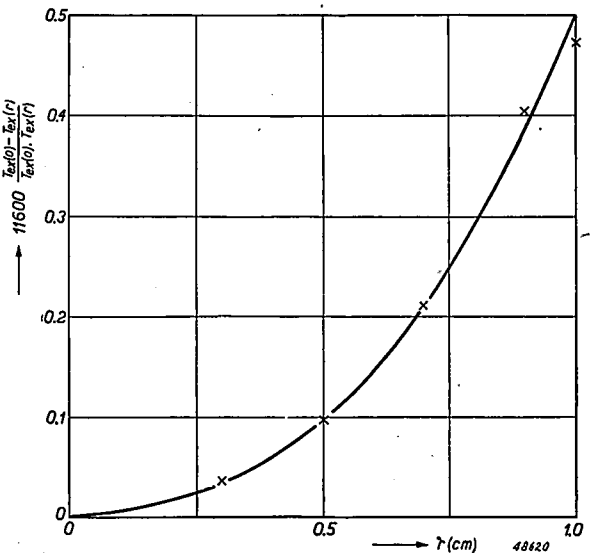


Fig. 4. The trend of the excitation temperature as a function of the distance to the axis for a discharge in a tube of 41 mm inner diameter with 12 mg of *Hg* vaporized per cm of length at an input of 35 W/cm (fig. 6 of [6])).

If T_g (1) and T_g (2·05) are known, the gas temperature T_g (r) between $r = 1$ and $r = 2{\cdot}05$ (the wall) may be estimated, by assuming that a constant amount of energy W is conducted through this cylinder:

$$W = -2\pi r\, \lambda(T_g)\, dT_g/dr\,. \tag{4}$$

Writing:

$$\lambda(T_g) = \lambda_0\, T_g^{\,k} \tag{5}$$

we obtain from (4):

$$\frac{2\pi\lambda_0}{W(k+1)}\,[T_g^{\,k+1}(r_1) - T_g^{\,k+1}(r_2)] = \ln\frac{r_2}{r_1}. \tag{6}$$

With $r_1 = 1$ cm, $r_2 = 2{\cdot}05$ cm, T_g (2·05) = 764 °K and T_g (1·0) from

fig. 4, the factor λ_0/W becomes known and the temperature between $r = 1$ and $r = 2{\cdot}05$ may be derived from (6). The value of λ_0 is only needed to determine W but not to derive the temperature trend.

We choose λ as much as possible in accordance with measurements of Schleiermacher [15]) and Braune, Basch and Wentzel [16]), as well as with the theory of Hasse and Cook [17]). Braune, Basch and Wentzel measured the viscosity η of mercury vapour between 491 and 883 °K. The coefficient of heat conductivity λ may be deduced from η by the equation $\lambda = 2{\cdot}52\ C_v\eta$. In this temperature range a value of 1·07 is found for k. If we calculate λ between 750 °K and 5000 °K using the theory of Hasse and Cook [17]), we find for k a value of 1·19. As an average we may thuse use $k = 9/8$. We do not need the value of λ_0, in order to determine the temperature trend between $r = 1$ and $r = 2{\cdot}05$, when T (1) and T (2·05) are known. As, however, we need λ_0 in 5A, we will consider this quantity too. Schleiermacher [15]) found at 476 °K, $\lambda = 1{\cdot}85.10^{-5}$, whereas Braune Basch and Wentzel [16]) give $1{\cdot}71.10^{-5}$ at this temperature. The mean value of $1{\cdot}78.10^{-5}$ leads to $\lambda_0 = 1{\cdot}75.10^{-8}$. We shall therefore assume that the heat conductivity λ of mercury vapour follows the equation:

$$\lambda = \tfrac{7}{4}.10^{-8}\ T_g^{9/8} \quad (T_g \text{ in } °\text{K}) . \tag{7}$$

The increase of the heat conductivity as a result of the energy carried away by the electrons is of no importance for our considerations, because in the discharge path where the conductivity also depends on the concentration of electrons, the temperature trend is determined experimentally with the aid of fig. 4, whereas in the surrounding space between $r = 1$ and $r = 2{\cdot}05$ the degree of ionization is so small that the contribution of the electrons to the heat conduction is negligible.

4. *The consequence of* $T_{ex} = T_g$

We now assume that the situation of fig. 1*a* applies, so that the excitation temperature and the gas temperature are the same and the electron temperature is higher. We shall consider the result for: a) the excitation and the gas temperatures, b) the cross-section of the mercury atom which in combination with Saha's equation has to be used in Langevin's formula, in order to achieve the right current density, and c) the pressure as deduced from the gas temperature and the amount of mercury vaporized.

a) *Excitation temperature and gas temperature*

We calculate these temperatures on the basis of the 2·4 W/cm energy radiated in the three yellow lines (2C), the temperature trend in the discharge path as measured in [6]), the value of $\Sigma A.g = 1{\cdot}35.10^9$ for these three lines as measured by Schouten and Smit [11]), the pressure as

measured under 2A and assuming a Boltzmann distribution for the population of the mercury levels. The number of transitions per sec is given by:

$$N = n_a \, \Sigma \, A.g. \exp(-eV/kT_{ex}), \tag{8}$$

where n_a represents the number of atoms in the ground state and V the height of the initial level. As the number of atoms per cm^3 at p atm and T_g °K equals $7{\cdot}4.10^{21}\, p/T_g$ and this equals practically n_a, as the number of atoms that are not in the ground state is negligible, we find for the total number of transitions per cm of arc length:

$$\int_0^{2{\cdot}05} \frac{2\pi r \times 0{\cdot}87 \times 7{\cdot}4.10^{21} \times 1{\cdot}35.10^{9} \exp(-102500/T_{ex})}{T_g} \, dr. \tag{9}$$

As the energy of 1 quantum of 5780 Å radiation is $3{\cdot}40.10^{-12}$ erg, we find from (9) for the radiation S of the yellow lines in watts per cm of arc length:

$$S = 1{\cdot}86.10^{13} \int_0^{2{\cdot}05} \frac{r \exp(-102500/T_{ex})}{T_g} \, dr. \tag{10}$$

With the trend of $T_{ex}(r)$ from fig. 4 and $T_g = T_{ex}$, we may determine the integral of (10) for every value of T_{ex} (0), and thus find the radiation as a function of T_{ex} (0). The value of T_{ex} (0) giving a radiation of 2·4 W/cm appears to be 5370 °K. In *fig. 5a* we have drawn $r \exp(-102500/T_{ex})/T_g$ as a function of r for T_{ex} (0) = 5370 °K. The value of the integral is 13.10^{-14}, so that with (10) : S = 2·4 W/cm. In table II the temperature is given for several values of r, derived from fig. 4 with T_{ex} (0) = 5370 °K.

TABLE II

r (cm)	0·0	0·2	0·4	0·6	0·8	1·0
T_{ex} (°K) = T_g	5370	5325	5220	5035	4720	4365

b) *The cross-section of the mercury atom*

Assuming temperature equilibrium, the temperature figuring in the Saha equation must be the same as that occurring in the Boltzmann equation. With the temperature derived from the intensity of the yellow lines, we may therefore calculate the cross-section σ of the mercury atom, which must be used in the mobility equation of Langevin [18]), in order to find the right current density i at the measured electric gradient G = 5·81 V/cm.

With:

$$i = e\,u\,G\,n_e/300\,, \tag{11}$$

$$u = \frac{0{\cdot}75\,e\,\lambda_e}{\sqrt{8kT_e\,m_e/\pi}} = 64.10^{10}\,n_a^{-1}\,\sigma^{-1}\,T_{ex}^{-1/2}, \qquad \text{(Langevin)} \tag{12}$$

and

$$n_e/n_a = x = 1{\cdot}15.10^{-3}\,p^{-1/2}\,T_{ex}^{3/4}\exp\,(-eV_i/2kT_{ex})\,, \qquad \text{(Saha)} \tag{13}$$

we find:

$$i = 2{\cdot}45.10^{6}\,eG\,T_{ex}^{3/4}\,p^{-1/2}\,\sigma^{-1}\exp\,(-60000/T_{ex})\,, \tag{14}$$

(in the Langevin equation we put $T_{ex}^{1/2}$ for $T_e^{1/2}$. This causes only a small error).

With $p = 0{\cdot}87$ atm and $G = 5{\cdot}81$ V/cm, we have, as the current amounted to 6 A:

$$2{\cdot}42.10^{-12}\,\sigma^{-1}\int_0^{2{\cdot}05} 2\pi r\,T_{ex}^{3/4}\exp\,(-60000/T_{ex})\,\mathrm{d}r = 6{\cdot}0\,. \tag{15}$$

For σ we thus obtain:

$$\sigma = 2{\cdot}53.10^{-12}\int_0^{2{\cdot}05} r\,T_{ex}^{3/4}\exp\,(-60000/T_{ex})\,\mathrm{d}r\,. \tag{16}$$

With T_{ex} from table II we may determine the value of the integral. In *fig. 5b*, $rT_{ex}^{3/4}\exp\,(-60000/T_{ex})$ is drawn as a function of r and the integral is found to be 17.10^{-4} so that: $\sigma = 4{\cdot}3.10^{-15}$ cm².

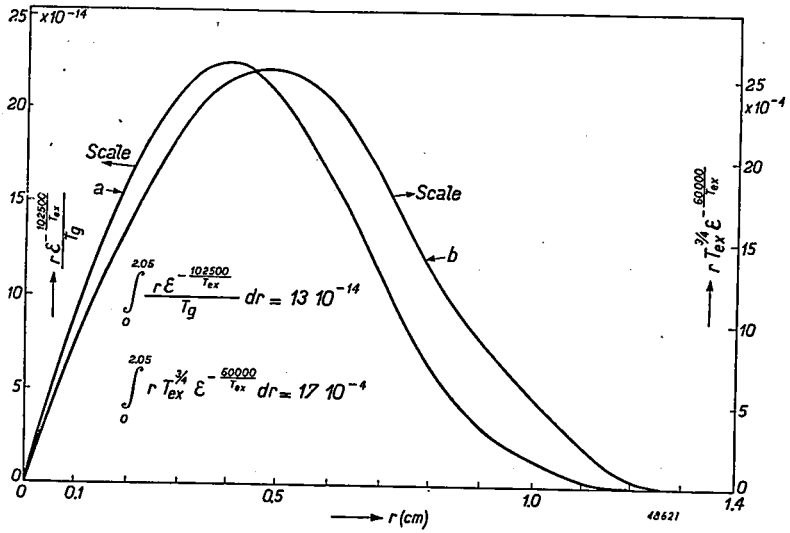


Fig. 5. Determination of the integrals of (10) (curve *a*) and (16) (curve *b*) for the case that the excitation temperature and the gas temperature are the same, and have values as given in table II.

This is a reasonable value. From data of different authors [20]) we have assumed in previous papers [19]) $\sigma = 5.10^{-15}$ cm^2 as the most probable value (with an uncertainty of a factor 2 or 3). If we had proceeded by assuming $\sigma = 5.10^{-15}$ and calculated T_{ex} (0), we would have found the right current at T_{ex} (0) = 5430 °K in close agreement with $T_{ex}(0)$ = 5370 °K deduced from the intensity of the yellow lines. We may therefore conclude that the temperature figuring in the Boltzmann equation is, within the errors of the measurements, the same as the temperature figuring in the Saha formula.

c) *The pressure*

As we assumed the gas temperature to be equal to the excitation temperature, the gas temperature is known between $r = 0$ and $r = 1$ cm (table II). Between $r = 1$ and $r = 2{\cdot}05$ the gas temperature is calculated as indicated in section 3 with T_g (1·0) = 4365 °K (table II) and T_g (2·05) = 764 °K (see 2B).

As the number of particles per cm of length is known since 12 mg of mercury is vaporized per cm of tube length, and the degree of ionization is so small that the number of electrons may be neglected compared with the number of atoms, the pressure may be calculated as follows: 1 cm^3 of mercury vapour of T_g °K and 1 atm weighs 200·6 × 273/(22·41 T_g) = 2440/T_g mg. Integrating over the cross-section of the tube we thus have:

$$\int_0^{2\cdot05} \frac{2\,\pi\, r\, p\, 2440}{T_g}\,\mathrm{d}r = 12\,,$$

or

$$p = \frac{78{\cdot}3.10^{-5}}{\int_0^{2\cdot05} \frac{r\,\mathrm{d}r}{T_g}}\,. \qquad (17)$$

In *fig. 6* the value of r/T_g is drawn as a function of r. The area amounts to 83.10^{-5} so that we obtain: $p = 78{\cdot}3/83 = 0{\cdot}94$ atm.

Experimentally (2A) we found a pressure of 0·87 atm. Although the discrepancy of 8% is not very large, we believe that it is larger than may be expected from the possible errors.

If we assume the radiation of the yellow lines to be 2·2 W/cm instead of 2·4, the value of $\Sigma A.g$ for these lines to be $1{\cdot}5.10^9$ instead of $1{\cdot}35.10^9$, the temperature of the inner side of the tube wall to be 750 °K instead of 764 °K, and the heat conductivity proportional to T instead of $T^{9/8}$, all these alterations reduce the calculated pressure, and yet we arrive at a pressure that is too high, namely 0·915 atm. The difference with the

measured pressure is thus reduced to 5%, whereas we estimate the possible error of the measured pressure to be about 3%. If we assume λ to be proportional to $T^{7/8}$ and maintain the other alterations, the deviation is reduced to 3%. Although these alterations cannot be excluded, we believe there is certainly an indication that the gas temperature is lower than the excitation temperature.

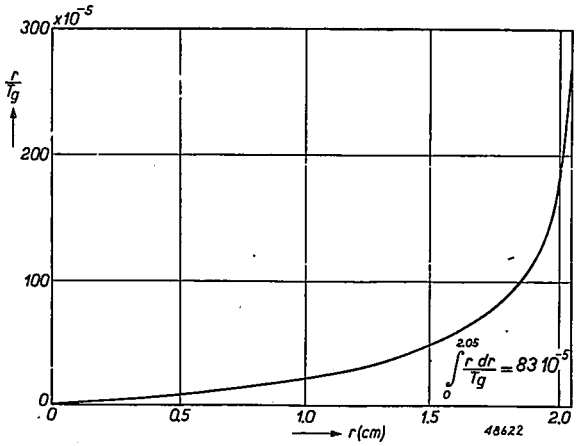


Fig. 6. Determination of the integral of (17), if T_g is given by table II and by equation (6) with T_g (2·05) = 764 °K.

Assuming $T_g = T_{ex}$, so that the trend of T_g may be drawn from fig. 4 we may calculate that value of T_g (0) which gives the measured pressure of 0·87 atm. This gives T_g (0) = 4700 °K (using T_g (2·05) = 764 °K and $k = \frac{9}{8}$), which is 12·5% smaller than the excitation temperature in the axis.

5. *The consequences of* $T_{ex} = T_e$

We now assume that the excitation is caused by collisions between atoms and electrons and that collisions of the second kind also occur mainly between excited atoms and electrons, so that $T_{ex} = T_e$.

5A. *The difference between the gas temperature and the excitation temperature*

If the atoms are not excited by mutual impacts the only energy which has to be transferred by the electrons to the atoms is that which the atoms lose through conduction of heat. According to Cravath [21]) the mean energy Q_1 transferred from the electrons to the atoms through elastic collisions is:

$$Q_1 = 2\, n_e\, n_a\, \delta^2\, (2\pi\, m_e)^{1/2}\, (kT_e)^{3/2}\, m_a^{-1}\, (1 - T_g/T_e)\,, \qquad (18)$$

where Q_1 is expressed in erg/cm^3 sec, and δ is the diameter of the mercury atom. The energy transferred is thus proportional to $(T_e - T_g)$. If we imagine a low gas temperature, Q_1 is large so that T_g will increase. As a result $T_e - T_g$ decreases, so that Q_1 will decrease and as the loss by conduction of heat Q_2 will increase, as a result of the increasing gas temperature a state of equilibrium will be reached where $Q_1 = Q_2$. We shall determine at what value of T_g this equilibrium is reached. As n_e occurs in (18), we have to know the degree of ionization when the gas temperature and the excitation temperature are not the same.

With:

$$\frac{p_i p_e}{p_a} \sim T_e^{5/2} \exp(-eV_i/kT_e),$$

where p_i, p_e and p_a represent the pressure of the ions, the electrons, and the atoms, respectively, we find with $p_e = x p_a\ T_e/T_g$ and $p_i \approx x p_a\ (T_e/T_g)^{1/2}$:

$$x = 1{\cdot}15.10^{-3}\ T_g^{3/4} T_e^{1/2}\ p_a^{-1/2} \exp(-eV_i/2kT_e). \qquad (19)$$

Introducing (19) in (18) we obtain for Q_1 in cal/cm^3 sec for 0·87 atm of mercury:

$$Q_1 = 4{\cdot}6.10^3\ T_e T_g^{-5/4}\ (T_e - T_g) \exp(-60000/T_e) \qquad (20)$$

(for $\frac{1}{4}\pi\delta^2$ we used $4{\cdot}1.10^{-15}$ in accordance with 5B).

In a cylinder of 1 cm height between r and $r + dr$, the energy $2\pi r Q_1 dr$ is transferred to the gas. In the equilibrium state this energy is the difference between the heat conducted through the rings at $r + dr$ and at r. Thus:

$$2\pi r\ Q_1\ dr = -\frac{d}{dr}(2\pi r\ \lambda\ dT_g/dr)\ dr. \qquad (21)$$

From (20) and (21) follows:

$$-\frac{1}{r}\frac{d}{dr}(r\ \lambda\ dT_g/dr) = 4{\cdot}6.10^3\ T_e T_g^{-5/4}\ (T_e - T_g) \exp(-60000/T_e). \qquad (22)$$

T_g is determined by the trend of T_e as given by fig. 4 (as $T_e = T_{ex}$) and by (22). The boundary conditions are that $(dT_g/dr)_{r=0} = 0$ and $(2\pi r\lambda\ dT_g/dr)_{r=1} = 2{\cdot}4$ cal/sec (= 10 watts). This latter condition follows from the interpretation of (2): Of the input L an amount of 10 watts is lost by conduction of heat, whilst the remaining $(L - 10)$ watts is radiated by the discharge. Of this radiation, however, 28% is absorbed in the discharge itself, in the surrounding vapour, in the quartz wall, and in the air in the immediate vicinity of the tube (formation of ozone). The main part of this absorbed energy originates from the two resonance lines [22]). If we neglect the absorption in the discharge itself, the flow of energy by conduction of heat at the edge of the discharge ($r = 1{\cdot}0$ cm) is 10 watts per cm of length.

To solve (22), we integrate once:

$$-2\pi r\lambda \mathrm{d}T_g/\mathrm{d}r = \int_0 2\pi r\, 4{\cdot}6.10^3\, T_e T_g^{-1/4} (T_e - T_g) \exp(-60000/T_e)\, \mathrm{d}r = W(r)\,. \tag{23}$$

If we choose r_1 and r_2 so near to each other that $W(r_1)$ is practically equal to $W(r_2)$, we obtain from (23):

$$T_g(r_1) - T_g(r_2) = \frac{W\left(\frac{r_1 + r_2}{2}\right)}{2\pi\lambda} \ln \frac{r_2}{r_1}\,. \tag{24}$$

We now proceed as follows:
We assume $T_e(0)$ and $T_g(0)$ provisionally ($T_g(0) < T_e(0)$) and determine $T_e(r)$ with the aid of fig. 4. The interval between $r = 0$ and $r = 1{\cdot}0$ cm is divided into 25 equal parts of 0·04 cm. As $(\mathrm{d}T_g/\mathrm{d}r)_{r=0} = 0$, we assume $T_g(0{\cdot}02) = T_g(0)$. With (23) we may determine $W(0{\cdot}02)$ since all the data are known. The energy produced between $r = 0{\cdot}02$ and 0·03 is determined provisionally, as T_g is not yet known between $r = 0{\cdot}02$ and 0·04. This energy added to $W(0{\cdot}02)$ gives the energy flow $W(0{\cdot}03)$ through the cylinder with radius $r = 0{\cdot}03$. Eq. (24) than gives $T_g(0{\cdot}02) - T_g(0{\cdot}04)$. For this we have to know the value of λ. We used $\lambda = \frac{7}{4}.10^{-8}\, T_g^{0{\cdot}8}$ as indicated in section 3. With the knowledge of $T_g(0{\cdot}04)$, the value of $W(0{\cdot}03)$ is determined a second time and $T_g(0{\cdot}04)$ determined definitely. Then $T_g(0{\cdot}08)$ is determined, and so step by step $T_g(0{\cdot}12)$, *etc.* Finally we find at $r = 1{\cdot}0$ the values $T_g(1{\cdot}0)$ and $W(1{\cdot}0)$. If $W(1{\cdot}0)$ is less than 10 watts, we assume a smaller value for $T_g(0)$ and the unalterated value for $T_e(0)$. According to (23) the values of W become larger now, because $(T_e - T_g)$ is larger. There is thus a value of $T_g(0)$ to be found that with the assumed $T_e(0)$ value gives $W(1{\cdot}0) = 2{\cdot}4$ cal/sec. With $T_e(r)$ and $T_g(r)$ known, we may calculate the radiation of the three yellow lines with (10) ($T_e = T_{ex}$). If we find that this radiation is smaller than 2·4 watts per cm of arc length, we assume $T_e(0)$ higher, determine $T_e(r)$ and $T_g(r)$ as pointed out above and finally arrive at $T_e(r)$ and $T_g(r)$ which: *a*) satisfy (23); *b*) give a radiation of the yellow lines of 2·4 W/cm, and *c*) lead to an energy of 10 W/cm transferred to the atoms by elastic collisions with the electrons. In table II these temperatures are given in the second and third columns.

We thus find a difference of 170° between electron temperature and gas temperature in the axis, whilst at the edge of the discharge this difference is doubled.

The fact that $T_{ex}(0)$ is 20° lower than in 4A (table II) is due to the greater number of atoms in the ground state as a result of the lower gas temperature (in (10) $T_g < T_{ex}$).

TABLE III

r (cm)	$T_e = T_{exc}$ (°K)	T_g (°K)	$T_g(0)/T_g(r) = b$	$T_e(0)/T_e(r) = b'$
0	5350	5180	1·000	1·000
0·2	5305	5130	1·010	1·008[5]
0·4	5205	4985	1·039	1·028
0·6	5020	4740	1·093	1·065
0·8	4705	4395	1·179	1·137
1·0	4350	4010	1·292	1·230

5B. *The cross-section of the mercury atom*

With this temperature trend of T_e and T_g (table III), we may again determine the cross-section σ of the mercury atom that gives the right current at the gradient observed, if we assume that the degree of ionization may be determined with the Saha equation. Quite analogous to 4B we find (now using (19) instead of (13)): $\sigma = 4{\cdot}1.10^{-15}$, which value was used in (18) and (20). This value is just as probable as the value $4{\cdot}3.10^{-15}$ found in 4B, so that from this no conclusion may be drawn as to the question which of the figures 1*a* and 1*b* is the more probable.

5C. *The pressure*

Adopting T_g between $r = 0$ and $r = 1{\cdot}0$ from table III and calculating T_g between $r = 1{\cdot}0$ and $r = 2{\cdot}05$ as indicated in section 3, we may compute the pressure in the same way as was done in 4C. The result is a calculated pressure of 0·89 atm. which lies within the possible experimental error of the measured value of 0·87 atm.

From (6) we may also determine the quantity W, that is the energy flowing between $r = 1$ and $r = 2{\cdot}05$ trough conduction of heat. With λ according to (7), we find with T_g (1·0) = 4010 (table III) and T_g (2·05) = 764° for W a value of 3·18 cal/sec or 13·3 watts. That this is larger than the energy we assumed to flow through the cylinder at $r = 1{\cdot}0$ (namely 10 watts) is not astonishing, for, between $r = 1{\cdot}0$ and 2·05 a part of the radiated energy is absorbed and continues its way to the wall as conducted heat. In 2B we estimated that 16 watts arrives at the tube wall as conduction of heat, so that 13·3 watts is a fair mean value between the energies conducted at $r = 1{\cdot}0$ and $r = 2{\cdot}05$. That this energy of 13·3 watts is in good agreement with the expected value means that equation (7) is a good approximation for λ.

If we determine the value of W in 4C, we find 15·8, a rather high value.

5D. *The excitation temperature derived from the intensity decrease from the axis to the wall of lines with initial levels at different heights.*

In [6]) a value of 8400 ± 1800 °K was found for the excitation temperature in the axis of the same discharge for which we now have found a value of 5350-5370 °K. The determination in [6]) followed from the equation

$$\ln \frac{I_1(0)}{I_1(r)} = 11600\, V_1 \frac{T_{ex}(0) - T_{ex}(r)}{T_{ex}(0) \,.\, T_{ex}(r)} - \ln \frac{T_g(0)}{T_g(r)} . \qquad (25)$$

We determined the values of

$$11600\ (T_{ex}(0) - T_{ex}(r))/T_{ex}(0) \,.\, T_{ex}(r) = a(r)$$

$$\text{and of } T_g(0)/T_g(r) = b(r) .$$

$T_{ex}(0)$ was calculated by assuming:

$$T_g(0)/T_g(r) = T_{ex}(0)/T_{ex}(r) . \qquad (26)$$

With this assumption we have:

$$T_{ex}(0) = \frac{11600\, b(r_0) - 1}{a(r_0)} . \qquad (27)$$

The value of $a(r)$ is given in fig. 4, whereas $b(r)$ was very uncertain and caused the large scattering of the temperatures derived from different values of r.

According to table III the assumption (26) does not hold. We may, however, apply a correction. For this purpose we have plotted in *fig. 7* the values of $T_g(0)/T_g(r) = b$ against $T_{ex}(0)/T_{ex}(r) = b'$ according to table III. For each measured value of b we may thus read the value of b'. Using in (27) this value of b' instead of b, we find the corrected value for $T_{ex}(0)$. In table IV we give the data for the same values of r_0 which were used in [6]):

TABLE IV

r_0	a	b	b'	$\frac{11600\,(b'-1)}{a} = T_{ex}(0)$
0·3	0·0349 ± 0·0003	1·0175	1·012	4000
0·5	0·0969 ± 0·0042	1·0445	1·030	3600
0·7	0·2115 ± 0·0005	1·197	1·151	8270
0·9	0·403 ± 0·014	1·501	(1·38)	10900
1·0	0·471 ± 0·016	1·228	1·177	4350
				6250 ± 1450 (mean error)

The value of b' for $r = 0{\cdot}9$ could not be read from fig. 7, because with the temperatures of table III the largest value of b is 1·23. The value 1·38 was estimated by analogy with the values of b and b' for $r_0 = 0{\cdot}7$ and $r_0 = 1{\cdot}0$.

The excitation temperature of 6250 ± 1450 °K, which is now found from the measurements in [6]), is consistent with the value of 5350 °K derived from the intensity of the yellow lines. The method of [6]) remains, of course, as regards the accuracy, much inferior to the method used in this paper, owing to the uncertainty of the determination of b. An advantage of the first method, however, is that it is independent of other measurements, whereas the second method depends on the value of $\Sigma A.g$ for the yellow lines, which is determined with the aid of the temperature of a carbon arc [11]).

We may thus conclude that the calculated pressure (4C *and* 5C) *and the temperature determination from the intensity decrease* (5D) *are better described with the situation of fig. 1b than with that of fig. 1a.*

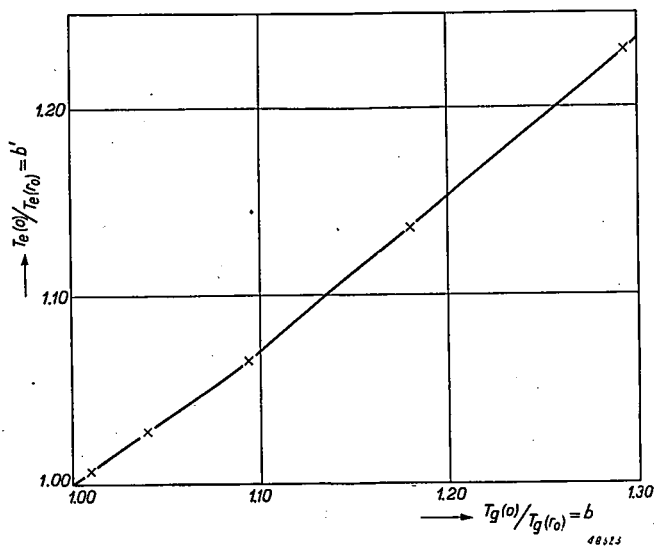


Fig. 7. b' as a function of b according to table III.

6 *The excitation process*

6A. *The relative excitation probability by electrons and by atoms*

In order to ascertain whether we may draw conclusions from the excitation process as to which of the situations of fig. 1*a* and 1*b* is the more probable, we will consider the excitation probabilities of the $3D$ and $3d$ levels (the initial levels of the yellow lines) from the ground state by mutual collisions of the atoms and by collisions of atoms with electrons.

The number of collisions N_{aa} per cm^3 and per sec between atoms with a relative energy between η and $\eta + d\eta$ is given by: [23])

$$N_{aa} = \frac{2n_a^2\delta^2}{(kT)^{1/2}}\left(\frac{\pi}{m_a}\right)^{1/2} \exp(-\eta/kT)\,\eta d\eta\,, \tag{28}$$

where

$$\eta = \tfrac{1}{4}\, m_a V^2\,. \tag{29}$$

The number of collisions N_{ae} between atoms and electrons with relative energies between η' and $\eta' + d\eta'$ is:

$$N_{ae} = \frac{n_a n_e \delta^2}{(kT)^{1/2}}\left(\frac{\pi}{2m_e}\right)^{1/2} \exp(-\eta'/kT)\,\eta' d\eta' \tag{30}$$

where

$$\eta' = \tfrac{1}{2}\, m_e V^2\,. \tag{31}$$

η and η' are the maximum quantities of energy that may be transferred by inelastic collisions. Supposing the excitation probability W to be proportional to $\eta - \eta_0$, if η_0 represents the excitation energy, we approximate:

$$W_a = q_a(\eta - \eta_0) \tag{32}$$

for collisions between atoms, and

$$W_e = q_e(\eta' - \eta_0) \tag{33}$$

for collisions between electrons and atoms. For the ratio of the number of excitations by collisions with atoms, to the number of excitations by collisions with electrons, we may thus write:

$$\frac{N_{aa}W_a}{N_{ae}W_e} = \frac{2\sqrt{2}\, m_e^{1/2} n_a q_a \int_{\eta_0}^{\infty} \eta(\eta-\eta_0)\exp(-\eta/kT)\,d\eta}{m_a^{1/2} n_e q_e \int_{\eta_0}^{\infty} \eta'(\eta'-\eta_0)\exp(-\eta'/kT)\,d\eta'} = 2\sqrt{2}\left(\frac{m_e}{m_a}\right)^{1/2}\frac{n_a}{n_e}\frac{q_a}{q_e} \tag{34}$$

For mercury $(m_e/m_a)^{1/2} = 1/600$. n_a/n_e varies from fully 10^3 in the axis to fully 10^4 at $r = 1$ cm. The value of q_a/q_e is uncertain because q_a is not known for mercury vapour. The excitation probability for the rare gases by collisions with atoms of rare gases or alkali atoms is, however, very small compared with the excitation probability for impacts with electrons [24]). If the same holds for mercury we estimate q_a/q_e at 10^{-3} to 10^{-4}, so that $n_a q_a/n_e q_e \approx 1$. According to (34) the excitation by electron impact would then be about 200 times more probable compared with excitation by mutual collisions of the atoms. Assuming detailed balance, the same holds for the collisions of the second kind, forming atoms in the ground state. From these

considerations one would therefore expect that the temperature figuring in the Boltzmann equation is the electron temperature.

Thus fig. 1b would be more probable than fig. 1a, which conclusion was also reached at the end of section 5.

6B. *The number of excitations from the ground state*

We will now consider whether the excitation of the 3 ^3D and 3*d* levels from the ground state by electron impact is frequent enough to make equilibrium probable. As found in 6A the collisions between the atoms give rise to far less excitations, so that these collisions have not to be considered for this purpose. To acquire equilibrium, the number of excitations has to be several times as large as the number of spontaneous transitions, the difference being compensated by collisions of the second kind. This being the case the radiation, which is not in equilibrium because it escapes through the tube wall, represents only a small leakage.

With (30) and (33) we find for the number of excitations:

$$N_{ae}W_e = \frac{n_a n_e \delta^2}{(kT)^{3/2}}\left(\frac{\pi}{2m_e}\right)^{1/2} q_e \int_{\eta_0}^{\infty} \eta'\,(\eta'-\eta_0)\exp\,(-\eta'/kT)\,\mathrm{d}\eta' =$$

$$= \frac{n_a n_e \delta^2}{(kT)^{3/2}}\left(\frac{\pi}{2m_e}\right)^{1/2} q_e(k^2T^2\eta_0 + 2k^3T^3)\exp\,(-\eta_0/kT)\,. \qquad (35)$$

The excitation probability of the 3 ^3D and 3*d* levels for electron impact has been measured by Hanle and Schaffernicht [25]). For the three yellow lines together the excitation probability for $\eta' - \eta_0 = 1$ eV is of the order of 1/100 if for δ the value $3{\cdot}5.10^{-8}$ is used. As 1 eV corresponds to $1/(6{\cdot}24.10^{11})$ erg, we find from (33): $q_e \approx 6\frac{1}{4}.10^9$. Considering the axis of the discharge, where $T = 5350$ °K, we find from (35) with $\eta_0 = 8{\cdot}845$ eV $N_{ae}W_e = 2{\cdot}3.10^{1}$/cm^3 sec. The number of transitions per cm^3 and per sec in the axis of the discharge, needed to arrive at the observed intensity of the yellow lines, is, however, according to (8):

$$1{\cdot}24.10^{18} \times 1{\cdot}35.10^{9}\exp\,(-102500/5350) = 9{\cdot}35.10^{18}\,.$$

The excitation from the ground level by electron impact is thus — at least for such high levels as the 3 ^3D and 3*d* levels — quite insufficient even to compensate the spontaneous transitions. In addition to the excitation from the ground level there will also occur excitation from levels situated between the ground level and the 3 ^3D levels. Each of these levels will supply a contribution of the same order as the ground level, as in (35) n_a becomes smaller by the same factor as the exponential function becomes larger. *There remains thus a discrepancy of a factor 100 to account for the energy radiated in the yellow lines, and of a factor 1000 to explain equilibrium.*

As according to 6A the excitation from the ground level by mutual collisions of the atoms is even 100 times smaller, a direct excitation from the ground level by these collisions is not important for the higher levels. An excitation by recombination of ions and electrons only transposes the difficulty, as the formation of ions by impacts with electrons and atoms is not sufficient to provide the higher levels with the required number of excitations.

7 *Discussion*

From the considerations of section 6 we cannot decide whether the excitation is carried out by collisions with electrons or with atoms. We do not need to consider the excitation by absorption of radiation, because this is no primary excitation and, moreover, does not occur for the higher levels.

The way in which the excitation takes place to such an extent that a Boltzmann equilibrium already exists at this pressure, we cannot explain .That this Boltzmann distribution exists at this pressure is, however, beyond doubt and is proved *e.g.* by the experiments of [6]). The fact that we find straight lines in the figures 3 and 4 of [6]) proves that the population of the levels is proportional to exp $(-eV/kT)$ where T is the same for all levels. It is not proved that the number of excited levels equals $g\ n_a$ exp $(-eV/kT)$, where n_a is the number of the atoms in the ground state. The number of excited atoms might be a certain factor larger or smaller than the Boltzmann equation requires; this factor should, however, be independent of T and have the same value for all levels. It is therefore very probable that the Boltzmann equation will be satisfied totally. The fact that the current density is correctly calculated with the aid of the Saha equation also points in the direction of temperature equilibrium, as does the fact that many conclusions derived from the equilibrium theory are in agreement with the experiments [26]).

We thus have only experimental evidence as to the probability whether fig. 1*a* or fig. 1*b* approximates closer to reality. The experiments indicate that fig. 1*b* is more probable than fig. 1*a*, although we cannot exclude fig. 1*a*. A measurement of T_e would give valuable information. A reliable method to determine T_e in high-pressure arcs is not yet known [27]).

From our experiments we may, however, draw the conclusion that if the gas temperature and the excitation temperature of the high-pressure mercury discharge under consideration are not the same, the difference between them is anyhow small, especially in the axis, where it very probably does not exceed 200°.

Jagodzinsky [28]) found a difference of 0·1° between the gas temperature and the electron temperature in a carbon arc of atmospheric pressure.

In this investigation the gas temperature has obviously been assumed to be independent of the pressure (1-4 cm) and of the input (240-324 W/cm). The variation of the degree of ionization is only taken into account in so far as it is caused by a change of pressure, whereas the variation caused by a change of temperature has been ignored. The conclusion that $T_e - T_g = 0{\cdot}1°$ may therefore be quite wrong.

Although the difference $T_e - T_g$ in the carbon arc may be smaller than in the mercury arc, as a result of the presence of rotation and vibration levels, giving the possibility of exchanges of small energy quantities by inelastic impacts, we believe that the estimations of $(T_e - T_g)$ of 20° and 50° by Mannkopff[10]) and Witte[29]), respectively, are more acceptable than the value of 0·1° estimated by Jagodzinsky.

Mr J. Riemens assisted in performing the measurements.

Eindhoven, July 1944.

REFERENCES

1) A. von Engel & M. Steenbeck, Wiss. Veröff. Siemens W. **10**, 155, 1931; **12**, 74, 1933.
2) W. Elenbaas, Physica **1**, 211, 1934.
3) C. G. Suits, Physics **6**, 190, 315, 1935; Gen. elect. Rev. **39**, 194, 1936.
4) L. S. Ornstein & W. R. van Wijk, Proc. roy. Acad. Amsterdam **33**, 44. 1930 L. S. Ornstein & H. Brinkman, Physica **1**, 797, 19, 1934; H. Brinkman Diss. Utrecht 1937.
5) I. Langmuir & H. Mott-Smith, Gen. elect. Rev. **27**, 449, 538, 616, 762, 810, 1924. I. Langmuir, Z. Phys. **46**, 271, 1928.
6) W. Elenbaas, Physica **9**, 53, 1942.
7) L. S. Ornstein & H. Brinkman, Physica **1**, 817, 1934.
8) P. Schulz, Z. Phys. **114**, 435, 1939.
9) A. Unsöld, Z. Astrophys. **12**, 56, 1936.
10) R. Mannkopff, Z. Phys. **86**, 161, 1933.
11) W. Schouten & J. A. Smit, Physica **10**, 661, 1943.
12) W. Elenbaas, Physica **4**, 413, 1937.
13) H. E. Seeman, Phys. Rev. **33**, 1094, 1929.
14) H. Krefft, F. Rössler & A. Rüttenauer, Z. tech. Phys. **18**, 20, 1937; F. Rössler, Ann. Phys., Lpz. **34**, 1, 1939.
15) A. Schleiermacher, Wied. Ann. **36**, 346, 1889.
16) H. Braune, R. Basch & W. Wentzel, Z. phys. Chem. (A) **137**, 447, 1928.
17) H. R. Hasse & W. R. Cook, Proc. roy. Soc. (A) (London), **125**, 196, 1929.
18) P. Langevin, Ann. d. Chem. et d. Phys. **81**, 5, 245, 1905.
19) W. Elenbaas, Physica **5**, 568, 1938; Physica **6**, 299, 1939.
20) R. R. Brode, Proc. roy. Soc. (A) (London), **125**, 134, 1929; T. J. Killian, Phys. Rev. **35**, 1238, 1930; B. Klarfeld, Techn. Phys. U.S.S.R. **4**, 44, 1937.
21) A. M. Cravath, Phys. Rev. **36**, 248, 1930.
22) R. Mannkopff, Z. Phys. **120**, 301, 1943.
23) R. H. Fowler, Statistical Mechanics, Cambridge (1929)p. 423 & 424.
24) W. Maurer, Phys. Zs. **40**, 161, 1939; V. Neumann, Diss. Jena. 1938; Ann. Phys., Lpz. **34**, 603, 1939.
25) W. Hanle & W. Schaffernicht, Ann. Phys., Lpz. **6**, 905, 1930; W. Hanle & K. Larché, Ergebn. d. ex. Naturw. Bd. 10, p. 308.
26) W. Elenbaas, Physica **2**, 169, 757, 1935; **3**, 12, 1936; **4**, 747, 1937.
27) R. C. Mason, Phys. Rev. **51**, 28, 1937; K. Sommermeyer, Z. Phys. **90**, 232, 1934; S. Gvosdover, Techn. Phys. U.S.S.R. **4**, 717, 1937; F. Borgnis, Ann. Phys., Lpz. **31**, 497, 1938; C. von Fragestein & M. Arndt, Ann. Phys., Lpz. **33**, 532, 1938.
28) H. Jagodzinsky, Z. Phys. **120**, 318, 1943.
29) H. Witte, Z. Phys. **88**, 415, 1934.

R 34 Philips Res. Rep. 2, 42-54, 1947

ON THE THEORY OF EDDY CURRENTS IN FERROMAGNETIC MATERIALS

by H. B. G. CASIMIR 538.541:538.221

Summary

The theory of eddy currents in ferromagnetic materials is developed for the limiting case that the depth of penetration d is small although the product μd is not necessarily small.
First the rigorous solution for the sphere is discussed for this limiting case. Next it is shown that the solution can be obtained from Laplace's equation with a new type of boundary condition. This boundary condition is then applied to a discussion of eddy currents in spheroids. Explicit formulae are found both for small and for very larges values of $\mu d/R$.
Special attention is given to the limits for very long and very flat spheroids, respectively.

1. *Introduction*

It is well known that the theoretical treatment of the eddy currents induced in a conductor by an alternating magnetic field becomes comparatively simple in two limiting cases, which are usually characterized by the condition that the depth of penetration, d, is either very small or very large compared with the dimensions of the conductor. This penetration depth is given by

$$d = \sqrt{\frac{c^2}{2\pi\mu\omega\sigma_e}}, \tag{1}$$

where ω is the angular frequency, μ the permeability and σ_e the conductivity in electrostatic units. If d is very large, the change of magnetic field due to eddy currents is entirely negligible; if it is very small the field distribution outside the conductor can be found from the boundary condition $H_n = 0$ (where H_n is the component normal to the surface of the conductor); the field distribution is identical with that for a diamagnetic body with $\mu = 0$. In the case of a ferromagnetic conductor with a high value of μ however, the condition that d be small is not sufficient: the essential condition is that the total magnetic flux through the body be small, and this flux is proportional to μd rather than to d. If μd is large, even though d is small, the body as a whole will still appear to be paramagnetic. This fact was pointed out to the present author by Dr Snoek, who applied this idea for giving a qualitative explanation of certain peculiarities in the behaviour of iron objects in a high-frequency field.

In the present note we shall study in some detail the limiting case when d is small, but when μd is not necessarily small. We shall study the subject

entirely from a mathematical point of view and we shall not discuss any practical applications. Perhaps our formulae may prove to be also of some practical use.

2. *Fundamental equations*

Neglecting the displacement current — this amounts to assuming that the wavelength of the high-frequency field is large compared with the dimensions of the conductor — we have

$$\operatorname{curl} \mathbf{H} = \frac{4\pi}{c}\,\mathbf{i} = \frac{4\pi\sigma}{c}\,\mathbf{E}\,,$$

$$\mathbf{B} = \mu\mathbf{H}\,,$$

$$\operatorname{div} \mathbf{B} = 0\,,$$

$$\operatorname{curl} \mathbf{E} = -\frac{1}{c}\,\dot{\mathbf{B}}\,,$$

from which it follows

$$\Delta\mathbf{H} = \frac{4\pi\mu\sigma}{c^2}\,\dot{\mathbf{H}} \qquad \text{inside the conductor,}$$

$$\Delta\mathbf{H} = 0 \qquad \text{outside the conductor.}$$

Let us assume that the exterior field $\mathbf{H}_0$ is homogeneous in the region occupied by the conductor and that it is periodic with frequency $\omega/2\pi$:

$$\mathbf{H}_0 = \mathbf{H}_0\,(t = 0)\,e^{j\omega t}\,.$$

Putting

$$d = \sqrt{\frac{c^2}{2\pi\mu\omega\sigma}}\,, \qquad K = (1 + j)/d\,,$$

we have

$$\Delta\mathbf{H} = \varkappa^2\,\mathbf{H} \qquad \text{inside the conductor,}$$

$$\Delta\mathbf{H} = 0 \qquad \text{outside the conductor,}$$

and

$$\operatorname{div} \mathbf{H} = 0 \qquad \text{both inside and outside.}$$

For large distances from the conductor $\mathbf{H}$ must approach $\mathbf{H}_0$:

$$\mathbf{H} \to \mathbf{H}_0 \qquad \text{at infinity,}$$

and to this we must add the usual boundary conditions at the surface of the conductor:

$$(H_{\text{tang}})_{\text{interior}} = (H_{\text{tang}})_{\text{exterior}}$$

$$\mu(H_n)_{\text{interior}} = (H_n)_{\text{exterior}}\,.$$

All these equations remain valid when the common time-factor $\exp(j\omega t)$ is omitted. In the following we shall omit this factor without introducing a new notation.

3. *Solution for a sphere* .

We shall first discuss a case where the complete solution is well known and can be expressed in terms of elementary functions. This will afford a useful example of the various limiting cases and will also serve as a useful check on the results to be obtained later on by a method of approximation. The solution is conveniently expressed by means of a vector potential **A**, from which **B** can be derived by means of the equation

$$\mathbf{B} = \operatorname{curl} \mathbf{A} ,$$

and which satisfies

$$\operatorname{div} \mathbf{A} = 0 .$$

Outside the sphere we have

$$\Delta \mathbf{A} = 0 ,$$

inside the sphere

$$\Delta \mathbf{A} = \varkappa^2 \mathbf{A} .$$

Outside the sphere the field is that of a magnetic dipole, just as in magnetostatics:

$$A_x = -\tfrac{1}{2} H_0 y - \frac{y}{r^3} P ,$$

$$A_y = +\tfrac{1}{2} H_0 x + \frac{x}{r^3} P .$$

Inside the sphere we put

$$A_x = -f(r)\, y ,$$

$$A_y = +f(r)\, x ;$$

this leads to the equation

$$f'' + \frac{4}{r} f' = \varkappa^2 f ,$$

with the solution

$$f = \frac{a}{r} \frac{\mathrm{d}}{\mathrm{d}r} \left(\frac{\sinh \varkappa r}{r} \right) = a\, u(r) ,$$

where a is a constant to be determined from the boundary conditions. As a matter of fact the boundary conditions lead to

$$H_0 + 2p = 2\, a\, u(R) ,$$

$$H_0 - p = \frac{2a}{\mu} \left\{ u(R) + \frac{R}{2} u'(R) \right\} ,$$

where

$$p = P/R^3 ,$$

and R is the radius of the sphere, from which both p and a can be determined. For p we find

$$p = \frac{\mu v(R) - 1}{\mu v(R) + 2} H_0 , \tag{2}$$

where

$$v(R) = \frac{2u(R)}{2u(R) + R\, u'(R)} .$$

It is interesting to notice that (2) is identical with the corresponding static formula, if we interpret $\mu\, v(R)$ as an effective permeability.

We are mainly interested in the case $d << R$, that is $|\varkappa R| >> 1$. Then

$$u(R) \approx \frac{\varkappa e^{\varkappa R}}{R^2}$$

and

$$u'(R) \approx \frac{\varkappa^2 e^{\varkappa R}}{R^2},$$

so that

$$v(R) \approx \frac{2}{\varkappa R}$$

and

$$p = \frac{(1-j)\,(\mu d/R) - 1}{(1-j)\,(\mu d/R) + 2} H_0. \tag{3}$$

The energy dissipation is proportional to the imaginary part of p, given by

$$\mathrm{Im}\,(p) = \frac{-3\,\mu d/2R}{(1+\mu d/R)^2 + 1} H_0.$$

As a matter of fact we find for W, the heat developed per unit time,

$$W = -\tfrac{1}{2}\omega\,\mathrm{Im}\,(p)\,H_0 = \tfrac{3}{4} H_0^2 R^3 \omega \frac{\mu d/R}{(1+\mu d/R)^2 + 1}. \tag{4}$$

It is now evident that even if d/R is small there are two entirely different limiting cases. If $\mu d/R$ is small we have

$$p = -\tfrac{1}{2} H_0$$

and

$$W = \tfrac{3}{8}\omega R^3 H_0^2 \cdot \frac{\mu d}{R}. \tag{5}$$

In this case the field distribution outside the sphere is that for a diamagnetic body, with $\mu = 0$. On the other hand if $\mu d >> R$ we have

$$p = H_0$$

and

$$W = \tfrac{3}{4}\omega R^3 H_0^2 \cdot \frac{R}{\mu d}. \tag{6}$$

The field outside the sphere agrees with that for a paramagnetic body with $\mu >> 1$. In this case an increase of μ leads to a decrease of the heat developed, but an increase of σ leads to an increase of W. In the other limiting case these things are just the other way round. For given values of σ and ω the quantity W considered as a function of μ passes trough a maximum value.

4. *Boundary conditions for the case of small penetration depth*

We shall now show that assuming $d \ll R$ we can find the field outside the conductor from the equation $\Delta \mathbf{H} = 0$ together with an appropriate boundary condition. This condition we shall now proceed to derive.

Let us consider a conductor bounded by a surface with radii of curvature much larger than d but otherwise arbitrary. Then inside this conductor the following equation holds along any normal to this surface:

$$\mathbf{H} = \mathbf{H}^{(s)}\, e^{-(1+j)n/d},$$

where $\mathbf{H}^{(s)}$ is the value at the surface and n the distance from the surface measured along the normal. It follows that

$$\left(\frac{\partial \mathbf{H}}{\partial n}\right)_{n=0} = -\frac{1+j}{d}\,\mathbf{H}^{(s)}.$$

On the surface we draw a closed curve C enclosing a surface element ΔS. We erect the normal in the inward direction in all points of ΔS and plot on each normal a distance $\mathrm{d}n$. In this way we obtain a surface element $\Delta S'$. To the volume bounded by ΔS, $\Delta S'$ and the surface F, generated by the normals in the points of C, we apply the condition of zero total flux. The flux entering through ΔS is given by

$$-\int\!\!\int_{(\Delta S)} H_n^{(s,e)}\, \mathrm{d}\sigma,$$

where $H_n^{(s,e)}$ is the value at the surface of the component of the field outside the conductor in the direction of the outward normal. The flux entering through $\Delta S'$ is

$$+\int\!\!\int_{(\Delta S')} B_n'\, \mathrm{d}\sigma,$$

where B_n' is the value of B_n at a distance $\mathrm{d}n$ from the surface. We have

$$\int\!\!\int_{(\Delta S')} B_n'\, \mathrm{d}\sigma - \int\!\!\int_{(\Delta S)} H_n^{(s,e)}\, \mathrm{d}\sigma \approx -\frac{1+j}{d}\, H_n^{(s,e)}\, \Delta S \cdot \mathrm{d}n + H_n^{(s,e)}\,(\Delta S' - \Delta S):$$

Since d is very small compared with the radii of curvature the last term is negligible. The flux entering through F is given by

$$-\mathrm{d}n \int_{(C)} B^{t,n}\, \mathrm{d}s = -\mu\, \mathrm{d}n \int_{(C)} H_{t,n}^{(s,e)}\, \mathrm{d}s,$$

where $H_{t,n}^{(s,e)}$ is the component of the tangential field outside the conductor in the direction of the outward normal of C. So we obtain the equation

$$\frac{1}{\Delta S}\int_{(C)} H_{t,n}^{(s,e)}\, \mathrm{d}s + \frac{1+j}{\mu d}\, H_n^{(s,e)} = 0.$$

Introducing the surface divergence, defined by

$$\operatorname{Lim}_{\Delta S \to 0} \frac{1}{\Delta S} \int_{(C)} H_{t,n}^{(s,e)} \, \mathrm{d}s = \operatorname{div}\, {}^{(S)}H_t^{(s,e)} ,$$

and simplifying the notation, we find

$$\frac{1+j}{\mu d} H_n + \operatorname{div}\, {}^{(S)}H_t = 0 . \tag{7}$$

This equation contains only the boundary values of the field outside the conductor. In order to derive an expression for div ${}^{(S)}H_t$ we erect in all points of ΔS the outward normal and plot a distance $\mathrm{d}n$ to obtain a surface element $\Delta S''$ and a surface F' generated by the normals through C. Applying the flux condition we obtain

$$-\iint_{(\Delta S'')} H_n \, \mathrm{d}\sigma + \iint_{(\Delta S)} H_n \, \mathrm{d}\sigma - \mathrm{d}n \int_{(C)} H_{t,n} \, \mathrm{d}s = 0$$

or, for small ΔS,

$$-\Delta S \frac{\partial H_n}{\partial n} - H_n \frac{\Delta S'' - \Delta S}{\mathrm{d}n} - \int_{(C)} H_{t,nC} \, \mathrm{d}s = 0$$

where all values are again taken at the surface of the conductor. Now

$$\frac{\Delta S'' - \Delta S}{\Delta S \cdot \mathrm{d}n} = \frac{1}{R_1} + \frac{1}{R_2}$$

where R_1, R_2 are the main radii of curvature, and in this case the term containing this expression is not negligible. So we find

$$\operatorname{div}\, {}^{(S)}H_t = -\left[\frac{\partial H_n}{\partial n} + \left(\frac{1}{R_1} + \frac{1}{R_2}\right) H_n\right],$$

whence

$$H - \frac{\mu d}{1+j}\left[\frac{\partial H_n}{\partial n} + \left(\frac{1}{R_1} + \frac{1}{R_2}\right) H_n\right] = 0 . \tag{8}$$

Introducing a magnetic potential Φ we find that Φ satisfies Laplace's equation

$$\Delta \Phi = 0$$

with the boundary condition

$$\frac{\partial \Phi}{\partial n} - \frac{\mu d}{1+j}\left[\frac{\partial^2 \Phi}{\partial n^2} + \left(\frac{1}{R_1} + \frac{1}{R_2}\right)\frac{\partial \Phi}{\partial n}\right] = 0 . \tag{9}$$

An alternative form of the boundary condition is obtained by introducing surface co-ordinates u_1, u_2, so that the line-element on S is given by

$$\mathrm{d}s^2 = \Sigma\, g_{ik} \, \mathrm{d}u^i \, \mathrm{d}u^k .$$

Then

$$\operatorname{div}\, {}^{(S)}H_t = -\frac{1}{\sqrt{g}} \Sigma\, \partial_i \, g^{ik} \sqrt{g}\, \partial_k \, \Phi$$

where

$$g = \mathrm{Det}\,(g_{ik})$$

and

$$\partial_i = \frac{\partial}{\partial u_i}.$$

The boundary condition becomes

$$\frac{\partial \Phi}{\partial n} + \frac{\mu d}{1+j}\frac{1}{\sqrt{g}} \Sigma\, \partial_i\, g^{ik}\, \sqrt{g}\, \partial_k \Phi = 0\,.$$

So far as the author knows, boundary conditions of this type have not been discussed previously.

5. *Application to a sphere*

It is now easy to arrive at the solution for a sphere. We assume tentatively

$$\Phi = -H_0\, r \cos\vartheta + \frac{p\cos\vartheta}{r^2}.$$

This satisfies Laplace's equation and the boundary condition for $r \to \infty$ Inserting this into (9) we find

$$\cos\vartheta \left[-H_0 - \frac{2p}{R^3} - \frac{\mu d}{1+j}\left\{-\frac{2}{R}H_0 - \frac{4p}{R^4} + \frac{6p}{R^4}\right\}\right] = 0$$

and hence

$$p = \frac{(1-j)\,(\mu d/R) - 1}{(1-j)\,(\mu d/R) + 2} H_0,$$

in agreement with (3). In a similar way we can obtain the solution for a circular cylinder in a field perpendicular to the axis.

6. *Application to spheroids; general considerations*

We shall now try to discuss the case of a spheroid in a field parallel to the axis of symmetry. Whereas in the case of a sphere the solution of Laplace's equation, satisfying the boundary condition (9), is of the same simple analytical form as the magnetostatic solution this is not true for a spheroid, and in order to avoid involved mathematics we shall content ourselves with a discussion of the two limiting cases $\mu d \ll R$ and $\mu d \gg R$.

First we shall derive an expression for the heat developed per surface element $d\sigma$. Be H_t the tangential component of the field and let us assume that all other components of the field inside the conductor are small. We have

$$\mathrm{curl}\,\mathbf{H} = \frac{4\pi}{c}\,\mathbf{i}$$

and therefore

$$i_t = \frac{c}{4\pi}\frac{1+j}{d}\, e^{-(1+j)\,(n,d)}\, H_t\,.$$

For the energy developed per second and per surface element it follows

$$\mathrm{d}W = \tfrac{1}{2}\frac{1}{\sigma_e}|i^2|\,\mathrm{d}\sigma = \frac{c^2}{32\pi^2\sigma_e d}|H_t^2|\,\mathrm{d}\sigma\,.$$

Using eq. (1) we find

$$\mathrm{d}W = \frac{1}{16\pi}\omega\mu d\,|H_t^2|\,\mathrm{d}\sigma\,,$$

so that W can be found by integration once H_t is known.

Now in the case $\mu d << R$ we have $H_n = 0$, and from the solution of Laplace's equation we find at once the value of H_t. For $\mu d >> R$ on the other hand we have to a first approximation

$$\mathrm{div}\ {}^{(S)}H_t^{(1)} = 0$$

from which it follows

$$H_t^{(1)} = 0\,.$$

A second approximation is found by integrating the equation

$$\mathrm{div}\ {}^{(S)}H_t^{(2)} = -\frac{1+j}{\mu d}H_n^{(1)}\,.$$

Outside the conductor we have

$$|H_t^{(2)}/H_n^{(1)}| = O\,(R/\mu d)\,,$$

which is small, but inside the conductor

$$|H_t^{(2)}/H_n| = O\,(R/d)\,,$$

which is large. Although outside the conductor the lines of force are nearly perpendicular to the surface, they are nearly parallel to the surface inside the conductor.

7. *Spheroids; the diamagnetic case*

If a spheroid consisting of a substance with susceptibility χ is placed in a static exterior field $\mathbf{H}_0$, the internal field $\mathbf{H}_i$ is given by

$$\mathbf{H}_i = \mathbf{H}_0 - \alpha\mathbf{M}$$

where α is the coefficient of demagnetization and $\mathbf{M} = \chi\,\mathbf{H}_i$ so that

$$\mathbf{H}_i = \frac{1}{1+\alpha\chi}\mathbf{H}_0$$

and

$$H_t = \frac{1}{1+\alpha\chi}H_0\sin\vartheta$$

where ϑ is the angle between the axis of symmetry and the normal to the

surface. In the limiting case $\mu \to 0$ or $\chi \to (-1/4\,\pi)$ the normal component of **H** tends to zero. Hence for the solution in the case $\mu d \ll R$

$$H_t = \frac{1}{1 - \alpha/4\pi} H_0 \sin \vartheta$$

and

$$W = \left(\frac{1}{1-\alpha/4\pi}\right)^2 H_0^2 \frac{\omega \mu d}{16\pi} \iint \sin^2 \vartheta \, d\sigma \,.$$

Writing the equation of the spheroid in the form

$$\frac{z^2}{c^2} + \frac{\varrho^2}{a^2} = 1 \,,$$

we find

$$\iint \sin^2 \vartheta \, d\sigma = 2\pi \,.\, 2 \int_0^c \sin^2 \vartheta \, \varrho \, \frac{dz}{\sin \vartheta} = 4\pi a c \int_0^1 \frac{1 - \zeta^2}{\sqrt{1 - \left(1 - \frac{a^2}{c^2}\right)\zeta^2}} \, d\zeta \,.$$

The integral is

$$I_1 = \left(1 + \frac{1}{2\varepsilon}\right) F(\varepsilon) - \frac{1}{2\varepsilon} \sqrt{(1+\varepsilon)}$$

where we have put

$$\varepsilon = \left(\frac{a^2}{c^2} - 1\right),$$

and where

$$F(\varepsilon) = \frac{1}{\sqrt{\varepsilon}} \ln \left(\sqrt{\varepsilon} + \sqrt{1+\varepsilon}\right) \qquad \text{for } \varepsilon > 0 \,,$$

and

$$F(\varepsilon) = \frac{1}{\sqrt{-\varepsilon}} \arcsin \sqrt{-\varepsilon} \qquad \text{for } \varepsilon < 0 \,.$$

Our final expression is

$$W = \left(\frac{1}{1-\alpha/4\pi}\right)^2 H_0^2 \cdot \frac{\omega \mu d}{4} \, ac \, I_1 \,. \tag{10}$$

For $\varepsilon = 0$ we find $I_1 = \frac{2}{3}$, and eq. (10) reduces to the expression for a sphere in the same limiting case. For small values of ε we can put

$$I_1 = \tfrac{2}{3}\left[1 + \tfrac{1}{10}\left(1 - \frac{a^2}{c^2}\right)\right]. \tag{11}$$

It is interesting to study the limiting cases $c \gg a$ (long needle) and $c \ll a$ (flat disc).

A) $c \gg a$. In this case α tends to zero and I_1 becomes $\pi/4$ (which incidentally differs less than 7 % from the values given by (11)). Our formula becomes

$$W = \frac{\pi}{16} \mu d \, ac \, \omega \, H_0^2 \,. \tag{12}$$

Comparing this with the heat developed per unit length in an infinitely long cylinder, which is given by

$$W = \tfrac{1}{8}\,\omega\, H_0^2\,\mu d\, R\,, \qquad (13$$

we arrive at the conclusion that the heat developed in our spheroid is equal to that developed in a length $\pi c/2$ of infinite cylinder.

B) $c \ll a$. We find

$$I_1 \approx \frac{c}{a} \ln \frac{2a}{\sqrt{e_0}\,.\,c}$$

(where $e_0 = 2.718\ldots$) and

$$\frac{a}{4\pi} = 1 - \frac{\pi}{2}\frac{c}{a}$$

so that

$$W = \frac{1}{\pi^2}\, a^2 \ln \frac{2a}{c\sqrt{e_0}} \cdot \omega\mu d\, H_0^2\,. \qquad (14)$$

So we see that the heat developed in a very thin disc perpendicular to the lines of force becomes very large. This is due to the compression of the lines of force at the edge of the disc.

8. *Spheroids; ferromagnetic case.*

To obtain our first approximation we put again

$$\mathbf{H}_i = \frac{1}{1 + a\chi}\,\mathbf{H}_0\,.$$

The corresponding value of the induction is

$$\mathbf{B}_i = \frac{4\pi\chi}{1 + a\chi}\,\mathbf{H}_0\,.$$

In the limit of very large χ this become:

$$\mathbf{B}_i = \frac{4\pi}{a}\,\mathbf{H}_0$$

and the normal component of H outside the conductor i

$$H_n^{(1)} = \frac{4\pi}{a}\, H_0 \cos\vartheta\,.$$

For ds, the line-element of the surface, we have in a meridian plane

$$\mathrm{d}s = \mathrm{d}\varrho/\cos\vartheta$$

and the equation for $H_t^{(2)}$ becomes

$$\frac{1}{\varrho}\frac{\mathrm{d}}{\mathrm{d}s}\,(\varrho\, H_t^{(2)}) = -\frac{1+j}{\mu d}\, H_n^{(1)}\,,$$

so that

$$H_t^{(2)} = -\frac{1+j}{\mu d}\cdot\frac{1}{\varrho}\cdot\frac{4\pi}{a}\, H_0 \int \varrho\cos\vartheta\,\frac{\mathrm{d}\varrho}{\cos\vartheta}\,,$$

or

$$H_t^{(2)} = -\frac{1+j}{\mu d} \cdot \frac{2\pi}{a} \cdot \varrho\, H_0 .$$

For W we find

$$W = \frac{\pi}{2a^2} \frac{\omega}{\mu d} H_0^2 \int\int \varrho \mathrm{d}\sigma = \frac{\pi^2}{a^2} \frac{\omega}{\mu d} H_0^2 \int \varrho^3 \mathrm{d}s ,$$

or

$$W = 2 \frac{\pi^2}{a^2} \frac{\omega}{\mu d} H_0^2 a^3 c\, I_2 \tag{15}$$

where

$$I_2 = \int_0^1 (1-\zeta^2) \sqrt{1+\varepsilon\zeta^2}\, \mathrm{d}\zeta .$$

One finds

$$I_2 = \tfrac{1}{4}\sqrt{1+\varepsilon} - \frac{1}{8\varepsilon}\sqrt{1+\varepsilon} + \tfrac{1}{2}\left(1+\frac{1}{4\varepsilon}\right) F(\varepsilon)$$

where the symbols have the same meaning as before.
For small ε this becomes

$$I_2 \approx \tfrac{2}{3}\left\{1 + \tfrac{1}{10}\left(\frac{a^2}{c^2} - 1\right)\right\}, \tag{16}$$

and for $\varepsilon = 0$ we find again the correct expression for a sphere. We shall discuss the same limiting cases as before:

A) $c \gg a$

We find

$$I_2 = 3\pi/16$$

which, again, is not very different from the value given by (16). The coefficient of demagnetization α approaches zero and we have to study its asymptotic behaviour. We have

$$\alpha = 4\pi \frac{1-e^2}{e^2}\left\{\frac{1}{2e} \ln \frac{1+e}{1-e} - 1\right\}$$

where

$$e = \sqrt{1 - a^2/c^2}$$

whence

$$\alpha = 4\pi \frac{a^2/c^2}{e^2}\left\{\frac{1}{2e} \ln \frac{(1+e)^2}{a^2/c^2} - 1\right\} \approx 4\pi \frac{a^2}{c^2} \ln \frac{2c}{e_0 a} .$$

So we find:

$$W = \frac{3\pi}{128} \frac{c^5}{a} \frac{1}{\left(\ln \frac{2c}{e_0 a}\right)^2} H_0^2\, \omega/\mu d .$$

B) $c \ll a$

We have

$$I_2 = \tfrac{1}{4}\frac{a}{c}$$

and

$$\alpha = 4\pi ;$$

hence for a disc perpendicular to the lines of force

$$W = \tfrac{1}{3}\tfrac{1}{2}\, a^4 \frac{\omega}{\mu d} H_0^2 . \tag{17}$$

It is interesting to note that this result is independent of c.

9. *The case* $d >> R$

For completeness's sake we shall also give the formula for the case of very large penetration depth. Since μ is supposed to be large the interior field is given by

$$B_i = \frac{4\pi}{a} H_0 .$$

The electromotive force in a ring between ϱ and $\varrho + \mathrm{d}\varrho$, and z and $z + \mathrm{d}z$ is

$$\pi\varrho^2\, \omega\, B_i/c$$

and its resistance is

$$\frac{1}{\sigma_e} 2\pi\varrho/\mathrm{d}\varrho\mathrm{d}z$$

so that the heat becomes

$$W = \tfrac{1}{2}\iint \frac{\sigma_e \pi^2 \varrho^4 \omega^2}{c^2} \left(\frac{4\pi}{a}\right)^2 H_0^2 \frac{\mathrm{d}\varrho\mathrm{d}z}{2\pi\varrho} = \left(\frac{4\pi}{a}\right)^2 \frac{1}{8\,\mu d^2}\, \omega H_0^2 \iint \varrho^3 \mathrm{d}\varrho\mathrm{d}z =$$

$$= \left(\frac{4\pi}{a}\right)^2 \frac{1}{30\,\mu d^2}\, a^4\, c\, \omega\, H_0^2 . \tag{18}$$

10. *Concluding remarks*

The general solution of the boundary value problem stated in section 2 has been tackled by several authors but the results are rather complicated It might be worth while to investigate whether the solution of the boundary value problem for $d << R$ but arbitrary μd leads to simpler results. Further it should be emphasized that in our formulae for a thin disc we always suppose that the penetration depth is small compared with the thickness of the disc. It might be worth while to give a special discussion of the case where d is large compared with this thickness, though small compared with the radius of the disc. Finally we should like to point out that by graphical interpolation on a logarithmic plot it is possible to obtain an estimate of W also in cases where our formulae do not apply.

Eindhoven, February 1946.

REFERENCES

[1]) J. L. Snoek, New developments in ferromagnetic materials, Elsevier's Publishing Company, Amsterdam - New-York 1946.
[2]) M. Divilkovsky, Le problème d'une sphère mètallique dans un champ magnétique alternatif homogène et son application à la théorie des fours à induction, J. Phys., U.S.S.R. **1**, 471-478, 1939.
[3]) Leigh Page, Magnetic moments at high frequencies, Phys. Rev. **60**, 675, 1941.
[4]) M. Jouguet, Courants de Foucault et fours à induction, Paris 1944.

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

Reprints of the majority of these papers can be obtained on application to the Administration of the Research Laboratory, Kastanjelaan, Eindhoven, Netherlands. Those papers of which no reprints are available in sufficient number are marked with an asterisk (*).

1694: Balth. van der Pol, Merkwaardige eigenschappen van geheele getallen, Natuurkundige voordrachten, Diligentia, band **23**, Den Haag, van Stockum, 1946 (Remarkable properties of whole numbers).

This lecture treats some remarkable properties of whole numbers: prime numbers, Goldbach's conjecture, complex prime numbers, perfect numbers, the number of prime numbers in a given interval, Skewes's number.

1695: P. J. Bouma, Colour equations, Physica 's-Grav. **12**, 189-194, 1946.

Two types of equations occur in colorimetry: algebraic equations and colour equations. The difference between these two kinds of equations is explained, and a special symbol ($\leftarrow\rightarrow$) is introduced in order to avoid the danger of confusion between the two types. This danger is illustrated by an example, where the same problem is solved without and with the use of colour equations.

(Continued on page 67)

R 35 *Philips Res. Rep. 2, 55-67, 1947*

RADIATION AND HEAT CONDUCTION IN LIGHT-SCATTERING MATERIAL

by H. C. HAMAKER

536.24:536.33

I. REFLECTION AND TRANSMISSION

Summary

On the basis of a set of simultaneous differential equations originally due to Schuster the transmission and reflection of light in light-scattering layers is discussed. Formulae previously developed by Kubelka and Munk are briefly recapitulated; they are extended so as to describe the luminescence of fluorescent screens excited by X-rays or electron bombardment. Likewise formulae are derived that include temperature radiation.

1. *Introduction*

In 1905 Schuster [1]) developed a set of simultaneous differential equations to describe the transmission of light through fog. Many years later the same set of equations was applied by Dreosti [2]) in dealing with transmission and reflection phenomena in opal glass, and by Kubelka and Munk [3]) in discussing the coating properties of layers of paint.

The original purpose of the present investigations was to extend this theory so as to include temperature radiation and the combined effect of radiation and heat conduction. These more complicated problems, however, will be reserved for a future paper. In the course of this work we came to the conclusion that the equations of Kubelka and Munk could, with advantage, be written in a simpler form and that they could be extended so as to apply to the radiation emitted by a fluorescent screen irradiated by X-rays *). These problems will form the main topic discussed below.

2. *Basic equations*

Only a one-dimensional case is considered (*fig. 1*) and the total radiant flux is plit up into two parts, *viz*:

I = the flux in the direction of the positive x axis and
J = the flux in the direction of the negative x axis.

On passing through the infinitesimal layer dx a fraction $a.I.dx$ of the

*) My attention was drawn to this point by Dr Klasens of this laboratory. Practical applications of various equations developed below will be found elsewhere in this issue.

flux I will be absorbed and a fraction $s.I.dx$ will be lost by scattering in a backward direction. On the other hand a quantity $s.J.dx$ will be added by scattering from the flux J, so that we obtain

$$\frac{dI}{dx} = -(a + s)\, I + sJ, \tag{1a}$$

and by the same arguments *)

$$\frac{dJ}{dx} = (a + s)\, J - sI\,. \tag{1b}$$

These are Schuster's equations. Scattering sideways has been disregarded, it being supposed that any loss of radiant energy in a sideward direction is compensated by an equal contribution from the neighbouring parts of the layer; this implies that the dimensions of the area investigated must be large compared with the thickness of the layer, a condition nearly always fulfilled in practical cases. It will also be evident that observations to which equations (1a) and (1b) are applied must be made under completely diffuse illumination.

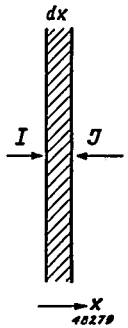


Fig. 1. The radiant flux is split up into two parts: I in the direction of positive x and J in the direction of negative x.

This last supposition is perhaps not quite correct. It has been observed by Spijkerboer [4]) and Dreosti [2]) that the intensity distribution over various angles after transmission through a light-scattering layer is independent of the intensity distribution of the incident light and of the thickness of the layer, provided this thickness exceeds a certain limit. The intensity distribution of the transmitted light for a thick layer may be designated as the specific distribution of the material in question; it will mainly depend on the ratio of absorption to scattering (a/s). The author inclines to the opinion that when the incident light happens to possess the specific intensity distribution then equations (1a) and (1b) will strictly hold. If the distribution of the incident beam is different, it will be

*) It should be noted that equation (1b) can be deduced from (1a) by interchanging I and J and changing the sign of dx. This is a principle which also applies in more complicated cases (see for instance equations (28) and (32)).

gradually transformed into the specific distribution as the light penetrates the layer, so that from a certain thickness onwards equations (1a) and (1b) can be applied. For very thin layers these equations may break down owing to the changes taking place in the intensity distribution.

Whether we must preferably carry out our observations with completely diffuse light or not depends on whether the specific distribution is near to the completely diffuse distribution or not. For instance, in the case of pure absorbtion the specific distribution is evidently a parallel bundle incident at right angles, and equations (1a) and (1b) apply if s is put equal to zero.

To decide whether these views are correct will require a theoretical investigation on the basis of such integral equations as developed by Schwarzschild [5]), King [6]) and Spijkerboer [4]). This would appear to the author a profitable way of attacking the problem.

In his original paper Schuster assumed isotropic scattering, the fraction s being exactly one half of the total amount scattered. As pointed out by Dreosti, this is an unnecessary restriction; whether scattering is symmetric with respect to the y-z plane or not is immaterial, provided s measures the fraction scattered backwards.

To conclude this section it may be mentioned that according to a theoretical investigation by Schwarzschild [5]) equations (1a) and (1b) furnish a satisfactory approximation to actual conditions. This is corroborated by Judd's [7]) experimental verification of the equations of Kubelka and Munk.

Consequently equations (1a) and (1b) may be adopted as a satisfactory basis for the discussion of such phenomena as we have in view.

3. *The general solution*

Equations (1a) and (1b) are solved by putting

$$I = C_1 \cdot e^{\sigma x} + C_2 \cdot e^{-\sigma x}; \quad J = C_3 \cdot e^{\sigma x} + C_4 \cdot e^{-\sigma x}, \tag{2}$$

only two of the four constants $C_1 \ldots . C_4$ being arbitrary. The complete solution reduced to its simplest form is

$$I = A(1 - \beta_0)\, e^{\sigma_0 x} + B(1 + \beta_0)\, e^{-\sigma_0 x}, \tag{3a}$$

$$J = A(1 + \beta_0)\, e^{\sigma_0 x} + B(1 - \beta_0)\, e^{-\sigma_0 x}, \tag{3b}$$

where

$$\sigma_0 = \sqrt{a(a + 2s)}, \tag{4}$$

$$\beta_0 = \sqrt{\frac{a}{(a + 2s)}} = \frac{\sigma_0}{a + 2s}, \tag{5}$$

both roots being taken with the positive sign *). In these equations A and B are two constants determined by the boundary conditions.

*) This is of course a pure convention. It is only essential that both roots should have the same sign.
The suffix 0 added to σ and β might have been omitted in the present case, but has been added in view of extensions which will be developed in a later paper.

For instance in the simplest case, a layer infinitely thick stretching from $x = 0$ to $x = +\infty$ with an incident beam of intensity I_0, we have

$$\begin{aligned} &I = I_0 \quad \text{for } x = 0 \quad \text{and} \\ &I = J = 0 \quad \text{for } x = +\infty \end{aligned} \tag{6}$$

from which we obtain

$$A = 0 \text{ and } B = \frac{I_0}{1 + \beta_0}. \tag{7}$$

The intensity of the reflected beam is then

$$J(0) = B(1 - \beta_0) = I_0 \frac{1 - \beta_0}{1 + \beta_0}$$

and the reflectivity is repesented by

$$R_\infty = \frac{J(0)}{I_0} = \frac{1 - \beta_0}{1 + \beta_0}, \tag{8}$$

an equation which may be of service in estimating the value of the constant β_0 from observations or estimates of the reflectivity of an infinitely thick layer.

From (3) and (7) we find

$$I = I_0 \,.\, e^{-\sigma_0 x}$$

so that σ_0 measures the rate of decay of the incident beam as we penetrate into the interior of the layer; I is reduced to about 1% of its original value when $\sigma_0\, x = 4{\cdot}5$. Consequently if we observe that certain specimens of porcelain are visibly transparent when 0·2 cm thick we may conclude that σ_0 (for visible light at least) must be of the order of $4{\cdot}5/0{\cdot}2 = 22{\cdot}5$ cm^{-1}; likewise for a paint which in a layer of 0·01 cm completely covers the underlayer, σ_0 must be of the order of $4{\cdot}5/{\cdot}001 = 450$ cm^{-1}. In such a way σ_0 may be estimated, when, as in most cases, no data are available from which it can be calculated with accuracy.

When considering a layer of finite thickness D applied to an underlayer of reflectivity r (*fig. 2*) the boundary conditions will be:

$$\begin{aligned} &I = I_0 \text{ when } x = 0\,, \\ &J = r.I \text{ when } x = +D\,. \end{aligned} \tag{9}$$

It has been found, however, that practically all our formulae are considerably simplified by introducing instead of r the related quantity

$$\varrho = \frac{1 - r}{1 + r}. \tag{10}$$

Then

$$r = \frac{1-\varrho}{1+\varrho}, \tag{11}$$

and the boundary conditions (9) become

$$\begin{aligned} I &= I_0 \text{ when } x = 0, \\ (1+\varrho)\, J &= (1-\varrho)\, I \text{ when } x = +D. \end{aligned} \tag{12}$$

Whenever similar situations arise, the boundary conditions will consistently be used in this form.

After inserting equations (3) in (12) and solving for A and B we obtain

$$A = \frac{-I_0 \cdot (\varrho - \beta_0)\, e^{-\sigma_0 D}}{(1+\beta_0)(\varrho+\beta_0)\, e^{\sigma_0 D} - (1-\beta_0)(\varrho-\beta_0)\, e^{-\sigma_0 D}} \tag{13a}$$

$$B = \frac{I_0 \cdot (\varrho + \beta_0)\, e^{\sigma_0 D}}{(1+\beta_0)(\varrho+\beta_0)\, e^{\sigma_0 D} - (1-\beta_0)(\varrho-\beta_0)\, e^{-\sigma_0 D}} \tag{13b}$$

and consequently for the reflectivity of the layer

$$\begin{aligned} R = \frac{J(0)}{I_0} &= \frac{A(1+\beta_0) + B(1-\beta_0)}{I_0} = \\ &= \frac{(1-\beta_0)(\varrho+\beta_0)\, e^{\sigma_0 D} - (1+\beta_0)(\varrho-\beta_0)\, e^{-\sigma_0 D}}{(1+\beta_0)(\varrho+\beta_0)\, e^{\sigma_0 D} - (1-\beta_0)(\varrho-\beta_0)\, e^{-\sigma_0 D}}. \end{aligned} \tag{14}$$

This is Kubelka and Munk's equation, which has been experimentally verified by Judd and colloborators.

When D is large equation (14) reduces to (8), as we should expect. The same simplification is introduced if the underlayer possesses the same reflectivity as an infinite layer of the light-scattering material. This is

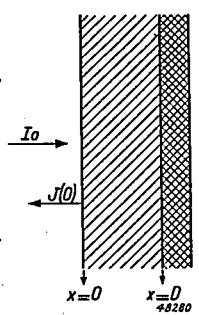


Fig. 2. A light-scattering layer (thickness D) applied to a reflecting underlayer (reflectivity r).

mathematically realized by putting $\varrho = \beta_0$ (compare equations (8) and (17)), which again reduces (14) to (8).

In order to simplify our equations as much as possible the constants a and s figuring in the differential equations (1) have above been replaced by σ_0 and β_0. Kubelka and Munk used R_∞ and s and these two were computed by Judd from his observations. The three sets (a, s), (β_0, σ_0), or (R_∞, s) are completely equivalent and the mutual relations between them are covered by equations (4), (5) and (8). In addition we note that the following relations, which are easily derived from the equations just mentioned, may sometimes be useful in calculating one constant from another:

$$\sigma_0 = \tfrac{1}{2}\, s \left(\frac{1}{R_\infty} - R_\infty \right), \tag{15}$$

$$\beta_0 = \frac{1 - R_\infty}{1 + R_\infty}, \tag{16}$$

$$\frac{s}{a} = \frac{2\, R_\infty}{(1 - R_\infty)^2}. \tag{17}$$

By these relations and (11) equation (14) can readily be transformed into the form given by Kubelka and Munk.

A still more general case occurs when the light-scattering layer is enclosed between two transparent plates as indicated in *fig. 3*, for instance when we investigate a powder compressed between two glass plates. Let r_1 be the reflectivity and t_1 the transmissivity of the front plate (on the side of negative x), the same quantities for the back being r_2 and t_2. Then of the incident intensity I_0 the fraction $t_1 \cdot I_0$ will enter the light-scattering material and the boundary conditions will be

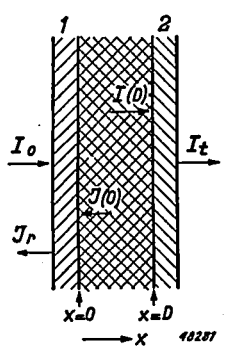


Fig. 3. A light-scattering layer (thickness D) between two transparent plates (1 and 2) I_0 is the radiation incident from the left; J_r is the amount reflected and I_t the amount transmitted.

$$I = t_1 \cdot I_0 + r_1 J \quad \text{or} \quad (1 + \varrho_1) I - (1 - \varrho_1) J = (1 + \varrho_1) t_1 \cdot I_0 \tag{18a}$$

when $x = 0$, and

$$J = r_2 \cdot I \quad \text{or} \quad (1 + \varrho_2) J - (1 - \varrho_2) I = 0 \tag{18b}$$

when $x = D$, ϱ_1 and ϱ_2 being defined by equation (10).

Inserting (3) in (18), A and B may be computed, but it is not essential to give their expressions in detail. From them we obtain

$$J(0) = \tfrac{1}{2} t_1 (1 + \varrho_1) \cdot I_0 \frac{(\varrho_2 + \beta_0)(1 - \beta_0) e^{\sigma_0 D} - (\varrho_2 - \beta_0)(1 + \beta_0) e^{-\sigma_0 D}}{(\varrho_1 + \beta_0)(\varrho_2 + \beta_0) e^{\sigma_0 D} - (\varrho_1 - \beta_0)(\varrho_2 - \beta_0) e^{-\sigma_0 D}} \tag{19}$$

which expresses the radiant energy incident from the right on the front plate 1. The total amount reflected will be

$$I_r = t_1 \cdot J(0) + r_1 \cdot I_0 .$$

When there is no plate in front we may put $r_1 = 0$, $\varrho_1 = 1$, $t_1 = 1$ and the expressions (19) or (20) simplify to (14), as we would expect.

In the same manner the intensity of the transmitted light can be computed, with the result:

$$I_t = t_2 \cdot I(D) = I_0 \frac{\beta_0 \cdot t_1 \cdot t_2 (1 + \varrho_1)(1 + \varrho_2)}{(\varrho_1 + \beta_0)(\varrho_2 + \beta_0) e^{\sigma_0 D} - (\varrho_1 - \beta_0)(\varrho_2 - \beta_0) e^{-\sigma_0 D}}, \tag{20}$$

an expression symmetric with respect to the suffixes 1 and 2, as it ought to be.

4 *Radiation excited by X-rays*

In X-ray photography a common arrangement is that of the intensifying back screen; the action of the X-rays on the photographic emulsion is strongly intensified by the light excited in a fluorescent screen pressed against the film at the back (see *fig. 4*).

Now let N_0 be the intensity of the X-ray beam at $x = 0$, that is after passing the photo-sensitive layer. On passing through the fluorescent material the X-ray intensity will decrease as

$$N = N_0 \cdot e^{-\mu x}$$

where μ is the total coefficient of X-ray absorption. The amount of Röntgen radiation absorbed between the plane $x = X$ and $x = X + \mathrm{d}X$ is

$$-\mathrm{d}N = \mu \cdot N_0 \cdot e^{-\mu X} \, \mathrm{d}X$$

and an equivalent amount of fluorescent light will be excited in this infinitesimal layer, *viz.*

$$\mathrm{d}i = k \,.\, \mu \cdot N_0 \,.\, e^{-\mu X}\, \mathrm{d}X \tag{21}$$

the constant k measuring the efficiency of the fluorescence.

Transmission of this visible light through the fluorescent layer will take place according to equations (1a, b), and the resulting intensity in the plane $x = 0$, that is the light incident on the photographic plate, will now be computed.

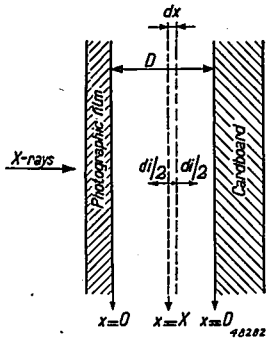


Fig. 4. A layer of fluorescent material (thickness D) applied to a piece of cardboard, used as an intensifying back screen in X-ray photography.

Let us suppose an amount of light $\mathrm{d}i$ to be created in the layer $\mathrm{d}x$ at $x = X$, half of which is emitted to the right, the other half to the left. To find the resulting distribution of light intensities the layer must be divided into two parts, *viz*: $0 < x < X$ and $X < x < D$. Assuming

$$I = I_1 = A_1(1 - \beta_0)\, e^{\sigma_0 x} + B_1(1 + \beta_0)\, e^{-\sigma_0 x} \tag{22a}$$

$$J = J_1 = A_1(1 + \beta_0)\, e^{\sigma_0 x} + B_1(1 - \beta_0)\, e^{-\sigma_0 x} \tag{22b}$$

when $0 < x < X$, and

$$I = I_2 = A_2(1 - \beta_0)\, e^{\sigma_0 x} + B_2(1 + \beta_0)\, e^{-\sigma_0 x} \tag{22c}$$

$$J = J_2 = A_2(1 + \beta_0)\, e^{\sigma_0 x} + B_2(1 - \beta_0)\, e^{-\sigma_0 x} \tag{22d}$$

when $X < x < D$,

the boundary conditions will be

$$\left.\begin{array}{ll} I_1 = r_1 \,.\, J_1 \text{ or } (1 + \varrho_1)\, I_1 - (1 - \varrho_1)\, J_1 = 0 & \text{for } x = 0 \\ J_1 = J_2 + \tfrac{1}{2}\,\mathrm{d}i,\ I_2 = I_1 + \tfrac{1}{2}\,\mathrm{d}i & \text{for } x = X \\ J_2 = r_2 \,.\, I_2 \text{ or } (1 + \varrho_2)\, J_2 - (1 - \varrho_2)\, I_2 = 0 & \text{for } x = D \end{array}\right\} \tag{23}$$

r_1, ϱ_1 referring to the film and r_2, ϱ_2 to the cardboard.

By combining (22) and (23) A_1, B_1, A_2, and B_2 may be calculated, and with these we subsequently obtain

$$J_1(0) = A_1(1+\beta_0) + B_1(1-\beta_0) =$$
$$= \tfrac{1}{2}\,\mathrm{d}i\,(1+\varrho_1)\,\frac{(\varrho_2+\beta_0)\,e^{\sigma_0(D-X)} - (\varrho_2-\beta_0)\,e^{-\sigma_0(D-X)}}{(\varrho_1+\beta_0)(\varrho_2+\beta_0)\,e^{\sigma_0 D} - (\varrho_1-\beta_0)(\varrho_2-\beta_0)\,e^{-\sigma_0 D}} \tag{24}$$

and similarly

$$I_2(D) = A_2(1-\beta_0)\,e^{\sigma_0 D} + B_2(1+\beta_0)\,e^{-\sigma_0 D} =$$
$$= \tfrac{1}{2}\,\mathrm{d}i\,(1+\varrho_2)\,\frac{(\varrho_1+\beta_0)\,e^{\sigma_0 X} - (\varrho_1-\beta_0)\,e^{-\sigma_0 X}}{(\varrho_1+\beta_0)(\varrho_2+\beta_0)\,e^{\sigma_0 D} - (\varrho_1-\beta_0)(\varrho_2-\beta_0)\,e^{-\sigma_0 D}} \tag{25}$$

for the radiation incident from the right on the photographic plate and from the left on the cardboard underlayer respectively. The latter quantity is not of essential interest in our present problem but has been computed for the sake of completeness.

To find the resultant intensities under X-ray irradiation we have to replace $\mathrm{d}i$ in (24) and (25) by the expression (21) and to integrate from $X = 0$ to $X = D$, with the result:

$$J_1(0) = N_0 \cdot k \cdot \mu \cdot \frac{(1+\varrho_1)}{2(\mu^2-\sigma_0^2)} \times$$
$$\times \frac{(\mu-\sigma_0)(\varrho_2+\beta_0)\,e^{\sigma_0 D} - (\mu+\sigma_0)(\varrho_2-\beta_0)\,e^{-\sigma_0 D} + 2(\sigma_0\varrho_2 - \mu\beta_0)\,e^{-\mu D}}{(\varrho_1+\beta_0)(\varrho_2+\beta_0)\,e^{\sigma_0 D} - (\varrho_1-\beta_0)(\varrho_2-\beta_0)\,e^{-\sigma_0 D}} \tag{26}$$

$$I_2(D) = N_0 \cdot k \cdot \mu \cdot \frac{(1+\varrho_2)}{2(\mu^2-\sigma_0^2)} \cdot e^{-\mu D} \times$$
$$\times \frac{-(\mu+\sigma_0)(\varrho_1+\beta_0)\,e^{\sigma_0 D} + (\mu-\sigma_0)(\varrho_1-\beta_0)\,e^{-\sigma_0 D} + 2(\sigma_0\varrho_1 + \mu\beta_0)\,e^{\mu D}}{(\varrho_1+\beta_0)(\varrho_2+\beta_0)\,e^{\sigma_0 D} - (\varrho_1-\beta_0)(\varrho_2-\beta_0)\,e^{-\sigma_0 D}} \tag{27}$$

According to circumstances these formulae can sometimes be simplified. For instance when D is large the terms $e^{-\sigma_0 D}$ can be disregarded with respect to the term $e^{\sigma_0 D}$. We shall not consider these questions in detail here; for practical applications we may refer to a paper by Dr Klasens.

An alternative method to treat the same problem is to start from the differential equations

$$\begin{aligned} \frac{\mathrm{d}I}{\mathrm{d}x} &= -(a+s)\,I + sJ + \tfrac{1}{2}\,C\,e^{-\mu x} \\ \frac{\mathrm{d}J}{\mathrm{d}x} &= (a+s)\,J - sI - \tfrac{1}{2}\,C\,e^{-\mu x} \end{aligned} \tag{28}$$

which are deduced from equations (1) by adding the amount of radiation excited by the X-rays, half of which contributes to I and the other half to J.

By common principles the general solution of (28) is found to be

$$\begin{aligned} I &= A(1-\beta_0)\, e^{\sigma_0 x} + B(1+\beta_0)\, e^{-\sigma_0 x} - \tfrac{1}{2} C \frac{\mu+\sigma_0}{(\mu^2-\sigma_0^2)} e^{-\mu x}, \\ J &= A(1+\beta_0)\, e^{\sigma_0 x} + B(1-\beta_0)\, e^{-\sigma_0 x} + \tfrac{1}{2} C \frac{\mu-\sigma_0}{(\mu^2-\sigma_0^2)} e^{-\mu x}. \end{aligned} \tag{92}$$

A and B must now be determined by the boundary conditions:

$$\begin{aligned} (1+\varrho_1)\, J - (1-\varrho_1)\, J &= 0 \quad \text{when } x = 0, \\ (1+\varrho_2)\, J - (1-\varrho_2)\, I &= 0 \quad \text{when } x = D. \end{aligned} \tag{30}$$

That the expressions (26) and (27) result, may readily be verified.

It should be mentioned that equations (28) are practically identical with the equations used by Dreosti [2]) in discussing the scattering of light when the layer is illuminated by parallel rays of visible light. Owing to scattering and absorption the incident beam will decrease in intensity as $C \,.\, e^{-kx}$ from left to right. When a parallel beam of intensity H passes the layer dx a fraction $p.H.dx$ will be scattered in the forward direction (positive x) and a fraction $q.H.dx$ in the backward direction (negative x). If to this scattered light we apply the principles adopted throughout this paper we must have

$$\begin{aligned} \frac{dI}{dx} &= -(a+s)\, I + sJ + qCe^{-kx}, \\ \frac{dJ}{dx} &= (a+s)\, J - sI - pCe^{-kx}, \end{aligned} \tag{31}$$

the only difference from (28) being that the coefficients p and q are unequal. The solution of (31) will describe the amount of diffuse light emanating from both sides of the layer. According to Dreosti's observations this theory is in good accord with experiments.

5. *Temperature radiation*

As a last problem we will consider a light-scattering layer heated to a uniform temperature and investigate the resulting radiation.

From the total radiant flux I in the direction of positive x a fraction $a.I.dx$ is absorbed in the layer dx (fig. 1). Consequently to conform to Kirchhoff's law this layer dx will contribute by radiation the amount $a.E_0.dx$ in the direction of negative x, where E_0 designates the black-body radiation at the temperature and wavelength in question. The same amount will of course be radiated in the opposite direction. By adding these quantities to equations (1) we are led to

$$\begin{aligned} \frac{dI}{dx} &= -(a+s)\,I + sJ + aE_0 \\ \frac{dJ}{dx} &= (a+s)\,J - sI - aE_0 \end{aligned} \tag{32}$$

which are solved by

$$\begin{aligned} I &= A(1-\beta_0)\,e^{\sigma_0 x} + B(1+\beta_0)\,e^{-\sigma_0 x} + E_0 \\ J &= A(1+\beta_0)\,e^{\sigma_0 x} + B(1-\beta_0)\,e^{-\sigma_0 x} + E_0 . \end{aligned} \tag{33}$$

If the light-scattering material is applied as a coating on an underlayer of reflectivity r the boundary conditions will read:

$$I = 0 \text{ for } x = 0$$
$$J = r.I + (1-r)\,E_0 \text{ or } (1+\varrho)\,J - (1-\varrho)\,I = 2\varrho E_0 \text{ for } x = D\,, \tag{34}$$

the term $(1-r)\,E_0$ representing the radiation contributed by the underlayer according to Kirchhoff's rule. After computing A and B we find for the temperature radiation

$$J(0) = E_0 \cdot \frac{2\beta_0\{(\varrho+\beta_0)\,e^{\sigma_0 D} + (\varrho-\beta_0)\,e^{-\sigma_0 D}\}}{(1+\beta_0)\,(\varrho+\beta_0)\,e^{\sigma_0 D} - (1-\beta_0)\,(\beta-\beta_0)\,e^{-\sigma_0 D}} = E_0\,(1-R) \tag{35}$$

where R is the reflectivity expressed by formula (14). This is again in keeping with Kirchhoff's law.

Other cases may be worked out in the same manner.

6. *The constants a, s, etc.*

To conclude this paper we shall briefly recapitulate the data available concerning the magnitude of the constants a and s and the derived quantities σ_0 and R_∞. The most extensive observations were made by Judd and collaborators [7]) who checked equation (14) for various materials. They computed R_∞ and s, from which, however, a and σ can easily be calculated with the aid of equations (15) and (17). A few characteristic data have been collected in table I. The dental silicate cements, having the lowest value of σ_0, are most transparent; next come the vitreous enamels, whereas the cold-water paints are very opaque; a coating of these of 0·01 to 0·02 cm will suffice to mask the colour of the underlayer almost completely.

Some further data have been deduced by Dreosti for opal glasses and mastix emulsions. His results are reproduced in table II. The opal glass has optical properties of the same order as the dental silicate cements of Judd, though they are somewhat more opaque. The mastix emulsions

TABLE I

R_∞, a and s according to observations by Judd and collaborators

	R_∞	s cm^{-1}	a cm^{-1}	σ_0 cm^{-1}	$\frac{s}{a}$
I. Vitreous enamels					
	0·89	100	0·68	11·7	147
	0·90	90	0·50	9·5	180
	0·89	64	0·435	7·5	147
	0·82	56	0·875	9·9	64
II. Cold-water paints (white)					
	0·88	900	7·35	115	122
	0·86	720	8·20	109	88
	0·93	650	1·72	47·5	380
III. Dental silicate cements					
White	0·76	6·2	0·235	1·72	26·4
White	0·50	3·6	0·90	2·70	4·0
Light yellow	0·40	3·8	1·71	4·00	2·2
Light yellow grey	0·59	2·1	0·30	1·16	7·0
Dark grey	0·16	1·7	3·74	5·15	0·45

TABLE II

R_∞, a and s observed by Dreosti, $\lambda = 5370$ Å

	R_∞	s cm^{-1}	a cm^{-1}	σ_0 cm^{-1}	$\frac{s}{a}$
I. Opal glasses; various samples					
	0·807	7·8	0·18	1·69	43
	0·750	18·3	0·76	5·33	24
	0·771	13·6	0·46	3·57	29·5
	0·56	8·0	1·40	4·95	5·7
II. Mastix emulsions					
G_0	0·39	0·31	0·147	0·336	2·1
G_1	0·091	0·22	0·915	1·11	0·24
G_2	0·020	0·06	1·12	2·21	0·03
K_0	0·338	0·66	0·42	0·85	1·6
K_1	0·176	0·60	1·15	1·64	0·52
K_2	0·030	0·20	3·00	3·19	0·06

Emulsions K and G both contained 5 g mastix per litre. The average particle diameter was 0·67 μ for G and 0·49 μ for K. In the sequence K_0, K_1, K_2 acid fuchsine has been added to the intermicellar fluid in increasing quantities. The exact concentrations have not been stated, and the description of the emulsions is, on the whole, rather incomplete.

are much more transparent. Unfortunately the description of these emulsions is incomplete, but it is of interest to note that adding a dye to the fluid not only increases the absorption a but at the same time diminishes the coefficient of scattering s. This clearly shows how closely these two constants are connected; they are not independent of one another.

Eindhoven, October 1942.

REFERENCES

1) A. Schuster, Astrophys. J. **21**, 1, 1905.
2) G. M. Dreosti, Thesis, Utrecht 1930.
3) P. Kubelka & F. Munk, Z. tech. Phys. **12**, 593, 1931.
4) J. Spijkerboer, Thesis, Utrecht 1917.
5) K. Schwarzschild, Berl. Ber. 1914, p. 1183.
6) L. V. King, Phil. Trans. Roy. Soc. London A **212**, 375, 1912.
7) D. B. Judd & collaborators, Bur. Stand. J. Res. **19**, 317, 1937.

ABSTRACTS

(Continued from page 54)

1696: J. F. H. Custers and J. C. Riemersma, The texture of straight-rolled and of cross-rolled molybdenum, Physica 's-Grav. **12**, 195-208, 1946.

With the aid of pole figures the textures of straight-rolled and of cross-rolled molybdenum are determined. The pole figures thus obtained show that earlier authors described these textures in a too simple way; they are at least twofold.

For example, after cross-rolling, there is found besides the so-called (100) [110] tecture ((100) parallel to the rolling plane, and [110] parallel to the rolling direction) a second texture, which is rotary symmetrical around the normal to the rolling plane, and which has a (111) plane parallel to this plane.

The texture of straight-rolled molybdenum turns out to be in good agreement with the texture of straight-rolled iron, as determined by Kurdjumow and Sachs; the texture of cross-rolled iron is not known.

1697: N. G. de Bruijn, A combinatorial problem, Proc. Kon. Ned. Akad. Wetenschappen Amsterdam **49**, 758-764, 1946.

A P_n-cycle is defined as an ordered cycle of 2^n digits 0 or 1 (*i.e.* a series of such digits placed on the circumference of a circle), such that the 2^n possible sets of n consecutive digits of that cycle are all different (as a consequence, any ordered set of n digits 0 or 1 occurs exactly once in that cycle). Posthumus studied these cycles in connection with a practical problem of telecommunication, and was led to the conjecture, that the number of P_n-cycles be equal to 2^N where $N = 2^{n-1}-n$. This conjecture is proved to be correct, as follows from a theorem concerning a special type of networks. Another application of this theorem is mentioned too.

(Continued on page 79)

 Philips Res. Rep. 2, 68-78, 1947

THE LIGHT EMISSION FROM FLUORESCENT SCREENS IRRADIATED BY X-RAYS

by H. A. KLASENS 535.371.07

Summary

Applying Schuster's theory as extended by Hamaker for the scattering and absorption of light, general equations are deduced for the amount of light emitted by fluorescent screens, irradiated by X-rays. Some commercial screens are examined to measure the "absorption" coefficient σ of the fluorescent light. Several means to increase the brightness of a screen are discussed.

The radiant flux from a fluorescent screen irradiated by X-rays is mainly determined by the amount of X-ray energy absorbed and by the efficiency of the conversion of this energy into light.

The total amount of X-ray energy absorbed in the screen is

$$N_0 (1 - e^{-\mu D}) \qquad (1)$$

where N_0 = intensity of the X-rays incident on the screen,
μ = total absorption coefficient of the X-rays,
D = thickness of the fluorescent layer.

For X-rays corresponding to a voltage of 50 kV the total absorption coefficient is about 30 cm^{-1}. In a layer of 0·03 cm 40% of the X-ray energy will thus be absorbed, most of which is lost as heat, only a small percentage being transformed into light. According to measurements of Widmann [1]) the efficiency of this transformation is about 3%. But even if we know exactly the amount of visible light excited we cannot calculate the light emitted from the surface, for a part of the light is lost by extinction in the fluorescent layer. The light radiated by the grains of the phosphor will be scattered again and again, so that it may have to travel a long way before reaching the surface, a certain amount being absorbed in the meantime.

This absorption has been taken into account by Glocker and co-workers [2]) when calculating the radiant power of the uncovered side of the screen facing the X-ray tube.

In the layer dx (*fig. 1*) an amount dN of X-ray energy is absorbed

$$dN = \mu \,.\, N_x \,.\, dx = \mu \,.\, N_0 \,.\, e^{-x\mu}\, dx \,.$$

A proportionate part dJ will be transformed into visible light

$$dJ = k \,.\, \mu \,.\, N_0 \,.\, e^{-\mu x}\, dx \,.$$

On its way to the surface facing the X-ray tube the intensity of this fluorescent light will be reduced in the ratio $e^{-\tau x}$, where τ is the absorption coefficient of the fluorescent light. Hence the total intensity on this side of the screen (used as an intensifying back-screen) is

$$J_B = \int_0^D k \,.\, \mu \,.\, N_0 \,.\, e^{-(\mu+\tau)x}\, dx = \frac{k \,.\, \mu \,.\, N_0}{\mu + \tau} \left\{1 - e^{-(\mu+\tau)D}\right\}. \qquad (2)$$

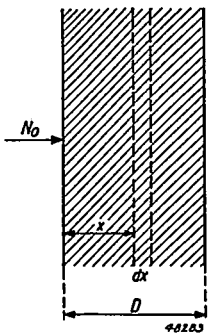


Fig. 1. A fluorescent layer (thickness D) irradiated by X-rays incident from the left.

In the same way the radiant flux from an intensifying front-screen is found to be

$$J_F = \int_0^D k \,.\, \mu \,.\, N_0 \,.\, e^{-\mu x}\, e^{-\tau(D-x)}\, dx = \frac{k \,.\, \mu \,.\, N_0}{\mu - \tau} \left\{e^{-\tau D} - e^{-\mu D}\right\}. \qquad (3)$$

These formulae, however, do not take into account the reflection at the boundaries of the layer against the cardboard or the photographic film. The reflecting power of the underlayer for instance influences strongly the brightness of the screen. The light emitted from a layer uncovered at both sides is much lower than the flux from an intensifying screen where the fluorescent layer is laid on a piece of white cardboard.

On the basis of Schuster's differential equations for the diffusion of light Hamaker [3]) has calculated the radiant flux incident on a photographic film in close contact with a fluorescent screen. In the case of an intensifying back-screen equation (26) of Hamaker's paper applies. His formula reads *)

$$J_B = \frac{k \mu N_0 (1 + \varrho_1)}{2(\mu^2 - \sigma^2)} \times \frac{(\varrho_2 + \beta)(\mu - \sigma) e^{\sigma D} - (\varrho_2 - \beta)(\mu + \sigma) e^{-\sigma D} + 2(\varrho_2 \sigma - \beta \mu) e^{-\mu D}}{(\varrho_1 + \beta)(\varrho_2 + \beta) e^{\sigma D} - (\varrho_1 - \beta)(\varrho_2 - \beta) e^{-\sigma D}} \qquad (4)$$

*) Instead of β and σ Hamaker used the symbols β_0 and σ_0, with a view to later extensions of the theory. As, however, these extensions do not refer to the X-ray problems considered here, the indices may be dropped in our case, thereby rendering the formulae somewhat simpler.

where the various symbols have the following meanings:

$\varrho_1 = \dfrac{1-r_1}{1+r_1}$, r_1 being the reflection coefficient of the film (under diffuse illumination);

$\varrho_2 = \dfrac{1-r_2}{1+r_2}$, r_2 being the reflection coefficient of the cardboard layer;

$\beta = \dfrac{1-R}{1+R}$, R designating the reflection coefficient for an infinitely thick layer of the phosphor;

$\sigma = \sqrt{a\,(a+2s)}$ where a is the coefficient of absorption and s the coefficient of scattering of the fluorescent material.

In the case of an intensifying front-screen (*fig. 2*) Hamaker's equation (27) may be used. If, however, we adhere to the principle that the index 1 (in ϱ_1, r_1) refers to the photographic film and the index 2 to the cardboard, then these indices must be interchanged in Hamaker's formula, so that we obtain *)

$$J_F = \frac{k\mu N_0 (1+\varrho_1)}{2(\mu^2-\sigma^2)} \times$$

$$\times \frac{2(\beta\mu + \varrho_2\sigma) - (\mu+\sigma)(\varrho_2+\beta)\,e^{-(\mu-\sigma)D} + (\mu-\sigma)(\varrho_2-\beta)\,e^{-(\mu+\sigma)D}}{(\varrho_1+\beta)(\varrho_2+\beta)\,e^{\sigma D} - (\varrho_1-\beta)(\varrho_2-\beta)\,e^{\sigma D}}. \quad (5)$$

For a screen not covered by a photosensitive film, as used in X-ray diagnosis and camera photography, we have

$$r_1 = 0,\ \varrho_1 = 1.$$

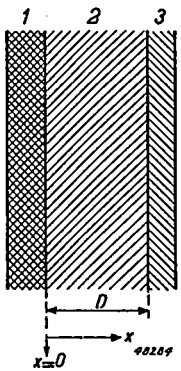


Fig. 2. An intensifying screen used as front-screen; 1 cardboard, 2 fluorescent layer, 3 photographic film. X-rays incident from the left.

*) It should be noted that equations (4) and (5) are not entirely correct in that N_0 measures the intensity of the X-ray beam in the plane $x = 0$ (fig. 2), which may be slightly different in the two cases considered owing to a difference between the absorption in the film and that in the cardboard. This absorption, however, is only a few per cent and can safely be ignored in comparison with the much greater errors in the observations by which the theory is verified below.

Another important simplification occurs when the cardboard underlayer has the same reflectivity as a fluorescent layer of infinite thickness, that is if we have $r_2 = R_\infty$ or $\varrho_2 = \beta$. In that case we are led to

$$J_B = \frac{k\mu N_0 (1+\varrho_1)}{2(\varrho_1+\beta)(\mu+\sigma)} \left\{1 - e^{-(\mu+\sigma)D}\right\}, \tag{6}$$

$$J_F = \frac{k\mu N_0 (1+\varrho_1)}{2(\varrho_1+\beta)(\mu-\sigma)} \left\{e^{-\sigma D} - e^{-\mu D}\right\}. \tag{7}$$

Except for a constant factor $(1+\varrho_1)/2\ (\varrho_1+\beta)$ these formulae are equivalent to those of Glocker and his co-workers. Their equations remain valid if $r_2 = R_\infty$. It is to be noted, however, that whereas τ in (2) and (3) was interpreted as an absorption coefficient, σ in (6) and (7) is a coefficient in which absorption and scattering both play a part.

To apply these equations to practical cases we have to know the value of the various constants occurring in them. The reflectivity for diffuse light of the photographic film and the cardboard to which the fluorescent layer has been applied were both determined with an Ulbricht integrating sphere, the blue mercury line ($\lambda = 4358$ Å) being used as corresponding closely in colour with the fluorescence of calcium tungstate (max. intenxity at 4300 Å). The reflectivity R_∞ was determined in the same way; several films of calcium tungstate were piled upon each other until the reflecting power had reached a maximum. From these observations ϱ_1, ϱ_2 and β were computed.

The thickness D was measured with a micrometer and the absorption coefficient μ of the fluorescent material for X-rays was determined with the aid of a dosimeter by the equation

$$\mu = \frac{\log N_0 - \log N}{D \log e},$$

where N_0 = intensity of the X-rays at a given distance from the tube, and N = intensity of the X-rays at the same distance but filtered through the fluorescent layer.

It remains to fix the value of σ. Dividing (4) by (5) we have

$$\frac{J_B}{J_F} = \frac{(\varrho_2+\beta)(\mu-\sigma)e^{\sigma D} - (\varrho_2-\beta)(\mu+\sigma)e^{-\mu D} + 2(\varrho_2\sigma - \mu\beta)e^{-\mu D}}{2(\beta\mu+\varrho_2\sigma) - (\mu+\sigma)(\varrho_2+\beta)e^{-(\mu-\sigma)D} + (\mu-\sigma)(\varrho_2-\beta)e^{-(\mu+\sigma)D}} \tag{8}$$

or if we use (6) and (7)

$$\frac{J_B}{J_F} = \frac{\mu-\sigma}{\mu+\sigma} \cdot \frac{1-e^{-(\mu+\sigma)D}}{e^{-\sigma D} - e^{-\mu D}}. \tag{9}$$

The ratio J_B/J_F was determined by a photographic method. On different areas of one film two parts of the same screen were used as front-screen and back-screen respectively. This assembly was irradiated by X-rays until suitable blackenings were obtained after developing. To ensure sufficient homogeneity the X-rays were filtered through 5-15 mm aluminium.

To deduce the ratio J_B/J_F from these data we have to know the density-log. (intensity) curve for the film in question. This curve was determined by photographing an aluminium wedge. In order to obtain perfect proportionality between the logarithm of the intensity of the X-rays transmitted and the thickness A of the wedge, the X-rays had to be more thoroughly filtered by means of 0·85 mm copper. In this way a satisfactory homogeneity was obtained and for rays corresponding to a voltage of 70 kV $\mathrm{d} \log N/\mathrm{d}A$ was determined at 0·457 cm^{-1}. Since $\mathrm{d} \log N/\mathrm{d}A$ is equal to $\mathrm{d} \log J/\mathrm{d}A$ the photograph of the wedge gives us the desired density curve. From this curve the ratio J_B/J_F can be read off at once *). Now the value of σ can be found graphically. Therefore we calculate for different values of σ the corresponding value of J_B/J_F using formula (8) or (9). Having done this for a sufficient number of different values to be able to draw a curve giving the relation between J_B/J_F and σ, the required value of σ can be read off at once from this curve.

A commercial intensifying back-screen of calcium tungstate examined in this way gave the following results:

grain diameter: 5-15 μ, thickness of the layer: 0·056 cm
$\mu = 25$ cm^{-1} for a voltage of 55 kV, the X-rays being filtered first through 15 mm aluminium

$$J_B/J_F = 1{\cdot}92$$

R_∞	= 0·69	β	= 0·19 **)
r_2	= 0·39	ϱ_2	= 0·43
r_1	= 0·18	ϱ_1	= 0·70

If we plot J_B/J_F against σ according to (8) we obtain curve 1 of *fig. 3*. The value $J_B/J_F = 1{\cdot}92$ corresponds to $\sigma = 58$ cm^{-1}.

Using (9) instead of (8) the curve 1a is obtained. We see that for high values of σ this formula gives a satisfactory approximation, and for σ a value 60 cm^{-1} is found.

The same measurement was carried out with X-rays corresponding to the higher voltage of 90 kV and filtered through 0.5 mm molybdenum

$$\mu = 10 \text{ cm}^{-1}$$
$$J_B/J_F = 1{\cdot}29 \text{ (calculated: } 1{\cdot}32)$$

*) A correction has to be made for the blackening caused by the X-rays themselves.

**) The various reflection coefficients were measured by Dr J. Voogd, for which I should like to express my gratitude.

The back-screen just investigated is commonly used in combination with a front-screen of the same material but differing in thickness, D being 0·030 cm.

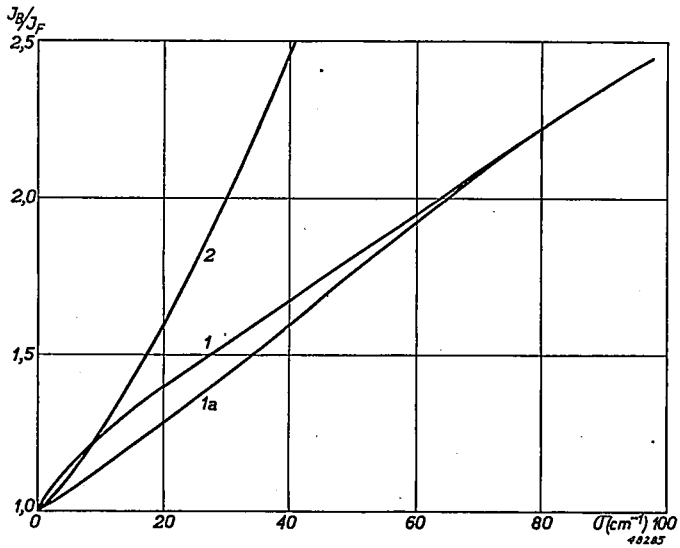


Fig. 3. Relation between J_B/J_F and σ.
1. according to (8); $\mu = 25$ cm^{-1}; $D = 0{\cdot}056$ cm. $\varrho_2 = 0{\cdot}43$; $\beta = 0{\cdot}49$
1a. according to (9); $\mu = 25$ cm^{-1}; $D = 0{\cdot}056$ cm
2. according to (9); $\mu = 26$ cm^{-1}; $D = 0{\cdot}078$ cm

J_B/J_F was in this case observed to be 1·28, against a computed value of 1·29 ($\mu = 25$ cm^{-1}), which illustrates the applicability of equation (8).

Finally the following test was carried out. Both screens ($D = 0{\cdot}056$ cm and $D = 0{\cdot}030$ cm) were used either as back-screen (B) or as front-screen (F), and the ratio of the intensities observed was compared with that calculated from equations (4) and (6) respectively and (5) and (7) respectively

The following results were obtained

voltage	μ (filter 10 mm Al)	$\frac{J(0{\cdot}056)}{J(0{\cdot}030)}_B$	$\frac{J(0{\cdot}056)}{J(0{\cdot}030)}_F$
80 kV	19 cm^{-1}	1·18	0·95
70 kV	24 cm^{-1}	1·11	0·85
60 kV	26 cm^{-1}	1·10	0·82
50 kV	35 cm^{-1}	1·11	0·72
40 kV	41 cm^{-1}	1·12	0·64

In *fig. 4* these values are plotted against μ. Curves IIa and IIIa give the

theoretical values according to (4) and (5) respectively, while IIb and IIIb correspond to (6) and (7) respectively. Moreover curve I gives the ratio of the amounts of light energy excited in both screens.

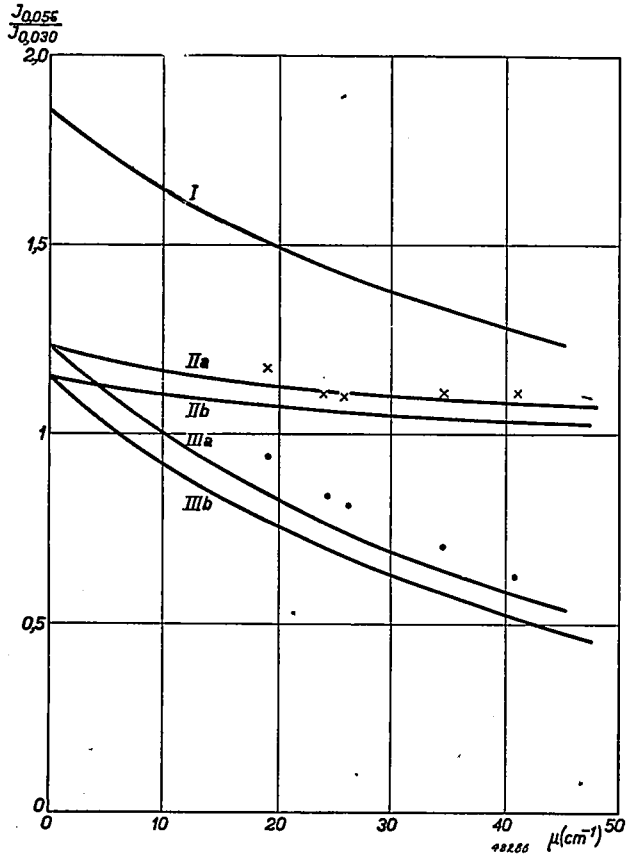


Fig. 4. Relation between $J(0{\cdot}056)/J(0{\cdot}030)$ and μ.

I for $\sigma = 0$
II theoretical curve for $\sigma = 58$ cm^{-1} both screens used as back-screen
a) according to (4)
b) according to (6)
III Theoretical curve for $\sigma = 58$ cm^{-1} both screens used as front-screen
a) according to (5)
b) according to (7)
×) measured values, both screens used as back-screen
•) measured values, both screens used as front-screen

This curve should represent also the ratio of the radiant power of both screens if there were no absorption of the light in the fluorescent layer ($\sigma = 0$).

The experimental results agree fairly well with the theoretical values,

having regard to the rather inaccurate photographic intensity measurements. Moreover there are other sources of error, hardly to be avoided, such as inhomogeneity of the fluorescent layer and of the X-rays and variation in the thickness of the layer, which may have produced inaccuracies both in D and in μ.

According to Hamaker

$$\beta = \sqrt{\frac{a}{a+2s}}$$

$$\sigma = \sqrt{a\,(a+2s)}\,.$$

Hence it follows

$$a = \beta\sigma\,, \qquad s = \frac{1-\beta^2}{2\beta}\,\sigma\,. \tag{10}$$

Knowing β and σ, we can now calculate the coefficients a and s:

$$a = 11\ \text{cm}^{-1}, \qquad s = 146\ \text{cm}^{-1}.$$

The coefficient of scattering s is high, so that on its way to the surface the light is scattered to a large extent in all directions, which evidently causes an increased absorption. This can be improved by reducing s which can be effected in two different ways, *viz.*

1) by soaking the fluorescent layer in oil; an oil stain on a thick layer of calcium tungstate is marked by a greater brightness, especially when used as front-screen;

2) by using grains of larger dimensions.

The screen investigated above had particles of 5-15 μ in size. To prove the influence of particle size a commercial screen with particles of 10-90 μ in diameter was examined in the same way with the following results:

$$D = 0{\cdot}078\ \text{cm}, \qquad R_\infty = 0{\cdot}77, \quad \beta = 0{\cdot}13\,.$$

For $\mu = 26\ \text{cm}^{-1}$, J_B/J_F was found to be 1·75.

Formula (9) can be applied in view of the high value of D. The theoretical curve is shown in fig. 3 (curve 2). From this curve we derive for σ the value 24 cm^{-1} and consequently according to (10)

$$a = 3{\cdot}1\ \text{cm}^{-1}, \qquad s = 90\ \text{cm}^{-1}.$$

Thus both a and s were found to be lower for the coarse-grained powder.

The curve showing the relation between J_F and D reaches a maximum J_m for a definite value D_m, which is about 0·2 mm for normal intensifying screens made of fine-grained calcium tungstate (grain diameter 3-15 μ).

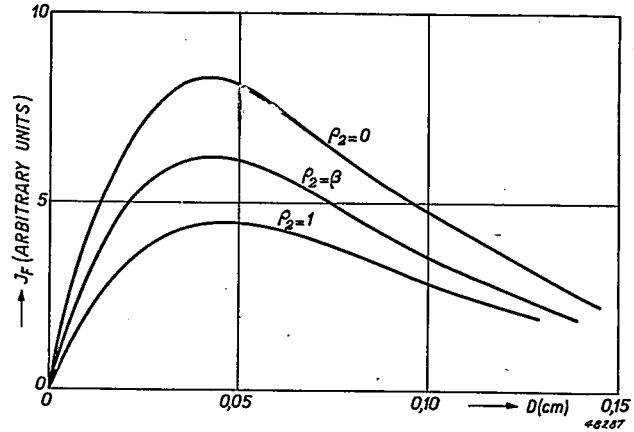


Fig. 5. Relation between J_F and D for different values of ϱ_2 ($\mu = 30$ cm^{-1}; $\sigma = 20$ cm^{-1}; $\beta = 0{\cdot}2$).

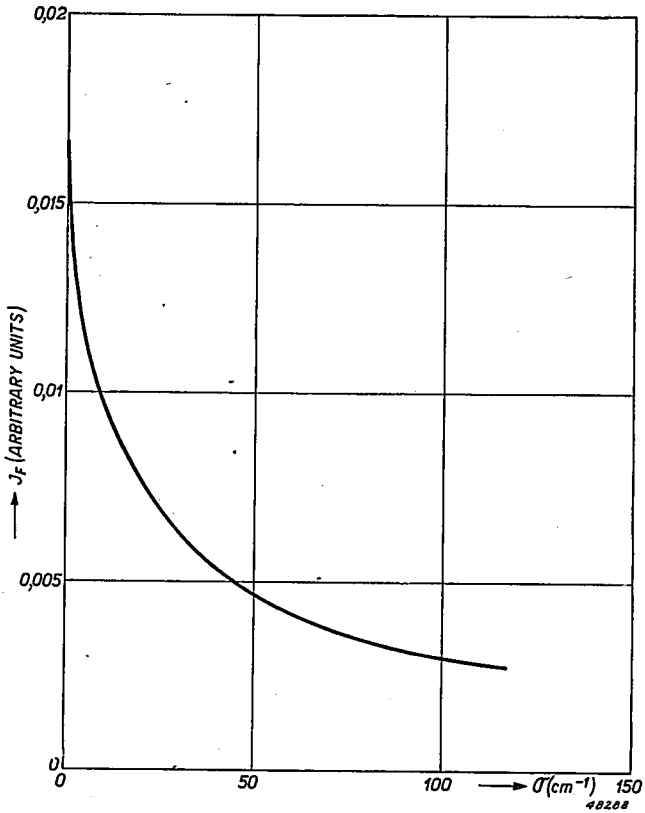


Fig. 6. Relation between the maximum value of J_F and σ ($\mu = 30$ cm^{-1}; $\varrho_2 = \beta$).

From equation (7) the value of D_m can be deduced, putting $dJ/dD = 0$.

$$\frac{dJ}{dD} = \frac{k\,\mu\,E_0}{(\varrho_1 + \beta)\,(\mu - \sigma)}\,(\mu e^{-\mu D} - \sigma e^{-\sigma D})$$

$$D_m = \frac{\log \mu - \log \sigma}{(\mu - \sigma)\log e}\,^{4)}.$$

This relation holds only when β and ϱ_2 have the same value, but *fig.* 5 shows that D_m varies very little for different values of ϱ_2. The maximum value of J on the other hand is greatly influenced by ϱ_2 and is much higher if the fluorescent layer is fixed upon an underlayer of high reflecting power. In *fig.* 6 the maximum value of J_F is plotted against σ for $\mu = 30$ cm^{-1}. We see that with increasing σ the value of J_m decreases very strongly. Therefore when aiming at screens with great brightness σ must be kept low and coarse-grained calcium tungstate should be used.

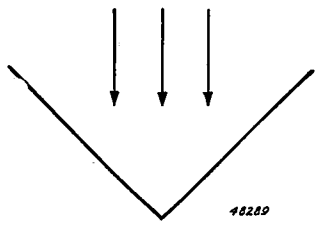


Fig. 7. Demonstration of the validity of Lambert's law in the case of fluorescence by X-rays.

The formulae for the radiant power of a fluorescent screen do not contain any information about the distribution in space of the radiation emitted. It can be shown by a simple experiment that generally this distribution corresponds exactly to Lambert's law . When irradiating two parts of a screen as shown in *fig.* 7, the X-rays incident along the bisector of the angle between the two parts, both show the same brightness no matter from which direction we look at them.

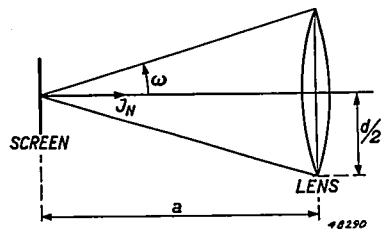


Fig. 8. The amount of light received by the camera lens in X-ray screen-photography.

For X-ray camera-photography it is of importance to know the fraction f of the radiated light contributing to the blackening of the film. The amount of light received by the lens of the photographic apparatus is (*fig. 8*):

$$\int_{\omega=0}^{\omega=\arcsin(d/2a)} \frac{J_N \cos\alpha}{a^2} 2\pi a \sin\omega \, a\,d\omega = \pi J_N \frac{d^2}{4a^2}.$$

Thus f is equal to $d^2/4a^2$, which is twice as high as the value calculated by Bouwers and Burger [5]), who assume an isotropic space distribution of the radiated light.

In conclusion I would like to thank Dr Hamaker for the many valuable discussions on the subject of this paper.

Eindhoven, November 1942

REFERENCES

1) M. Widmann, Z. tech. Phys. **22**, 27, 1941.
2) R. Glocker, E. Kaupp and H. Widmann, Ann. Phys., Lpz. **85**, 313, 1928.
3) H. C. Hamaker, Philips Res. Rep. **2**, 55, 1947.
4) Deduced already by H. Klug, Fortschr. Rö. **55**, 191, 1937.
5) A. Bouwers and G. C. E. Burger, Philips tech. Rev. **5**, 258, 1940.

ABSTRACTS

(Continued from page 67)

1698: D. Polder and J. H. van Santen, The effective permeability of mixtures of solids, Physica 's-Grav. **12**, 257-270, 1946.

For several purposes it is of importance to know how the dielectric or magnetic permeability of a mixture of substances that show mutually different permeabilities in pure state depends on the composition of the mixture. In this paper the case is dealt with where the different substances are powders of solids, packed in a medium like oil or wax or air. The individual particles or holes are assumed to be ellipsoidal. The calculation is based on an approximation which for spherical particles proves to give results identical with those obtained by Böttcher's method. The effect of the shape of the ellipsoids is discussed. Considerations are given concerning the nature of the approximations used in this paper and in other theories on this subject.

1699: M. J. O. Strutt and A. van der Ziel, Signal-to-noise ratio at v.h.f., Wireless Engineer **23**, 241-249, 1946.

The two basic concepts of correllation between fluctuation currents and voltages and of the decomposition of fluctuating quantities into singly periodic components are discussed briefly. The noise ratio of a grounded-cathode amplifier is evaluated and the conditions for minimum noise ratio are stated. Experimental evidence is shown to confirm the theoretical results. Further grounded-grid amplifier stages and velocity-modulation valves are considered and the requirements as to valve construction (uniform electron paths and low dielectric and other losses) are dealt with. It is shown that the proposed reduction of noise ratio applies also to wide-band reception. In order to compensate the loss or gain incurred by the reduction of noise ratio suitable feed-back may be applied. The interrelation of noise figures introduced by various authors and the present noise ratio are discussed.

1700: B. D. H. Tellegen, Het bepalen van de integratieconstanten bij de berekening van in- en uitschakelverschijnselen, T. Ned. Radiogenootschap **11**, 173-188, 1946 (The evaluation of the integration constants occurring in the computation of transient phenomena).

A network is considered containing an e.m.f. v under the influence of which a current i flows in a certain branch. A method is given to calculate from the differential equation connecting i and v the discontinuities in i and its derivatives resulting from discontinuities in v and its derivatives.

1701*: H. A. Klasens, Transfer of energy between centres in zinc sulphide phosphors, Nature, London, **158**, 306, 1946.

The transfer of energy between centres in zinc sulphide phosphors containing two kinds of centres (*e.g.* blue and green) is explained by the supposition that after bringing an electron from one centre (*e.g.* blue) to an excited state (by lifting it in the empty so-called conduction band of energy), the hole in the blue centre may be transferred to a filled band and thus may travel to a green centre. The reverse process (green to blue) is supposed to have a negligible probability. The above hypothesis is expressed in the form of two simultaneous differential equations giving the number of excited centres as a function of time.

From these equations formulae may be derived for the intensity ratio of the blue and the green emission during irradiation as a function of the irradiation intensity and the temperature, and for the temperature dependence of the emission of *e.g.* the blue centres. The effect of "killers" on the after-glow (green emission) is explained on similar lines.

1702: J. Boeke, Ontwikkeling van de electrische meetapparaten in de chemie gedurende de oorlog, Chem. Weekblad **42**, 230-273, 1946 (Development of electrical measuring instruments for chemical purposes during the war).

This article contains a survey of recent improvements in electrical methods as used in chemistry; especially in the following fields: electrochemistry, dielectric constants, measurement of moisture, spectrometry (absorption and emission), chemical analysis, industrial apparatus.

1703: W. de Groot, The influence of irradiation with light on the dielectric proporties of ZnS phosphors, Physica 's-Grav. **12**, 402-404, 1946.

Experiments are described, whereby the influence of irradiation with light on the light emission and the dielectric behaviour of ZnS phosphors are examined simultaneously. It is proved that the change of capacity of a condenser containing ZnS, on irradiation, and the decay of this change, on stopping irradiation, are much slower than the rise and decay of luminescence. Some arguments are given in favour of the supposition that the change in dielectric behaviour must be ascribed to free electrons and not to trapped electrons. Some phenomena, as *e.g.* a steep initial decay of the ΔC-effect in a phosphor containing a "killer", remain unexplained.

VOL. 2 NO. 2 APRIL 1947

Philips Research Reports

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

R 37 *Philips Res. Rep. 2, 81-102, 1947*

HARDENING OF METALS BY INTERNAL OXIDATION

by J. L. MEIJERING and M. J. DRUYVESTEYN

PART I 621.785.613

Summary

Alloys of silver, copper and nickel can be dispersion-hardened by diffusing oxygen into an alloy with e.g. 1-2 atomic % of an element having a sufficient affinity for O. Too small an affinity leads to a coarser distribution of the oxide formed, because conglomeration must take place via the atoms, and dissociation occurs more frequently when the oxide is not very stable. Thermodynamical considerations are given, and calculations of the penetration of the reaction front. More experiments (including X-ray, electrical and mechanical investigations of the internally oxidized alloys) will follow in Part II.

1. *General aspects*

Hardening a ductile metal (or alloy) without altering its structure or composition fundamentally can only be accomplished by blocking the glide planes with minute particles, so-called "dispersion hardening". Generally these particles are formed by ageing a supersaturated solid solution, obtained by quenching from higher temperatures. The basic condition for this "precipitation hardening" is an appreciable fall in solid solubility with decreasing temperature. The hardness obtained is more or less destroyed by annealing at temperatures higher than the most favourable ageing temperature; this is caused by coarsening or partial resolving of the precipate. It will be possible to obtain more stable precipitates by a chemical reaction of the solute B with another element X diffusing into the solid solution A + B. Obviously this will require two things:

1st: X must diffuse more rapidly in A than B does, otherwise the compound (*e.g.* BX) will not be formed in the interior, but as a surface layer;

2nd: the affinity of X for B must be sufficiently greater than for A.

Now *oxygen* diffuses through silver [1]), copper [2]) and nickel [3]) with a speed considerably greater than metallic elements.

Its affinity for Ag is so small that silver cannot form (solid) Ag_2O when

heated above 190 °C in O_2 of normal pressure. This affinity is larger for Cu and Ni, but remains considerably smaller than for such metals as Mg, Be, Al.

It therefore seemed quite possible to harden an alloy of Ag with *e.g.* 1 or 2 atomic % Mg, simply by heating it in air or O_2 at 800 °. This indeed proved to be the case.

The following results were obtained with strips 1 mm thick (machined from castings); after 2 hours heating in air the hardness was:

0·3 weight % Mg ± 170 Vickers (kg/mm², load: $2^1/_2$ kg),
0·4 weight % Al ± 160 Vickers,
0·2 weight % Be ± 135 Vickers.

Similar strips when annealed in H_2 or N_2 showed hardnesses of 30 to 50 Vickers only.

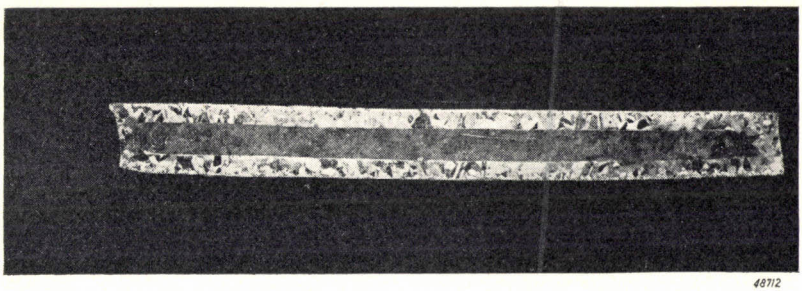


Fig. 1. Ag with 0·3 % Mg, 50 minutes 800° air; etched $NH_4OH + H_2O_2$. Magn. 10×.

With an alloy of Ag with 0·7 % Mn a hardness of 140 was reached after oxidation in air, whilst Ag with 0·6 % Ti, although not homogeneous, gave 100 Vickers, but Zn (0·8 %) and Sn (1·5 %) yielded much smaller hardnesses (± 60), while Cd (1 %) and Al (1 %) with only 40 *) and 30 Vickers respectively did not harden at all **).

With the exception of the last mentioned all these alloys were deeply penetrated by oxygen, which transformed the solute into oxide. This was shown by the appearance of a rim in the cross-section of the strips, increa-

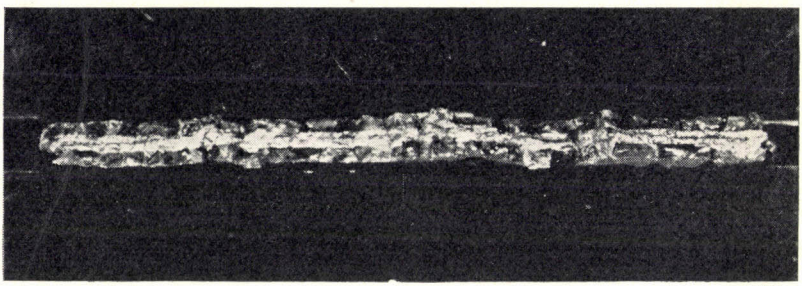
*) This low hardness cannot be due to loss of CdO by evaporation, as it was found also 0·1 mm beneath the original surface.

**) We found this hardening of Ag (and also that of Cu and Ni) in 1942. It was not until 1945 that we learned from the Chem. Abstr. that Rhines and Grobe [4]) had examined the internal oxidation of silver alloys metallographically. We have not yet seen the paper itself, but apparently no hardening effects were mentioned. This is confirmed by the paper of Chaston [5]), which we saw while ours was being written. Chaston describes the hardening of Ag with several concentrations of Al. There is very good agreement with our results.

sing in depth with time of oxidation. This rim (with a sharp boundary, see section 6) becomes visible after polishing and etching (*e.g.* with aqueous ammonia + H_2O_2) because it etches more rapidly than the untransformed interior. In those cases where oxidation causes considerable hardening this rim is also shown by breaking the strip, its fracture being brittle (crystalline), whilst the interior has a tough fracture. On thicker specimens it was verified that the Vickers hardness was only abnormally high in the rim.

The absence of a good rim in the alloy with 1 weight % Al accounts for its unsuitability for hardening by oxidation; the formation of a surface layer rich in Al_2O_3 evidently makes this alloy resistant against its further oxidation.

The absence (or smallness) of the hardening by oxidation in the alloys



48713

Fig. 2. Cu with 0·35 % Be, $1^1/_2$ h. 950° air. Crystalline fracture of internally oxidized rim. Magn. ± 10×.

with Cd, Zn and Sn appears to be correlated with their affinities for O being less than those of Al, Be, Mg and Mn. This point will be considered in greater detail in section 6.

Several alloys of *copper* were then also succesfully hardened by oxidation. Because of the higher melting point and the slower diffusion of oxygen in Cu most experiments were made at 950° in air. An essential difference with the oxidation of the Ag alloys is of course that a scale of copper oxide (mostly Cu_2O) is formed, which is easily removed. By heating in air only as long as is necessary for the formation of a Cu_2O layer sufficiently thick to supply the oxygen required, and completing the annealing in a neutral or weakly oxidizing atmosphere (N_2, CO_2), the loss of metal can be reduced.

In most cases the remaining metal shows a rim (when the strip is oxidized so long that only part of the cross-section is occupied by it), which in this case is sometimes already visible after polishing, its colour being more copper-red than the interior. The formation of these rims has already been

studied for a large number of Cu alloys [6] [7] [8], but we do not find any reference to the appreciable hardness obtained in some cases.

The highest values were obtained with Be (0·2-0·35 weight %): 160-170 Vickers. Mg (0·6 %), Ti (1 %) and Al (0·6-1·0 %) gave values ranging from 105 to 125 Vickers. Zr (0·8 %) too, in spite of its small solid solubility (± 0·3 *atomic* %) in Cu, gave a hardness slightly less than 100. Mn, Zn, Si, Sn and Sb gave rims too, but no or little hardening. Cu with 2·2 and 4·5 weight % Al gave no rims at all and therefore no hardening (50 Vickers). These last alloys are known to be fairly resistant to oxidation [7], the formation of copper oxide on the surface being incomparably less than on the 1 % alloy.

We see that while Mn enables oxidation hardening in silver, it does not so in copper with its greater affinity for oxygen.

Nickel, which has a still greater affinity, only gave oxidation hardening (at 1200°) when alloyed with Al. Be, Mg, Ti, Si, Mn and Cr gave no or only small effects.

In the next four sections calculations of the penetration of oxygen in these alloys are given. In section 6 the conglomeration of the oxide particles formed is discussed, it being shown that the difference between hardening and non-hardening alloys can be explained rather well on a thermodynamical basis. Measurements of lattice constants (section 8) and electrical resistivities (section 12) support the view that the MgO and Al_2O_3 particles that harden silver are very small.

2. *Calculation of the penetration of oxygen in a dilute solid solution*

We will now consider the first condition for hardening by oxidation (page 81) more closely. We take for example a thick plane strip of an alloy of Ag with Mg (atomic fraction c_2) *) heated in oxygen. We suppose the rate of solution of O at the surface to be so great that we can take its concentration here to be constant $= c_1$ (atomic fraction; a function of temperature and oxygen pressure). For Ag alloys at 800 °C this will be practically the case already after a short time **). The diffusion coefficients D_1 and D_2 of O and Mg respectviely we may take to be independent of the concentration of O, Mg and MgO, when c_2 is small (c_1 will always be small). The concentrations of O and Mg (which we will call u_1 and u_2 respectively) are functions of the distance from the surface of the strip x, and of time. They

*) For non-bivalent solutes this must be modified according to the formula of the oxide formed.

**) Johnson and Larose [1] found the quantity of O_2 passing through thin silver strips into a vacuum approximately inversely proportional to their thickness, at 600°-700 °C.

change not only by the diffusion but also by their mutual reaction, which we will assume to be bimolecular. We get the following differential equations:

$$\left.\begin{aligned} \frac{\partial u_1}{\partial t} &= D_1 \frac{\partial^2 u_1}{\partial x^2} - K u_1 u_2 \\ \frac{\partial u_2}{\partial t} &= D_2 \frac{\partial^2 u_2}{\partial x^2} - K u_1 u_2 \end{aligned}\right\} \tag{1}$$

with the conditions

$$\begin{aligned} u_1 &= c_1 \quad \text{for } x = 0 \text{ and all } t, \\ u_2 &= c_2 \quad \text{for } t = 0 \text{ and all } x. \end{aligned} \tag{2}$$

To enable us to obtain a solution we assume that the reaction is so fast that free Mg and O are nowhere together. This results in a sharp boundary (where the whole reaction is taking place) between an outer part containing O and MgO, and an inner part containing Mg. This supposition, suggested by the experiments (see section 1), will be examined later on.

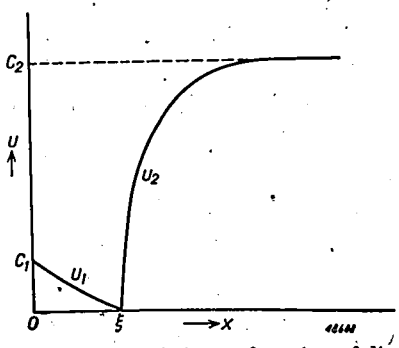


Fig. 3. Concentration of O (u_1) and Mg (u_2) as a function of distance from the surface of the strip. (Schematically).

Its effect is that the term $K u_1 u_2$ disappears in both equations (1) while three new boundary conditions arise:

$$\left.\begin{aligned} u_1 &= 0 \\ u_2 &= 0 \\ D_1 \frac{\partial u_1}{\partial x} &= -D_2 \frac{\partial u_2}{\partial x} \end{aligned}\right\} \text{for } x = \xi. \tag{3}$$

ξ, a function of time, is the distance of the boundary from the surface of the strip. The third condition expresses that equal numbers of O and Mg atoms are meeting at the boundary.

We now take u_1 and u_2 to be functions of $x/\sqrt{t}$ *). Then $\xi/\sqrt{t}$ is a constant, and further we can transform both our partial differential equations (without $K u_1 u_2$) into ordinary ones, which are easily integrated to:

$$u_1 = A_1 + B_1 \operatorname{erf}\left(\frac{x}{2\sqrt{D_1 t}}\right),$$

$$u_2 = A_2 + B_2 \operatorname{erf}\left(\frac{x}{2\sqrt{D_2 t}}\right),$$

$\left(\operatorname{erf} y \text{ is the normalized Gauss function } \frac{2}{\sqrt{\pi}} \int_0^y e^{-z^2}\, dz\right)$. From the conditions (2) it follows that $A_1 = c_1$ and $A_2 + B_2 = c_2$.

With the first two of the conditions (3) we complete our elimination of the four integration constants, $\xi/\sqrt{t}$ remaining the only unknown, which is then given implicitly by the remaining condition. Writing α for $\xi^2/4t$ the result is:

$$u_1 = c_1 \left\{1 - \frac{\operatorname{erf}\left(\frac{x}{2\sqrt{D_1 t}}\right)}{\operatorname{erf}\left(\sqrt{\frac{\alpha}{D_1}}\right)}\right\} \qquad \text{for } x < \xi,$$

$$u_2 = c_2 \left\{\frac{\operatorname{erf}\left(\frac{x}{2\sqrt{D_2 t}}\right) - \operatorname{erf}\left(\sqrt{\frac{\alpha}{D_2}}\right)}{1 - \operatorname{erf}\left(\sqrt{\frac{\alpha}{D_2}}\right)}\right\} \qquad \text{for } x > \xi,$$

where α is given by the parameter equation:

$$\frac{c_1 \sqrt{D_1}\, e^{-\frac{\alpha}{D_1}}}{\operatorname{erf}\left(\sqrt{\frac{\alpha}{D_1}}\right)} = \frac{c_2 \sqrt{D_2}\, e^{-\frac{\alpha}{D_2}}}{1 - \operatorname{erf}\left(\sqrt{\frac{\alpha}{D_2}}\right)}. \tag{4}$$

*) This can be proved *a priori* by a dimensional consideration, as in the case of an element diffusing into a strip of *pure* metal (infinitely thick) but with a diffusion coefficient varying with its concentration. Indeed, Mg can be treated as negative oxygen in our problem.

The problem of the penetration of *e.g.* C in Fe at 800°, with the formation of a second phase (austenite), can be solved in the same way (provided that the transference at the phase boundary is rapid enough). Only the concentration then shows not only a kink but a jump. This completes the analogy with a well-known heat-conduction problem: the penetration of frost in a moist soil. The only essential difference in the solution is that the parameter equation for α becomes rather more complicated than in our oxidation problem.

We can also calculate the concentration of the MgO formed, if it remains *in situ* after its formation, which is reasonable to assume. The number of molecules MgO formed is equal to that of the Mg atoms arriving at the boundary; to get the number of molecules in a certain fixed place we must divide by the velocity with which the boundary is moving. Calling the molecular fraction of the oxide c_{ox}, we find

$$c_{ox} = \frac{D_2 \left(\frac{\partial u_2}{\partial x}\right)_{x=\xi}}{\frac{d\xi}{dt}} = \frac{1}{\sqrt{\pi}} c_2 \frac{\sqrt{\frac{D_2}{a}} e^{-\frac{a}{D_2}}}{1 - \operatorname{erf}\left(\sqrt{\frac{a}{D_2}}\right)},$$

thus greater than c_2, and independent of the depth x.

Convenient expansions can be derived when $D_1 >> a >> D_2$, which is the case when $c_1 D_1 >> c_2 D_2$ but $c_1 << c_2$. These conditions are in most practical cases well fulfilled: for Ag in air at 800° c_1 is about 10^{-4} (0·01 at. %) while D_1 is $\pm$ 1 cm²/day *); c_2 will be normally about 1 at. % and D_2 for metallic solutes in silver at 800° is about 10^{-4} cm²/day [10]), so here $c_1/c_2 = 10^{-2}$ and $D_1/D_2 = 10^4$ approximately.

For $D_1 >> a >> D_2$ we use the approximations

$$\frac{e^{-y^2}}{y \operatorname{erf} y} = \frac{\sqrt{\pi}}{2y^2} \qquad \text{for } y << 1$$

and

$$\frac{e^{-y^2}}{y(1 - \operatorname{erf} y)} = \sqrt{\pi}\left(1 + \frac{1}{2y^2}\right) \qquad \text{for } y >> 1$$

and obtain: $\frac{c_{ox}}{c_2} = 1 + \frac{c_2 D_2}{c_1 D_1}$.

In the limiting case $\left(\frac{D_2}{D_1} = 0 \text{ and } \frac{c_1}{c_2} \to 0\right)$ we have $u_1 = c_1\left(1 - \frac{x}{\xi}\right)$ for $x < \xi$ and $u_2 = c_2$ for $x > \xi$, and $a = \frac{c_1 D_1}{2c_2}$ (which is easily derived directly from the differential equation $c_2 \frac{d\xi}{dt} = D_1 \frac{c_1}{\xi}$). In the next approximation

$$a = \frac{c_1 D_1}{2c_2}\left(1 - \frac{c_2 D_2}{c_1 D_1} - \tfrac{1}{3}\frac{c_1}{c_2}\right).$$

We will now try to examine how sharp the oxidation boundary is in the important case that D_2 is practically zero. We return to the differential equations (1) *with* the term Ku_1u_2 but with $D_2 = 0$. Then

*) Calculated from Johnson and Larose [1]) and Steacie and Johnson [9]). See also section 7.

$u_1 = -\frac{1}{K}\frac{\partial \ln u_2}{\partial t}$ and thus $\frac{\partial u_1}{\partial t} = -\frac{D_1}{K}\frac{\partial^3 \ln u_2}{\partial x^2 \partial t} + \frac{\partial u_2}{\partial t}$. Integration gives $u_2 - u_1 = \frac{D_1}{K}\frac{\partial^2 \ln u_2}{\partial x^2} + f(x)$. For $t = 0$ $u_2 \equiv c_2$ and $u_1 \equiv 0$, so $f(x) = c_2$.

In normal cases, when $c_1 \ll c_2$, it is to be expected that u_1 will everywhere be small in comparison with $c_2 - u_2$; we will at first ignore it. Putting $\ln \frac{c_2}{u_2} = y$ we then get $\frac{d^2y}{dx^2} = -\frac{K}{D_1} c_2 e^{-y} + \frac{K}{D_1} c_2$. After multiplying by $\frac{dy}{dx}$ this is integrated to $x = \sqrt{\frac{D_1}{2Kc_2}} \int_b^y \frac{dz}{\sqrt{e^{-z} + z + A}}$.

The first integration constant A is found to be $= -1$, given the condition that $dy/dx = 0$ for $y = 0$ ($x = \infty$), while the second constant b is of no importance.

We must now say something about K. Mg is substituted on the Ag lattice points while O will probably be dissolved in the equally numerous octahedral interstices, which are situated also on a face-centered cubic "lattice". We assume that the diffusion takes place only by jumps between neighbouring interstices. Then the jumping frequency of an O atom is 12 D_1/a^2, where a is the lattice constant (about 4×10^{-8} cm). Now it is plausible to assume that reaction takes place when an O atom jumps from an interstice to a neighbouring one, which has a (free) Mg atom on one of those four surrounding lattice points which are not shared with the first interstice. We are only considering dilute solutions, so the fraction of the jumps that give reaction is $4u_2$. We thus conclude that $K = 48\ D_1/a^2$.

A good measure for the sharpness of the oxidation boundary will be the distance between the plane where $c_{ox} = 0{\cdot}1\ c_2$ and the plane where $c_{ox} = 0{\cdot}9\ c_2$. In our limiting case $c_{ox} = c_2 - u_2$ (because the Mg atoms are immobile), so this measure, after substitution of A and K and graphical integration, is found to be:

$$x_{0{\cdot}1} - x_{0{\cdot}9} = \frac{a}{\sqrt{96c_2}} \int_{\ln 10/9}^{\ln 10} \frac{dz}{\sqrt{e^{-z} + z - 1}} = \frac{0{\cdot}50}{\sqrt{c_2}}\, a\,.$$

Further $x_{0{\cdot}01} - x_{0{\cdot}99} = \frac{0{\cdot}99}{\sqrt{c_2}}\, a$; $x_{0{\cdot}001} - x_{0{\cdot}999} = \frac{1{\cdot}43}{\sqrt{c_2}}\, a$; $x_{0{\cdot}0001} - x_{0{\cdot}9999} = \frac{1{\cdot}85}{\sqrt{c_2}}\, a$, *etc.* The latter integrals are easily calculated from the former ones because for $y \gg 1$ $\int^y \frac{dz}{\sqrt{e^{-z} + z - 1}}$ is $2\sqrt{y-1} + C'$ and for $y \ll 1$ it is

$\sqrt{2} \ln y + C''$. The curve of c_{ox} against x has a point of inflexion at $c_{ox} = 0{\cdot}58\ c_2$. In practice c_2 is about 0·01, so $x_{0{\cdot}1} - x_{0{\cdot}9}$ should be about 2×10^{-7} cm. Microscopical observation only indicates (section 6) that it is smaller than $\pm$ 100 times this value, thus it remains possible that K is even 10^4 times smaller than our theoretical value. This means that *if* the reaction has an activation energy (besides that of the diffusion), it must be less than 20 kcal per g.mol. MgO.

The term u_1 disregarded above in the equation for $\frac{\partial^2 \ln u_2}{\partial x^2}$ makes the absolute value of this differential quotient everywhere slightly greater, so that the sharpness of the boundary is increased a little. When D_2/D_1 is not zero we think that the gradient of c_{ox} (not of u_2!) remains very steep when $c_1 << c_2$ and $c_1 D_1 >> c_2 D_2$ (as is normally the case), provided the oxide formed is stable.

3. *Influence of scale formation*

When *e.g.* a copper alloy is oxidized, the metallic rim of Cu + *e.g.* BeO is formed beneath a scale of copper oxide (+ BeO). We will calculate the penetration of the oxidation boundary with time for the "normal" (limiting) case ($D_1 >> a >> D_2$), assuming that the depth of scaling *) η is proportional to $\sqrt{t}$. We put $\eta = 2\sqrt{\beta t}$ in analogy with $\xi = 2\sqrt{\alpha t}$ for $\beta = 0$ (section 2).

In all our experiments with Cu and Ni alloys $\beta < \alpha$, thus automatically $\beta << D_1$. We further assume that at the interface scale/metallic rim u_1 remains constant = the solubility concentration c_1. We get $\frac{d\xi}{dt} = \frac{2\alpha}{\xi - 2\sqrt{\beta t}}$ **)

which, after putting $\frac{\xi}{\sqrt{t}} = z$, is integrated to

$$C - \ln t = \left(1 + \sqrt{\frac{\beta}{4\alpha + \beta}}\right) \ln\left(z - \sqrt{\beta} + \sqrt{4\alpha + \beta}\right) + \\ + \left(1 - \sqrt{\frac{\beta}{4\alpha + \beta}}\right) \ln\left(z - \sqrt{\beta} - \sqrt{4\alpha + \beta}\right).$$

We have tacitly assumed that the scale formation begins at $t = 0$, but we will consider the case where a metallic oxidized rim (of thickness ξ_0) may already be present (from a preliminary oxidation with subsequent removal of the scale).

If $t \to 0$, $z \to \infty$ (for $\xi_0 \neq 0$) and $C - \ln t = 2 \ln z = \ln \xi^2 - \ln t$. So $C = \ln \xi_0^2$.

*) *i.e.* the thickness of the metal layer converted into oxide.

**) The zero point of ξ is situated at the original surface of the metal.

The general solution becomes:

$$\left.\begin{aligned} t &= \frac{\xi_0^2}{(4\alpha+\beta)\,(w+1)^{1+p}\,(w-1)^{1-p}} \\ w &= \frac{(\xi/\sqrt{t})-\sqrt{\beta}}{\sqrt{4\alpha+\beta}}\,; \quad p=\sqrt{\frac{\beta}{4\alpha+\beta}} \end{aligned}\right\} \quad \text{for } \xi_0 \neq 0\,.$$

When $\xi_0 = 0$ evidently w must be $= 1$ and thus $\xi = (\sqrt{4\alpha+\beta}+\sqrt{\beta})\sqrt{t}$. As is to be expected, this penetration is faster, but the thickness of the metallic rim $= (\sqrt{4\alpha+\beta}-\sqrt{\beta})\sqrt{t}$ is smaller than when no scale is formed ($\beta = 0$). In *fig. 4* the results of a calculation with $\beta = \frac{1}{6}\alpha$ (as is about the case with *e.g.* 0·25 weight % Be-copper at 950°) are given; the curve for $\xi_0 \neq 0$ is nearing asymptotically (for $\xi/\xi_0 \to \infty$) to the straight line that is the solution for $\xi_0 = 0$. The other, dashed, line gives the depth of the scaling.

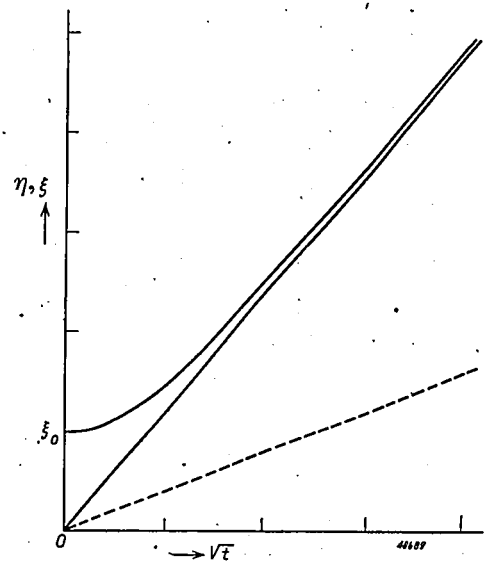


Fig. 4. Depth of internal oxidation (ξ) and scaling (η). $\beta = \frac{1}{6}\alpha$.

4. *Plane strips of finite thickness*

In section 2 we saw that c_{ox} is always greater than c_2. This, of course, is only possible when the strip is infinitely thick. In plane strips of finite thickness the results of section 2 remain quantitatively valid near the surface, but near the middle c_{ox} must become less than c_2, because the total amount of molecules MgO must be the same as the total amount of Mg atoms before the oxidation. The solution of this problem presents great

difficulties, except when $D_1 = D_2$ *), because then we can subtract the two differential equations (1) (section 2), which gives: $\frac{\partial(u_1 - u_2)}{\partial t} = \frac{D\,\partial^2(u_1 - u_2)}{\partial x^2}$.

The problem is now analogous to that of *e.g.* the carburation of a strip. The solution satisfying the initial and boundary conditions can be written:

$$u_1 - u_2 = -c_2 + (c_1 + c_2)\left[\left\{1 - \operatorname{erf}\left(\frac{x}{2\sqrt{Dt}}\right)\right\} + \right.$$

$$\left. + \left\{1 - \operatorname{erf}\left(\frac{d-x}{2\sqrt{Dt}}\right)\right\} - \left\{1 - \operatorname{erf}\left(\frac{d+x}{2\sqrt{Dt}}\right)\right\} - \left\{1 - \operatorname{erf}\left(\frac{2d-x}{2\sqrt{Dt}}\right)\right\} \ldots\right]$$

when d is the thickness of the strip. At the oxidation boundary $u_1 - u_2 = 0$.

We now solve the case when $D_1 = D_2$ and c_1 is so chosen that at the surface $c_{ox} = 1{\cdot}01\ c_2$ ($c_{ox} = 1{\cdot}01\ c_2$ corresponds to our standard case for Ag, see p. 87, where $D_1 = 10^4\ D_2$ and $c_1 = 10^{-2}\ c_2$ approximately). We thus replace the few fast O atoms by many slow ones. With the formulae derived in section 2 it is found that c_1/c_2 must be $2{\cdot}389.10^{22}$. Because of this extremely high value the terms $1 - \operatorname{erf}(x/2\sqrt{Dt})$ and $1 - \operatorname{erf}\{(d-x)/2\sqrt{Dt}\}$ are already very small at the oxidation boundary, and the next terms can be ignored.

We again call x at the boundary ξ, and put $\xi/2\sqrt{Dt} = y$ and $(d-\xi)/2\sqrt{Dt} = z$, which gives:

$$\begin{cases} y/z = \xi/(d-\xi) \\ 1 - \operatorname{erf} y + 1 - \operatorname{erf} z = 4{\cdot}186.10^{-23}. \end{cases}$$

When y and z are not approximately equal, one of the two terms $1 - \operatorname{erf}$ will predominate completely and the values of $\xi/\sqrt{t}$, $(d-\xi)/\sqrt{t}$ respectively will be the same as at the surface. Near the middle ($\xi = d/2$) the value is slightly greater. Even at $\xi = 0{\cdot}4833\ d$ the difference is not yet 0·001 %, exactly in the middle 0·7 %.

The calculation of c_{ox} as a function of x/d in our artificial case is also easy. With

$$c_{ox} = \frac{D\left(\frac{\partial u_2}{\partial x}\right)_{x=\xi}}{\frac{\mathrm{d}\xi}{\mathrm{d}t}} \text{ we get } \frac{c_{ox}}{c_2} = \frac{(e^{-y^2} - e^{-z^2})^2}{ye^{-y^2} + ze^{-z^2}} \cdot \frac{2{\cdot}389.10^{22}}{\sqrt{\pi}}.$$

This ratio becomes only appreciably different from 1·0100 near the middle of the strip. At $x = 0{\cdot}4833\ d$ it has fallen to 1·0064; its further course is pictured in *fig. 5*.

*) and when $D_2 = 0$ of course.

Now the case we have solved ($c_1/c_2 = 4 \cdot 10^{22}$!) has in itself no practical significance, but as c_{ox}/c_2 and thus α near the surface are the same as in our real standard case, u_2 as a function of x will also be the same there (section 2). The two "waves" of Mg deficiency will therefore begin to interfere in the middle in the same way. Furthermore c_{ox}/c_2 integrated over the total

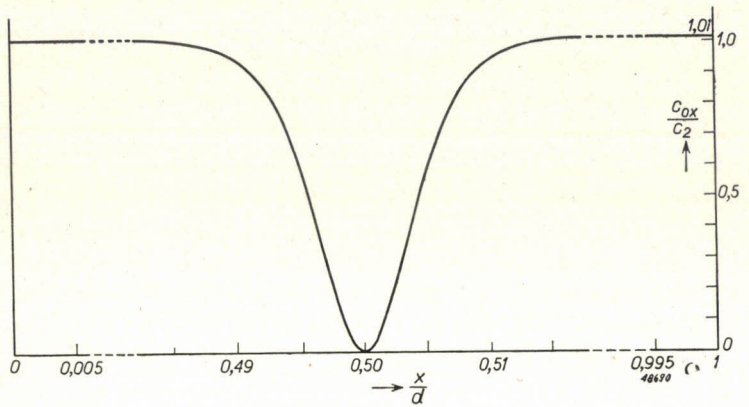


Fig. 5. Distribution of MgO over cross-section of Ag strip, after oxidation is complete, the enrichment at the surface by diffusion of Mg being 1 %.

diameter of the strip must yield the same value in both cases, and lastly it is easy to see that exactly in the middle c_{ox} and dc_{ox}/dx must be zero in our real case also. We therefore think that the curve in fig. 5 will not differ very much from the real one. This narrow zone poor in oxide can often be clearly seen at low magnifications in the fracture of 1 mm-thick hardened Ag or Cu strips (after complete oxidation is certain !), running as a thin line over the crystals. This "middle-line" can be more easily shown by polishing and etching; in Ag + 1% Cd, which does not harden, it was also found, though more vague.

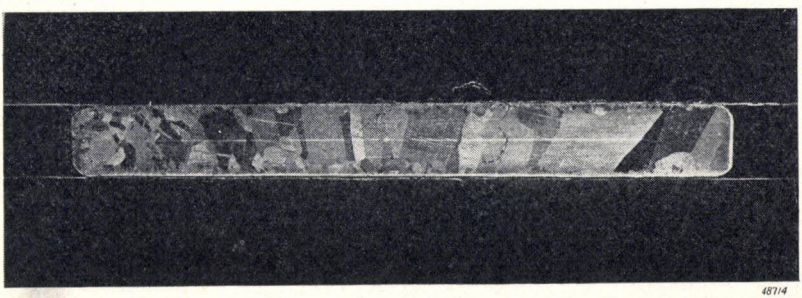


Fig. 6. Ag with 1·5 weight % Mn, 48 h. 800° air; etched $NH_4OH + H_2O_2$. Magn. 10×. Although oxidation is completed, a "middle-line", poor in oxide, persists (compare fig. 5).

5. *Cylindrical and spherical specimens*

We have experimented with wires too; calculations were accordingly made for a cylinder, and for the sake of completeness also for the other quasi one-dimensional case: the sphere. We can do this only for the "normal" Ag case, when $c_1 << c_2$ and $c_1 D_1 >> c_2 D_2$.

In the limiting case for a plane strip $\partial u_1/\partial x$ was constant, for a *cylinder* $r\,\partial u_1/\partial r$ must be constant. Thus $u_1 = c_1 \left(1 - \frac{\ln (R/r)}{\ln (R/\varrho)}\right)$ when R is the radius of the cylinder and ϱ the radial coordinate of the oxidation boundary.

$$D_1 \left(\frac{\partial u_1}{\partial r}\right)_{r=\varrho} = \frac{c_1 D_1}{\varrho \ln (R/\varrho)} = -c_2 \frac{\mathrm{d}\varrho}{\mathrm{d}t}; \quad \frac{c_1 D_1}{c_2} t = \tfrac{1}{4}\left(R^2 - \varrho^2 + 2\,\varrho^2 \ln \frac{\varrho}{R}\right).$$

Substituting α for $c_1 D_1/2c_2$ (see p. 87) and calling $(R-\varrho)/R = f$ we get

$$4\alpha t/R^2 = f - \tfrac{1}{2} f^2 + (1-f)^2 \ln (1-f) = f^2 - \tfrac{1}{3} f^3 - \tfrac{1}{12} f^4 \ldots$$

Similarly for a *sphere*, starting with $r^2\,\partial u_1/\partial r$ = constant, we get $4\alpha t/R^2 = f^2 - \frac{2}{3} f^3$.

For a plane strip we can write $4\alpha t/(\frac{1}{2}d)^2 = f^2$ if d is the thickness and $\frac{1}{2} f d = \xi$.

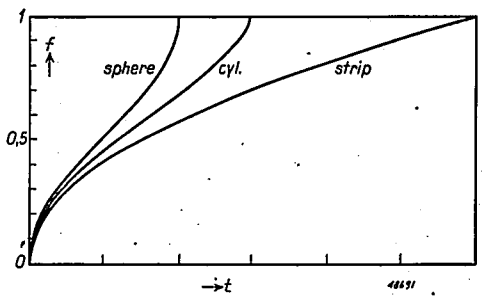


Fig. 7. Comparison of speed of internal oxidation of plane, cylindrical and spherical specimens.

The penetration of the oxidation boundary for the three cases is given in *fig. 7*. The times necessary for complete oxidation of a strip, a cylinder and a sphere of equal diameters stand in a proportion as 1: $\frac{1}{2}$: $\frac{1}{3}$. The same formulae are valid for the diffusion from a cylindrical or spherical *) *pore* into the surrounding alloy. Then r and ϱ are greater than R, the radius of the pore. If we now call $(\varrho - R)/R = g$ instead of $-f$ we get:

$$4\alpha t/R^2 = (1+g)^2 \ln (1+g) - g - \tfrac{1}{2} g^2 \qquad \text{(cyl. pore)}$$
$$4\alpha t/R^2 = g^2 + \tfrac{2}{3} g^3 \qquad \text{(spher. pore)}$$

*) but containing O_2 of constant pressure!

The smaller the radius of the pore, the slower (especially in the later stages) the penetration of the oxidation. For a spherical pore when $g >> 1$ we have $\varrho - R = \sqrt[3]{6\alpha R t}$. We also mention the case of a cylindrical *pipe* with external and internal radii R_1 and R_2 respectively. If we call $(R_1/R_2)^2 = n$ we find for the radial coordinate of the "middle-line" $\varrho_m = R_2 \sqrt{\frac{n-1}{\ln n}}$ and for the necessary oxidation time T:

$$\frac{8\alpha T}{R_2^2} = 1 - \frac{n-1}{\ln n}\left(1 - \ln \frac{n-1}{\ln n}\right).$$

For a *thin pipe* (or *bent strip*) with thickness d and radius R expansion for $(n-1) << 1$ gives the position of the "middle-line": $x_m = (\frac{1}{2} + \frac{1}{24} d/R)\, d$, reckoned from the exterior surface, and the oxidation time:

$$T = d^2/16\alpha \cdot (1 + \tfrac{1}{360}\, d^2/R^2).$$

We also calculated approximately c_{ox} as a function of ϱ for the cylinder and sphere, but as the quantitative result is still more uncertain than for the strip (section 4) we only give the qualitative result: when $c_2 D_2 << c_1 D_1$ in both cases c_{ox}, from its value $c_2\ (1 + c_2 D_2/c_1 D_1)$ at the surface, drops immediately but slowly with decreasing ϱ; no "middle-point" will be visible in practice. Indeed we were not able to observe it in a rod (Ag with 0·35 weight % Mg) of 4 mm diameter after complete oxidation.

Concave interfaces will cause an *in*crease of c_{ox} below the surface, owing to the extra slowing down of the penetration of oxygen.

6. *Coagulation of the oxide particles*

In a dilute alloy most of the oxide will be formed at first as single molecules (*e.g.* MgO). Hitherto we have assumed that the oxide formed is immobile. It is obvious that the molecules cannot move as a whole through the lattice. But when a molecule dissociates again after a while (and this will happen all the more according as the oxide is less stable) its constituent atoms will be able to move and form greater (and therefore more stable) complexes. This is borne out by experiment: in an oxidized silver alloy with 1 % Cd, which does not harden, at 1200 times magnification round oxide particles were visible, and the oxidation boundary (which at 15 × looked quite sharp) was vague; but in hardened alloys with about 1½ atomic % Mg or Be and 1 at. % Al (0·25 weight %) at the highest magnification no particles were visible and the oxidation boundary was as sharp as a crystal boundary.

In an oxidized copper alloy with 1 % Zn, oxide particles (short needles) were visible, in accordance with Fröhlich [7]) and Rhines [8]), but not in a hardened 2 at. % Be-Cu alloy. Rhines identified visible particles in practically all his copper alloys, including those with Be. His sentence,

however *): "A rather deep etch was usually required", seems to indicate that the etching was not stopped before spots were visible. Although these are caused by the internal oxidation (indicated by the different etching of the unoxidized interior), in our opinion they give only a relative picture of the size and number of the BeO particles. In our Cu alloys with ZnO and SiO_2 *e.g.* the particles were clearly visible in the *polished* state; by subsequent etching they became more or less enlarged.

As coarse particles will give practically no hardening, we understand better the second condition for hardening on page 81. We do not know whether the maximum hardness will be obtained when the oxide is in the form of single molecules, or whether there is an optimum molecular weight, which then, however, will be small. We know that the medium particle size in *e.g.* hardened Mg-Ag alloys cannot be great, because the lattice constant increases by the oxidation (section 8), so that we can say that the MgO is (at least partially) *in solution* in the Ag (this solution is of course strongly supersaturated).

Silver with more than about 0·4 weight % (2 at. %) Al or Mg after oxidation and polishing *did* show visible oxide, but in quite another form. When the concentration was not too high very few *streaks* were visible, often as much as $^1/_2$ millimeter long, slightly curved, mostly roughly parallel to the oxidation boundary. After etching they stood out in relief above the surface of the alloy. These "inner oxide films", as we will call them, were also found in copper with 1 weight % or more Al. In 0·5 % SiCu they were more irregularly curved. These films in oxidized SiCu alloys are already described by Smith [6]), Fröhlich [7]) and Rhines [8]); according to Fröhlich they are composed of a copper silicate. It is not impossible that the films in Al-copper consist of the spinel $CuAl_2O_4$, but in Al-silver and especially in Mg-silver such a double oxide looks very improbable.

In contrast with these silver alloys the oxidized Si-copper showed normal small (round) particles between the oxide films; this is in accordance with its very slight hardening.

A very curious phenomenon is exhibited by the 1·0 weight % AlCu alloy: in some parts of the oxidized zone a precipitate is visible, in other parts not (see *fig. 8*). The boundaries between these different areas are independent of the crystal boundaries. In 0·6% AlCu neither the precipitate nor the oxide films were found.

Fig. 9 shows that the oxide films form a barrier against the oxygen. When the Al (or Mg) content increases, their number grows and they begin closer under the surface, till at last the penetration of oxygen is so seriously impeded that the alloy is hardened only with difficulty. This is discussed for the silver alloys in section 10. An Al-copper alloy with many oxide

*) [8]) page 9.

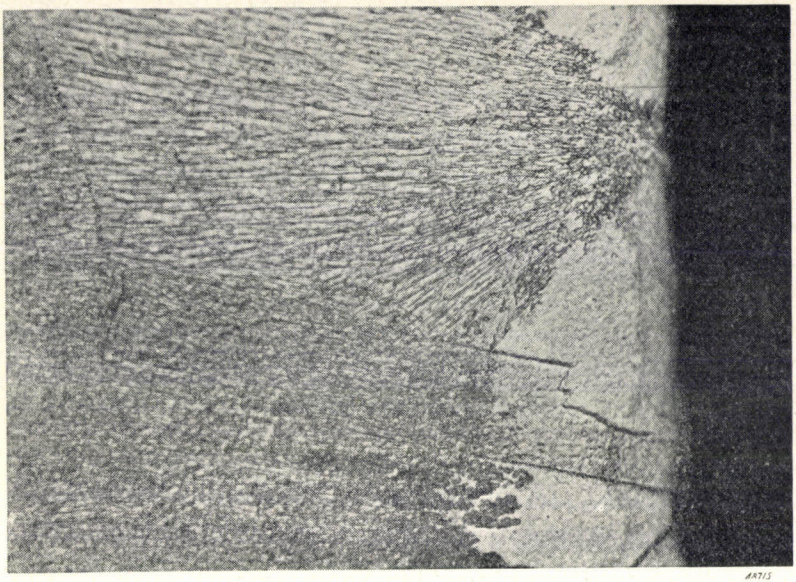
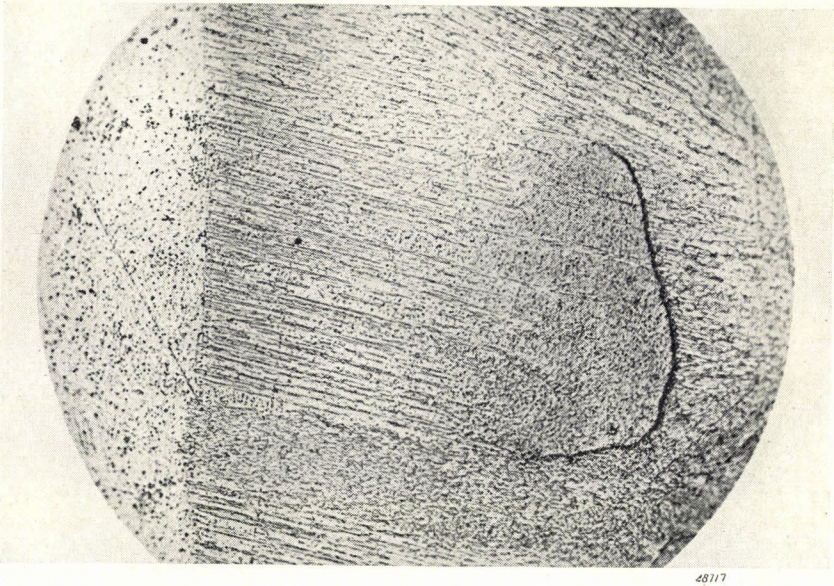


Fig. 8. Cu with 1·05 % Al, $2^1/_2$ h. 950° air; etched $NH_4OH + H_2O_2$, Magn. 500×. A precipitate is shown in some parts of the oxidized rim (preferentially near the oxidation boundary, see figs 9 and 10); their boundaries do not coincide with crystal boundaries. Most of the precipitate shows striations, with a diverging tendency in some rapidly broadening regions. Arrow indicates direction of penetration of O.

Fig. 9. Same specimen as of fig. 8. Magn. 1000×. Inner oxide film forming part of the oxidation boundary. The depth of oxidation close *beside* the oxide film appears to have profited from the oxygen barred by it. Arrow indicates direction of penetration of O.

films is shown by Rhines *); its concentration (0·72 weight % Al) is lower than that where they are only beginning to appear in our alloys (1%), but this is probably compensated by the higher oxidation temperature (1000° against 950° with us).

Nickel alloys with Al, Be, Ti, Mn, Si, Cr ($1\frac{1}{2}$ at. %) and Mg (1 at. %) were internally oxidized in air at 1200° (4 or 5 hours) but did not harden much. They all showed oxide inclusions in the polished state: round spots in



48117

Fig. 10. Same specimen as figs 8 and 9. Magn. 500×. Inner oxide film "behind the front". Arrow indicates direction of penetration of O.

the Be, Ti and Mg alloys and short needles (like those in oxidized 1 % ZnCu) in the Mn and Cr alloys. The oxidized Al-nickel showed an abnormal picture: long needles (or probably intersected plates) running preferentially in a direction perpendicular to the oxidation boundary. In our opinion they occupy a position between "normal" inclusions (which are short and have directions correlated with the orientation of the surrounding crystal) and "inner oxide films" (which are not so straight). It is clear that these "long needles" interfere only slightly with the penetration of oxygen, in contrast with the oxide films, which have just the most unfavourable position for it. Indeed the external oxidation of nickel is only moderately slowed down by 8 at. % Al, while that of copper is decreased several orders of magnitude by the same amount.

*) [8]) fig. 20.

The oxidized Si-nickel showed a few "long needles" among normal round spots, in analogy with oxidized Si-copper.

We also oxidized nickel with a much greater percentage of those elements that harden copper: 5 at. % Be respectively Ti, 4 and 8 at. % Al. (The solid solubility of Mg is very probably too small.) The first two alloys softened from ± 200 *) to 140 Vickers, but the Al alloys hardened: from 90 to 210 and from 100 to 250 Vickers respectively. Metallographic

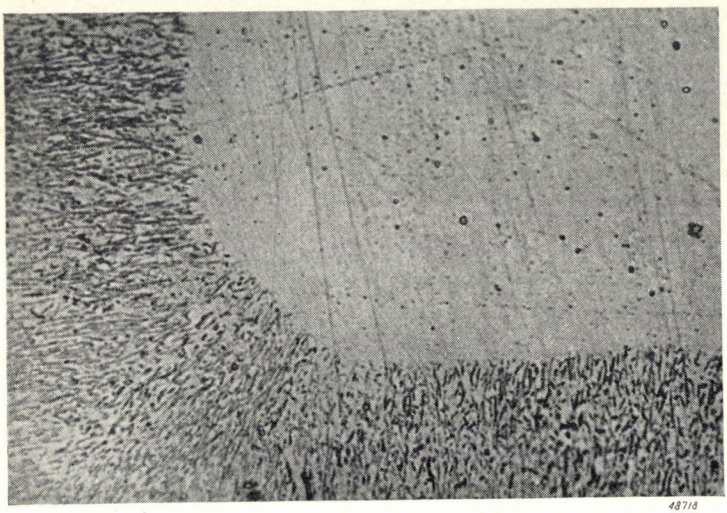


Fig. 11. Ni with 2 weight % Al, 5 h. 1200° air; not etched. Magn. 300×. Cross-section of strip near a corner. The oxide needles run substantially parallel to the direction of oxygen penetration. They differ from the striations in figs 8, 9 and 10 in the following respects:
1. more clear-cut, especially when compared in the polished state,
2. present *everywhere* in the rim,
3. no systematic deviations in directions, as e.g. in fig. 10.

examination (*fig. 11*) revealed the same "long needles" in greater number. The most concentrated alloy also showed a normal inner oxide film in a few places. Nickel with 20 at. % Al is fairly resistant to oxidation.

The hardening of the more concentrated Al-Ni alloys means apparently that a fair amount of the Al_2O_3 is in submicroscopical form. We do not think that the oxide plates can be responsible for the greater part of the hardening. With regard to this possibility we oxidized at 800° some more concentrated silver alloys, which could be expected to give coarse oxide. A 5 at. % Sn-Ag alloy was not penetrated by oxygen, 5% Cd-silver and 5% Cu-silver showed fairly coarse round inclusions and hardnesses of 60 and 35 Vickers respectively; the oxidized rim of 5 at. % Zn-silver had a serrated boundary, there were rather long and straight needles arranged fan-wise in little bunches, the hardness was 60 Vickers. Copper with 5% Mn and with 10% Zn yielded 70 and 60 Vickers respectively. All this, however, does not prove very much, in view of the great hardness of corundum.

*) The high hardnesses of these unoxidized alloys are apparently due to precipitation hardening.

We now will try to put the second condition on page 81 in a more quantitative form. The velocity of coagulation of *e.g.* BeO in copper depends for a good deal on the reaction $BeO \rightleftarrows Be + O$ (all dissolved in Cu). In an element of volume, lying well inside the oxidized zone, the chemical potential of the oxygen is about the same as (somewhat less than) that of O in the Cu_2O layer. Thus we can take also the reaction

$$2\ Cu(solid) + BeO(dissolved) \rightleftarrows Cu_2O(solid) + Be(dissolved).$$

For our rough calculation the entropy change in this reaction can be disregarded, and we find for the ratio of the concentrations of Be and BeO: $\frac{(Be)}{(BeO)} = e^{-\frac{\Delta H}{RT}}$, where ΔH is the reaction enthalpy, T the absolute temperature and R the gas constant. We can write ΔH as $(Q_{Cu_2O} - Q_{BeO}) - (\lambda_{Beo} - \lambda_{Be})$, where Q stands for the enthalpy ("heat") of formation of the solid oxide from the solid metal and oxygen gas, and λ for the enthalpy of solution in the solvent metal (*i.e.* Cu); the signs are in the usual thermodynamical sense, the Q's thus being negative. For trivalent and quadrivalent solutes obvious modifications are necessary: *e.g.* for Ti in Ni

$$\Delta H = (2\ Q_{NiO} - Q_{TiO_2}) - (\lambda_{TiO_2} - \lambda_{Ti}),$$

and for Al in Cu

$$\Delta H = (\tfrac{3}{2}\ Q_{Cu_2O} - \tfrac{1}{2}\ Q_{Al_2O_3}) - (\tfrac{1}{2}\lambda_{Al_2O_3} - \lambda_{Al}).$$

Ag as a solvent metal must be treated otherwise, because no Ag_2O is formed and the chemical potential of the dissolved O is governed by the oxygen pressure. It can be made to fit in the scheme, however, when we use (instead of —7) the values of Q tabulated for a few temperatures and oxygen pressures in the first column of table I. These stand for the heat of for-

TABLE I

—Q, negative enthalpy of formation, in kcal.

NiO 58	MgO 146	$\frac{1}{2}\ Al_2O_3$ 197	ZrO_2 258
Cu_2O 41	BeO 145	$\frac{1}{2}\ Cr_2O_3$ 140	TiO_2 225
(Ag_2O) 7	MnO 96	$\frac{1}{2}\ Mn_2O_3$ 116	SiO_2 204
Ag, 800 °C, 1 atm air 17	ZnO 83		SnO_2 138
Ag, 800 °C, 1 atm O_2 16	SnO 68		MnO_2 123
Ag, 650 °C, 1 atm air 15	CdO 62		
Ag, 650 °C, 1 atm O_2 14			

mation of a hypothetical Ag_2O which at the temperature in question would have a dissociation pressure equal to the oxygen pressure in question;

they are obtained by multiplying the true value of Q_{Ag_2O} by the ratio of the absolute temperature in question to the absolute temperature where Ag_2O has a dissociation pressure of 1 respectively 0·21 atm. Of Sn and Mn different oxides are possible; by determining the increase in weight of completely oxidized silver-alloy strips, as oxides formed were found SnO_2 and about $MnO_{1\cdot8}$ (thus between Mn_2O_3 and MnO_2). In bulk SnO_2 and Mn_2O_3 are the stable oxides at 800 °C and 0·21 or 1 atmosphere O_2. In oxidized Mg-silver and Al-silver the formulae of the oxides were found to agree within the limits of experimental accuracy (about 3%) with MgO and Al_2O_3. For the Cu and Ni alloys these experiments are hampered by the formation of an oxide layer.

TABLE II

ΔQ, obtained as follows: from the values in the 2nd, 3rd and 4th columns of table I, 1, $1^1/_2$ or 2 times the values of the 1st column are substracted respectively. Alloys corresponding to the values in brackets have not been examined *).

	Ag, 800° air **)	Cu_2O	NiO
MgO	129	105	88
BeO	128	104	87
MnO	79	55	38
ZnO	66	42	(25)
SnO	51	27	(10)
CdO	45	(21)	(4)
$^1/_2$ Al_2O_3	171·5	135·5	110
$^1/_2$ Cr_2O_3	(114·5)	(78·5)	53
$^1/_2$ Mn_2O_3	90·5	54·5	29
ZrO_2	(224)	176	(142)
TiO_2	191	143	109
SiO_2	(170)	122	88
SnO_2	104	56	(22)
MnO_2	89	41	7

Of the enthalpies of solution λ of the oxides we only know that they must be positive and fairly large, while those of the metallic solutes will generally be slightly negative (*e.g.* — 5 kcal), although for Al in Cu and Ni they are some — 15 and — 35 kcal per g atom respectively (for dilute alloys).

*) We think that the solid solubilities of Cr, Zr and Si in Ag and of Zr in Ni will be too small to give substantial effects.

**) For 650° or/and O_2 the values become some kcal greater.

Thus if we write $\Delta H = \Delta Q - \Delta\lambda$, the second term is not known, but comparison of ΔQ for solutes with the same valency (in all three solvent metals) with their ability to harden (compare section 1) is already instrutive. There is no sharp transition between hardening and non-hardening combinations; we will put the limit at 100 Vickers for Ag and Cu alloys and 150 Vickers for Ni alloys. So those combinations where these hardnesses are not reached by oxidation are reckoned as non-hardening in the rest of this section, provided the solid solubility exceeds $\frac{1}{2}$ at. % (thus ZrCu is not considered).

The bivalent solutes with $\Delta Q = 104$ (Be in Cu) or greater give hardening (Mn in Ag is not considered as bivalent solute now), while those with 88 (Mg in Ni) or less do not. For the trivalent solutes these borderline values are 90·5 (Mn in Ag) and 54·5 (Mn in Cu), while for the quadrivalent solutes they are 143 (Ti in Cu) and 122 (Si in Cu). The larger value of ΔQ required for hardening in the last case points to an extra large value of $\Delta\lambda$ for quadrivalent solutes. Now this is not surprising, because with the increasing distortion of the solvent lattice when the number of O atoms picked up by each dissolved metal atom increases from 1 over $1\frac{1}{2}$ to 2, an equivalent rise in λ for the oxides must be expected. It proved possible to arrive at a single criterion for the three valency groups together (taking also into account the different temperatures used for Ag, Cu and Ni alloys) by putting $\Delta\lambda = 30$ kcal for all bivalent solutes, 45 kcal for all trivalent and 75 kcal for all quadrivalent solutes. The result is that all alloys investigated split into two groups: those with $\frac{\Delta H}{T} > 42$ cal/degree do harden, those with $\frac{\Delta H}{T} < 40$ cal/degree do not *). This borderline value gives 10^{-9} for the factor $e^{-\frac{\Delta H}{RT}}$, which stands for the fraction of the time wherein a solute metal atom in the oxidized zone is free to move. As the diffusion coefficient D_2 (see section 2) of the metallic solutes at useful temperatures is of the order of 10^{-4} cm^2/day, one can roughly say that in one day a solute metal atom displaces itself $\sqrt{10^{-13}}$ cm, that is about 10 atom distances, sufficient to meet another oxide molecule (or complex) and combine with it, picking up a free O atom. When $\Delta H/T$ is only 5 cal/degree greater or smaller, the time becomes 10 days and 0·1 day respectively. Our rough treatment of the problem thus constitutes a good working hypothesis.

A lower temperature would favour hardening; indeed at 650 °C in air ZnO begins to harden silver (see section 10) and the corresponding $\Delta H/T = 41$, against 34 at 800 °C.

*) SnO, SnO_2 and MnO_2 would accordingly not harden, while Mn_2O_3 (and also MnO, if it were formed) would harden only silver.

Increasing concentration (c_2) and diffusion coefficient (D_2) of the metallic solute facilitate, of course, the conglomeration of the oxide.

When the oxidation boundary has only just passed the point in question, the chemical potential of the dissolved O is much less than at the surface of the strip. This means that at first the conglomeration of the oxide is much faster than later on. In some cases (see section 10) the hardening of silver alloys is much greater in oxygen than in air (both 1 atm); in our opinion this is due not only to the higher oxygen concentrations but also to the fact that the range of small initial concentrations is passed more rapidly at a certain point, the oxidation boundary moving faster in the first case.

Eindhoven, July 1946

REFERENCES

1) F. M. G. Johnson and P. Larose, J. Am. Chem. Soc. **46**, 1377, 1924; **49**, 312, 1927.
2) C. E. Ransley, J. Inst. Met. (London) **65**, 147, 1939.
3) C. J. Smithells and C. E. Ransley, Proc. Roy. Soc. A **155**, 195,1936.
4) F. N. Rhines and A. H. Grobe, Trans. A.I.M.M.E., T.P. 1439, 1942.
5) J. C. Chaston, J. Inst. Met. (London) **71**, 23, 1945.
6) C. S. Smith, Min. and Met. **11**, 213, 1930; **13**, 481, 1932.
7) K. W. Fröhlich, Z. Metallk. **28**, 368, 1936.
8) F. N. Rhines, Met. Technol. **7**, T.P. 1162, 1940. Of the two other papers of Rhines *et. al.* (Met. Technol. **7**, T.P. 1235 and **8**, T.P. 1368) we have so far only been able to read short abstracts in the Chem. Zentralblatt and Chem. Abstracts.
9) E. W. R. Steacie and F. M. G. Johnson, Proc. Roy. Soc. A **112**, 542, 1926.
10) W. Seith and E. Peretti, Z. Elektrochem. **42**, 570, 1936.

Philips Res. Rep. 2, 103-111, 1947

RADIATION AND HEAT CONDUCTION IN LIGHT-SCATTERING MATERIAL

by H. C. HAMAKER

536.24:536.33

II. GENERAL EQUATIONS INCLUDING HEAT CONDUCTION

Summary

The equations of Schuster describing the transmission of radiation through light-scattering material, and which have been recapitulated in a previous paper [1]), are now extended so as to include temperature radiation and heat conduction. After linearization we are led to a system of three simultaneous linear differential equations amongst which one of the second order. These are solved and certain general features of the solution are discussed. In particular an expression is deduced for the energy transport due to radiation.

7. *Introduction*

In the first paper of this series *) we have recapitulated a set of simultaneous differential equations describing the reflection and transmission of light in light-scattering material. Solutions applying to various conditions were discussed and the equations were subsequently extended so as to include the contribution of temperature radiation. In doing so the temperature was, however, supposed to be uniform throughout the light-scattering layer, a condition which is seldom satisfied in practical circumstances. Frequently light-scattering material is applied in the form of a coating on a metal core which is heated by an electric current or by a special heating element**). In that case a temperature gradient will prevail in the layer; heat will be transmitted both by conduction and by radiation, and the question arises which of these processes predominates.

Besides, radiation will be emitted not from a well-defined surface at a fixed temperature, but, owing to the slight transparency of the material, from a set of successive layers at slightly different temperatures; another problem which here arises is that of the accuracy with which the temperature of the outer surface of the layer can be determined by optical pyrometry.

In the present paper the theories given in (I) will be extended so as to

*) See reference [1]). This first paper will be quoted as (I); the numbering of sections and figures has been continued.

**) The most common example is no doubt the oxide-coated cathode or the tungsten filament coated with aluminium oxide used for heating such cathods. The theories given below were originally developed to discuss the temperature conditions in such layers.

include heat conduction and the influence of a temperature gradient. On the basis of the resulting equations the problems just mentioned will subsequently be discussed.

8. *Basic equations*

As before only one dimensional cases will be considered. We start from equations (32), *viz*:

$$\begin{aligned} \frac{dI}{dx} &= -(a+s)\,I + s\,J + a\,E(T)\,, \\ \frac{dJ}{dx} &= (a+s)\,J - s\,I - a\,E(T)\,, \end{aligned} \tag{36}$$

the meaning of which has been sufficiently explained in (I). In these expressions $E(T)$ represents the black-body radiation and $a\,E(T)$ is the amount of radiation contributed by the material at the temperature T according to Kirchoff's law. In solving equations (32) in (I), T and consequently E were assumed constant but this supposition must now be dropped. They will both be functions of x, and to solve our problem a third equation must be added which expresses that heat is nowhere accumulated. This equation reads

$$k\frac{d^2T}{dx^2} + a(I+J) = 2a\,E(T)\,, \tag{37}$$

k being the heat conductivity *). The first term on the left side represents the heat accumulated by conduction, the second term gives the heat absorbed, and the sum of these equals the heat loss by radiation.

In equations (36) and (37) the radiation is represented by simple symbols, no distinction being made between light of different wavelengths. This is not correct. In reality a, s, E, I and J will all vary with the wavelength and in that case the terms $a(I+J)$ and $2a\,E(T)$ in (37) must be replaced by corresponding integrals. This, however, introduces considerable complications and does not furnish a satisfactory basis for the discussion of elementary problems. We shall return to the complete equations in a later paper, restricting ourselves for the present to (36) and (37). In these equations a, s, E, I and J then refer to the total radiation. As a consequence the following discussion will only indicate the order of magnitude of various effects, but must not be interpreted to reproduce actual conditions in full detail.

*) In the Netherlands the normalized symbol for the heat conductivity is λ; as, however, in a later paper we shall develop equations containing both thermal conductivity and wavelength, λ has been reserved for the latter quantity and we have adopted k, the symbol recommended in Great Brittain, for the thermal conductivity.

$E(T)$, the total black-body radiation, is given by the Stefan-Boltzmann formula

$$E(T) = \sigma' \,.\, T^4\,; \quad \sigma' = 5{\cdot}7 \times 10^{-5} \text{ erg cm}^{-2} \text{ sec}^{-1} \text{ °K}^{-4}\,. \tag{38}$$

As a rule, however, the layers with which we have to deal in practice are very thin, and the temperature difference between both surfaces will be so small that we may put with sufficient accuracy

$$E = E_0 + b(T - T_0) = E_0 + 4\sigma' \,.\, T_0{}^3(T - T_0)\,, \tag{39}$$

where T_0 may be any temperature close to the temperatures actually occurring in the layer, E_0 being the corresponding total radiation.

When (39) holds the temperature may be fixed equally well by E as by T and, since this was found to simplify matters, E has been retained in our equations, *viz*:

$$\begin{aligned} \frac{\mathrm{d}I}{\mathrm{d}x} &= -(a+s)\,I + s\,J + a\,E\,, \\ \frac{\mathrm{d}J}{\mathrm{d}x} &= \quad (a+s)\,J - s\,I - a\,E\,, \\ -\frac{k}{b}\frac{\mathrm{d}^2E}{\mathrm{d}x^2} &+ a(I+J) = 2\,a\,E\,, \end{aligned} \tag{40}$$

which are solved by putting

$$I = A_1 \,.\, e^{\sigma x}, \quad J = A_2 \,.\, e^{\sigma x}, \quad E = A_3 \,.\, e^{\sigma x}, \tag{41}$$

leading to

$$\begin{vmatrix} \sigma + a + s & -s & -a \\ s & \sigma - a - s & a \\ a & a & \dfrac{k}{b}\sigma^2 - 2a \end{vmatrix} = 0 \tag{42}$$

for the determinantal equation from which σ is to be evaluated.
The solution is

$$\sigma^2 = 0\,, \tag{43}$$

$$\sigma^2 = 2a\frac{b}{k} + a(a+2s) = 2a\frac{b}{k} + \sigma_0{}^2\,, \tag{44}$$

where $\sigma_0 = \sqrt{a(a+2s)}$ is the value of the exponential coefficient which occurred in our earlier equations (I). By the theory of linear differential equations, the equation $\sigma^2 = 0$ leads to a solution of (40) of the form

$$I = C_1 x + D_1\,, \quad J = C_2 x + D_2\,, \quad E = C_3 x + D_3\,, \tag{45}$$

whereas (44) corresponds to a solution

$$I = A_1 e^{\sigma x} + B_1 e^{-\sigma x}, \quad J = A_2 e^{\sigma x} + B_2 e^{-\sigma x}, \quad E = A_3 e^{\sigma x} + B_3 e^{-\sigma x}. \tag{46}$$

The complete solution is obtained by adding (45) and (46).

Of the various constants A_k, B_k, C_k and D_k ($k = 1, 2, 3$) only four are arbitrary; the relations that must exist between them are obtained by inserting (45) and (46) in the original equations (40). It will not be necessary to go into details, but we may at once pass over to the complete solution which reads

$$\begin{aligned} I &= A(1-\beta)\, e^{\sigma x} + B(1+\beta)\, e^{-\sigma x} + C(\sigma x - \beta) + D, \\ J &= A(1+\beta)\, e^{\sigma x} + B(1-\beta)\, e^{-\sigma x} + C(\sigma x + \beta) + D, \\ E &= -A\,\alpha\, e^{\sigma x} - B\,\alpha\, e^{-x\sigma} + C\,\sigma x + D. \end{aligned} \tag{47}$$

where

$$\sigma = \text{positive root of (44)},$$

$$\beta = \frac{\sigma}{a + 2s}; \tag{48}$$

$$\alpha = \frac{2b}{k(a + 2s)} = \frac{2b\beta}{k\sigma}. \tag{49}$$

The constants A, B, C and D must be determined from the boundary conditions which are actually four in number; for at each boundary surface we have one condition to be satisfied by the temperature or the temperature gradient, and a second condition for the radiant energy. This will be amply examplified, when we are discussing special applications.

In equations (40) to (47) E_0 and T_0 from (39) do not occur. Nevertheless it will be found essential in any application of these equations to determine T_0 beforehand. For on T_0 depends the value of $b = 4\,\sigma'\,T_0^3$, and b influences the values of σ, β, and α.

It would of course be possible to write equations (47) in a different form, for instance by replacing the product $C\sigma$ by a single constant. The justification of the special form adopted lies, however, in the fact that in the equations that determine A, B, C, and D, only α and β occur, a mathematical advantage as will be fully demonstrated later on. Besides, it has been possible to relate the constants α and β to certain distinct physical features of the phenomena under consideration.

9. *General discussion*

Let us now consider equations (47) a little more in detail. The two exponential terms will rapidly diminish as we go in the direction of positive and negative x respectively. This suggest that these terms will be of impor-

tance only in the neighbourhood of the boundary planes where they provide certain corrections required to fulfil the boundary conditions. As a matter of fact, as we proceed towards the interior the influence of the specific boundary conditions rapidly decreases and throughout the greater part of a thick layer uniform conditions must be expected to prevail independently of the characteristic properties of the surfaces. These uniform conditions are represented by the linear terms in (47), $I = C(\sigma x - \beta) + D$, *etc.* The situation will be such as sketched in *fig. 5*, the deviations from linearity near the surfaces being due to the exponential terms in (47).

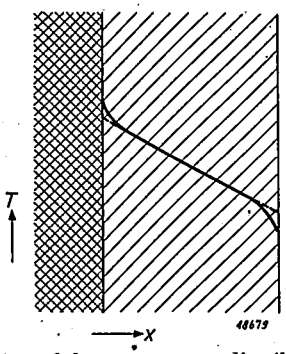


Fig. 5. Schematic representation of the temperature distribution as given by equation (47).

From this it may be inferred that the total transfer of heat, which is constant over the entire layer, will depend only on the temperature gradient in the interior, that is on the constant C. This can easily be verified; the total transfer of heat in the positive direction is evidently

$$W = -k\frac{dT}{dx} + (I - J) = -\frac{k}{b}\cdot\frac{dE}{dx} + (I - J) \tag{50}$$

which, after inserting (47), yields

$$W = -C\left(\frac{k\sigma}{b} + 2\beta\right) = -C\,.\,2\beta\frac{1+a}{a}, \tag{51}$$

the exponential terms cancelling out. In all practical applications, where the total transfer of heat in the direction of positive x is assumed to be known, C will be fixed by this fundamental equation.

In the region where I, J and E are linear functions of x we have from (47)

$$C = \frac{1}{\sigma}\cdot\frac{dE}{dx} = \frac{b}{\sigma}\cdot\frac{dT}{dx} \tag{52}$$

which introduced in (51) leads to

$$W = -\left(k + \frac{2b}{a + 2s}\right)\frac{dT}{dx} = -k\,(1 + \alpha)\,\frac{dT}{dx}. \tag{53}$$

Now

$$V = -k\,\frac{dT}{dx} \tag{54}$$

represents the amount of heat transmitted by pure conduction, and hence

$$U = -\frac{2b}{a + 2s}\cdot\frac{dT}{dx} = -\alpha k\,\frac{dT}{dx} \tag{55}$$

will measure the heat transfer due to radiation. As might have been expected, U is proportional to b, that is to the temperature gradient of total radiation, and U decreases when the scattering, measured by s, is increased. By Kirchoff's law an increase in the absorption a simultaneously entails an increase in radiation; the net result of these two counteracting influences appears to be a reduction in the radiative conduction of heat.

Dividing (55) by (54) we have

$$\alpha = \frac{U}{V} = \frac{\text{heat transfer by radiation}}{\text{heat transfer by pure conduction}}, \tag{56}$$

and this provides a simple physical interpretation of the constant α defined by (49).

10. *Numerical values of the various constants*

Before entering upon more detailed applications of equations (47) let us briefly discuss the order of magnitude of the various constants which have been introduced.

A. *a and s*

In two respects our theory is a crude one; firstly the radiation distribution has been replaced by two single fluxes I and J in one direction only; and secondly the spectral energy distribution has been completely neglected. Consequently we can expect to obtain only rough estimates of practical conditions and hence a precise evaluation of a and s will not be possible.

As a rough approximation it is even conceivable that equations (47) may be of use in cases where the scattering coefficient s is zero. Then a reduces to the extinction coefficient of transparent matter, defined by the formula

$$I = I_0\,e^{-ad},$$

and may assume values varying within extremely wide limits; for, substances such as quartz or rock salt show little absorbtion, even in the infra red, whereas other materials absorb radiation almost completely in layers only a few μ thick.

In the temperature range of practical interest, 1000-2000 °K, the radiative maximum lies at 1·5-3 μ, and to this region of the spectrum we should mainly direct our attention. Such values of a and s as quoted in (I) (tables I and II) relate to observations with visible light and are for that reason not representative for the purposes we now have in mind. The only information I have been able to detect is furnished by Coblentz's observations on the reflectivity of powdered oxides [2]), some of which have been reproduced in table III.

TABLE III

Reflectivities of powdered oxides observed by Coblentz:

λ	0·60 μ	0·95 μ	4·4 μ	8·8 μ
I. White oxides				
MgO	0·863	—	0·16	0·025
ZnO	0·822	0·864	0·085	0·026
SiO_2	—	0·810	0·415	0·11
ZrO_2	0·858	0·841	0·232	0·051
ThO_2	0·860	—	0·469	0·071
Al_2O_3	0·841	0·877	0·208	0·200
II. Coloured oxides				
PbO	0·518	—	0·506	0·256
Fe_2O_3	0·263	0·410	0·229	0·037
Cr_2O_3	0·270	0·446	0·329	0·05

Generally the reflectivity decreases with increasing wavelength and in the region of 1·5-3 μ values higher than 0·80 are hardly to be expected. From the reflectivity we may compute the ratio s/a by (17) and for $R_\infty \doteq 0{\cdot}80$ we find $s/a = 40$; in most cases this ratio will; however, be considerably less. Information concerning the absolute value of a and s is not available.

B. *The heat conductivity k*

From Landolt-Bornstein's numerous data on heat conductivity *) we learn that for such substances as glass, quartz, silica brick, zirconia brick

*) Especially Ergänzungsband II b page 1265 ff.

etc. k practically always lies between 1×10^{-3} and 5×10^{-3} cal cm^{-1} sec^{-1} $°C^{-1}$. In c.g.s. units we have

$$20 \times 10^4 > k > 4 \times 10^4 \text{ ergs } cm^{-1} \text{ sec}^{-1} °K^{-1},$$

and within these limits k will be assumed to vary.

C. *The constant b*

The constant b, the temperature gradient of total black-body radiation, is given by

$$b = 4\sigma' T_0^3 = 22{\cdot}8 \times 10^{-5} \times T_0^3 \text{ ergs } cm^{-2} \text{ sec}^{-1} °K^{-1}$$

which yields

T_0	800	1200	1600	2000 °K
b	11·7	39·4	93·5	182·4 × 10^4 ergs cm^{-2} sec^{-1} $°K^{-1}$

D. *The value of α*

By (40) we have

$$\alpha = \frac{2b}{k(a + 2s)} \text{ (dimensionless)}.$$

Taking for b the high value of 200×10^{-4} ergs cm^{-2} sec^{-1} $°K^{-1}$ and for k the low value of 4×10^4 ergs cm^{-1} sec^{-1} $°K^{-1}$ we obtain

$$\alpha = \frac{100}{(a + 2s)}. \tag{57}$$

With such values of a and s as recorded in tables I and II, α may reach values up to $\alpha = 0{\cdot}5$ which means that the heat transfer by radiation is one half of that by pure conduction, and one third of the total heat conduction.

For transparent matter ($s = 0$) α may become very much larger and it is quite conceivable that in some cases heat will be transmitted almost entirely by radiation. These are, however, cases in which the application of our equations is doubtful; in general computations from equations (47) $\alpha = 0{\cdot}25$ is the highest value we have used.

E. *The constant β*

The possible values of β might be fixed by the same methods but the following argument leads more directly to a conclusive result. By combining equations (4), (44) and (49) we have

$$\sigma^2 = \sigma_0^2 (1 + \alpha), \tag{58}$$

and consequently by (5), (16) and (48)

$$\beta = \beta_0 \cdot \sqrt{(1+\alpha)} = \frac{1-R_\infty}{1+R_\infty}\sqrt{(1+\alpha)}\,, \tag{59}$$

where R_∞ is the reflectivity (for total radiation) of an infinitely thick layer. In most cases it will be possible to formulate a reasonable assumption with respect to R_∞ and, if α has already been fixed, the value of β follows at once from (59).

F. *The value of* σ

After eliminating σ and $2b/k$ from (44), (48) and (49) it will be found that by α and β only the ratio s/a is fixed but that the absolute value of a and s, and hence of σ, can still be chosen at will. In (47), however, σ only occurs in the form of the product σx so that curves valid for different values of σ may be deduced from one another by plotting them on a different scale along the x-axis. In future discussions we will invariably plot our results against σx; it then only remains to discuss equations (47) for varying value of α and β which brings our problem within the scope of systematic treatment.

G. *Limiting cases*

Formulae deduced from (47) must reduce to the expressions given in (I) when the temperature is made uniform. This can be achieved by making $k \to \infty$ which gives

$$\alpha \to 0, \quad \beta \to \beta_0, \text{ and } \sigma \to \sigma_0\,.$$

Likewise it is evident that by keeping the ratio a/s constant but increasing a and s without limit we must obtain equations valid for opaque matter. As is easily verified, this gives

$$\alpha \to 0, \ \beta \to \beta_0, \ \sigma \to \infty, \text{ but } \alpha \,.\, \sigma \to 2\,\frac{b}{k}\,\beta_0\,.$$

These limiting cases will sometimes be of use in checking conclusions drawn from (47).

Eindhoven, July 1946

REFERENCES

[1]) H. C. Hamaker. Philips Res. Rep. 2, 55-68, 1947. This paper is quoted as (I). The enumeration of formulae and figures has been continued.

[2]) W. W. Coblentz, Bull. Bur. Standards 9, 283-325, 1913.

Philips Res. Rep. 2, 112-125, 1947

RADIATION AND HEAT CONDUCTION IN LIGHT-SCATTERING MATERIAL

by H. C. HAMAKER

536.24:536.33

III. APPLICATION OF THE THEORY

Summary

The set of equations developed in the foregoing paper and describing the combined transmission of energy by radiation and by thermal conduction in light-scattering material are applied to practical problems. In the first place the boundary conditions at the surface of a thick layer are investigated. Near the surface the temperature is no longer a linear function of the distance from the surface. Formulae for the deviation from linearity are deduced and the order of magnitude of these deviations is indicated. On the basis of these theories it is pointed out that observations on the thermal conductivity of oxide cathodes by various authors are unreliable since they are founded on unjustifiable assumptions.

Application of the present theory to layers of finite thicknes are only briefly discussed. On the whole the theory is a very crude one, so that it is not to be expected that detailed applications will be of great practical value.

11. *Layers of infinite thickness*

A. *A free surface*

Consider an infinite slab of material stretching from $x = 0$ to $x = -\infty$, the boundary at $x = 0$ being a free surface and heat being transferred from left to right (*fig. 6*). From the discussion of section 9 it is evident that in (47) we must have $B = 0$ so that we are left with the equations

$$I = A(1-\beta)\, e^{\sigma x} + C(\sigma x - \beta) + D\,, \tag{60a}$$

$$J = A(1+\beta)\, e^{\sigma x} + C(\sigma x + \beta) + D\,, \tag{60b}$$

$$E = -a\, A e^{\sigma x} + C\,\sigma x + D\,. \tag{60c}$$

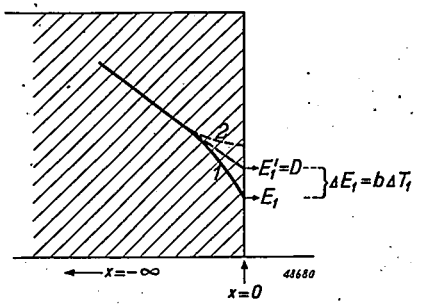


Fig. 6. Schematic temperature curves in the neighbourhood of a free surface.

Let us, moreover, suppose that from experiment we have determined the heat loss by radiation, U, and the heat loss by conduction, V, both at the surface $x = 0$, while there is no radiation incident from an external source. We then have the boundary conditions

$$I(0) = A(1-\beta) - C\beta + D = U\,, \tag{61}$$

$$J(0) = A(1+\beta) + C\beta + D = 0\,, \tag{62}$$

and

$$\left\{-\frac{k}{b}\frac{\mathrm{d}E}{\mathrm{d}x}\right\}_{x=0} = V.$$

This last equation can, however, according to equation (51) be replaced by

$$C = -\frac{\alpha}{2\beta\,(1+\alpha)}\{U+V\}. \tag{63}$$

From (61), (62) and (63) we obtain

$$A = \frac{1}{2\beta\,(1+\alpha)}\{\alpha V - U\}, \tag{64}$$

$$D = \tfrac{1}{2}U - A,$$

and consequently for the surface temperature E_1:

$$E_1 = E(0) = -\alpha A + D = \frac{1+\beta}{2\beta}U - \frac{\alpha}{2\beta}V, \tag{65}$$

an equation that can at once be applied to practical problems.

When the free surface is in contact with a vacuum, V, the heat loss by conduction must be zero, and consequently

$$U = \frac{2\beta}{1+\beta}E_1\,, \tag{66}$$

a relation between the radiation U and the surface temperature specified by E_1. From (I) equation (8) we know that, when the layer is of uniform temperature E_1, the outward radiation would be

$$U' = \frac{2\beta_0}{1+\beta_0}E_1\,. \tag{67}$$

The actual increase in temperature towards the interior brings in its tail an increased radiation and this effect is expressed by the difference between (66) and (67); since $\beta > \beta_0$ (see (59)) we will have $U > U'$.

To determine the surface temperature of an oxide coated cathode Clausing and Ludwig [1]) proceeded in the following way:

1) the cathode was heated to a uniform and known temperature, and the emissivity ε was computed from the observed radiation,
2) subsequently the cathode was heated in the normal way (a transport of energy taking place from the core to the outer surface), and from the radiation emitted under these circumstances, the surface temperature was computed by means of the emissivity as determined under 1).

To these pyrometer observations, made in a very limited region of the spectrum (red light), our theory does not, strictly, apply; their full discussion will require much more elaborate equations. But in principle an exactly similar experiment might be made with total radiation. In that case the emissivity computed from the first experiment would be

$$\varepsilon = \frac{1 + \beta_0}{2\beta_0}, \tag{68}$$

and the temperature deduced from the radiation U observed in the second experiment would be

$$E_1' = \frac{1 + \beta_0}{2\beta_0} U, \tag{69}$$

instead of E_1 as given by (66). The error made is

$$\Delta E_1 = b\Delta T = \left\{\frac{1 + \beta_0}{2\beta_0} - \frac{1+\beta}{2\beta}\right\} U. \tag{70}$$

For instance, the total energy radiated by an oxide-coated cathode at the normal temperature of 1100 °K is about $\frac{1}{3}$ that of a black body. Thus

$$U = \tfrac{1}{3}\sigma' (1100)^4 \text{ ergs cm}^{-2} \text{ sec}^{-1},$$

and since $b = 4\sigma' (1100)^3$ ergs cm^{-2} sec^{-1} °K^{-1} we have $\frac{U}{b} \approx 100$ °K and

$$\Delta T \approx 100 \left\{\frac{1 + \beta_0}{2\beta_0} - \frac{1+\beta}{2\beta}\right\} \text{°K}.$$

To a total emissivity of $\frac{1}{3}$ will correspond $R_\infty = \frac{2}{3}$ and, by (16), $\beta_0 = 0{\cdot}2$. Taking $\alpha = 0{\cdot}1$ which is not excessively large we have by (59) $\beta = 0{\cdot}21$ and

$$\Delta T \approx 12 \text{ °K}. \tag{71}$$

This indicates the order of magnitude of the errors that may occur.

By solving (65) for U we obtain

$$U = \frac{2\beta}{1+\beta} E_1 + \frac{\alpha}{1+\beta} V. \tag{72}$$

When V, the heat loss by conduction, differs from zero the temperature gradient in the layer will be steeper and, even if the surface temperature

remains the same, the radiation emitted will increase by the amount represented by the second term in (72).

Here again we come in conflict with experiments, this time those of Patai and Tomaschek [2]). To measure the heat conductivity they heated an oxide-coated cathode first in a vacuum and subsequently in an atmosphere of an inert gas, assuming that, when the radiation was kept constant, the temperature of the outer surface would be the same in both cases. By (66) and (72) we have

$$U = \frac{2\beta}{1+\beta} E_1 = \frac{\alpha}{1+\beta} V + \frac{2\beta}{1+\beta} E_1', \tag{73}$$

where E_1 is the surface temperature in vacuum, E_1' that in gas, and V the additional heat loss introduced by the gas atmosphere.

The change in surface temperature is consequently, using (39)

$$\Delta T = \frac{E_1 - E_1'}{b} = \frac{\alpha}{2\beta} \frac{V}{b}. \tag{74}$$

Assuming that V is about equal to the heat loss by radiation, and adopting the same data used in the numerical example dealt with above ($U = \frac{1}{3} E_1$, $U/b = 100$ °K) we find

$$\Delta T \approx 25 \text{ °K}. \tag{75}$$

By inserting the gas the surface temperature may undergo changes of this order of magnitude, even if the radiation is kept constant. Though Patai and Tomaschek only observed the radiation emitted at about 6500 Å (with a pyrometer), it is likely that the amount of the error will be correctly estimated by the above arguments *).

Both observations of Clausing and Ludwig, and of Patai and Tomaschek were made to measure the thermal conductivity of an oxide coating; it follows from our theory, however, that conductivity observations of this kind may be untrustworthy.

At some distance from the surface the exponential terms in (60) will be negligibly small and the temperature will be a linear function of x:

$$E = Cx + D. \tag{76}$$

Hence D fixes the surface temperature, if this straight line is extrapolated up to the surface (fig. 6) and the difference

$$E_1 - D = b\,\Delta T_1 = -\alpha A = \frac{\alpha}{2\beta\,(1+\alpha)} \{U - \alpha V\} \tag{77}$$

is a measure of the deviation from linearity.

*) Another error, which will not be discussed here, is introduced by the fact that the thermal conductivity of a porous oxide layer may possess different values in vacuum and in a gas atmosfere.

We see that ΔT_1 will be zero when $U/V = \alpha$, that is when the heat loss by radiation and by conduction at the surface are in the same ratio to each other as the heat transfer by radiation and by conduction in the bulk of the material (see (56)). When $U/V < \alpha$, $\Delta T_1 < 0$ and the temperature curve will have the shape of curve 1 in fig. 6 (p. 112); in the opposite case $U/V > \alpha$ we must have a curve as type 2. In vacuum ($V = 0$) the temperature difference will be of the same order as the estimate (75) (compare (77) and (74)).

B. *Contact with a metallic surface*

Let us next discuss the case that the light-scattering layer is, at $x = 0$, in contact with a metallic plate to which heat is transferred (*fig. 7a*). In this

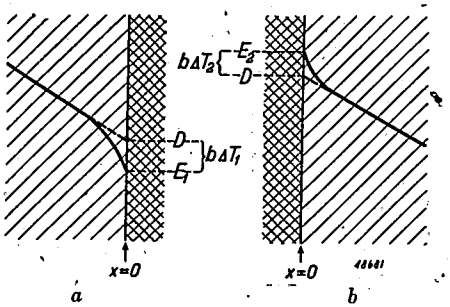


Fig. 7. Temperature curves in the neighbourhood of the contact with a metal underlayer.

case it is not of special interest to distinguish between the heat losses by radiation, U, and by conduction, V, and instead we shall assume that we have fixed by experiment both the surface temperature E_1 and the total heat transport W. This gives (see (51))

$$C = \frac{-\alpha}{2\beta\,(1+\alpha)}\,W \tag{78}$$

as before and from (60c)

$$-\alpha A + D_1 = E_1\,. \tag{79}$$

In addition we must have (as in (I) eq. (34))

$$J = r_1 I + (1 - r_1)\,E_1 \qquad \text{for } x = 0 \tag{80}$$

which, with

$$r_1 = \frac{1-\varrho_1}{1+\varrho_1}, \tag{81}$$

transforms into

$$(1+\varrho_1)\,J - (1-\varrho_1)\,I = 2\varrho_1 E \qquad \text{for } x = 0. \tag{82}$$

(Compare (11) and (34)).

Inserting (60) in (82) we obtain

$$A = \frac{\alpha W}{2\,(1+\alpha)\,\{\beta + (1+\alpha)\,\varrho_1\}}, \tag{83}$$

and for the deviation from linearity

$$\Delta E_1 = b\Delta T_1 = -\alpha A = \frac{-\alpha^2 W}{2\,(1+\alpha)\,\{\beta + (1+\alpha)\,\varrho_1\}}. \tag{84}$$

This is essentially negative, so that the temperature curve will always be of the type drawn in fig. 7*a*. Conditions in which the temperature curve is a straight line up to the surface do not exist. On the other hand, for values of α of the order 0·1 or less the deviation from linearity is considerably less than in the case of a free surface, ΔT as given by (84) being roughly proportional to α^2, whereas ΔT in equation (77) varies with α. It should also be noted that ΔT_1 in the present case only depends on W and not on E_1.

If the layer is in touch with a metal plate from which heat is transported away (as in *fig.* 7*b*) we will have exactly the same equation

$$\Delta E_2 = b\,\Delta T_2 = \frac{\alpha^2 W}{2\,(1+\alpha)\,\{\beta + (1+\alpha)\,\varrho_2\}} \tag{85}$$

but with the sign reversed; ΔT_2 will be essentially positive.

C. *Other cases*

Alternative cases of interest are those in which there is a certain amount of radiation incident from outside, or in which a reflecting surface at room temperature is mounted at some distance from a free surface, the corresponding boundary conditions being $J = J_1$ for $x = 0$, or $J = r_1 I$ for $x = 0$ respectively. Cases like these may be dealt with in a similar manner, but we will not consider them in detail.

12. *Finite layers. The complete formulae*

When dealing with layers of finite dimensions we have to apply the complete formulae (47) and four conditions are required to determine the four constants A, B, C and D.

A. *One free surface*

Let us first consider a layer with a free surface at $x = 0$, and in contact with a heated metallic underlayer at $x = -X$ (*fig.* 5). Assuming no radiation incident from outside we must have $J = 0$ when $x = 0$, which gives

$$A(1+\beta) + B(1-\beta) + C\beta + D = 0. \tag{86}$$

Likewise we have $I = rJ + (1-r)\,E$ for $x = -X$, which yields

$$A\{\beta - (1+\alpha)\,\varrho\{e^{-\xi} - B\{\beta + (1+\alpha)\,\varrho\{e^{\xi} + \beta C = 0\,, \qquad (87)$$

ϱ having the same meaning as before (see (81)) and the product σX being replaced by the single symbol ξ

$$\sigma X = \xi\,. \qquad (88)$$

Assuming moreover that from experiment we know the heat losses by radiation and by conduction, U and V, at the free surface, we obtain the additional conditions

$$I(0) = A(1-\beta) + B(1+\beta) - \beta C + D = U\,, \qquad (89)$$

and in the same way as in (63)

$$C = \frac{\alpha}{2\beta\,(1+\alpha)}\{U + V\{\,. \qquad (63)$$

The four conditions (86), (87), (89) and (63) suffice to determine A, B, C and D in terms of U and V and from them to compute the temperature E_1 at $x = 0$ and E_2 at $x = -X$. Since all our equations are linear and homogeneous in A, B, C, D, E_1, E_2, U and V, it is at once evident that the final result will be of the form

$$E_1 = E_{1U}\,.\,U + E_{1V}\,.\,V\,, \qquad (90)$$
$$E_2 = E_{2U}\,.\,U + E_{2V}\,.\,V\,, \qquad (91)$$

in which E_{1U}, E_{1V}, E_{2U}, E_{2U} are functions of the properties of the layer (α and β), of the underlayer (ϱ) and of the thickness ξ which are given by the following expressions:

$$E_{1U} = \frac{1}{\Delta} \times \frac{1}{2\beta}\Big[(1+\beta)\{\beta + (1+\alpha)\,\varrho\{e^{\xi} + (1-\beta)\{\beta - (1+\alpha)\,\varrho\{e^{-\xi} + 2\alpha\beta\Big], \qquad (92)$$

$$E_{1V} = \frac{1}{\Delta} \times -\frac{\alpha}{2\beta}\Big[\{\beta + (1+\alpha)\,\varrho\{e^{\xi} + \{\beta - (1+\alpha)\,\varrho\{e^{-\xi} - 2\beta\Big], \qquad (93)$$

$$E_{2U} = \frac{1}{\Delta} \times \frac{\alpha}{2\beta(1+\alpha)}\Big[\{[1+\beta(1+\alpha)]\,[\beta + (1+\alpha)\,\varrho] + \beta\alpha^2\{e^{\xi} + \{[1-\beta(1+\alpha)]\,[\beta - (1+\alpha)\,\varrho] + \beta\alpha^2\{e^{-\xi} + 4\alpha\beta\Big] + \frac{\alpha}{2\beta(1+\alpha)}\,\xi, \qquad (94)$$

$$E_{2V} = \frac{1}{\Delta} \times \frac{\alpha}{2\beta(1+\alpha)}\Big[\{\beta(1-\alpha) + \varrho(1+\alpha)\{e^{\xi} + \{\beta(1-\alpha) - \varrho(1+\alpha)\{e^{-\xi} - 2\beta(1-\alpha)\Big] + \frac{\alpha}{2\beta(1+\alpha)}\,\xi, \qquad (95)$$

where $$\Delta = \{\beta + (1+\alpha)\,\varrho\{e^{\xi} - \{\beta - (1+\alpha)\,\varrho\{e^{-\xi}\,. \qquad (96)$$

We have seen earlier that in the interior of a thick layer the change of temperature with the distance is given by the linear term

$$C\xi = \frac{-\alpha}{2\beta(1+\alpha)}\xi(U+V)$$

and, as was to be expected, E_{2U} and E_{2V} both contain the corresponding term $\frac{\alpha}{2\beta(1+\alpha)}\xi$. It will also easily be verified that for $\xi \to \infty$ equation (90) reduces to (65) and, neglecting the linear part, the difference of (91) and (90) reduces to the differences of (85) and (84).

It is also of practical importance to have expressions for E_1 and E_2 when ξ is very small. When $\xi \ll 1$

$$E_{1U} = \frac{1+\varrho}{2\varrho} + \frac{(1+\alpha)\varrho^2 - \beta^2}{2\beta\varrho^2(1+\alpha)}\xi + \frac{2\beta^2 - \varrho^2(1+\alpha)(2+\alpha)}{4\varrho^3(1+\alpha)^2}\xi^2, \tag{97}$$

$$E_{1V} = -\frac{\alpha}{2\beta}\xi + \frac{\alpha}{4\varrho(1+\alpha)}\xi^2, \tag{98}$$

$$E_{2U} = \frac{1+\varrho}{2\varrho} + \frac{(1+\alpha)\varrho^2 - \beta^2}{2\beta\varrho^2(1+\alpha)}\xi + \frac{\beta^2 - \varrho^2(1+\alpha)}{2\varrho^3(1+\alpha)^2}\xi^2, \tag{99}$$

$$E_{2V} = \frac{\alpha}{4\varrho(1+\alpha)}\xi^2. \tag{100}$$

When $V = 0$ equations (90) and (91) reduce to the single relation

$$E_1 = E_2 = \frac{1+\varrho}{2\varrho}U,$$

which expresses that the radiation is that emitted by the underlayer, and that the intensity is no longer affected by the amount of heat lost by conduction.

Further applications of these expressions will be made in the next section.

13. *Approximations for thin layers*

Using the approximations (93) to (96) we get for very thin layers ($\xi \ll 1$)

$$E_2 - E_1 = b(T_2 - T_1) = \frac{\alpha}{2\beta}\xi V + \frac{\alpha}{4\varrho(1+\alpha)}\xi^2 U \tag{101}$$

which by (48), (49), (58) and (88) can be transformed into

$$T_2 - T_1 = \frac{X}{k}V + \frac{a}{2k\varrho}X^2 U, \tag{102}$$

X being the thickness of the layer.

This is a result allowing a simple explanation. For k is the thermal conductivity, radiation excluded, and XV/k is the temperature difference that will occur if the heat V alone is transported through the layer by conduction. The additional increase produced by the loss of heat by radiation is of the second order; it is indeed inherent to our basic assumptions that radiation will play a part in heat conduction, only over finite distances; and if the layer is thin enough radiation will not interfere.

If the layer is heated in vacuum, V will be zero and the temperature difference will, to a first approximation, be negligible. This is contrary to the interpretation of various experiments. For in the papers of Clausing and Ludwig [1]) and of Patai and Tomaschek [2]) it has been assumed that the temperature difference across the layer will be XU/k, even if the energy loss is of a radiative nature only.

14. *The complete equations. General applications*

The approximations for thick and for thin layers dealt with in the foregoing sections will suffice in most practical cases; on the whole, the theory developed in these papers is such a crude one that more detailed applications will hardly ever be justified. We shall therefore not discuss the com-

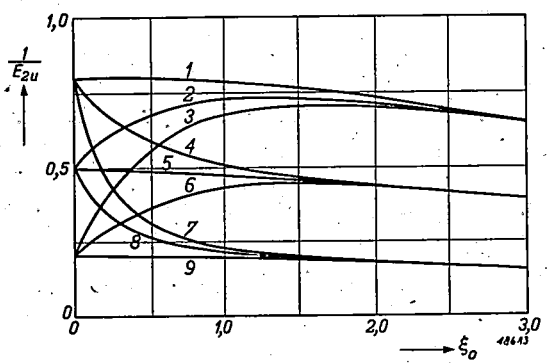


Fig. 8. The function $1/E_{2U}$. $\alpha = 0{\cdot}25$

Curve	R_∞	β	r	ϱ
1	0·20	0·7453	0·20	0·6667
2	0·20	0·6453	0·50	0·3333
3	0·20	0·6453	0·80	0·1111
4	0·50	0·3727	0·20	0·6667
5	0·50	0·3727	0·50	0·3333
6	0·50	0·3727	0·80	0·1111
7	0·80	0·1214	0·20	0·6667
8	0·80	0·1214	0·50	0·3333
9	0·80	0·1214	0·80	0·1111

plete equations, (92) to (96), *in extenso* but restrict ourselves to indicating a few of the problems that may occur in practice; as we shall show, the four functions (92) to (95) will contain all the information that may be required.

1) In a vacuum the heat loss by conduction at the surface is zero, $V = 0$, and if under these conditions the temperature of the underlayer E_2 is kept constant, the heat loss by radiation at the surface, U, will evidently vary as (see (91) and (94))

$$U = \frac{E_2}{E_{2U}} \tag{103}$$

and we get a set of curves as shown in *fig. 8*.

2) Alternatively it may experimentally be possible to keep the outward radiation U constant, in which case E_2 will be a function of layer thickness represented by

$$E_2 = E_{2U} \,.\, U \,. \tag{104}$$

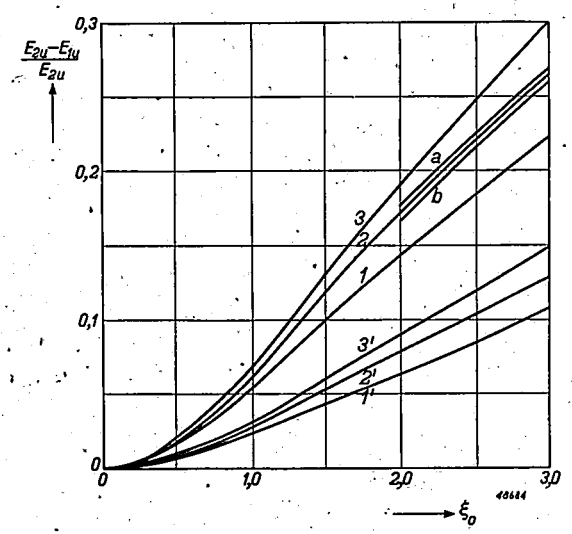


Fig. 9

The function $\frac{E_{2U}-E_{1U}}{E_{2U}}$.

Curve	a	R_∞	β	r
1	0·25	0·20	0·7453	0·50
2a	0·25	0·50	0·3727	0·20
2	0·25	0·50	0·3727	0·50
2b	0·25	0·50	0·3727	0·80
3	0·25	0·80	0·1242	0·50
1′	0·10	0·20	0·6992	0·50
2′	0·10	0·50	0·3496	0·50
3′	0·10	0·80	0·1165	0·50

It should be noted, however, that under these circumstances the average temperature of the layer may undergo very marked changes so that the constant b (eq. (39)), and consequently α, β and ϱ (eqs (49), (48) and (44)), must be altered with ξ, which renders the problem a more complicated one.

3) If, again in a vacuum, we should like to know the temperature difference across the layer as a function of the thickness we have from equations (90) and (91)

$$E_2 - E_1 = (E_{2U} - E_{1U}) \cdot U = E_2 \frac{E_{2U} - E_{1U}}{E_{2U}}, \qquad (105)$$

a functional relation represented in *fig. 9*. As indicated by the curves 2, 2a and 2b the difference $E_2 - E_1$ varies only in a very slight degree with the reflectivity r of the underlayer; these variations are so small in fact that in view of the approximate character of the present theory, they may safely be neglected for all practical purposes. The same holds for other types of curves which will presently come up for discussion. The reflectivity of the coating material R_∞ plays a more important part; besides, the function $(E_{2U} - E_{1U})/E_{2U}$ is approximately proportional to α which is again a feature exhibited by other curves as well (see below).

4) The treatment of the total radiation analogue of Clausing and Ludwig's experiment (see page 114) can now also easily be extended to layers of finite thickness. First we investigate the radiation U as a function of the thickness of the coating for $V = 0$ and for a the temperature that is uniform throughout the layer. We obtain

$$U^0 = \frac{E_1^0}{E_{1U}^0}, \qquad (106)$$

where E_1^0 represents the temperature and E_{1U}^0 is the function E_{1U} for a uniform temperature, that is for $\alpha = 0$, $\beta = \beta_0$, and $\sigma = \sigma_0$.

In a subsequent series of experiments the underlayer is heated to such a temperature that $U = U_0$; what will be the temperature of the outer surface? We have

$$E_1 = E_{1U} \cdot U = E_1^0 \frac{E_{1U}}{E_{1U}^0},$$

so that the relative changes in surface temperature are given by

$$\frac{\Delta T}{T} = \frac{E_1 - E_1^0}{E_1^0} = \frac{E_{1U} - E_{1U}^0}{E_{1U}^0}, \qquad (107)$$

a function represented in *fig. 10*. In connection with this expression it should be noted that the variable ξ has a different value for different values of α.

By (88) and (58) we have in fact

$$\xi = \sqrt{(1+\alpha)}\,\xi_0 \tag{108}$$

where ξ_0 is ξ for $\alpha = 0$. In (107) E_{1U} and E^0_{1U} must be evaluated for the same values of ξ_0 and not for the same values of ξ.

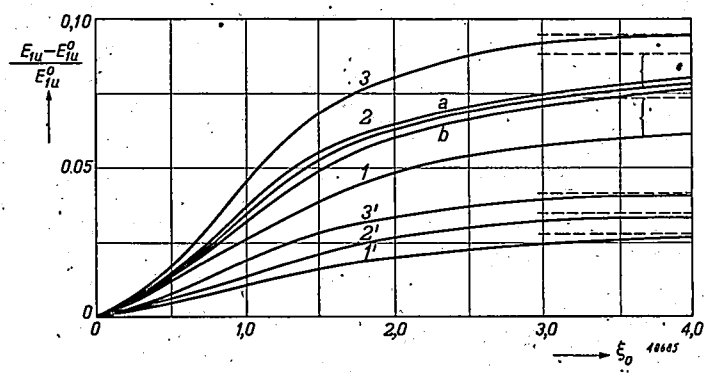


Fig. 10. The function $\frac{E_{1U}-E^0_{1U}}{E^0_{1U}}$, notation of curves as in fig. 9.

5) Finally in such experiments as those of Patai and Tomaschek discussed on page 115 the problem is, to what extent the surface temperature changes if U is kept constant and V is made to vary. Equation (90) yields in this case

$$E_1' - E_1 = E_{1V} \cdot V \tag{109}$$

and similarly for the temperature of the underlayer

$$E_2' - E_2 = E_{2V} \cdot V, \tag{110}$$

where E_1' and E_2' are the temperatures in a gas atmosphere, and E_1 and E_2 are the temperatures before introducing gas into the system. The functions E_{1V} and E_{2V} have been plotted in *figs 11* and *12*.

With respect to the dependence on r, R_∞, and α the remarks made on page 122 with reference to fig. 10 also apply to figs 10, 11 and 12.

Problems differing from those considered above may be dealt with along similar lines.

15. *Thermal conductivity*

Observations on thermal conductivity may be treated in a similar manner. Let us suppose the layer to stretch from $x = -X$ to $x = 0$ being

at both ends in contact with a metal surface at temperatures E_2 and E_1 respectively. In analogy to equation (82), the boundary conditions will now be

$$\left.\begin{aligned}(1+\varrho_2)\,I-(1-\varrho_2)\,J &= 2\varrho_2 E \text{ and } E=E_2 \text{ for } x=-X,\\ (1+\varrho_1)\,J-(1-\varrho_1)\,I &= 2\varrho_1 E \text{ and } E=E_1 \text{ for } x=0.\end{aligned}\right\} \quad (111)$$

Inserting the full expressions (47) and making use of (78) we obtain after elimination of the four constants A, B, C and D

$$E_2-E_1 = W\cdot\frac{a^2}{2(1+a)}\,\frac{[2\beta+(1+a)(\varrho_1+\varrho_2)]\,e^{\xi}+[2\beta-(1+a)(\varrho_1+\varrho_2)]\,e^{-\xi}-4\beta}{[\beta+(1+a)\varrho_1][\beta+(1+a)\varrho_2]\,e^{\xi}-[\beta-(1+a)\varrho_1][\beta-(1+a)\varrho_2]\,e^{-\xi}}+$$
$$+\,W\frac{a}{2\beta(1+a)}\,\xi. \quad (112)$$

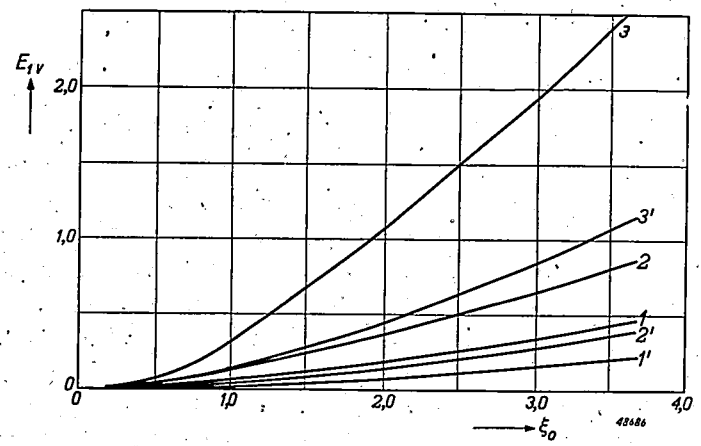


Fig. 11. The function E_{1V}; notation of curves as in fig. 9.

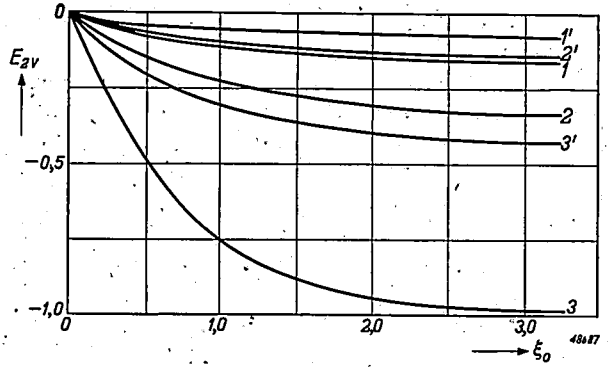


Fig. 12. The function E_{2V}; notation of curves as in fig. 9.

The two boundary conditions (111) are symmetric in E_1 and E_2 and we see that, as a consequence, the difference $E_2 - E_1$ is now proportional to the total heat transfer W. It is no longer necessary to distinguish between heat transported by conduction and by radiation, as it was in the case previously dicussed.

The approximation applicable to very thin layers may be written in the form

$$W = (E_2 - E_1)\left\{\frac{2\beta}{a\xi} + \frac{2\varrho_1\varrho_2}{\varrho_1 + \varrho_2}\right\}, \tag{113}$$

which by (39), (48), (49), (88) and (81) can be transformed into

$$W = (T_2 - T_1)\frac{k}{X} + (E_2 - E_1)\frac{(1-r_1)(1-r_2)}{r_1 r_2}. \tag{114}$$

This is again a comprehensible result; for the first term on the right-hand side represents the energy transport by pure conduction and, as is easily verified, the second term equals the heat transported by radiation through a vacuum between two surfaces opposite one another with total radiations E_2, E_1 and total reflectivities r_2 and r_1 respectively; indeed, if the layer is thin enough, it will not interfere with the radiation. For layers that cannot be considered as very thin, the relation between the energy transport and the temperature difference may be computed from equation (112). We will not work out this equation in further detail in the present paper.

It should be noted, however, that observations on the thermal conductivity of quartz, glass or similar substances are generally restricted to comparatively low temperatures for fear that radiation will produce systematic errors. Even though they furnish only very rough approximations and do not directly bear on such transparent substances, our present equations may help in forming a picture of the interplay of conduction and radiation, and in estimating the errors that may be introduced by radiation if this is no longer negligible.

Eindhoven, July 1946

REFERENCES

1) P. Clausing und J. B. Ludwig. Physica **13**, 193-205, 1933.
2) E. Patai und Z. Tomaschek. Kolloid. Z. **74**, 253-265, 1936.

Philips Res. Rep. **2**, 126-135, 1947

AN IMPROVED METHOD FOR COUPLING VALVES AT ULTRA-SHORT WAVES

by A. VAN WEEL

621.396.621.53.029.6

Summary

A method for coupling two electron valves, or one valve with an antenna is described, by which method the difficulties due to the finite inductance of the internal electrode leads of a valve can be eliminated up to very high frequencies. In addition to this the new system provides a very simple way to realize matching of the valve impedances.

One of the main difficulties encountered in designing receivers or transmitters for wavelengths in the metre or decimetre range, is the coupling of valves (or of one valve with an antenna) in such a way that maximum signal power is transferred. At these extremely high frequencies, the numerical value of the product $LC = 1/\omega^2$ for tuned circuits becomes very small. As the capacity cannot be reduced below the value of the internal valve capacity, the only way of further reducing the product LC is by cutting down the inductance (it does not matter whether this is a lumped inductance or a parallel-wire system).

However, the inductance cannot be reduced below a practical limit, which is reached when the total circuit inductance is of the same magnitude as the inductance of the internal electrode leads of the valve. To illustrate this, we first consider *fig. 1*, in which one stage of a conventional H.F. amplifier is drawn (only the A.C. elements, like in most of the following figures).

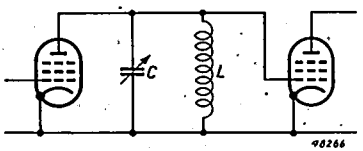


Fig. 1. A conventional H.F. amplifier stage.

The circuit consists of a tuned LC-circuit between the earth and the output and input electrodes of the two valves which are connected in parallel. In *fig. 2* the same circuit is shown, but now the internal capacities C_a and C_g and the internal inductances L_a and L_g of the valves have been taken into account. (We have omitted the inductances of the cathode leads;

this is allowable in push-pull circuits using double valves in one envelope, when the inductance of the intercathode lead can be neglected. We may therefore consider fig. 2 as one half of a push-pull circuit.)

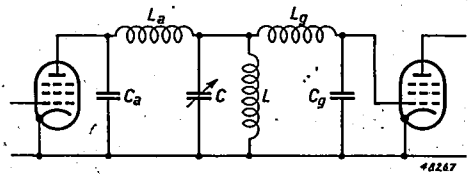


Fig. 2. A conventional H.F. amplifier stage with internal valve inductances and capacities taken into account.

It is clear from fig. 2 that at those frequencies where the magnitude of L_a, L and L_g are comparable, this circuit becomes complicated, and practice has shown that its efficiency diminishes rapidly. At very high frequencies C_a and L_a or C_g and L_g may even be tuned without any external inductance at all.

One way to overcome this difficulty is to insert capacities in series with L_a and L_g and by chosing the magnitude of these capacities so as to tune the inductances L_a and L_g. Together with the actual (parallel) tuning condenser, three condensers in each stage have then to be tuned (in the complete push-pull circuit even five variable condensers per stage!) and practice has again shown that it is very difficult to reach satisfactory results in this way.

Some time ago we published a method to overcome these difficulties [1]) and we will now go into more details about this improved circuit, which is shown in *fig. 3*!

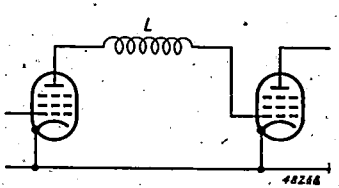


Fig. 3. The new H.F. amplifier stage.

As this circuit looks rather unconventional, we have drawn a more complete circuit in *fig. 4*, taking into account the internal capacities and inductances of the valves. It can be seen from this figure that the tuned circuit consists of $(L_a + L + L_g)$ in series with C_g and with C_a. The input

[1]) A. van Weel, Thesis, Delft 1943, page 85.

current is supplied to the "terminals" of C_a while the output voltage is taken from the "terminals" of C_g. *Fig. 5* gives the same circuit drawn somewhat differently to show the principle more clearly.

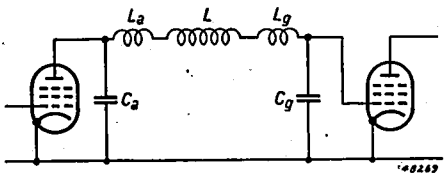


Fig. 4. The new H.F. amplifier stage with the internal valve inductances and capacities taken into account.

One advantage of this circuit compared with the conventional circuits is the fact that the internal inductances now are connected in series with, and therefore form part of, the total circuit inductance. As a consequence their presence does not complicate the circuit. Moreover, the total circuit inductance is larger than in conventional circuits because the circuit capacity is smaller, the series connection of C_a and C_g forming the latter being smaller than the parallel connection of these capacities in the circuit drawn in fig. 1.

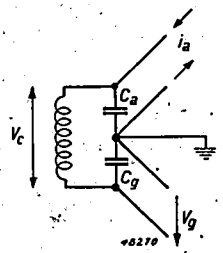


Fig. 5. The circuit of fig. 4 drawn in a somewhat different manner.

The circuit can be tuned either by varying the magnitude of the inductance, but this is often difficult to realize, or by the adjustment of a variable trimmer condenser. This trimmer may be connected in series with L; the total capacity of the circuit then consists of the series connection of C_a, C_g and the trimmer. Another possibility is to connect the trimmer "in parallel". To make it quite clear to what part of the circuit the trimmer is connected in parallel and what voltage may be expected across the trimmer, it is necessary to describe a special property of this circuit.

From fig. 5 it can be seen that the connection of C_a and C_g which is supposed to be connected to "earth", is tapped on the actual circuit voltage V_c. On the inductance a point can be found that is in the same way tapped on the circuit voltage. This point divides the inductance in inverse proportion to C_a/C_g.

The main fact about this point is that it carries practically no (high-frequency) voltage against earth. Short-circuiting this point to earth should not make any difference in the performance of the circuit. However, this is only true if all the windings of the inductance are fully coupled. In practice there is always some stray inductance and this applies especially to inductances of the magnitude of the internal inductances of the valves. What happens in this case if the two points mentioned are short-circuited, can be seen from the circuit of *fig. 6*, in which the two parts of the inductances are supposed to show no mutual coupling at all.

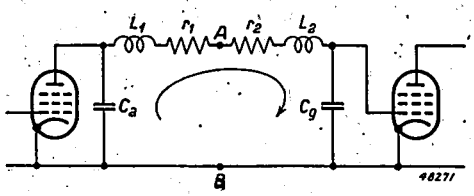


Fig. 6. The new circuit with the inductance split into two separate parts on the place of the voltage node. Short-circuiting of A and B makes the right half of the circuit currentless.

In the figure the losses of the inductances are taken into account by inserting the series resistances r_1 and r_2. The inductances fulfil the equation:

$$\frac{L_1}{L_2} = \frac{C_g}{C_a}. \qquad (1)$$

From fig. 6 it is clear that when the points A and B are short-circuited, no current whatever will flow in the right half of the circuit. Short-circuiting is therefore not allowed in this case.

If there is a certain amount of coupling between both inductances, short-circuiting of the points A and B will give the circuit the properties of two coupled tuned circuits. We mentioned above that at very high frequencies the mutual coupling of the two parts of the total inductance will be very small as a consequence of the influence of the internal inductances. The efficiency of a bandpass filter that could be formed in this way, will therefore be very poor.

As short-circuiting of the points A and B is not allowable, it remains to

be seen what impedance may be inserted between A and B without altering the properties of the circuit. This question is important in view of the D.C. feeding of the valves. To answer the question it is important to note that looking from the point A to both sides, a series-resonance circuit is seen (equation (1) can also be written:

$$L_1 C_a = L_2 C_g = (L_1 + L_2) \frac{C_g C_a}{C_g + C_a} = \frac{1}{\omega_0^2},$$

ω_0 = resonance frequency of the whole circuit).

Looking from A to the left, a resistance r_1 is measured, while to the right this resistance is r_2. As the current circulates in the circuit in the way indicated by the arrow in fig. 6, it is clear that if an impedance, large in comparison with r_1 and with r_2, is connected between A and B, the current distribution in the circuit will not seriously be influenced. The resistances r_1 and r_2 are normally small, so this condition is easy to fulfil. For receiving valves and a wavelength of one metre the series resistances are usually not more then 10 ohms. A resistance of some hundred ohms will not affect the properties of the circuit then.

From the foregoing it is clear that in the circuit given in fig. 6 some A.C. voltage remains between the points A and B. However, this voltage is smaller than the voltage between any other point of the inductance and the earth. We will call this point therefore the voltage node.

Turning back to the problem of tuning the circuit with a parallel trimmer, we notice first that this trimmer can be connected in parallel to the inductance of the circuit, *i.e.* between anode and grid. When the internal inductances are not negligible, the trimmer can only be connected in parallel to a part of the total inductance: the trimmer is tapped on the circuit.

Another way of connecting the trimmer is to connect it directly in parallel to either C_a or C_g (only possible if the internal inductances can be neglected). From fig. 5 it is clear that in this case too, only part of the circuit voltage comes across the trimmer. The voltage across the trimmer diminishes further when the trimmer cannot be connected directly in parallel to C_a or C_g because of the influence of the internal inductances. As described above, the smallest voltage across the trimmer will be measured when the timmer is connected between the voltage node on the inductance and the earth.

However, this last case is rather abnormal and requires closer inspection. In *fig. 7* the circuit concerned is drawn. L_a is supposed to be tuned with C_a, and L_g with C_g. Now it is clear that by connecting the capacity between this point of the inductance and the earth, a second mesh in the

network is formed and therefore a second resonance must be expected. The insertion of the trimmer capacity does not influence the current that circulates in the way indicated by the full-drawn arrow. But the dashed arrows indicate the second resonance system. We have in this way formed a bandfilter consisting of two capacitively coupled circuits.

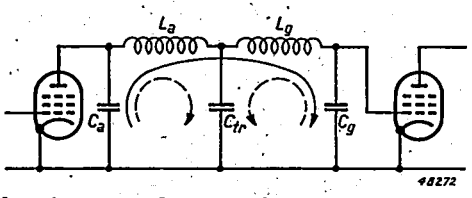


Fig. 7. Connecting the trimmer condenser at the voltage node gives the circuit the properties of a capacitively coupled bandfilter.

As stated above, when the impedance of the trimmer capacity is high compared with the series resistance of the original circuit, this trimmer capacity will not influence the circuit properties *in the neighbourhood of its (original) resonance frequency*. But well away from this frequency, a second resonance frequency will be found.

The same theory applies to the case that the trimmer is connected between any other point of the inductance, apart from its terminals, and the earth. As the magnitude of the trimmer capacity is usually rather small, this second resonance frequency is quite different from the original frequency (it is much higher) and will not be of any importance in most circumstances. However, it should be taken into account because it could do harm, for instance in receivers when it could by bad luck equal the image frequency.

A distinct advantage of connecting the trimmer in parallel to one of the valve capacities (we neglect the internal inductances now, but the following holds also in cases we should take them into account) is that it provides a simple way of matching the valve impedances. *Fig. 8* shows the circuit where now all losses are concentrated in the parallel resistances

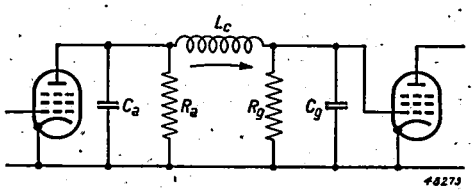


Fig. 8. Matching the valve resistances R_a and R_g can be realized by establishing the proper relation between the end capacities C_a and C_g.

R_a and R_g (we leave it undecided whether they are due to copper, dielectric, transition-time, or other, losses).

Optimum matching is realized when the same amount of power is dissipated in R_a and in R_g:

$$\frac{V_a^2}{R_a} = \frac{V_g^2}{R_g} . \tag{2}$$

As the current circulates in the way indicated by the arrow, equation (2) may also be written:

$$|Z_a|^2/R_a = |Z_g|^2/R_g , \tag{3}$$
$$Z_a = R_a/(1 + j\omega R_a C_a) , \tag{4}$$
$$Z_g = R_g/(1 + j\omega R_g C_g) . \tag{5}$$

From (3), (4) and (5) follows:

$$R_a/(1 + \omega^2 R_a^2 C_a^2) = R_g/(1 + \omega^2 R_g^2 C_g^2) . \tag{6}$$

If R_a and R_g fulfil the conditions

$$\omega^2 R_a^2 C_a^2 >> 1 , \tag{7}$$
$$\omega^2 R_g^2 C_g^2 >> 1 , \tag{8}$$

equation (6) may be written

$$\frac{R_a}{R_g} = \frac{C_g^2}{C_a^2} . \tag{9}$$

Given the valves, the value of R_a/R_g is fixed. But the ratio C_g^2/C_a^2 can also be brought to this same value by increasing the magnitude of the capacity that is too small to fulfil equation (9).

As condition (9) is not very critical for optimum matching, the same capacity which is used for increasing one valve capacity, may be used for the final tuning of the circuit (coarse tuning can be accomplished by varying the inductance).

The same circuit can be used to match a resistance (in practice mostly an antenna resistance) to a valve (*fig. 9*). In parallel to the antenna

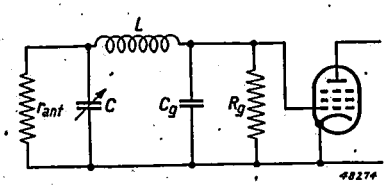


Fig. 9. Matching an antenna resistance to a valve input resistance with the aid of the new circuit.

resistance r_{ant}, a variable condenser C is connected. The circuit is now the same as that drawn in fig. 8, and the same theory holds for this case.

After the theoretical considerations given above, we will now go a little more into details about the practical side of this circuit. As its main advantages show themselves especially at very high frequences, it is important to state that where we mentioned an inductance in the foregoing, this inductance may be either a lumped inductance or a part of a parallel-wire system. The length of this system can easily be found from the consideration that, if at the place of the voltage node both wires are short-circuited (*fig. 10*), the parts of the parallel-wire system to the left and to the right are tuned with the valve capacities. The total length of the system is therefore always smaller than $2 \times \frac{1}{4} \lambda = \frac{1}{2} \lambda$.

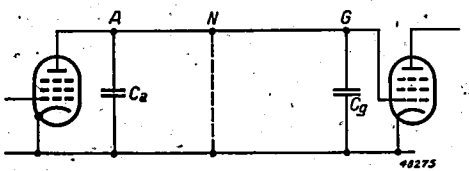


Fig. 10. A parallel-wire system may be used in the circuit; again a voltage node can be found.

Sometimes the place of the voltage node is well inside the envelope of one of the valves, usually on the grid lead, because the input capacity of a valve is normally considerably larger than the output capacity. This does not influence the performance of the circuit, provided of course that the output capacity of the preceding valve is so small, that the length *A-N* in fig. 10 can span the length of the internal anode lead, the lead connecting the two valves and the required length of internal grid lead of the second valve.

At those frequencies where the total length *A-G* in fig. 10 is not enough to span the physical distance between anode and grid, a capacity C may be inserted so as to lengthen the parallel-wire system (*fig. 11*). It is

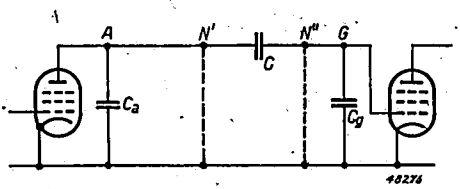


Fig. 11. Lengthening the parallel-wire system with a series condenser. Two voltage nodes can be found in this case.

important to note that in this case two voltage nodes (N' and N'' in fig. 11) will be found on the wires. The part of the parallel-wire system between A and N' is tuned with C_a, that between N'' and G is tuned with C_g, and that between N' and N'' with C.

A capacity C is necessary in all circuits to separate the D.C. voltages of anode and grid. If there is no necessity for lengthening the line, this capacity should be given a large value. If the direct currents are fed to the electrodes over high impedances (chokes or high resistances), these impedances may be connected to any convenient point of the upper line. However, if the large capacity C is situated just at the place of the voltage node, the D.C. may be fed over small impedances to the two terminals of this condenser.

Fig. 12 shows a complete circuit of a push-pull amplifier stage, based on the principles described. The grid resistances R_g are supposed to be large compared with the input resistance of the second valve; therefore they may be connected to the most convenient points, in this case to the grid terminals of the second valve.

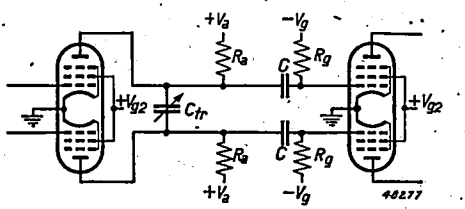


Fig. 12. Complete circuit for a push-pull amplifier stage.

The resistances R_a cannot be made as large as R_g because of the D.C. voltage drop across these resistances as a consequence of the anode currents. For this reason, the resistances R_a are connected to the wires at the place of the voltage node.

A variable tuning condenser C_{tr} is connected between the two wires at a place outside this voltage node. Tuning could be accomplished without the extra trimmer if the condensers C were made variable. To maintain the symmetry of the circuit, both capacities C should in that case be varied simultaneously and to the same extent. It is clear that the use of one separate trimmer C_{tr} is easier. Apart from this, the trimmer may be used to realize optimum coupling conditions as outlined above.

Under certain circumstances the same circuit is useful at much lower frequencies (using lumped inductances for the parallel-wire system). If a receiver or a transmitter is divided into separate panels on a rack, the

transfer of the high- or intermediate-frequency voltage from one panel to another sometimes proves difficult because of the possibility of radiation or unwanted coupling by the connection between the two panels, when this connection cannot be screened sufficiently. However, by using a circuit as described in the foregoing, the inductance can be split up into two parts, each situated in a different panel, such that the voltage node coincides with the unscreened connection point between the two panels. Now, the H.F. voltage at this point is very small, so the possibility of radiation or unwanted coupling is greatly reduced.

In conclusion we give a description of a practical circuit, used to match an antenna to the input resistance of a valve. *Fig. 13* shows a push-pull valve to the grid terminals of which a parallel-wire system has been connected. The surge impedance of this parallel-wire system is made equal to the antenna resistance r_{ant}. Eventually, the antenna feeder itself could be used for this purpose.

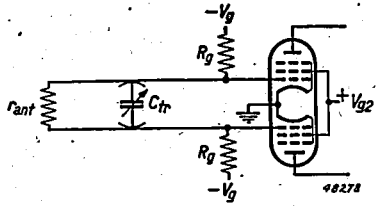


Fig. 13. Antenna-coupling circuit for a push-pull receiving valve.

Between the wires, a movable and variable condenser C_{tr} is connected. Looking from the terminals of C_{tr} to the left, the line impedance always equals the resistance r_{ant}, independently of the place of the condenser. The matching of r_{ant} to the input resistance of the valve can now be found experimentally by placing the capacity in a certain (arbitrary) point and tuning the capacity to maximum transfer of energy. Next, the condenser is moved a little, and again tuned. In this way, optimum coupling conditions are quite easily established.

The circuits described in this article are successfully used in receivers as well as in transmitters.

Eindhoven, July 1946

Philips Res. Rep. 2, 136-160, 1947

INTERFERENCE PROBLEMS IN FREQUENCY MODULATION *)

by F. L. H. M. STUMPERS

621.396.82:621.396.216

Summary

After a survey of definitions, the general problem of interference with frequency-modulated signals is treated. Special attention is paid to the pauses of the desired signal. The case of equal amplitudes gives rise to some interesting mathematical relations. The loudness level of disturbances is computed. In the last two sections the interference caused by synchronized transmitters (or by two-path transmission of one signal) is extensively dealt with. Many numerical examples illustrate the theory.

1. *Introduction*

When two frequency-modulated transmitters are stationed so near to each other in the frequency band as to cause partial or complete overlapping of the spectra, interference will result. A special case is when both transmitters are on the same wavelength, and this we can specialize again by giving them the same programme (synchronized transmitters). Similar problems occur when signals from one transmitter reach the receiver along two different paths.

Crosby [1)] [2)] was the first to note the detrimental effect of multipath transmission on frequency-modulation reception over long distances, and he published photos of distorted low-frequency signals. He also gave a theoretical consideration explaining the most important facts. It should be remarked here that as long ago as 1930 Eckersley [3)] predicted that echoes of 2 or 3 msec comparable in amplitude to the original signal would jeopardize f.m. reception (we use the abbreviations f.m. and a.m. for frequency modulation and amplitude modulation, respectively). Thanks to the inactivity of the ionosphere at wavelengths below 10 m the effect of echoes is not so important here.

With regard to the permissible amplitude of an interfering transmitter most writers take it for granted, following early experiments by Armstrong [4)] and Weir [5)], that a ratio of 2 : 1 in amplitude is sufficient to suppress the disturbance. No quantitative criterion is given, however, and later on Guy and Morris [6)] deduced from their experiments that a ratio of 10 : 1, or at least of 4 : 1, is required. Vellat [7)] concluded,

*) First published as part of a Dutch thesis for the doctor's degree at Delft, May 1946, on "Some investigations on frequency modulation".

though erroneously, that the mutual disturbance of synchronized transmitters is only due to imperfect limiting in the receivers. Wheeler [8]) does not fully justify his remark that no cross-talk occurs with a perfect limiter. In addition to the logical method of limiting first, both the last-mentioned writers also investigated frequency detection without limiting. In this case Vellat overestimated the disturbance of a frequency-modulated signal by an unmodulated carrier by a factor 2.

As is well known, with amplitude-modulation receivers we can have "cross-modulation" as a result of curvature of valve characteristics: the programme of a transmitter on a different wavelength comes through in the pauses of the desired programme. The third-degree term in the valve characteristic, which causes this effect in a.m., does not produce the same result in the case of f.m. It is possible, however, that with two transmitters having central frequencies f_1 and f_2, new frequencies $2f_1-f_2$ or $2f_2-f_1$ appear in the desired band and consequently cause disturbances. According to Wheeler [9]) these effects are not so great as to require a selectivity device before the first valve.

Bell [10]) starts from the formula for the interference between two carrier waves $\cos \omega_1 t$ and $A \cos \omega_2 t$, the instantaneous frequency of the resultant being:

$$\omega_1 + (\omega_2-\omega_1)\frac{A^2 + A\cos(\omega_2-\omega_1)t}{1 + A^2 + 2A\cos(\omega_2-\omega_1)t}. \qquad (1)$$

He continues with the remark that when $\omega_2-\omega_1$ is super-audible the effective frequency will be $\omega_1 + (\omega_2-\omega_1)A^2/(1 + A^2)$. The value of this remark seems questionable as the next statement, that the modulation is suppressed by a factor $A^2/(1 + A^2)$, is erroneous. Wundt and Hoffmann [11]) give a series of excellent curves for distorted signals and for special cases a computation of the effect on the fundamental tone of the desired modulation caused by a disturbing unmodulated transmitter on the same wavelength. Kulp's [12]) suggestion that a perfect detector may cause enlargement of the harmonics is not justified. Some general considerations on interference in relation to amplitude-, phase- and frequency-modulated signals were given by Keall [13]).

After our thesis had been printed, an experimental and theoretical paper by Corrington [14]) reached us. This author comes to formulae, partly similar to those at which we arrived independently, but he failed to notice how the limiting case of equality can be computed elementarily without the help of Bessel functions. We will return to this later on.

In the following the spectra will be regarded systematically, and the energy of disturbances will be computed, if necessary accounting for

pre-emphasis. No reference to this effect was found in the literature. Our mathematical treatment, especially for equal amplitudes, will also be different.

2. *Definitions*

Before proceeding to deal with our problem we would stress the importance of unambiguous definitions of amplitude and frequency for any given function of time (see also Van der Pol[15])). As is well known, a harmonic vibration can be represented by the motion of the projection on a diameter, of a point running along a circle with uniform velocity. The direction of motion will be chosen counter-clockwise. The vector connecting centre and moving point is called the characteristic vector. It is defined by its length: the *amplitude*, and the angle it makes with a fixed direction: the *phase angle*. As fixed direction we choose the direction of the diameter on which the motion is projected. The motion of the projection can be represented by the formulae:

$$f_0(t) = A \,.\, \cos \varphi(t), \quad \varphi(t) = \varphi\,(0) + \omega t\,. \tag{2}$$

In like manner we can represent a given function of time by the projection of a moving vector of varying length $A(t)$ making a time-dependent angle $\varphi(t)$ with a fixed direction. Thus:

$$f(t) = A(t) \,.\, \cos \varphi(t)\,. \tag{2a}$$

Now $\varphi(t)$ is not necessarily a linear function of time. Moreover, one is still free in the choice of the splitting. Neither is the best choice always self-evident. Take, for instance, as given function $\cos \omega_1 t + \cos \omega_2 t$. There are different ways of defining A and φ:

$$\begin{aligned} A_1(t) &= 2 \cos \tfrac{1}{2}(\omega_1 + \omega_2)\, t, & \varphi_1(t) &= \tfrac{1}{2}(\omega_1 - \omega_2)\, t, \\ A_2(t) &= 2 \cos \tfrac{1}{2}(\omega_1 - \omega_2)\, t, & \varphi_2(t) &= \tfrac{1}{2}(\omega_1 + \omega_2)\, t, \\ A_3(t) &= 1 + \frac{\cos \omega_1 t}{\cos \omega_2 t}, & \varphi_3(t) &= \omega_2 t\,. \end{aligned}$$

To make the splitting unambiguous, we require that:

1) the new definitions of amplitude and phase when applied to harmonic vibrations should give the same result as the classical definitions;
2) the characteristic vector of a sum of functions coincide with the resultant of the vectors representing the separate functions (see Cocci and Sartori[16])).

These demands fix the characteristic vector for the sum of two or more harmonic vibrations. The generalization to a convergent infinite series of components is then simple.

Under proper conditions, we can represent any function of time given in the interval from $-T$ to T by a Fourier series. We thus have:

$$f(t) = \sum_0^\infty A_n \cos(n\omega t + \varphi_n); \quad \varphi_0 = 0 . \tag{3}$$

The separate components are again harmonic vibrations, which we may now consider as being performed by the projections on a common diameter, of points moving with uniform speed on concentric circles. According to the condition 2), the amplitude of $f(t)$ is defined by the length of the resulting characteristic vector. By projecting also on a line perpendicular to the diameter used before we get:

$$A(t) = [\{\sum_0^\infty A_n . \cos(n\omega t + \varphi_n)\}^2 + \{\sum_1^\infty A_n . \sin(n\omega t + \varphi_n)\}^2]^{1/2} . \tag{4}$$

The phase angle is again defined as the angle between characteristic vector and fixed direction. Thus:

$$\varphi(t) = \arctan \frac{\Sigma A_n . \sin(n\omega t + \varphi_n)}{\Sigma A_n . \cos(n\omega t + \varphi_n)} = \arcsin \frac{\Sigma A_n . \sin(n\omega t + \varphi_n)}{A(t)} .$$

Of course the phase is defined modulo 2π only.

In this way, in any time interval every function can be split into amplitude and phase. At every moment we therefore know the *instantaneous amplitude* and the *instantaneous phase*.

The next important factor is the frequency. In case of a harmonic vibration the end point of the characteristic vector moves at a constant angular velocity along the circle. This angular velocity is called the angular or radial frequency. It is expressed in radians per second. Quite analogously in the general case the angular velocity of the characteristic vector is the (angular) *frequency* for any function of time. It is the derivative of the phase with respect to time: $d\varphi/dt$.

Thus the *instantaneous frequency* is also defined (the definition of frequency as derivative of phase was introduced in radio by Van der Pol[17]). As one period or cycle is 2π radians, we may also use cycles or periods per second. However, unless otherwise stated, we will take the frequency in radians per second.

With periodic signals one will choose the time interval corresponding to one period. In calculations with impulses occurring only once Fourier integrals may be preferred. Of course, the time interval should be long in comparison with the greatest period of the frequencies under discussion.

The instantaneous frequency of a signal resulting from decreasing

all frequencies of the original components by a constant amount (superheterodyne principle) is not necessarily the difference of the instantaneous frequencies. This holds only if all component frequencies are originally larger than the difference frequency. In most practical cases all frequencies in which we are interested are in a comparatively narrow band around a central frequency. This makes the application much simpler.

3. *The general problem*

We compute the instantaneous frequency of the resultant vibration caused by two frequency-modulated components, the amplitude of the stronger being 1 and that of the weaker A (in suitable units). The frequency of the modulating signal for the first vibration is p and for the second q rad/sec. The frequency sweeps are $\Delta\omega_1$ and $\Delta\omega_2$ rad/sec, respectively (although we limit ourselves here to one low-frequency signal on each transmitter, it is not difficult to take more frequencies into account).

The components are therefore:

$$\cos \varphi_1(t) = \cos (\omega_1 t + m_1 \sin pt),$$

and

$$A \cos \varphi_2(t) = A \cos (\omega_2 t + m_2 \sin qt + \varphi_0),$$

$$m_1 = \frac{\Delta\omega_1}{p}; \; m_2 = \frac{\Delta\omega_2}{q}.$$

The modulation index m is the quotient of the frequency sweep and the modulating frequency.

The instantaneous amplitude is $\{1 + A^2 + 2A \cos \varphi(t)\}^{1/2}$, where $\varphi(t) = \varphi_2(t) - \varphi_1(t)$ (see *fig. 1*). The phase angle is defined by:

$$\Psi = \arctan \frac{\sin \varphi_1 + A \sin \varphi_2}{\cos \varphi_1 + A \cos \varphi_2} = \varphi_1 + \arctan \frac{A \sin \varphi}{1 + A \cos \varphi}. \qquad (5)$$

Now we use the well-known formula $\arctan x = \frac{1}{2i} \ln \frac{1 + ix}{1 - ix}$. The result is

$$\Psi = \varphi_1 + \frac{1}{2i} \ln \frac{1 + Ae^{i\varphi}}{1 + A e^{-i\varphi}}.$$

Development of the logarithm is permitted, as $A < 1$, and we get in this way:

$$\Psi = \varphi_1 + A \sin \varphi - \frac{A^2}{2} \sin 2\varphi + \frac{A^3}{3} \sin 3\varphi - \ldots . \qquad (6)$$

By differentiating (5) the definition of instantaneous frequency gives:

$$\omega_m = \frac{d\Psi}{dt} = \omega_1 + \Delta\omega_1 \cos pt + \frac{A\cos\varphi + A^2}{1 + 2A\cos\varphi + A^2} \cdot \frac{d\varphi}{dt}.$$

The first terms in this result give the required signal in the absence of a second transmitter.

Now the function $f(A, \varphi) = \dfrac{A\cos\varphi + A^2}{1 + 2A\cos\varphi + A^2}$ is so important that it is worth while considering it in some detail. It is a periodic function of φ of period 2π, and can be represented by a Fourier series. If $A < 1$, we get, *e.g.* differentiating (6), the development:

$$f(A, \varphi) = A\cos\varphi - A^2\cos 2\varphi + A^3\cos 3\varphi - \ldots . \tag{7}$$

Furthermore:

$$f(A, 0) = \frac{A}{1+A}; \quad f\left(A, \frac{\pi}{2}\right) = \frac{A^2}{1+A^2}; \quad f(A, \pi) = \frac{-A}{1-A}.$$

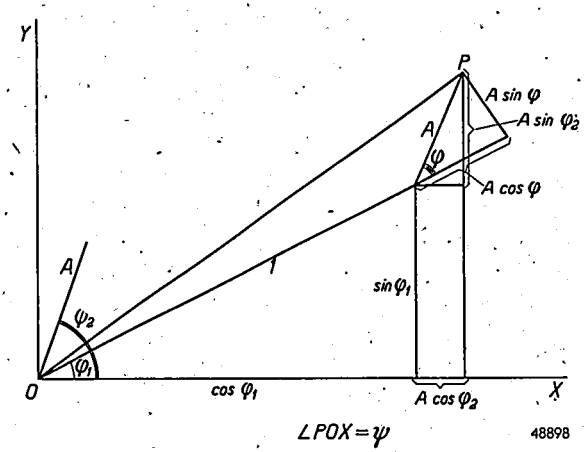


Fig. 1. Addition of characteristic vectors.

The last relation is valid when $A \neq 1$. As A approaches 1 the peak at π grows sharper and the absolute value of $f(A, \pi)$ exceeds any given number. Then the following relations are valid: (all φ mod 2π)

$$\lim_{A=1} f(A, \varphi) = \tfrac{1}{2}, \quad \varphi \neq \pi \tag{8a}$$

$$\lim_{A=1-0} f(A, \pi) = -\infty \tag{8b}$$

$$\lim_{\varepsilon=0} \lim_{A=1-0} \int_{\pi-\varepsilon}^{\pi+\varepsilon} f(A, \varphi)\, d\varphi = -\pi \tag{8c}$$

and finally

$$\lim_{\varepsilon=0} \lim_{A=1-0} \int_{\pi-\varepsilon}^{\pi+\varepsilon} f(A, \varphi) \cdot g(\varphi) \cdot \mathrm{d}\varphi = -\pi g(\pi) \qquad (8d)$$

if $g(\varphi)$ is continuous at $\varphi = \pi$.
With the δ-functions commonly used in operational calculus we can at once write equations (8a, b, c, d) (modulo 2π included):

$$\lim_{A=1-0} f(A, \varphi) \doteq \tfrac{1}{2} - \frac{\pi}{2}\,\delta\left(\cos\frac{\varphi}{2}\right) \qquad (9)$$

(in operational calculus $\delta(x) = 0$, if $x \neq 0$; $\delta(0) = \infty$; $\int_{-\infty}^{+\infty} \delta(x)\,\mathrm{d}x = 1$).

It is sufficient to consider $A \leqq 1$, as all other values follow from the relation:

$$f(A, \varphi) + f\left(\frac{1}{A}, \varphi\right) = 1\,. \qquad (10)$$

In *fig. 2* $f(A, \varphi)$ is drawn for various values of A.
We have seen in formula (6):

$$\Psi = \varphi_1 - \sum_1^{\infty}{}_s \frac{(-1)^s A^s}{s} \sin s\varphi\,.$$

Here

$$\sin s\varphi = \operatorname{Im} e^{is\varphi} = \operatorname{Im} e^{is(\omega_s t - \omega_1 t + \varphi_0 + m_1 \sin qt - m_1 \sin pt)}\,.$$

Now we use the development:

$$e^{im \sin pt} = \sum_{-\infty}^{+\infty}{}_n J_n(m)\, e^{i npt}\,. \qquad (11)$$

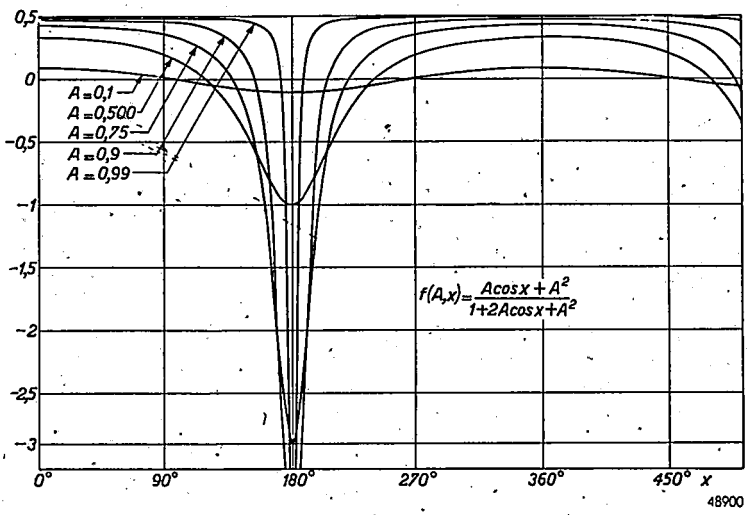


Fig. 2. $f(A, x)$ for various values of A and x.

Thus:

$$\Psi = \varphi_1 - \sum_1^{\infty}{}_s \frac{(-1)^s A^s}{s} \sum_{-\infty}^{+\infty}{}_n J_n(sm_2) \sum_{-\infty}^{+\infty}{}_m J_m(sm_1) \sin\{(s\omega_2 - s\omega_1 - mp + nq)\, t + s\varphi_0\}.$$

The resultant of the two frequency-modulated components therefore has the instantaneous frequency:

$$\omega_m = \frac{d\Psi}{dt} = \omega_0 + \Delta\omega_1 \cos pt - \sum_1^{\infty}{}_s \frac{(-1)^s A^s}{s} \sum_{-\infty}^{+\infty}{}_m J_m(sm_1) \sum_{-\infty}^{+\infty}{}_n J_n(sm_2) \cdot$$
$$\cdot (s\omega_2 - s\omega_1 - mp + nq) \cos\{(s\omega_2 - s\omega_1 - mp + nq)\, t + s\varphi_0\}. \quad (12)$$

We see that if $\omega_2 \neq \omega_1$, the detected signal in general does not contain the frequency q. Hence there is no understandable cross-talk, and the second term gives only noise. When the audio-frequency bandwidth is ω_a, and we wish to compute the audible disturbance, we have to add the energy of all components for which m, n and s satisfy the relation:

$$-\omega_a < s(\omega_2 - \omega_1) - mp + nq < \omega_a.$$

From the symmetric character of the expression (12) it is clear that when we exchange the modulations the disturbance remains the same. Therefore every computation covers two cases. When neither the desired nor the disturbing transmitter is modulated, the disturbance energy is:

$$\sum_1^{n}{}_s \frac{A^{2s}}{s^2} \frac{s^2(\omega_2 - \omega_1)^2}{2} = \frac{A^2(1 - A^{2n})}{1 - A^2} \frac{(\omega_2 - \omega_1)^2}{2}, \quad (13a)$$

in which n is to be chosen such as to satisfy

$$n(\omega_2 - \omega_1) \leqq \omega_a < (n + 1)\,(\omega_2 - \omega_1).$$

In *fig. 3*, instead of the energy the effective sweep of the disturbance is put along the ordinate. We have reckoned with an audio-frequency bandwidth of 15 kc/sec. The amplitude of the disturbance increases linearly with the frequency difference and the proportionality factor changes with the number of components in the audio band.

In a normal music or voice spectrum the higher tones have less energy than the lower ones. A better use of the frequency band is made when before modulation the amplitude of every frequency ω in the low-frequency signal is multiplied by a factor $(1 + R^2C^2\omega^2)^{1/2}$. This is called pre-emphasis. In the receiver the reverse operation has to be applied; de-emphasis. One has to choose a suitable RC time. In our computations we will use 75.10^{-6} sec as agreed in the U.S.A. for broadcasting purposes. The advantages of pre-emphasis will be manifest, if the transmitted low-frequency spectrum possesses so little energy in its high tones that after pre-emphasis

the corresponding maximum frequency sweep is not larger than it is for the low tones. Otherwise one has to reduce the whole volume. This is the case when the RC time is chosen too high (as the originally adopted time of 10^{-4} sec seems to have been), and will be so too when there is much extra energy in the high tones (cymbals, triangle). The RC time has of course been chosen for the average spectrum and not for exceptional circumstances.

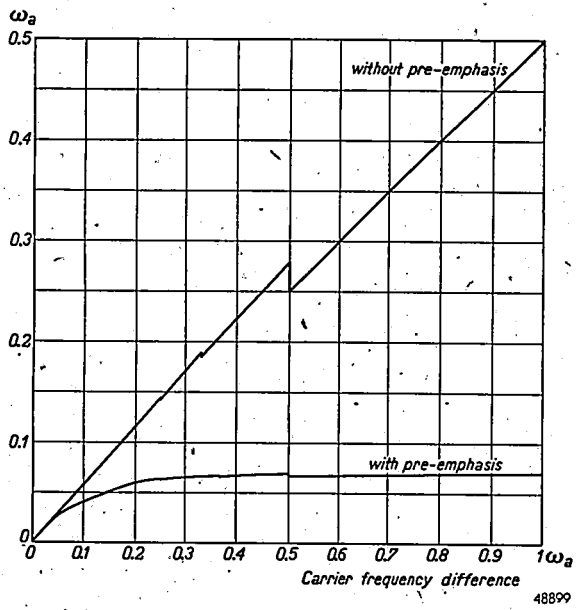


Fig. 3. Effective sweep of disturbance caused by a transmitter with given carrier frequency difference in the pauses of the desired isgnal. Interfering transmitter non-modulated. $A = 0{\cdot}5$.

With pre-emphasis and no transmitter modulated, the energy of the disturbance will be:

$$\sum_{1}^{n} \frac{A^{2s}(\omega_2 - \omega_1)^2}{2\{1 + R^2 C^2(\omega_2 - \omega_1)^2\}} \qquad (13b)$$

(n has to be chosen as above).

Usually the energy of disturbances is compared with the energy of full modulation. In our computations we will use a full-modulation sweep of 75 kc/sec. Taking 10 times the logarithm (of base 10) of the quotient of the energies we get the relative energy in decibels (db). When in the following we give the disturbance energy in db, the energy in db below

full modulation will be meant. The equations (13) apply to the following case: We are listening to a transmitter; there is a pause in the desired programme, and we hear the disturbance caused by another, not modulated, transmitter. If the carrier-frequency difference is 5 kc/sec, and $A = 0{\cdot}1$, the disturbance energy is 43·5 db below full modulation. When pre-emphasis is applied (and de-emphasis) the disturbance energy is 51·7 db. If $A = 0{\cdot}25$ we get 35·6 db without, and 43·6 db with pre-emphasis. If $A = 0{\cdot}5$ the results are 28·4 and 37·4 db respectively.

Now the energy of a disturbance is not the best measure of the interference as the sensitiveness of the ear is not taken into account. We have also computed the loudness level in phons. Use was made of the curves in a paper by Fletcher [24]). The listener is supposed to be in a quiet room, and the masking effect of room noise is not reckoned with. If full modulation corresponds to 80 phons the loudness level of the disturbances is (no pre-emphasis): 29 ($A = 0{\cdot}1$), 39 ($A = 0{\cdot}25$), or 47 phons ($A = 0{\cdot}5$). With pre-emphasis those levels are 22, 29 and 36 phons respectively. When full modulation corresponds to 90 phons the six numbers, just given, will become: 39, 49, 59 or 30, 41, 47 phons.

If, with the same carrier-frequency difference the disturbing transmitter is modulated by 2000 cycles/sec with a sweep of 12 kc/sec, the energy levels of the disturbances are 39·1, 31·0 and 24·7 db below full modulation without pre-emphasis ($A = 0{\cdot}1$, $0{\cdot}25$, $0{\cdot}5$) and 52·3, 44·2, 37·8 db with pre-emphasis. With full modulation corresponding to 80 phons the respective loudness levels are 36, 47 and 58 phons without pre-emphasis and 22, 33, 43 phons with pre-emphasis. With full modulation corresponding to 90 phons the figures are in the same order 51, 62, 71 phons and 34, 47, 58 phons. Now we increase the sweep of the disturbing transmitter to 48 kc/sec. For $A = 0{\cdot}1$ the disturbance energy drops to 50·9 db without, and to 61·0 db with pre-emphasis. Again with a full-modulation loudness level of 80 phons the loudness level of the disturbances in this case is 26 or 17 phons, and when full modulation is 90 phons these figures are 38 or 28 phons.

4. *Two transmitters with the same carrier wavelength*

With both transmitters using the same carrier frequency, the instantaneous frequency of the resultant according to (12) is:

$$\omega_0 + \Delta\omega_1 \cos pt - \\ -\sum_1^{\infty} \frac{(-1)^s A^s}{s} \sum_{-\infty}^{\infty} J_m(sm_1) \sum_{-\infty}^{+\infty} J_n(sm_2) \cdot (nq - mp) \cos \{(nq - mp)\, t + s\varphi_0\} \quad (14)$$

Here we may have cross-talk, as the frequency q occurs in the output

spectrum. Moreover, if the modulation indices are small the important term in this respect is

$$A\,J_0(m_1)\,J_1(m_2)\,.\,q\,.\cos(qt+\varphi_0) \approx \tfrac{1}{2}\,A\,\Delta\omega_2 \cos(qt+\varphi_0)\,.$$

Thus the disturbing spectrum is heard at $\frac{1}{2}A$ times its original strength. This effect only plays a rôle when the indices are small. Otherwise many other components of equal strength make reception unintelligible. When, however, one wishes to account for large differences in level without special artifices, such as compression or expansion, the transmitters may well be modulated with a small sweep during a large percentage of the time.

We now consider a few special cases.

a) One transmitter not modulated.

If the desired signal is not modulated the instantaneous frequency is:

$$\omega_0 - 2q\sum_0^\infty{}^n (2n+1)\cos(2n+1)\,qt \sum_1^\infty{}^s \frac{(-1)^s A^s}{s} J_{2n+1}(sm)\,.\cos s\varphi_0$$
$$+\,2q\sum_0^\infty{}^n 2n \sin 2nqt \sum_1^\infty{}^s \frac{(-1)^s A^s}{s} J_{2n}(sm)\sin s\varphi_0\,. \qquad (15)$$

With this formula we computed the disturbance energy for a few cases. Audio-frequency band 15 kc/sec, maximum sweep of desired transmitter 75 kc/sec. We will take $\varphi_0 = 0$, and $m = 10$ (thus the sweep of the disturbing transmitter varies). We will also give the disturbance in phons for full modulation corresponding to 80 or 90 phons.

q	No pre-emphasis			With pre-emphasis		
cycles/sec	$A = 0{\cdot}1$	$A = 0{\cdot}25$	$A = 0{\cdot}5$	$A = 0{\cdot}1$	$A = 0{\cdot}25$	$A = 0{\cdot}5$
7500	62·0	53·9	58·8	73·3	65·2	70·1 db
max. mod. 80 ph.	10	15	12	0	6	2 phons
max. mod. 90ph.	17	23	21	8	14	10 phons
4500	52·6	43·5	35·3	67·7	59·1	51·4 db
	15	23	29	3	11	13 phons
	26	33	42	13	22	25 phons
2800	40·8	32·3	25·2	57·0	48·5	41·4 db
	26	35	42	13	18	27 phons
	37	47	53	21	31	37 phons
2100	39·0	30·5	23·3	54·5	46·0	38.8 db
	33	43	49	14	23	33 phons
	43	54	63	26	36	44 phons
1600	37·1	29·2	23·0	52·2	44·0	38·5 db
	33	45	53	16	27	36 phons
	47	57	64	29	42	49 phons
1300	38·3	30·4	24·3	52·0	43·9	37·4 db
	35	47	55	20	32	40 phons
	49	59	67	33	43	53 phons

From our formula we expect a large disturbance in the high audio-frequencies, and in the maximum of J_1. This maximum is 0·5819, and it is reached at $m = 1·84$. With this value of m we get the disturbance energy, as follows:

q	No pre-emphasis			With pre-emphasis		
cycles/sec	$A = 0·1$	$A = 0·25$	$A = 0·5$	$A = 0·1$	$A = 0·25$	$A = 0·5$
15000	32·7	24·9	19·4	47·8	42·0	36·4 db
	31	39	44	12	22	27 phons
	42	48	56	26	38	37 phons
8000	38·2	30·4	24·8	53·4	45·6	40·0 db
	29	37	43	16	23	28 phons
	39	47	54	25	33	39 phons

Now we will consider what happens if the disturbing transmitter becomes as strong as the desired one. We can also write the instantaneous frequency in the form:

$$\omega_0 + \Delta\omega \cos qt \, \frac{A \cos (m \sin qt + \varphi_0) + A^2}{1 + 2A \cos (m \sin qt + \varphi_0) + A^2} . \qquad (16)$$

We have seen earlier that $f(x, A)$ gets a peak for $x = (2k + 1)\pi$, which is sharper the more A approaches 1. Here $x = m \sin qt + \varphi_0$, and thus, in half a low-frequency cycle x runs from $-m + \varphi_0$ to $m + \varphi_0$, to return in the next half cycle. The number of disturbance peaks in such a half cycle depends on the phase angle φ_0, and is equal to $[\Delta\omega/\pi q]$ or 1 more ($[x]$ means the largest integer contained in x). The peaks occur here in a product form with $\cos qt$. Their direction turns over when A passes 1. The situation is easily plotted graphically (in *fig. 4* the peaks are drawn for the limiting case $A = 1—0$. In other cases one has to replace them by the less sharp peaks corresponding to the value of A, as in fig. 2).

We now take the limit $A = 1—0$ in formula (16), and use equation (9). The result is:

$$\frac{\Delta\omega}{2} \cos qt - \frac{\pi}{2} \Delta\omega \cos qt \, . \, \delta \left\{ \cos \frac{m \sin qt + \varphi_0}{2} \right\} . \qquad (17)$$

Returning to formula (15), we take here also the limit for $A = 1—0$. We apply the well-known integrals of Fourier analysis to equation (17), and, using the peculiarities of δ-functions, we find the harmonics in closed form. In this way we get two different forms for the sweep of the fundamental tone:

$$2q \left\{ J_1(m) \cos \varphi_0 - J_1(2m) \, . \, \frac{\cos 2\varphi_0}{2} + J_1(3m) \, . \, \frac{\cos 3\varphi_0}{3} + \ldots \right\} =$$

$$2q \left[\frac{m}{4} - \Sigma k \cos \left\{ \arcsin \frac{(2k + 1)\pi - \varphi_0}{m} \right\} \right] , \qquad (18a)$$

in which the summation must be taken over all values of k satisfying:

$$-m < (2k+1)\,\pi - \varphi_0 \leqq m.$$

The definition domain of arcsin runs from $-\pi/2$ to $\pi/2$.

In the same way the sweep corresponding to the second harmonic is:

$$2q \,.\, 2\left\{J_2(m) \,.\, \sin\varphi_0 - J_2(2m) \,.\, \frac{\sin 2\varphi_0}{2} + J_2(3m) \,.\, \frac{\sin 3\varphi_0}{3} - \ldots\right\} =$$

$$2q\ \Sigma k \sin 2\left[\arcsin\frac{(2k+1)\,\pi-\varphi_0}{m}\right], \qquad (18b)$$

where the summation must be taken over the same values of k. Likewise the sweep of the third harmonic is:

$$2q \,.\, 3\left\{J_3(m) \,.\, \cos\varphi_0 - J_3(2m) \,.\, \frac{\cos 2\varphi_0}{2} + J_3(3m) \,.\, \frac{\cos 3\varphi_0}{3} + \ldots\right\} =$$

$$-\,2q\ \Sigma k \cos 3\left[\arcsin\frac{(2k+1)\,\pi-\varphi_0}{m}\right]. \qquad (18c)$$

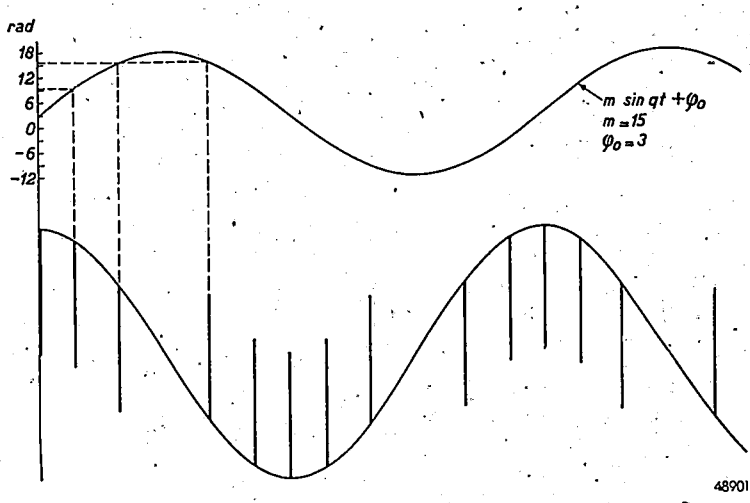


Fig. 4. Output of the receiver in the pauses of the desired signal. The interfering signa is modulated and has nearly the same amplitude. Graphical construction for the peaks.

When we divide the left and right terms by $2q$ in all three equations (18), remarkable identities for series of Bessel functions are obtained. Furthermore we see that, if there lies no odd multiple of π between $-m+\varphi_0$ and $m+\varphi_0$, the modulation comes through undistorted but with half the original amplitude.

5. *On synchronized transmitters and two-path transmission. One modulating frequency.*

We will now consider the interference spectra resulting from two synchronized transmitters modulated with a single tone. The same formulae apply to the case where the signals of one transmitter reach the receive along different paths. We will again give the largest signal the amplitude 1, and the other $A < 1$. We will ascribe the phase difference of the two signals entirely to a time delay. One receiver gets two signals:

$$\cos(\omega_0 t + m \sin pt) \quad \text{and} \quad A \cos\{\omega_0 t + m \sin p(t+\tau) + \omega_0 \tau\};$$

ω_0 = carrier frequency; $m = \Delta\omega/p$ = modulation index; $\Delta\omega$ = frequency sweep; p = audio signal frequency; τ = time delay.
The resultant amplitude B of the two interfering signals is given by:

$$B^2 = 1 + A^2 + 2A \cos\left\{2m \sin p\frac{\tau}{2} \cdot \cos p\left(t + \frac{\tau}{2}\right) + \omega_0\tau\right\}.$$

If the phase difference between the two signals is an odd multiple of π, then $B = 1-A$. Difficulties will arise if the amplitude B drops below the effective local-noise amplitude, or below the threshold of the limiter. For small values of τ we can replace $2\ m \sin p\ (\tau/2)$ by $\Delta\omega\tau$. If $\Delta\omega\tau > \pi$ the minimum value of the amplitude is reached at least once in every low-frequency cycle. When however $\Delta\omega\tau$ is smaller and we travel in the domain of equal amplitudes, $\omega_0\tau$ is the cardinal quantity.

If the total amplitude is permitted to drop as low as one tenth of the amplitude of one transmitter alone, the danger zone is only from 174°16′ to 185°44′ in every 360°. Travelling from one station towards the other in the equality area this means only 3% of time. If the total amplitude is only allowed to drop to half the amplitude of one signal, difficulties will arise between 151°3′ and 208°57′, but even this is only 16% of the time. It is clear that conditions become worse, if one travels on a path shaped like part of a hyperbola, such that the phase difference between the signals is continually π or $(2k-1)\ \pi$. Even in this case, however, it should be possible to get better conditions by moving one metre aside. From considerations of amplitudes only, we can thus deduce that with $\Delta\omega\tau \ll \pi$ there are places (near $\omega_0\tau = \pi$) in the region of equal field strength where the noise overrules the signals all the time but there is more space where the signals are always stronger. If, however, in this region $\Delta\omega\tau \gg \pi$ we have to expect some noise everywhere not very much dependent on the place or the value of $\omega_0\tau$. The effective value of the noise depends on the ratio of the field strength of one transmitter alone to the local-noise level.

Now we will investigate what effect those two synchronized transmitters produce in an ideal frequency detector (no amplitude effect). As stressed in the preceding paragraph this will give a fair approximation of actual conditions, if the amplitude of one transmitter is sufficiently above local noise. Be $m_1 = 2\,m \sin p(\tau/2)$, then we find with the help of formulae (5), (7), and (11) for the instantaneous frequency:

$$\omega_m = \omega_0 + \Delta\omega \cos pt$$
$$-2p \sum_0^\infty{}^n (-1)^n (2n+1) \sin (2n+1) p\left(t+\frac{\tau}{2}\right) . \sum_1^\infty{}^s \frac{(-1)^s A^s}{s} . J_{2n+1}(sm_1) \cos s\omega_0\tau$$
$$+2p \sum_1^\infty{}^n (-1)^n 2n \sin 2n\,p\left(t+\frac{\tau}{2}\right) . \sum_1^\infty{}^s \frac{(-1)^s A^s}{s} J_{2n}(sm_1) . \sin s\omega_0\tau . \quad (19)$$

In the case of limit $A = 1\text{—}0$, we may also write:

$$\omega = \omega_0 + \Delta\omega \cos\frac{p\tau}{2} . \cos p\left(t+\frac{\tau}{2}\right) +$$
$$+ \pi\delta\left[\cos\tfrac{1}{2}\left\{m_1 \cos p\left(t+\frac{\tau}{2}\right)+\omega_0\tau\right\}\right] . \Delta\omega \sin\frac{p\tau}{2} . \sin p\left(t+\frac{\tau}{2}\right) . \quad (20)$$

Comparing formulae (19) and (15), we see that the disturbance is the same as that caused in the pauses of the desired signal by a disturbing transmitter on the same wavelength, with frequency sweep $2\Delta\omega \sin (p\tau/2)$ and low-frequency signal proportional to $\cos\left(pt + p\,\frac{\tau}{2}+\frac{\pi}{2}\right)$. The situation of the disturbing peaks can therefore be computed in the same way as in the preceding section.

To get an impression of the amplitude relations, we will give the situation for two transmitters 72 km distant from each other. The distances are given along the direct connection line, and the figures are computed from a map, based on calculations by Van der Pol and Bremmer[18]). It is seen there that, if the transmitter aerials are placed at the same height, the ratio between the amplitudes does not vary largely with this height.

Distances	Amplitude ratio
36 km-36 km	A = 100%
38 km-34 km	A = 75%
40 km-32 km	A = 55%
42 km-30 km	A = 42%
46 km-26 km	A = 22%
50 km-22 km	A = 12%

At a frequency sweep of 10 kc/s, a path difference of 5 km corresponds to $\Delta\omega\tau = 1$. In case of equality of signals, with formula (20), the amplitude of the n^{th} harmonic of the original modulating tone is found to be:

$$C_n = -2p \, \Sigma^k \sin n \left\{ \arccos \frac{(2k+1)\,\pi - \omega_0\tau}{m_1} \right\}, \qquad (21)$$

where the summation has to be taken over all values of k satisfying:

$$-m_1 < (2k+1)\,\pi - \omega_0\tau < m_1$$

(by comparing this result with the direct result of formula (19) we can easily produce identities similar to formulae (18)). The sum reduces to one term as long as $\Delta\omega\tau < \pi$, and all harmonics are smaller than $2p$. In case $\Delta\omega\tau = 1$ distortion only occurs if $\pi - 1 < \omega_0\tau < \pi + 1$; the sum in formula (21) reduces to one term, and we can write the instantaneous frequency:

$$\omega_m = \omega_0 + \Delta\omega \cos\frac{p\tau}{2} \cdot \cos p\left(t + \frac{\tau}{2}\right) + \Sigma\, C_n \sin np\left(t + \frac{\tau}{2}\right).$$

The values of C_n depend on $\omega_0\tau$, as shown in the following table ($\omega_0\tau$ has been given so as to make arccos $(\pi - \omega_0\tau)$ a simple angle).

$\beta_0\tau = \pi - 0{\cdot}965$		$\pi - 0{\cdot}866$	$\pi - 0{\cdot}707$	$\pi - 0{\cdot}5$	$\pi - 0{\cdot}259$	π
C_1	$-0{\cdot}518\ p$	$-p$	$-1{\cdot}414\ p$	$-1{\cdot}732\ p$	$-1{\cdot}93\ p$	$-2p$
C_2	$-p$	$-1{\cdot}732\ p$	$-2\ p$	$-1{\cdot}732\ p$	$-p$	0
C_3	$-1{\cdot}414\ p$	$-2\ p$	$-1{\cdot}414\ p$	0	$1{\cdot}414\ p$	$2p$
C_4	$-1{\cdot}732\ p$	$-1{\cdot}732\ p$	0	$1{\cdot}732\ p$	$1{\cdot}732\ p$	0
C_5	$-1{\cdot}93\ p$	$-p$	$1{\cdot}414\ p$	$1{\cdot}732\ p$	$-0{\cdot}518\ p$	$-2p$
C_6	$-2\ p$	0	$2\ p$	0	$-2\ p$	0

On the other side of π the same absolute value of C_n returns symmetrically, but the sign changes for the even terms. For large values of m_1 the sum (21) can be replaced by an integral for all lower harmonics ($\pi n \ll m_1$). This integral is:

$$-\frac{pm_1}{\pi} \int_{-1}^{+1} \sin n(\arccos x)\, dx\,,$$

and the result of the integration is $\frac{1}{2}\, pm_1$ if $n = 1$, and 0 if $n = 2, 3, \ldots$ $\left(\ll \frac{m_1}{\pi}\right)$. The resultant frequency, apart from very high harmonics, is therefore:

$$\omega_0 + \Delta\omega \cos\frac{p\tau}{2} \cdot \cos p\left(t + \frac{\tau}{2}\right) + \Delta\omega \sin\frac{p\tau}{2} \cdot \sin p\left(t + \frac{\tau}{2}\right) = \omega_0 + \Delta\omega \cos pt\,.$$

In addition to the fact that a high sweep asks much more energy (if the sweep is 5 times higher, the energy has to be 25 times larger) for the same coverage those high harmonics may correspond to audible intermodulation to be discussed in the next section.

Even with a very small time difference, beginning with $\omega_0\tau = \pi$, harmonics of the same order, as given above for $\Delta\omega\tau = 1$, can be noted if, with equality of signals the phase difference remains between very narrow limits. The situation is, however, very much different if one signal is at least twice as strong as the other. As this has an application in f.m. broadcasting (two-path reception) we will give the maximum value of the disturbance in table form for $\Delta\omega_{max} = 75$ kc/sec, $\omega_a = 15$ kc/sec. The energy of the harmonics (mainly second harmonic) is given in db below full modulation for $p = 7500$ cycles/sec and $p = 1000$ cycles/sec, and $A = 0{\cdot}1$ or 0.5.

$\Delta\omega\tau$	No pre-emphasis				Pre-emphasis			
	7500 c/s		1000 c/s		7500 c/s		1500 c/s	
	0·1	0·5	0·1	0·5	0·1	0·5	0·1	0·5
0·1	86·3	66·0	103·8	83·5	103·4	83·0	106·5	83·8 db
0·5	58·5	37·8	76·0	56·3	75·5	54·9	81·3	59·0 db
1	46·9	28·1	64·4	45·6	64·0	45·1	67·1	48·3 db
2	37·0	22·5	54·5	40·0	54·1	39·5	57·3	42·7 db
3	34·4	14·3	51·9	31·3	51·5	31·4	54·7	34·2 db

As we have only chosen $\Delta\omega\tau$, and not $\Delta\omega$, we cannot give the loudness level of harmonics and fundamental together. We will however give the loudness of the harmonics in loudness units for a full-modulation level of 90 phons.

0·1	—	10	—	8	—	—	—	4 l.u.
0·5	44	600	40	485	1	68	2	360 l.u.
1	240	1500	195	1380	12	330	135	1150 l.u.
2	700	2200	580	2000	79	510	460	1700 l.u.
3	800	4000	770	5300	125	1000	575	4300 l.u.

Full modulation corresponds to 20000 loudness units at 7500 cycles/sec, and 41000 l.u. at 1000 c/s without pre-emphasis. With pre-emphasis these figures are 8600 and 40000 loudness units respectively. With full modulation corresponding to 80 phons the results are slightly more favourable.

From measurements by Waynick [19]) over a distance of 73 km, and by Englund, Crawford and Mumford [20]) over 120 km (sea traject), it follows that one signal is always much stronger over distances of this order. Under special conditions, however, such as occur in large concrete buildings with steel frames, Corrington [14]) found strong echoes at short distances.

Crosby's experiments over very long distances have already been mentioned. In judging the results one may notice that, according to Guy and Morris, a difference between full modulation and noise level of 50 db is necessary for excellent reception, 40 db is sufficient for good reception, 30 db is not satisfactory, and 20 db is decidedly bad. If readability is the only point of interest then, according to Crosby [21]), a difference in level of about 25 db is sufficient. For this purpose smaller audio-frequency bands and lower deviation ratios can be used (the deviation ratio is the ratio of frequency sweep to audio-frequency bandwidth).

Though synchronized transmitters can hardly satisfy the severe demands of broadcasting, they can still be considered for communication purposes, and for this reason we will consider some other details in the last section.

5. *Synchronized transmitters. Intermodulation*

Intermodulation products are the sum and differences frequencies caused in a non-linear system when originally only two frequencies are present. Generally a computation of intermodulation products is more useful than the calculation of harmonics only. In our case two synchronized transmitters are modulated in the same way by two low-frequency signals. The instantaneous frequency of the resultant can be derived in the same way as formula (19):

$$\omega_m = \omega_0 + \Delta\omega_1 \cos pt + \Delta\omega_2 \cos qt. + \sum_1^\infty{}_s \frac{(-1)^{s+1}A^s}{s} \sum_0^\infty{}_l \sum_0^\infty{}_k 2\, a_{lk}(lp+kq) \cdot J_l(sm_1)\, J_k(sm_2) \cdot \sin(lp+kq)\left(t+\frac{\tau}{2}\right) \quad (22)$$

where
$$\begin{aligned} a_{lk} &= -\sin s\omega_0\tau, \text{ if } l+k = 0, \bmod 4, \\ &= \sin s\omega_0\tau, \text{ if } l+k = 2, \bmod 4, \\ &= -\cos s\omega_0\tau, \text{ if } l+k = 1, \bmod 4, \\ &= \cos s\omega_0\tau, \text{ if } l+k = 3, \bmod 4, \end{aligned}$$

$$m_1 = 2\frac{\Delta\omega_1}{p}\sin\frac{p\tau}{2};\quad m_2 = 2\frac{\Delta\omega_2}{q}\sin\frac{q\tau}{2}.$$

It may be noted that the frequencies $ap + bq$ and $aq - bq$ have sweeps proportional to their frequencies, and if $m_1 = m_2$ the same applies to $bp + aq$ and $bp - aq$.

Here again we can put the instantaneous frequency in another form, if $\lim A = 1-0$, corresponding to (20):

$$\omega_m = \omega_0 + \Delta\omega_1 \cos(p\tau/2)\cos pu + \Delta\omega_2 \cos(q\tau/2)\cos qu + \pi\delta\left[\cos\tfrac{1}{2}(m_1\cos pu + m_2\cos qu + \omega_0\tau)\right] \left\{\Delta\omega_1 \sin(p\tau/2)\sin pu + \Delta\omega_2 \sin(q\tau/2)\sin qu\right\}, \quad (23)$$

$$u = t + \tau/2.$$

This makes clear that, with equal amplitudes, distortion can only occur if:

$$\pi - (m_1 + m_2) < \omega_0\tau < \pi + m_1 + m_2 .$$

For a small time difference, $m_1 + m_2 = (\Delta\omega_1 + \Delta\omega_2)\tau$; therefore the product of total frequency sweep and time delay is important.

The computation of intermodulation products with formula (23) is very easy if p and q are proportional to simple numbers, *e.g.* $p = 2s$, $q = 3s$. As an example we take $m_1 = m_2 = 1$. The Fourier analysis gives the amplitude of $\sin ns(t + \tau/2)$ in integral form:

$$C_n = s \int_{-\pi}^{\pi} \delta\{\cos \tfrac{1}{2}(\omega_0\tau + \cos 2x + \cos 3x)\}(2 \sin 2x + 3 \sin 3x) \sin nx \, dx .$$

The peculiarities of δ-functions make it easy to evaluate the integral, as we have only to look for such values of x as make $\cos 2x + \cos 3x + \omega_0\tau = \pi$. Take for instance $\omega_0\tau = \pi$, then the values of x satisfying $\cos 2x + \cos 3x = 0$ are $x = -180°, -108°, -54°$. Beginning at $-180°$ the value of $\cos 2x + \cos 3x$ is 0, then it rises to 0·708, falls to −1·634, and rises again to 2, which is the value for $x = 0$; from there onwards the function is symmetric. We thus find:

$C_1 = 2s\,(\sin 180° - \sin 108° + \sin 54°) = -0{\cdot}284\,s$,
$C_2 = 2s\,(-\sin 216° + \sin 108°) = 3{\cdot}078s$,
$C_3 = 2s\,(-\sin 324° + \sin 162°) = -1{\cdot}794s$,
$C_4 = 2s\,(-\sin 432° + \sin 216°) = -0{\cdot}727s$,
$C_5 = 2s\,(-\sin 540° + \sin 270°) = -2s$, *etc.*

For other values of $\omega_0\tau$ the calculation can be made in an analogous way. Travelling from one station towards the other, in the area where both transmitters give equal fieldstrength, one would find no distortion at all during 37% of the time ($2 - \pi < \omega_0\tau < \pi - 2$). The calculations correspond to a total sweep of 10 kc/sec and a time delay of 33 μsec (path difference 10 km).

We have evaluated the series in formula (22) with the help of Jahnke-Emde's tables [22]), and the results are given in *figs 5* to *12*. Though the maximum values of the distortion are not very much different for $(\Delta\omega_1 + \Delta\omega_2)\,\tau = 1$, 2, or 4, the time during which large distortion occurs changes. In the figures for the even distortion coefficients the curves should be continued anti-symmetrically from π onwards. Although the values of the distortion products can be evaluated from the drawings, the reader may like to have a numerical example. This is given in the following table. The two audio frequencies are 1000 and 1300 cycles/sec. In the first two columns $\Delta\omega_1\tau = \Delta\omega_2\tau = 1$ (*e.g.* $\Delta\omega_1 = \Delta\omega_2 = 10$ kc/sec and $\tau = 16{\cdot}7$ μsec). In the last two columns either the time delay or the sweep is doubled.

If we call $k + l$ the order of the intermodulation product $kp + lq$, we have computed in the table the maximum values of the components up to the fifth order in the band of 4000 c/s. The maxima of different components do not occur at the same place or for the same phase difference.

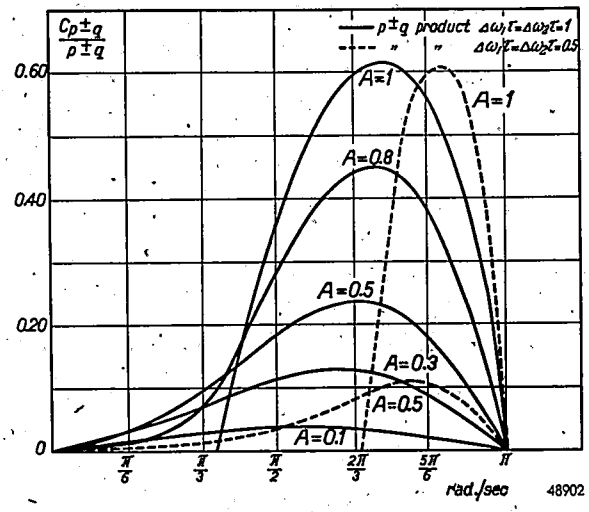


Fig. 5. Intermodulation products with frequencies $p \pm q$. ($\Delta\omega_1\tau = \Delta\omega_2\tau = 1$, or 0·5). Abscissa: $\omega_0\tau$ = product of time difference and carrier frequency. Ordinate: $C_{p\pm q}/(p\pm q)$, where $C_{p\pm q}$ is the sweep of the component of frequency $p \pm q$.

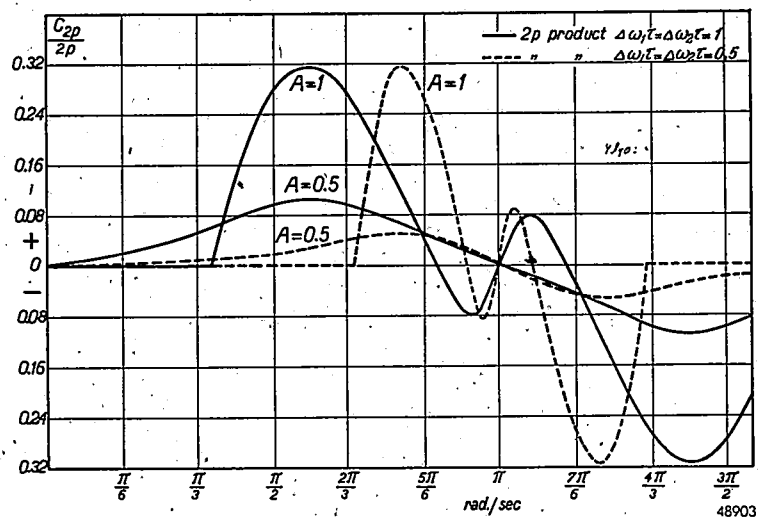


Fig. 6. Intermodulation products with frequencies $2p$ or $2q$. ($\Delta\omega_1\tau = \Delta\omega_2\tau = 1$, or 0·5). C_{2p} = sweep of the component with frequency $2p$.

These intermodulation components will not affect readability too much, if the original components are 10 or even 5 kc/sec. If however in the product $2\omega\Delta\tau$, $\Delta\omega$ is only 2kc/sec and $\tau = 83{\cdot}3$ μsec (25 km), then the distortion will be too large.

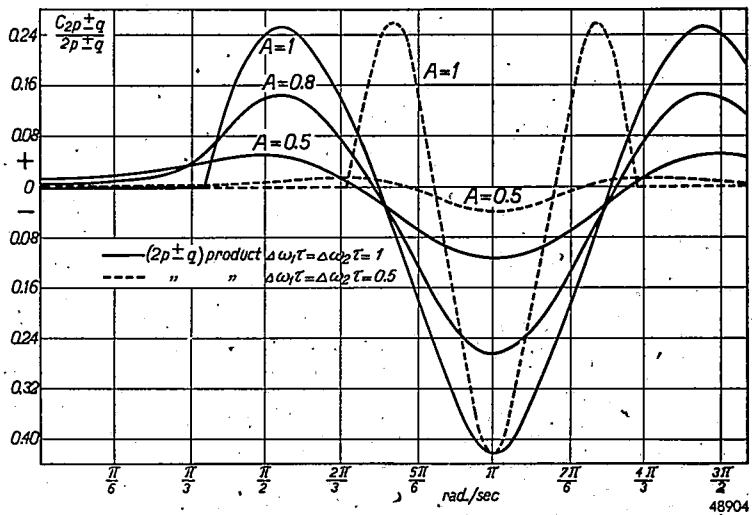


Fig. 7. Intermodulation products with frequencies $2p \pm q$. ($\Delta\omega_1\tau = \Delta\omega_2\tau = 1$, or 0·5).

Intermodulation products in c/sec	$A = 1$	$A = 0{\cdot}5$	$A = 1$	$A = 0{\cdot}5$
300	0·18	0·07	0·18	0·10 kc/sec
400	0·10	0·01	0·10	0·02
600	0·09	0·02	0·09	0·04
700	0·30	0·08	0·34	0·15
1600	0·68	0·18	0·78	0·35
1700	0·12	0·02	0·18	0·07
1900	0·48	0·06	0·48	0·13
2000	0·63	0·21	0·63	0·21
2300	1·42	0·55	1·42	0·75
2600	0·82	0·28	0·82	0·28
2700	0·12	0·01	0·13	0·03
2900	0·19	0·04	0·27	0·11
3000	0·42	0·07	0·36	0·12
3300	1·40	0·38	1·60	0·71
3600	1·53	0·41	1·75	0·78
3900	0·55	0·09	0·47	0·16
4000	0·20	0·02	0·36	0·07

The advantages of frequency modulation are well known, and it therefore seems worth while to consider how difficulties that may arise in the region of equal field strengths of two synchronized transmitters, can best be met. In this region the amplitude of one transmitter should be kept well above the effective local-noise level, and the value of $\Delta\omega\tau$ should be small. Both demands point to a low deviation ratio. If one transmitter is much stronger than the other and has, for instance, the same field at 50 km as the other at 20 km, the low-frequency signal in the weaker transmitter should

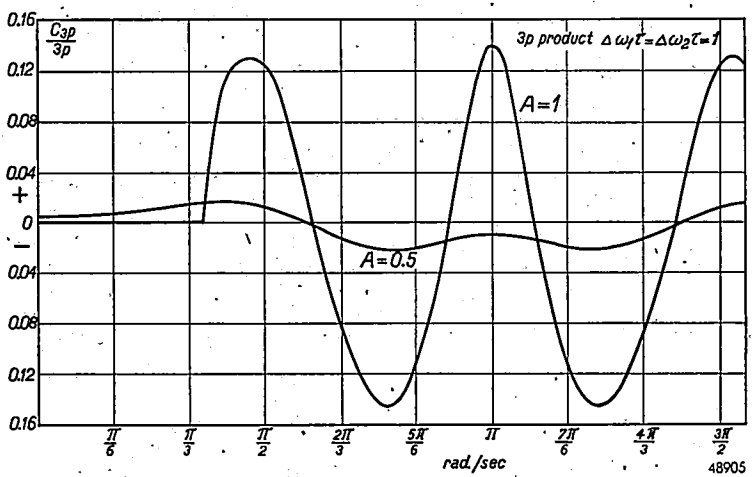


Fig. 8. Intermodulation products with frequencies $3p$ or $3q$. ($\Delta\omega_1\tau = \Delta\omega_2\tau = 1$).

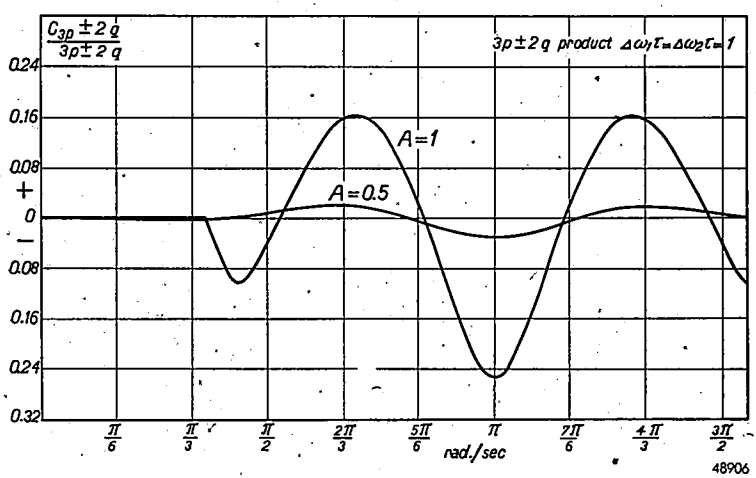


Fig. 9. Intermodulation products with frequencies $3p \pm 2q$. ($\Delta\omega_1\tau = \Delta\omega_2\tau = 1$).

be given a corresponding time delay of 100 μsec. This applies where the coverage of one transmitter is good except for a small area, screened by hills. With equal transmitters feeding from a central point would be the simplest solution. As other systems, *e.g.* the interesting one put forward by Brinkley [23]), require that the product of p and τ must not exceed $\pi/2$,

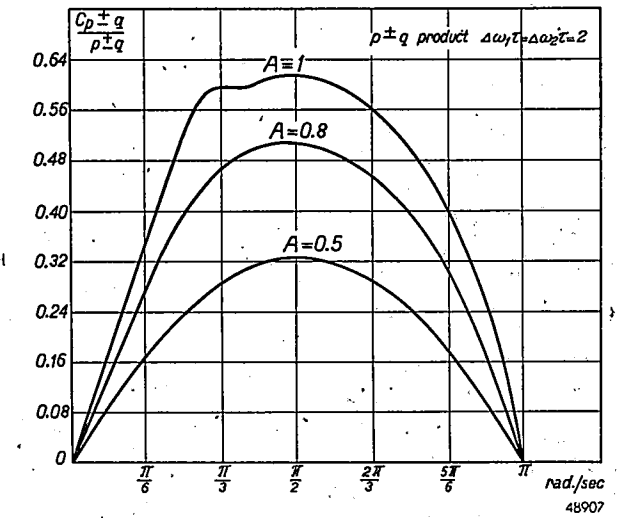


Fig. 10. Intermodulation products with frequencies $p \pm q$. ($\Delta\omega_1\tau = \Delta\omega_2\tau = 2$).

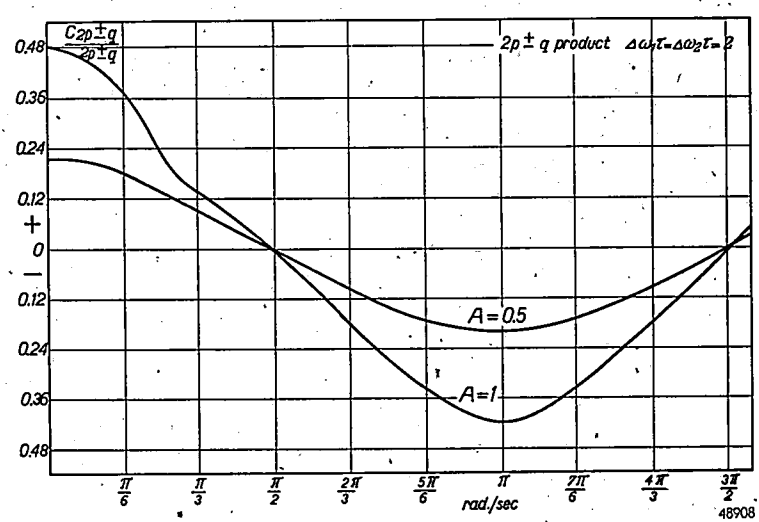


Fig. 11. Intermodulation products with frequencies $2p \pm q$. ($\Delta\omega_1\tau = \Delta\omega_2\tau = 2$).

this equalization seems to be no special disadvantage of frequency modulation. As in ordinary f.m., pre-emphasis can improve the reception as long as there is less energy in the high frequencies than in those around 800 c/s.

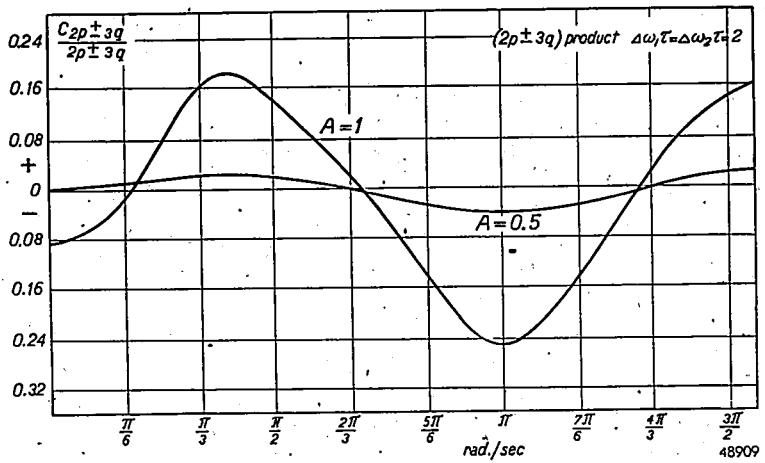


Fig. 12. Intermodulation products with frequencies $3p \pm 2q$. ($\Delta\omega_1\tau = \Delta\omega_2\tau = 2$).

For communications purposes the RC time chosen for broadcasting might be less suitable. We have already seen that a large frequency sweep can be helpful if the transmitters have so much power that local noise is un important, but that this would be a very expensive solution from the point of view of energy dissipation.

The relative merits of systems with two wavelengths, demanding change-over of the receivers, as compared with synchronized systems are not discussed here.

Finally I wish to thank Prof. Balth. van der Pol and Prof. H. Bremekamp for many helpful discussions.

Eindhoven, July 1946

REFERENCES

1) M. G. Crosby, Frequency-modulation propagation characteristics, Proc. Inst. Radio Engrs **24**, 898-913, 1936.

2) M. G. Crosby, Observations of frequency-modulation propagation on 26 megacycles, Proc. Inst. Radio Engrs **29**, 398-403, 1941.

3) T. L. Eckersley, Frequency modulation and distortion, Wireless Engr **7**, 482-484, 1930.

4) E. H. Armstrong, A method of reducing disturbances in radiosignalling by a method of frequency modulation. Proc. Inst. Radio Engrs **24**, 689-740, 1936.

5) I. R. Weir, Field test of frequency and amplitude modulation with ultra-high frequencies, Gen. elect. Rev. **42**, 188-191, 270-273, 1939.

6) N. Guy and H. Morris, N.B.C. frequency modulation field test, RCA Rev. **5**, 190-225, 1940.

7) T. Vellat, Der Empfang frequenzmodulierter Wellen, Elekt. Nach. Tech. **18**, 61-96, 1941.

8) H. A. Wheeler, Common channel interferences between two frequency modulated signals. Proc. Inst. Radio Engrs **30**, 34-50, 1942.

9) H. A. Wheeler, Two-signal cross modulation in a frequency-modulation receiver, Proc. Inst. Radio Engrs **28**, 537-541, 1940.

10) D. A. Bell, F.M. communication systems, interference and propagation characteristics. Wireless Engr **20**, 233-242, 1943.

11) R. Wundt and C. Hoffmann, Ueberlagerungserscheinungen beim Empfang frequenzmodulierter Wellen, Lorenz Berichte April 1941, pp. 63-96.

12) R. Kulp, Spektra und Klirrfaktoren frequenz- und amplitudenmodulierter Wellen, Elekt. Nach. Tech. **19**, 72-81, 96-106, 126-135, 1941.

13) O. E. Keall, Interference in relation to amplitude, phase and frequency modulated systems, Wireless Engr **18**, 6-17, 56-63, 1941.

14) M. S. Corrington, Frequency modulation distortion caused by multipath transmission, Proc. Inst. Radio Engrs **33**, 878-891, 1945.

15) Balth. van der Pol, The fundamental principles of frequency modulation, J. Instn elect. Engrs, part III, **93**, 153-158, 1946.

16) C. Cocci and R. Sartori, Funzione modulate, Alta Frequenza **10**, 67-98, 1941.

17) B. van der Pol, Frequency modulation, Proc. Inst. Radio Engrs **18**, 1194-1205, 1930.

18) B. van der Pol and H. Bremmer, The diffraction of electromagnetic waves from an electrical point source round a finitely conductive sphere with applications to radiotelegraphy and the theory of the rainbow, Phil. Mag. **24**, 141-176, 825-864, 1937; **25**, 817-834, 1938.

19) A. H. Waynick, Experiments on the propagation of ultra-short radio waves, Proc. Inst. Radio Engrs **28**, 468-475, 1940.

20) C. Englund, A. Crawford and W. Mumford. Some results of a study of ultra-short-wave transmission phenomena, Bell Syst. tech. J. **17**, 489-519, 1938.

21) M. G. Crosby, Bandwidth and readability in frequency modulation, RCA Rev. **5**, 363-370, 1941.

22) Jahnke and Emde, Tables of functions, 2nd edition, Leipzig 1933.

23) J. R. Brinkley, A method of increasing the range of u.h.f. communication systems by multi-carrier amplitude modulation, J. Instn elect. Engrs, part III, **93**, 159-176, 1946.

24) H. Fletcher, Loudness, masking and their relation to hearing and noise measurement, J. acoust. Soc. Amer. **9**, 275-293, 1938.

VOL. 2 NO. 3 JUNE 1947

Philips Research Reports

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

R 42 *Philips Res. Rep. 2, 161-170, 1947*

INFLUENCE OF COOLING CONDITIONS ON HIGH-PRESSURE DISCHARGES

by W. ELENBAAS

537.527:536.2

Summary

The influence of cooling on high-pressure discharges in tubes and in free air is theoretically discussed and for the case of discharges in tubes verified experimentally. The agreement between experiment and theory is satisfactory.

The influence of cooling conditions on high-pressure discharges, in closed tubes as well as in the open air, has been investigated experimentally and theoretically in many cases. In this paper we shall deal with some new experiments in the case of closed tubes, and with the theory regarding closed tubes as well as open arcs.

I. *High-pressure discharges in closed tubes*

A. *Experiments*

a. Cooling of the discharge by adding a gas of a high thermal conductivity has been investigated before in the case of the high-pressure mercury arc [1]). The added gases, He, Ne and A are practically not excited (as a result of their high excitation potentials), so that they only diminish the mean free path of the electrons and increase the coefficient of heat conduction of the gas. As the mercury radiation decreases at increasing percentage of the rare gas and for constant pressure of the rare gas decreases in the order A, Ne, He, we concluded that the temperature of the discharge decreased at increasing heat conductivity of the mercury-rare gas mixture. The behaviour of the electric gradient was in accordance with this decreasing temperature. Addition of 16 atom per cent of He to the mercury vapour caused a decrease of 4% of the temperature of the discharge.

b. Apart from a single observation that the gradient increases at increasing temperature of the tube wall [2]), no experiments are known about the influence of cooling or heating of the tube wall. We therefore carried out some measurements to ascertain the influence of the tempe-

rature of the tube wall on the high-pressure mercury discharge. As we shall see further on, the influence on the radiation is practically nil, whilst the influence on the electric gradient is measurable. We therefore measured the latter.

A tube having an inner diameter of 22 mm, an over-all length of 32·2 cm and a distance between the electrodes of 30·5 cm, contained 130 mg of Hg and A at a pressure of 2·25 mm. The tube was placed in an oven and operated at 30 W/cm (D.C.) in the vertical position. All the mercury is then vaporized. Half-way between the electrodes a *Chromel-Alumel* thermo-element of 200 μ was bound on the tube wall by means of a wire of 150 μ. The discharge was operated during some hours at maximum temperature of the furnace after which the arc voltage was constant. The furnace temperature was then gradually lowered and the temperature of the tube wall and the arc voltage read at intervals. The current was varied slightly in order to keep the input constant at 30 W/cm. After reaching the lowest temperature of the tube wall, the temperature of the furnace was increased again and readings were taken at increasing temperature. In *fig. 1* the variation of the arc voltage is shown as a function of the temperature of the tube wall, for two tubes of practically the same dimensions. The readings at decreasing temperature are marked × and those at increasing temperature ○.

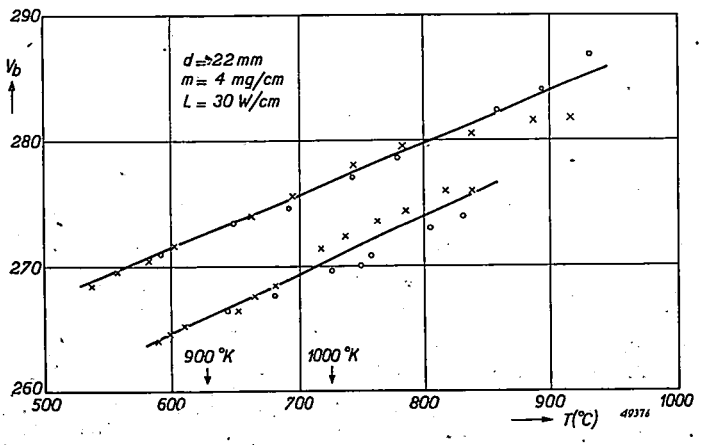


Fig. 1. Arc voltage of two high-pressure mercury discharges as a function of the wall temperature.
○ increasing wall temperature.
× decreasing wall temperature.

Assuming 15 volts for anode and cathode fall, we find for these two tubes an increase of the gradient of 1·7 and 2·0 % respectively for an increase of 100° of the temperature of the tube wall.

B. *Theory*

a. The case where the cooling is increased by an increase of the heat conductivity of the gas mixture is sufficiently treated in [1]).

b. The influence of the temperature of the tube wall on the high-pressure mercury discharge was considered by Schmitz[3]), who used a graphical solution of the differential equation that governs the high-pressure discharge if operated in a tube. At constant mean density (which also applies to our experiments) he finds for an increase of 100° of the temperature of the tube wall a decrease of 1% of the temperature in the axis and an increase of $2^1/_2$% of the electric gradient (fig. 10 in [3])). This increase of $2^1/_2$% agrees well with the measured value of 1·7 - 2% as found above, especially if one takes into account the elaborate graphical integration of the differential equation.

It is possible to calculate the influence of the wall temperature without using the Schmitz solution, and as this method gives a better insight as to why the temperature of the arc increases at decreasing wall temperature, we shall give these considerations here. We shall distinguish between two cases: constant pressure and constant mean density.

1. *Constant pressure*

For our purpose the temperature course from the axis to the wall may be supposed to be parabolic: $T = T_0 - A\,r^2$. Now we suppose two cases 1 and 2 with the same tube but with wall temperatures of 900 and 1000 °K respectively and equal axis temperatures. With $T_0 \approx 5500$ °K, we have $A_1/A_2 = 4600/4500 = 1{\cdot}022$. We have the same temperature in case 1 at a distance r_1 from the axis as in case 2 at a distance r_2 if: $A_1 r_1^2 = A_2 r_2^2$, or if:

$$r_2/r_1 = (A_1/A_2)^{1/2} = 1{\cdot}011\,. \qquad (1)$$

The discharge 2 thus can be derived from discharge 1 by a broadening of 1·1%. That the tube diameter of 2 is not 1·1% larger than that of 1 is not important, because the zones of low temperature near the wall do not contribute to the discharge. As regards the energy balance of discharge 2 compared with that of 1, the radiation of 2 will be larger because the zones $2\pi r_2\,dr_2$ and $2\pi r_1\,dr_1$, have the same temperatures and the ratio of their areas is as r_2^2 to r_1^2. The total radiation for the discharge 2 is thus 2·2% (eq. (1)) higher than for the discharge 1. Apart from the radiation loss, the discharge has an energy loss by conduction of heat, amounting to about 10 watts per cm of arc length [4]). Will this energy loss be the same for our two discharges 1 and 2? In *fig. 2* we have drawn the two temperature curves $T = T_0 - A\,r^2$ with A_1 and A_2 according to (1). As $dT/dr = -2\,Ar$, we find for the energy conducted to the wall in case 2: $W_2 = -2\pi R\lambda(dT/dr)_R = 4\pi R^2 A_2 \lambda$. In case 1 the temperature at P is the same as

the wall temperature in 2 if $r_1 = R\,(A_2/A_1)^{1/2}$ (eq. (1)). As the temperatures are the same, the coefficients of heat conduction are likewise the same, so that we find:

$$\frac{(W_2)_R}{(W_1)_P} = \frac{R^2\,A_2}{R^2(A_2/A_1)\,A_1} = 1{\cdot}0\,. \qquad (2)$$

The heat conducted through the cylinder at P in case 1 is thus equal to the heat conducted to the wall in case 2, and as between P and the wall no heat is developed, the heat arriving at the wall in case 1 is the same as that in case 2.

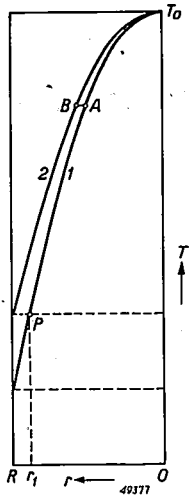


Fig. 2. Parabolic temperature-distribution curve for equal axis temperatures and different wall temperatures.

One might imagine that these equal conduction losses are a consequence of the parabolic course of the temperature. In the space between discharge and wall the actual course of the temperature is determined by:

$$W = -2\pi\,r\,\lambda(T)\,\mathrm{d}T/\mathrm{d}r\,. \qquad (3)$$

The dependence of λ on temperature has been estimated in a previous paper [5]) as:

$$\lambda(T) = \lambda_0\,T^{9/8}. \qquad (4)$$

From (3) and (4) we obtain:

$$W = \frac{2\pi\,\lambda_0}{17/8}\,\frac{T(r)^{17/8} - T(r')^{17/8}}{\ln\,(r/r')}. \qquad (5)$$

Assuming that the edge of the discharge lies at 4000° and that in case 1 this temperature is reached at $r = \frac{1}{2}\,R$ (point A in fig. 2), this temperature

is reached in case 2 at B, for which $r_2 = \frac{1}{2} \times 1{\cdot}011\ R$. As the wall temperatures in case 1 and 2 are 900 and 1000° respectively, we obtain:

$$\frac{W_2}{W_1} = \frac{4000^{17/8} - 1000^{17/8}}{4000^{17/8} - 900^{17/8}} \cdot \frac{\ln 0{\cdot}5}{\ln 0{\cdot}5055} = 0{\cdot}989 \times 1{\cdot}016 = 1{\cdot}005\,. \qquad (6)$$

Although we do not find exactly the same conduction loss for the two cases as indicated by (2), the difference is so small as to be negligible, the more so as the conduction loss is only a small part of the total input (at 30 W/cm only one third).

The conduction losses of the two discharges being the same, the temperature of discharge 2 has to decrease to such an extent that the 2·2% increase of radiation caused by the broadening of the discharge is compensated. The radiation of the two discharges is then again the same and the energy balances are satisfied at equal inputs. The influence of the wall temperature on the radiation is thus practically zero in contrast with case a where the conduction loss increased by the increasing heat conductivity and as a result the radiation decreased.

To reach energy balance again, the temperature of our discharge 2 has to decrease so far that the radiation loss decreases by 2·2%. (The variation of the conduction loss for this small temperature variation may be ignored). We can easily calculate this temperature decrease. The total radiation S may be written as:

$$S = C \int_0^R (pr/T) \exp(-11600\ V_m/T)\, \mathrm{d}r\,, \qquad (7)$$

where V_m represents the height of a hypothetical mean level at 7·8 volts above the ground level [6]). Denoting the effective temperature of the discharge giving a radiation, which is 2·2% too large, by T_2 and that giving the right radiation by T_2', we have, as the pressures are the same and the variation of T is small compared with the variation of the e-power:

$$\exp\left\{-11600 \times 7{\cdot}8 \left(\frac{1}{T_2} - \frac{1}{T_2'}\right)\right\} = 1{\cdot}022\,,$$

from which we obtain:

$$T_2 - T_2' = \Delta T = 2{\cdot}5 \times 10^{-7}\ T_2\ T_2'.$$

With $T_0 = 5500$ °K the effective temperature is approximately 5250, so that $\Delta T \approx 7°$. The decrease of the axis temperature will be about the same. The effect caused by 100° difference in tube-wall temperature is thus of the order of 0·15% in arc temperature.

As already stated we cannot observe this temperature variation in the radiation. The total radiation remains the same and the variation of the distribution over the spectral lines by a difference of 0·15% in temperature

is so small as to be undetectable. Apart from the broadening of the discharge by 1%, which will also be very difficult to observe, the electric gradient seems the only quantity in which the change of temperature may be manifested. We shall calculate this change of the gradient. Denoting the current by i, the gradient by G and the degree of ionization by x, we have:

$$i \backsim G \int_0^R rx\, T^{-1/2}\, \mathrm{d}r\,. \tag{8}$$

Multiplying both sides by G and remembering that $iG = L$ is constant, we obtain:

$$G \backsim \frac{1}{\sqrt{\int_0^R rx\, T^{-1/2}\, \mathrm{d}r}} \backsim \frac{p^{1/4}}{\sqrt{\int_0^R rT^{3/4} \exp(-\varepsilon_i/2kT)\, \mathrm{d}r}} = \frac{p^{1/4}}{\sqrt{\int_0^R rT^{3/4} \exp(-60000/T)\, \mathrm{d}r}} \tag{9}$$

The degree of ionization x was calculated from the Saha equation.

As the pressure is kept constant, we have in (9) only to consider the variation of the denominator with varying wall temperature. We compute the variation of G by first considering the variation due to the broadening of the discharge by 1·1% and after that we take into account the decrease of the temperature by which the radiation decreases by 2·2%. The broadening of the discharge causes the integral to increase by 2·2%, so that as a result G decreases by 1·1%. The decrease in temperature results in a smaller increase of G than the decrease of S. According to (7) the radiation is approximately proportional to $\exp(-90000/T)$, whereas the gradient according to (9) is about proportional to $\exp(30000/T)$. We thus estimate that the temperature decrease causing a radiation decrease of 2·2%, causes an increase of the gradient of $\frac{1}{3} \times 2{\cdot}2\% = 0{\cdot}7\%$. Together with the decrease of 1·1% found above, we thus obtain a decrease of 0·4% in the gradient by an increase of 100° in wall temperature. As it is very difficult to keep the pressure exactly constant, it will hardly be possible to verify this decrease experimentally. Comparing with the calculations of Schmitz [3]), we there find in fig. 7 for the gradient a decrease of 16% for an increase of 1200° in wall temperature or 1·3% per 100°. The temperature decrease found by Schmitz amounts to 6% for the same increase or $\frac{1}{2}$% per 100°. Both variations calculated by Schmitz are thus about 3 times larger than those calculated above. As experiments, as stated already, are very difficult at constant pressure, we will make the calculations for constant mean density, because we have experimental data for this case (see A/b).

2. *Constant mean density*

As the pressure occurs in (7) and (9), we have to know the variation of the

pressure when varying the temperature of the tube wall. If m means the quantity of gas per cm of tube length, we have:

$$p \sim \frac{m}{\int_0^R (r/T)\,\mathrm{d}r},$$

or with the parabolic temperature curve:

$$p \sim \frac{m}{\int_0^R r\,\mathrm{d}r/(T_0 - Ar^2)} = \frac{m(T_0 - T_w)}{R^2 \ln T_0/T_w}. \qquad (10)$$

For case 1 we have $T_w = 900$ and $T_0 = 5500$, whereas in case 2 : $T_w = 1000$ and as a first approximation $T_0 = 5500$. From (10) we obtain:

$$p_2/p_1 = 1{\cdot}039\,.$$

As a result of this higher pressure the radiation in 2 is according to (7) 3·9% higher. As the discharge 2 likewise as in the previous section is moreover 1·1% broader, the radiation is already for this reason 2·2% higher, so that a total of 6·1% increase of radiation results. As the conduction loss is independent of the pressure and we found a constant heat loss in the previous section, we have the same heat loss in the cases 1 and 2, so that in case 2 the temperature must decrease to such an extent that the radiation decreases by 6·1%.

In analogy to the previous section, we thus have:

$$\exp\left\{-90000\left(\frac{1}{T_2} - \frac{1}{T_2'}\right)\right\} = 1{\cdot}061$$

or

$$T_2 - T_2' = \Delta T = 6{\cdot}5 \times 10^{-7}\, T_2\, T_2' \approx 18°\,.$$

If we take this decrease of 18° into consideration for the calculation of the pressure with (10), we find:

$$p_2/p_1 = 1{\cdot}037\,.$$

To reach energy balance the radiation must thus decrease by 3·7 + 2·2 = 5·9%, from which we obtain: $\Delta T \approx 17\frac{1}{2}$ or about 0·3%.

For the gradient of case 2 compared with case 1, we obtain:

a. Through the variation of p according to (9) an increase of 3·7%/4 = 0·92%.

b. Through the variation of T_1 so that the radiation decreases 5·9%, an increase of $\frac{1}{3} \times 5{\cdot}9\% = 1{\cdot}97\%$.

c. Through the broadening of 1·1% of the discharge according to (9) a decrease of 1·1%.

The result is thus an increase of 1·8% in the gradient for an increase of 100° in wall temperature. This is in excellent agreement with the values of 1·7 and 2% found in our experiments (section A/b). Schmitz [3]) (fig. 10) obtains a decrease of 1% for the temperature (thus again a variation three times larger than that calculated by us) and an increase of $2\frac{1}{2}$% of the gradient in rather good agreement with our calculations and measurements.

II. *Convection-stabilized discharges* [7])

Suppose the discharges 1 and 2 have the same input L per cm of length, but the cooling of 2 is better.

A. The cooling of 2 is better because the coefficient of heat conduction (λ) is larger, the other properties of the gases being the same.

We introduce a discharge 2′ with the same cooling conditions as 2, but an input L', so that 2′ is similar to 1. Now we found [8]) that discharges under free convection in gases of which the coefficient of heat conductivity is different, are similar if the inputs are proportional to the coefficients of heat conductivity. We thus have:

$$L'/L = \lambda_2/\lambda_1 . \tag{11}$$

As $\lambda_2 > \lambda_1$, we have $L' > L$. Now we diminish the input of 2′ to that of 2 and see how the temperature changes. As its temperature was originally the same as that of 1, the temperature of 2 will be lower if a decreasing input results in a decreasing temperature. According to measurements of Suits [9]) this is indeed the case. Without these direct measurements we would expect this for the two following reasons:

1. Mannkopff [10]) found the diameter of the discharge proportional to L. As practically all the energy is carried away by conduction and convection this means that the temperature gradient on the cylinder which marks the diameter of the discharge is constant. If therefore in *fig. 3* curve *a* represents the temperature curve of a certain discharge, we know that a discharge

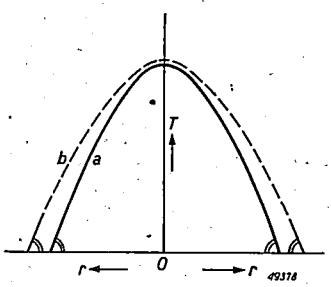


Fig. 3. If *a* represents the temperature-distribution curve of a convection-stabilized discharge, the discharge with a larger input has a larger diameter, whereas dT/dr at the edge is the same. The temperature at the axis for the latter discharge is therefore probably higher.

having a larger input has also a larger diameter but the same slope at the edge. It is therefore very likely that the temperature curve b will run as given by the dashed curve and the axis temperature will thus be higher at a larger input.

2. As the diameter δ of the discharge is proportional to L and the radiation of similar discharges is proportional to δ^2, we obtain for similar discharges the radiation proportional to L^2, so that the radiation per watt becomes proportional to L. If the temperature should rise with L, the efficiency η must rise as L^a with $a > 1$. This appears to be the case in a large region. Ornstein, Brinkman and Ter Horst [11]) found $\eta \backsim i^{1\cdot2}$, for $i < 15$A. As the current-voltage characteristic of the discharge is negative $G \backsim i^{-\beta}$, so that $L = iG \backsim i^{1-\beta}$ or $i \backsim L^{1+\gamma}$. Instead of $\eta \backsim i^{1\cdot2}$ we may therefore write $\eta \backsim (L^{1+\gamma})^{1\cdot2}$. We thus find $a > 1$, so that the temperature must have increased with increasing input. Holm and Lotz [12]) found at higher currents (100-2500 A) a slower increase of the efficiency or even a small decrease. As it is not known whether the diameter of the discharge in this region is still proportional to the current, we cannot draw conclusions for these high-current discharges. Of course it is impossible that η continues to increase proportionally to the input as soon as the radiation becomes an important part of the input. At moderate currents we may thus assume that the temperature rises with the input (at constant pressure). The temperature of discharge 2 is thus lower than that of 2′ and as 2′ was similar to 1, the temperature of 2 is lower than that of 1. Cooling of the convection-stabilized discharge by enlargement of the coefficient of heat conduction thus results in a decrease of the arc temperature.

B. *The discharge 2 is cooled by forced convection*

If the velocity of the forced convection is so high that the natural convection may be ignored, we found [8]) that if the radiation loss is small, the discharge remains similar at varying velocity and constant pressure when the electrode distance is varied proportionally to v^{-1}. The diameter then likewise varies proportionally to v^{-1}. If the electrode distance has no influence on the temperature and if the radiation loss may be disregarded, we thus find a constant temperature at varying velocity. This is contradictory to the statement of Kesselring [13]) that the temperature of the discharge is increased by cooling. This was concluded from the minimum theory and the experimental fact that the diameter of the discharge is decreased by cooling [14]). The decrease of the arc diameter by cooling is in agreement with our calculations. Whether the exact relation between diameter and velocity is satisfied ($\delta \backsim v^{-1}$) is not known.

Summarizing we may thus state that it is impossible to say in general whether cooling of a discharge at constant input increases or decreases the

arc temperature. An increase of the heat conductvity of the gas results in a decrease of the arc temperature both for the discharges in closed tubes and for the convection-stabilized discharges. Cooling of the tube wall causes a very small increase in arc temperature at constant pressure and a larger increase at constant average density. Variation of the forced convection has, in the first approximation, no influence on the arc temperature.

Eindhoven, October 1945

REFERENCES

1) W. Elenbaas, Physica **3**, 219, 1936.
2) W. Elenbaas, Physica **2**, 173, 1935.
3) G. Schmitz, Phys. Z. **45**, 65, 1944.
4) W. Elenbaas, Physica **4**, 413, 1937.
5) W. Elenbaas, Philips Res. Rep. **1**, 314, 1946.
6) W. Elenbaas, Physica **2**, 757, 1935.
7) W. Weizel, Z. techn. Phys. **24**, 90, 1943.
8) W. Elenbaas, Philips Res. Rep. **1**, 339, 1946.
9) C. G. Suits, J. applied Phys. **10**, 728, 1939.
10) R. Mannkopff, Z. Phys. **120**, 228, 1943.
11) L. S. Ornstein and H. Brinkman, Physica **1**, 821, 1934.
12) R. Holm and A. Lotz, Wiss. Ver. Siemens W. 13 II 87, 1934.
13) F. Kesselring, E.T.Z. **55**, 92, 116, 165, 1934.
14) F. Kesselring and F. Koppelmann, Elektrot. Z. **57**, 1498, 1936.

 Philips Res. Rep. 2, 171-176, 1947

ON THE ACTIVATION OF OXIDE-COATED CATHODES

by H. C. HAMAKER, H. BRUINING and A. H. W. ATEN Jr

621.385.1.032.216

1. *Introduction*

It is a common experience that, to obtain a satisfactory emission from an oxide-coated cathode, the outgassing of the bulb and the various parts inside it must be carried out in a special order and the correct pumping procedure has to be found out for each type of tube by a separate series of experiments. Apparently, during outgassing, various factors are in operation which afterwards influence the emission produced by the cathode.

One of these factors has been detected and has been made the subject of a special investigation which finally led to the following conclusions:

a) Glass heated to 400 °C evolves a small amount of hydrochloric acid.
b) In a vacuum tube with an oxide cathode this HCl reacts with the carbonate or the oxide on the cathode to give Ba- or Sr-chloride.
c) When the cathode is subsequently heated these chlorides evaporate and condense on the grid and the anode.
d) Under electron bombardment these chlorides decompose, thereby producing chlorine atoms or positive chlorine ions *) which poison the cathode emission in a very pronounced way.

2. *Initial observations*

We used the type of cathodes as applied in cathode-ray tubes (*fig. 1*). The surface *C* is sprayed with Ba-Sr carbonate and the cathode is heated by an isolated tungsten spiral *S*. The relation of the voltage applied and the temperature of the cathode observed by means of a thermocouple was as follows:

Voltage	3	4	5	6	volts
Temperature	770	900	1015	1105	°C

Opposite the cathode a nickel sheet anode *A* was mounted and, with the addition of a Ba-getter, the assembly was sealed in a radio-valve bulb of normal dimensions.

*) Since the electron affinity of a chlorine atom is about 4 eV and its ionization potential about 13 eV the formation of a positive ion from an ionic crystal of BaCl or SrCl is not very likely. The poisoning effect was observed when the impinging electrons had an energy of not more then 10 eV.

On the pump the bulb was first baked out at 400 °C for 30-60 minutes; subsequently the anode was outgassed by high-frequency induction and then the carbonates decomposed by applying up to 6 volts to the heating element of the cathode.

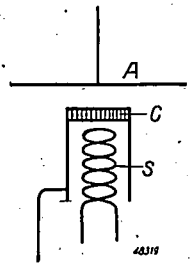


Fig. 1. Arrangement used.

Now it was observed that during this procedure, when the decomposition of the carbonates was practically completed, a minute quantity of some colourless substance evaporated from the cathode, forming on the anode a stain recognizable by a series of newtonian interference rings. This stain cannot be due to evaporation of Ba- or Sr-oxide since this requires a much higher temperature. Besides the rate of evaporation of BaO and Ni are approximately the same so that, if the cathode is heated to a sufficiently high temperature, a mixture of metal and oxide will be deposited on the anode giving a stain steel blue in colour with a single brownish rim. This, however, requires half an hour's heating at 6 volts whereas the formation of the colourless stain was complete in 5 minutes at 5 volts.

This colourless stain recognizable by interference rings will in the following simply be designated as the "stain".

After degassing on the pump at 6 volts the cathode is as a rule in a fairly active state, the anode current being for instance 10 mA at an anode voltage V_a of 5 volts. We observed, however, that when the anode voltage is further increases I_a abruptly dropped to a value of 0·1 or 0·2 mA as soon as V_a had reached a value of $\pm$ 10 volts as illustrated in *fig. 2*.

That this phenomenon is a direct consequence of the stain on the anode could be demonstrated by a tube with two different anodes which could be interchanged (*fig. 3*). While on the pump, a stain was formed on anode A_1, and subsequently activation was carried out opposite the clean anode A_2. When a final steady state had been reached the anode current would be, say, 20 mA at $V_a = 20$ volts, no drop in emission being observed on passing the critical value of 10 volts. When under these circumstances, keeping V_a

at 20 volts, the anodes where interchanged, the emission fell almost instantaneously to something like 1 mA or even less. By switching back to anode A_2 it could be shown that this drop in the emission was actually due to poisoning of the cathode. The emission opposite A_2 was low at first but by a process of activation lasting about 20 minutes the cathode would gradually recover its initial state of activity.

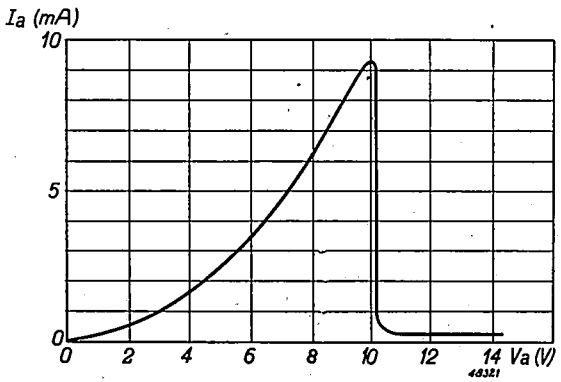


Fig. 2. Schematic representation of the anode current, I_a, when the voltage, V_a, is steadily increased.

Continued experiments brought to light that the "poisoning voltage" ($V_a = 10$ volts) was independent of the current density whereas the rate of poisoning is strongly dependent on this factor. If by lowering the cathode temperature the emission is brought down to 1 μA, poisoning opposite the stain can still be observed but now as a slow and gradual decline in the anode current.

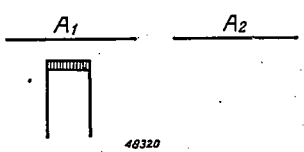


Fig. 3. Arrangement with two interchangable anodes.

The conclusions to be drawn from these observations is that the substance of the stain decomposes under electron bombardment, producing atoms (or positive ions) which poison the cathode emission. Evidently this decomposition sets in when the energy of the bombarding electrons has reached a value of 10 V.

3. *The stain and its origin*

The following experiments were successively carried out to detect the nature of the processes to which the formation of the stain is due.

a) A cathode was thoroughly degassed in vacuum at a heating tension of 8 volts, the oxides afterwards being reconverted into carbonates by letting in CO_2. This preliminary treatment did not, however, prevent the formation of the stain when the cathode was subsequently used in a normal tube as shown in fig. 1. The stain does not therefore result from impurities in the carbonates.

b) An important step forward was the discovery that formation of a stain only occurred when the bulb had been baked out at 400 °C but not after baking out at 200 °C. This suggests that the stain is produced by a chemical reaction of the carbonates with a gaseous substance evolved from the glass envelope at higher temperatures.

s) At the very outset endeavours to determine the chemical composition of the stain were not succesful. At a later stage it was noted, however, that the stain did not show chemical reactions when exposed to the atmosphere and was moreover soluble in water. As it was highly likely that the stain would consist of a Ba- or Sr-compound these observations strongly suggested Ba- or Sr-chloride, a hypothesis which was corroborated by chemical analysis. The presence of chlorine could be demonstrated by microanalysis.

This combination of observations has led to the following theory.

A common process for the production of hydrochloric acid consists in heating a mixture of NaCl and SiO_2 in the presence of water vapour

$$2\,NaCl + SiO_2 + H_2O = 2\,HCl + Na_2SiO_3$$

Similar reactions take place when the SiO_2 is replaced by Al_2O_3, Fe_2O_3, or various other oxides [1]). On a production scale these reactions occur at an appreciable rate at red to dull-red temperatures so that it is quite conceivable that under the much more sensitive conditions of our experiments they will be perceivable at 400 °C. Besides information received from the Philips Glass Factory confirmed that the Na_2CO_3 used in the manufacture of glass as a rule contains about 1% NaCl as impurity. Consequently reactions of the type mentioned may easily take place in glass when baked out at 400 °C and will lead to the evolution of a certain amount of HCl. This HCl reacts with the carbonates of the cathode and in this way produces the stain phenomenon.

These suppositions have been confirmed by various other experiments. It was ascertained that $BaCl_2$ actually evaporates at the same temperature at which the formation of the stain was observed and that the arti-

ficial stain thus prepared had an identical poisoning effect on the emission of the oxide cathode.

With the co-operation of the Philips Glass Factory chlorine-free glass was prepared and in bulbs of this glass the stain did not occur even after baking at 400 °C. Preparation of such a glass did not prove, however, a very simple matter.

The stain can also be avoided by taking care that during the baking-out procedure a sufficient amount of a HCl absorbent is present in the bulb. As such, a large patch of Ba-carbonate somewhere in the bulb proved effective. Likewise evaporation of some of the Ba-getter before baking out, or rinsing the interior of the bulb with KOH solution before sealing-in, prevented the stain.

It should incidentally also be noted that the phenomena described above explain the relative abundance of chlorine ions in the negative-ion spectrum of cathode-ray tubes [2]).

4. *The "poisoning potential"*.

In a separate series of experiments films of BaF_2, CaF_2, NaCl, *etc.*, evaporated onto the anode, were found to poison the emission in exactly the same manner as described above. With a tube of the construction indicated in fig. 1, the drop in the emission occurs at a sharply defined potential, the "*poisoning potential*". The question arose whether this poisoning potential could not be used as a measure of the "*critical potential*" at which decomposition of a substance under electron bombardmet sets in. This, however, is not the case.

As mentioned above, the poisoned cathode reactivates of its own accord when brought opposite a clean anode. Hence the rate of emission opposite a stain must be sonsidered as a dynamic equilibrium between the natural tendency of the cathode to reactivate and the rate at which poisonous material reaches the cathode surface. If this view is correct, we must expect the poisoning voltage to be higher, the higher the temperature of the ca- and the better its state of activation. Shifts in the poisoning voltage of 1 or 2 volts and in the direction expected were actually observed when these conditions were varied. Likewise the poisoning voltage increased and became less pronounced when the distance between cathode and anode was increased, a manipulation which obviously reduces the amount of poison incident on the cathode.

These difficulties might perhaps be partly overcome by working at very low temperatures and correspondingly low current densities. Even then, however, the problem remains to determine the contact potential of the anode which in view of the inhomogeneous contamination due to the stain will be practically impossible in our case.

In this connection it should be noted that Jacobs [3]) has recently made use of the stain phenomenon to determine the critical potential of various metal oxides. His observations where actually carried out at low current densities and on homogeneously oxidized pieces of metal, the contact potential being determined by the method of incipient anode current. According to Jacobs the critical potential equals the heat of formation.

Eindhoven, July 1946

REFERENCES

1) Gmelin-Kraut's Handbuch System No. 6, 99, 1927.
2) C. H. Bachman and D. W. Carnahan, Proc. Inst. Radio Engrs **26**, 529, 1938.
L. F. Broadway and A. F. Pearce, Proc. phys. Soc. (A) London, **51**, 335-348, 1939.
3) H. Jacobs, J. appl. Phys. **17**, 596-603, 1946.

Philips Res. Rep. 2, 177-182, 1947

PHOTOLUMINESCENCE IN THE QUATERNARY SYSTEM $MgWO_4$ - $ZnWO_4$ - $MgMoO_4$ - $ZnMoO_4$

by F. A. KRÖGER

535.37:541.123.6

Summary

In the quaternary system (Mg, Zn) (W, Mo)O_4 four different crystal structures appear. The fluorescence and absorption of products of these structures are studied.

Introduction

Nearly all tungstates and molybdates show photoluminescence upon excitation by an appropriate wavelength and at a sufficiently low temperature. The properties of the fluorescence are mainly determined by the tungstate and molybdate groups, and are only secondarily affected by the cations and the crystal structure. The influence of the structure can most favourably be studied in pure systems which may exist in several modifications, *e.g.* $MgWO_4$, whereas the influence of other factors can be determined in a series of solid solutions in which the structure remains the same; to investigate the influence of the cations, for instance, mixed crystals of the type $(A,B)XO_4$ may be used, in which A and B stand for different cations, and X for W or Mo. Finally, the specific character of the tungstate and molybdate can be demonstrated with the aid of solid solutions $A(W,Mo)O_4$. Having investigated the systems $Ca(W,Mo)O_4$, $(Ca,Pb)WO_4$, $(Sr,Pb)WO_4$ and $(Ba, Pb)WO_4$ in a previous paper [1]) we shall now report on some experimental studies of products belonging to the quaternary system $(Mg,Zn)(W,Mo)O_4$.

The phase diagram

Pure $MgWO_4$ is known to exist in three different modifications. A high-temperature α-phase is stable above 1250 °C, a second phase (β) is stable below this temperature while a third modification is obtained by synthesizing $MgWO_4$ at about 300 °C *) [2]). The structure of $ZnWO_4$ is identical with that of β $MgWO_4$. The structures of $ZnMoO_4$ and $MgMoO_4$ on the other hand, though showing some resemblance, are not identical; they differ markedly from the tungstates.

*) A fourth modification found by Hill, Dunning and Megaw [2]) is probably the α-modification, which is instable at this temperature, but which may be formed in consequence of Ostwald's law of successive states.

Products belonging to the quaternary system have been prepared by heating at 900 °C for 1-2 hours a mixture of MgO, $ZnCO_3$, H_2WO_4 and H_2MoO_4 in proportions corresponding to the stoichiometric formula AXO_4 of the final product (in which A stands for Mg,Zn, and X for W,Mo). *)

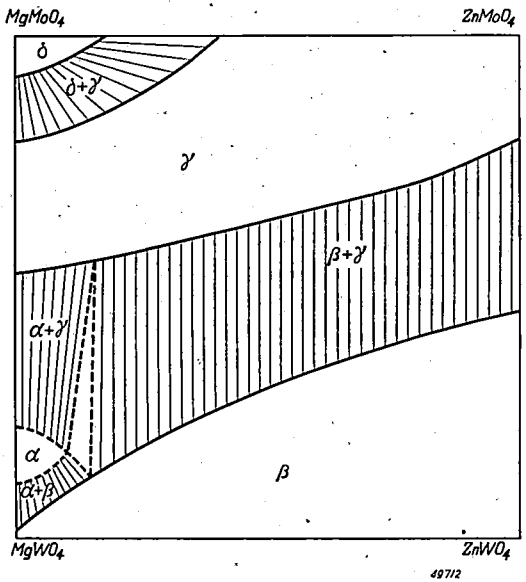


Fig. 1. Phase diagram of the quaternary system (Mg, Zn) (W, Mo)O_4 at 900 °C.

X-ray investigations revealed that in the quaternary system (Mg,Zn) (W,Mo)O_4 four different crystal structures appear. For a temperature of 900 °C the approximate boundaries of the various regions of solid solutions are shown in *fig. 1*, in which uncertain parts have been indicated by dashed lines. In the α-region the structure is that of the high-temperature modification of $MgWO_4$: obviously the transition point $\alpha \rightleftarrows \beta$ is lowered by the incorporation of molybdate; $MgWO_4$ and $ZnWO_4$ can be homogeneously mixed in the β-phase, in which, moreover, the tungstate can be replaced by molybdate up to a fraction increasing towards the zinc side. Another large region of solid solutions is covered by the γ-phase which has the structure of zinc molybdate. The magnesium molybdate structure covers a small range of compositions (δ) . X-ray diffraction lines of products of the system Mg(W,Mo)O_4 with the structures α, β, γ and δ are shown in *fig. 2* with estimated intensities.

*) The preparations have been carried out by Mr A. Bosman.

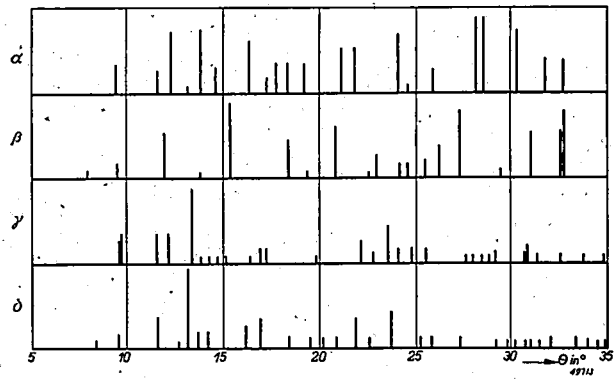


Fig. 2. X-ray diffraction patterns for the α-, β-, γ- and δ-structures of the system $Mg(W,Mo)O_4$. (Cu, Kα radiation).

Photoluminescence, absorption and excitation

For various products of the quaternary system the luminescence effects for excitation by short-wave ($\lambda = 2537$ Å) and long-wave ($\lambda = 3650$ Å) ultra-violet radiation, both at room temperature and at —180 °C, are summarized in table I.

At room temperature only products of the β-phase containing less than 10 % molybdate show a considerable fluorescence, the emission having the same spectral distribution in $MgWO_4$, $ZnWO_4$ and their solid solutions, and consisting of a broad band with a maximum at 4900 Å; this emission must be attributed to tungstate groups. At lower temperatures luminescence is also shown by products of the β-region containing more molybdate, and by products of the α- and γ-region. The δ-phase remains non-luminescent even at a temperature as low as —180 °C. Whether fluorescence will occur at a still lower temperature has not been investigated.

In the β-field the emission due to the tungstate groups is excited by ultra-violet radiation of $\lambda < 3200$ Å. In the products containing molybdate two effects occur:

1) a new emission occurs, situated at longer wavelengths than the tungstate fluorescence; it is orange-red at the zinc side, orange-yellow at the magnesium side; and
2) the excitation spectrum is extended towards longer wavelengths, excitation being possible by ultra violet of $\lambda < 3800$ Å; this is due to the appearance of a new absorption band.

Excitation in the long-wave absorption band gives rise to the orange fluorescence alone. Obviously the long-wave absorption band and the long-wave emission band(s) are closely related, both being due to the same configuration, probably the molybdate groups.

TABLE I

Composition				Crystal structure	Fluorescence			
Mg + Zn = 100		W + Mo = 100			Excitation by 3650 Å		Excitation by 2537 Å	
Mg	Zn	W	Mo		25 °C	−180 °C	25 °C	−180 °C
100	—	100	—	α	—	—	—	f blue
100	—	100	—	β	—	—	s bl.-green	s bl.-green
100	—	99	1	β	ff orange	f orange	m bl.-green	m bl.-green
100	—	95	5	$\beta(+\alpha)$	m orange	m orange	f yellowish	m bl.-green
100	—	90	10	$\alpha(+\beta)$	m orange	m orange	ff brown-yell.	f bl.-green
100	—	75	25	$\alpha(+\gamma)$	f orange	m orange	—	f yellow
100	—	50	50	$\gamma(+\alpha)$	ff orange	f orange	—	m yellow
100	—	25	75	γ	—	f or.-yell.	f yellow	s yellow
100	—	10	90	$\gamma + \delta$	—	f or.-yell.	f yellow	s yellow
100	—	5	95	α	—	—	—	—
100	—	—	100	δ	—	—	—	—
—	100	100	—	β	—	—	s bl.-green	s bl.-green
—	100	99	1	β	f orange-yell.	f orange-red	m or.-yell. + bl.-green	m bl.-green
—	100	95	5	β	m orange-yell.	m orange-red	f orange-yell. (+ bl.-gr.)	orange + bl.-green
—	100	90	10	β	m orange-yell.	m orange-red	m orange-yell.	m orange-red
—	100	75	25	β	f orange-yell.	s orange-red	f orange-yell.	s orange-red
—	100	50	50	$\beta(+\gamma)$	ff orange-yell.	s orange-red	ff orange-yell.	f yellow + m orange-red
—	100	25	75	$\gamma(+\beta)$	fff orange-yell.	s orange-red	—	m yellow
—	100	10	90	γ	—	m orange-red	fff yellow	m yellow
—	100	2	98	γ	—	m orange-red	ff yellow	m yellow
—	100	—	100	γ	—	—	ff yellow	s yellow
75	25	100	—	β	—	—	s bl.-green	s bl.-green
50	50	100	—	β	—	—	s bl.-green	s bl.-green
25	75	100	—	β	—	—	s bl.-green	s bl.-green
100	—	—	100	δ	—	—	—	—
75	25	—	100	$\gamma + \delta$	—	—	fff yellow	m yellow
50	50	—	100	γ	—	—	ff yellow	s yellow
25	75	100	100	γ	—	—	ff yellow	s yellow
—	100	—	100	γ	—	—	ff yellow	s yellow
50	50	95	5	β	f orange-yell.	f orange	m bl.-green + f orange	m bl.-green + f orange
50	50	75	25	β	m orange-yell.	m orange	m orange-yell.	f bl.-green + orange-yell.
50	50	50	50	$\beta + \gamma$	ff orange-yell.	ff orange	f orange-yell.	f or.-yel. + yellow
50	50	25	75	γ	—	fff orange	—	m yellow

s = strong, m = medium, f = faint, ff = very faint, fff = extremely faint.

For excitation by short-wave ultra violet in the region $\lambda < 3200$ Å in which tungstate and molybdate absorption bands overlap, the two kinds of emission appear together; *fig. 3* shows the absorption and the spectral distribution of the fluorescence of $ZnWO_4$ and the solid solution $ZnWO_4$-25

mol. % $ZnMoO_4$. The relationship between the position of the absorption edge and the maximum of the emission, as found for pure tungstates and molybdates [1]), holds good also for these mixed crystals.

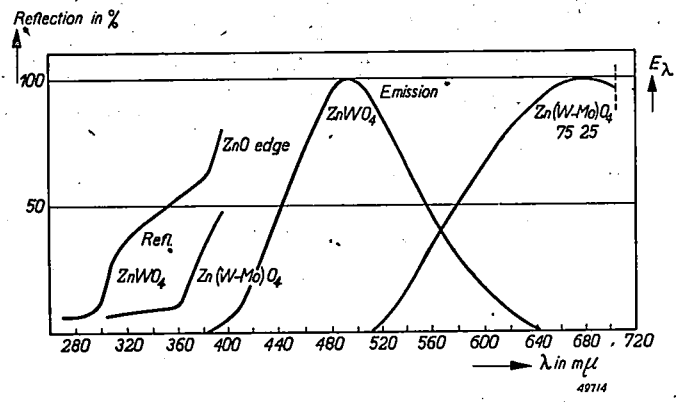


Fig. 3. Spectral distribution of reflection and fluorescence of $ZnWO_4$ and $ZnWO_4$ - 25 mol. % $ZnMoO_4$.

The luminescence of the α-phase is not quite certain, since the products of this structure always contained admixtures of one of the other phases. Probably, however, a long-wave absorption and a long-wave emission occur also in this case. The products of the γ-structure, excited by $\lambda = 2537$ Å, all show a yellow emission as found in $ZnMoO_4$. Excitation by long-wave ultra-violet radiation does not occur, while a long-wave emission has not been observed either.

Temperature dependence of the luminescence

In all products the fluorescence has maximum intensity at low temperatures. Above a certain critical temperature the efficiency decreases gradually with increasing temperature, the quenching being generally completed in a range of 100-150 °C. For products of the system $ZnWO_4$-$MgWO_4$ the limits of these quenching ranges are shown in *fig. 4*.

As in the case of the spectral distribution of the emission, the replacement of magnesium by zinc has again but little effect; yet decrease of the quenching range is observed, which is maximal at intermediate compositions.

Discussion

As has been observed with $Ca(W,Mo)O_4$, in the system now studied tungstate and molybdate groups also give rise to their own absorption and emission bands. The bands due to the molybdate groups are situated at longer wavelengths than those due to the tungstate groups. For this reason

the presence of small quantities of molybdate in tungstates is clearly indicated by the absorption and excitation spectrum, but conversely the presence of tungstate in molybdates can only be detected at higher concentrations.

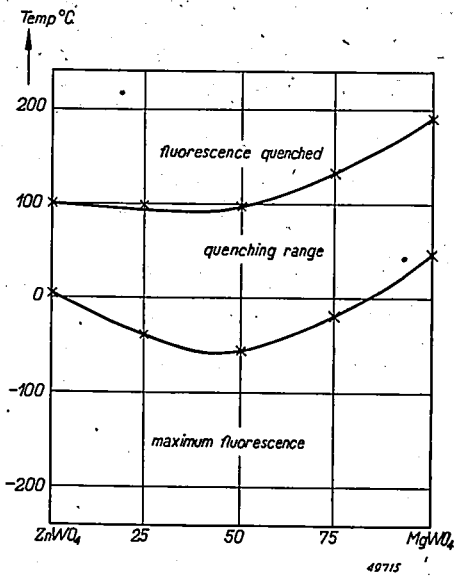


Fig. 4. Quenching ranges of the fluorescence of mixed crystals (Zn, Mg)WO_4.

The effect of the substitution of zinc by magnesium is markedly different from that caused by the substitution of calcium, strontium, or barium, by lead in tungstates, discussed in an earlier paper [1]). Lead causes the appearance of new bands with different spectral distributions and quenching ranges, while the interchange of magnesium and zinc has very little influence. This indicates that magnesium and zinc ions seem to have nearly identical fields around them; the slight difference which undoubtedly exist may be made responsible for the slightly different quenching points of the pure zinc and magnesium compounds. In mixed crystals containing both cations together, a number of tungstate or molydbate groups will be surrounded asymmetrically, and if this produces a lowering of the quenching point — as previously suggested [1]) — this effect may account for the minimum observed in fig. 4.

Eindhoven, January 1947

REFERENCES

[1]) F. A. Kröger, Some Aspects of the Luminescence of Solids, Amsterdam - New York, 1947, in print, especially Chapter III.

[2]) Hüniger, cited by K. Birus, Ergbn. Exakt. Naturw. **20**, 183-267, 1942; especially p. 221; C. G. A. Hill, Trans. Faraday Soc. **42**, 666-668, 1946; N. J. Dunning & H. D. Megaw, Trans. Faraday Soc. **42**, 705-709, 1946.

Philips Res. Rep. 2, 183-189, 1947

LUMINESCENCE OF SOLID SOLUTIONS OF THE SYSTEM $CaMoO_4$-$PbMoO_4$ AND OF SOME OTHER SYSTEMS

by F. A. KRÖGER

535.37

Summary

As shown in a previous paper, the fluorescence of solid solutions (Zn, Mg) WO_4 can be interpreted simply as the superposition of the fluorescence of its two components; but in the systems (Ca, Pb) WO_4, (Sr, Pb) WO_4 and (Ba, Pb) WO_4 new emission bands were observed which were attributed to tungstate groups with mixed surroundings of lead and calcium, strontium or barium ions. In this paper the systems (Ca, Sr) WO_4, (Ca, Sr) MoO_4, and (Ca, Mg)$_3$ WO_6 are shown to behave as (Zn, Mg) WO_4, whereas (Ca, Pb) MoO_4 behaves as (Ca, Pb) WO_4.

1. *Introduction*

Below a certain critical temperature practically all tungstates and molybdates show fluorescence which has been shown to depend on the crystal structure and on the cations that are present in the lattice [1]). The influence of the cations can best be studied in solid solutions containing different cations, but possessing throughout the same crystal structure. As discussed in a previous communication the fluorescence of solid solutions (Zn, Mg) WO_4 is nearly identical with that of the components [3]), but in the systems (Ca, Pb) WO_4, (Sr, Pb) WO_4 and (Ba, Pb) WO_4, the fluorescence of the solid solutions cannot be explained as a simple superposition of the bands of the components: in the latter systems new bands appear which must be attributed to tungstate groups having a mixed surroundings consisting of calcium, strontium or barium, together with lead [1]). Continuing these investigations we shall in the present paper report about the systems (Ca, Pb) MoO_4, (Ca, Sr) WO_4, (Ca, Sr) MoO_4 and (Ca, Mg)$_3$ WO_6.

2. (Ca, Pb) MoO_4

Calcium molybdate and lead molybdate crystallize in the scheelite structure with nearly the same lattice dimensions, and they form an uninterrupted series of solid solutions of this structure. Pure $CaMoO_4$ and $PbMoO_4$ were precipitated from solutions of calcium and lead nitrate by addition of a solution of ammonium molybdate, and the mixed crystals were made by heating the precipitates together in the required proportion at a temperature of about 850 °C for about one hour. *)

*) The preparations have been carried out by Mr A. Bosman in this laboratory.

2a. *Fluorescence*

Pure $CaMoO_4$ shows fluorescence in a broad yellow band with maximum at 5350 Å, pure $PbMoO_4$ in a green band with maximum at 5120 Å; in solid solutions of these two substances at least one new band appears with a maximum at longer wavelengths. *Fig. 1* shows the spectral distribution

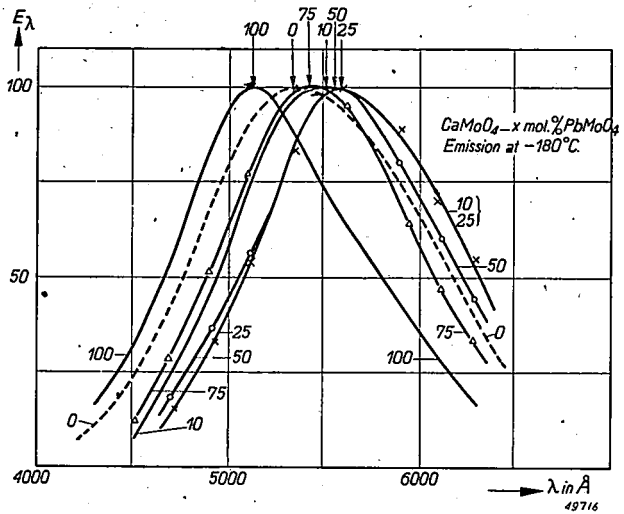


Fig. 1. Spectral distribution of the fluorescence of mixed crystals (Ca, Pb) MoO_4.

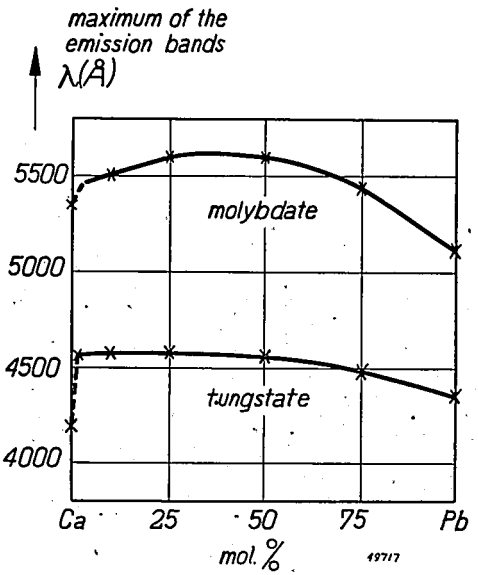


Fig. 2. Variation of the maximum of the fluorescence band with the composition for (Ca, Pb) WO_4 and (Ca, Pb) MoO_4.

of the emission at — 180 °C for various products, while the variation in the position of the maximum with the composition is shown in *fig. 2*. In these observations $CaMoO_4$ was excited by $\lambda = 2537$ Å, but all systems containing lead were excited by $\lambda = 3650$ Å.

For comparison the variation of the blue bands in the system $CaWO_4$-$PbWO_4$ has also been plotted.

2b. *Reflection and excitation*

The reflection spectrum of the pure compounds and of various solid solutions is shown in *fig. 3*. While for $CaMoO_4$ nearly total absorption is

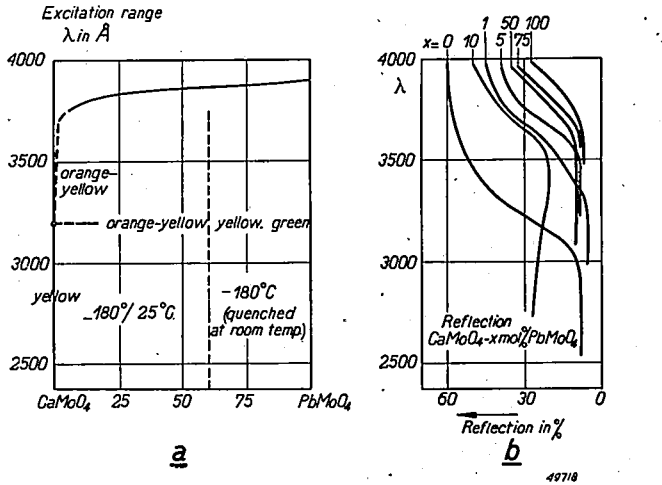


Fig. 3. Spectral distribution of the reflection and of the excitation of fluorescence in the system (Ca, Pb) MoO_4.

observed at wavelengths below 3000 Å, this limit is situated at longer wavelengths for products containing lead, shifting slightly towards the long-wave side with increasing lead concentration. At a sufficiently low temperature (*cf.* section 2c) irradiation into the absorption bands indicated by the reflection gives rise to fluorescence. In $CaMoO_4$, absorption of $\lambda <$ 3200 Å produces the well-known yellow fluorescence of this substance. For products containing a few per cent of lead molybdate the excitation spectrum consists of two parts. Excitation by $\lambda < 3200$ Å gives rise to the yellow $CaMoO_4$ band together with an orange-yellow band, but excitation by $3200 < \lambda < 3800$ Å produces the latter band only. With increasing lead content the yellow band disappears, while the other band gradually varies its position on the wavelength scale in the direction of shorter wavelengths. While products containing up to 60% lead molybdate are lumines-

cent at room temperature, are products of higher lead content only luminescent at low temperatures (*cf.* section 2c).

2c. *Temperature dependence*

As shown in *fig. 4* the temperature dependence of the fluorescence is different for excitation by $\lambda = 2537$ Å (*fig. 4b*) and $\lambda = 3650$ Å (*fig. 4a*).

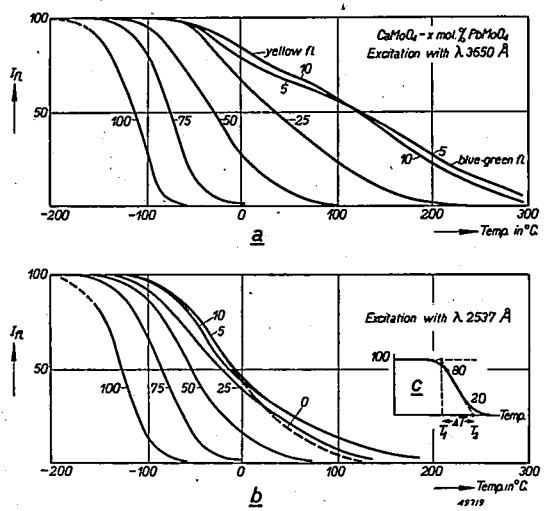


Fig. 4. Temperature dependence of the fluorescence of (Ca, Pb) MoO_4 for excitation by (a) $\lambda = 3650$ Å, (b) $\lambda = 2537$ Å; (c) definition of the quenching range $\Delta T = T_2 - T_1$.

In both cases the incorporation of lead in a concentration below 25 mol.% causes an increase in the temperature at which the quenching is completed, but in the latter case (3650 Å) this increase is much more pronounced. At higher lead concentrations the quenching points are lowered and gradually approach the ones of pure lead molybdate. The upper and lower limits of the quenching range T_1 and T_2 — determined as shown in *fig. 4c* — have been plotted in *fig. 5* as a function of the composition.

The spectral distribution of the fluorescence observed at the highest temperatures is slightly different from that appearing at lower temperatures. For instance the fluorescence of $CaMoO_4$-10% $PbMoO_4$, excitated by $\lambda = 3650$ Å, is orange-red below 100 °C, but is yellow at higher temperatures. We infer from this fact that the system is more complex than has been assumed in section 2b, the incorporation of lead in $CaMoO_4$ causing two bands at least. This may also explain the observed differences in temperature dependence for excitation by $\lambda = 2537$ Å and $\lambda = 3650$ Å. Each will probably be associated with its own absorption band. Excitation by

different wavelengths produces the sub-bands in different proportions. If each has its own characteristic quenching range, the total quenching curves will also be different.

A similar behaviour has previously been observed for the system $CaWO_4$-$PbWO_4$ (*fig. 5b*). In this case the complexity of the emission spectrum is not indicated by a variation of the quenching curves with the exciting wavelengths, but it is made probable by a comparison of the corresponding barium-lead and strontium-lead systems. [1])

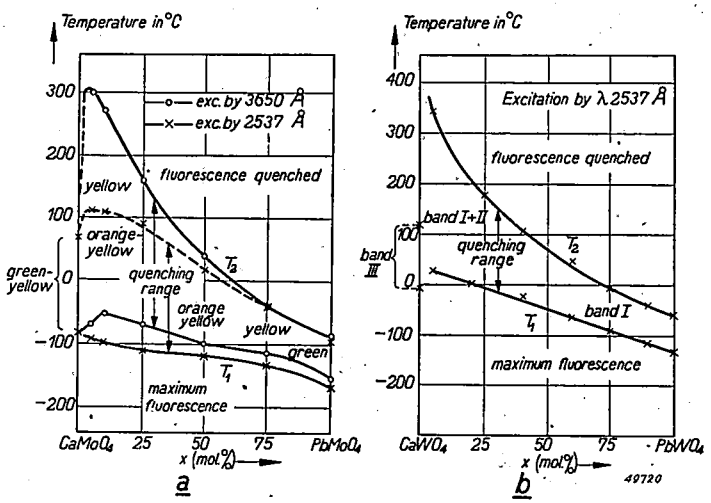


Fig. 5. Quenching ranges of the fluorescence in (Ca, Pb) MoO_4 and (Ca, Pb) WO_4.

3. *Other systems*

Calcium and strontium tungstates form uninterrupted series of solid solutions. The same applies to molybdates.

We have investigated the fluorescence in (Ca, Sr) WO_4 and (Ca, Sr) MoO_4. In both cases the emission of the components is nearly identical, and the emission of the solid solutions is intermediate between that of the components, while the quenching ranges vary regularly with the composition (*fig. 6*).

We have also studied the fluorescence of mixed crystals of the composition M_3WO_6 in which M stands for Ca,Mg, Ca,Sr, Ca,Ba [2]). Pure Mg_3WO_6 does not exist, but solid solutions $(Ca,Mg)_3WO_6$ can be prepared containing up to about 50% Mg_3WO_6. Over this range the emission remains unchanged, white-blue. As may be seen in *fig. 7*, the quenching temperatures pass through a minimum at 30% Mg_3WO_6, very similar to what has been observed in the system $ZnWO_4$-$MgWO_4$. [3])

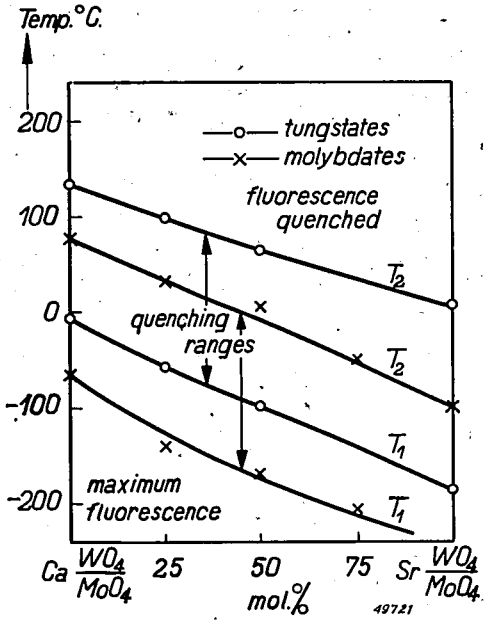


Fig. 6. Quenching ranges of the fluorescence of (Ca, Sr) WO_4 and (Ca, Sr)MoO_4, excited by $\lambda = 2537$ Å.

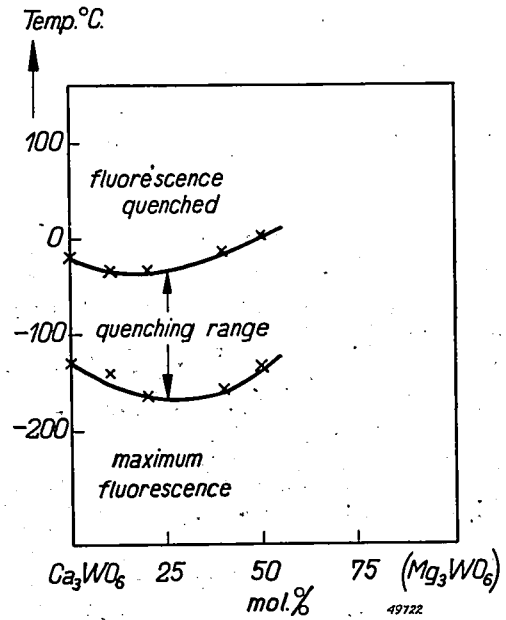


Fig. 7. Quenching ranges of the fluorescence of (Ca, Mg)$_3WO_6$.

It is difficult to determine the properties of the systems $(Ca,Sr)_3WO_6$ and $(Ca,Ba)_3WO_6$ because the quenching ranges for the barium and strontium compounds lie so low that at —180 °C the maximum efficiency is not reached. It could be ascertained, however, that the quenching range of the fluorescence of the calcium compound is decreased by the incorporation of strontium and barium.

4. *Discussion*

Formation of solid solutions of the type considered in this paper may have two effects:

1) a gradual change of the properties of a certain emission band,

2) the appearance of a new emission band.

Effects of the first kind occur when the luminescent centre is not disturbed to such an extent that the electronic transition underlying the emission is essentially changed. Variation of the quenching point will be caused by a change in the interaction of the excited centre and its surroundings. In the frame-work of a theory of the energy dissipation given by Mott and Seitz [4]), properties of the configuration formed by the nearest neighbours, such as the mass of the atoms, the electric charge, the symmetry of the surroundings, interatomic distances, must be of great importance. If, on the other hand, the dissipation is supposed to be a property of the lattice as a whole (Möglich and Rompe [5])), the formation of a solid solution must vary this property for the lattice as a whole. The first assumption seems to be the most acceptable one.

Under particular circumstances, the changed surroundings will not only cause a variation in the interaction, but they will also change the electronic transition in the centre, either by causing a transition involving entirely new energy levels, or by changing the relative probabilities for the transition between levels already present. In the molybdates and tungstates with mixed cations studied so far, only the systems containing lead show such a new emission at intermediate compositions. Systems containing calcium together with strontium, or zinc together with magnesium, do not show this effect.

Eindhoven, January 1947

REFERENCES

1) F. A. Kröger, Some Aspects of the Luminescence of Solids, Amsterdam - New York 1947, in print, especially chapter IV.
2) H. P. Rooksby & E. G. Steward, Nature **157**, 548-549, 1946.
3) F. A. Kröger, Philips Res. Rep. **2**, 177-182, 1947.
4) N. F. Mott, Proc. Roy. Soc. London (A) **167**, 384-391, 1938; R. W. Gurney & N. F. Mott, Trans. Faraday Soc. **35**, 69-73, 1939; F. Seitz, Trans. Faraday Soc. **35**, 74-85, 1939.
5) K. Birus, F. Möglich & R. Rompe, Phys. Z. **44**, 122-129, 1943; M. Schön, Forschungen und Fortschritte, **19**, 4-9, 1943.

Philips Res. Rep. **2**, 190-204, 1947

THE *i,V*-CHARACTERISTIC OF THE COATING OF OXIDE CATHODES DURING SHORT-TIME THERMIONIC EMISSION

by R. LOOSJES and H. J. VINK

537.583:621.385.1.032.216

Summary

The potential differences existing across an oxide coating during short-time emission (condenser discharge with an "*RC*-time" of 10^{-4} sec) were measured.
For this purpose a new measuring method was worked out. Using this method it was found that at current densities of about 5-10 A/cm^2 remarkably high potential differences exist across the oxide coating (50-200 V) at the normal working temperatures (900-1100 °K) at which the experiments were carried out.

1. *Introduction*

The study of the oxide-coated cathode under microsecond-pulse conditions has shown that the emission mechanism — and probably also the conduction mechanism — differs widely from that under conditions of direct current [1]). This fact accentuates our lack of knowledge about the oxide cathode under conditions between these two extremes (*e.g.* millisecond pulses). In this article we will present some data concerning this middle range.

For reasons outside the scope of this article we carried out some experiments in order to determine the *i,V*-characteristic of the oxide coating in this range. As the results of this investigation were rather unexpected it seemed worth while to publish these data. We made use of pulses with an exponential decay of the voltage, which were generated by a condenser discharge. The "*RC*-time" being of the order of 10^{-4} sec, one can assume that these pulses are equivalent to square pulses of the same order.

2. *Measuring method*

It has been our chief object to find a method of measuring the *i,V*-characteristic of the coating without disturbing its true working conditions in any respect. Because of this restriction all methods introducing a probe in the coating or applying an electrode to the surface of it were discarded.

Therefore, we made use of diodes in which the distance between anode and cathode, both of which had plane surfaces, could be varied. The voltage corresponding to a given emission current was observed as a function of the distance. By extrapolating to zero distance the potential difference across

the layer could be deduced without the anode coming into contact with the cathode surface.

In the following section the diodes and the electrical arrangement used will be described in greater detail, and in sections 4, 5 and 6 we will deal with the measurements and their results.

3. *Description of the diodes and the electrical apparatus*

The cathode F (*fig. 1*) was of the type as is used in some oscillograph tubes, with a plane surface of 8 mm^2. The plane Ni-sheet anode A could slide along the Ni poles D and E by means of the Ni tube B and the Ni ring C simply by tapping the tube. H is an insulated filament. K is a chromel-alumel thermocouple. The oxide coating is marked G. Two pointed Ni poles L and M are used for the determination of the cathode-anode distance (see later on).

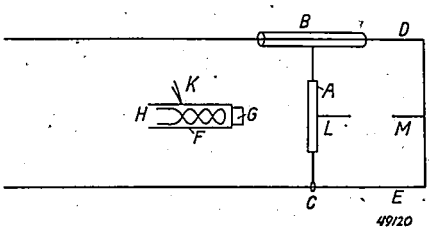


Fig. 1. Arrangement of a diode with moving anode in order to measure the diode characteristic with various cathode-anode distances.

The whole arrangement was built in a glass envelope and subjected to a baking and outgassing process. At the moment of sealing off the pressure was 10^{-5} mm Hg or less. For improving and maintaining the vacuum a barium getter was applied.

A pulse apparatus similar to that used by us has already been used earlier in this laboratory [2]) and elsewhere [3]). In this type of apparatus the electrical pulses needed to draw a current during a short time are produced by periodic discharge of a condenser. The voltages and corresponding currents are reproduced on an oscillograph.

The apparatus we used is schematically represented in *fig. 2*. The condenser C is charged from a D.C. source via a resistance R_1. The switching mechanism I periodically permits C to discharge via D, R_2 and R_3. It may be a purely mechanical mechanism, *e.g.* a commutator, but also an electrical one, *e.g.* a gas-filled triode. Both types of switches have been used in our experiments.

A commutator has the advantage of its simplicity although the frequent

occurrence of faulty contacts is a serious handicap. A gas-filled triode works much better in this respect. The frequency of switching can be regulated by an external A.C. voltage between cathode and grid, but also by means of the circuit indicated by broken lines in fig. 2.

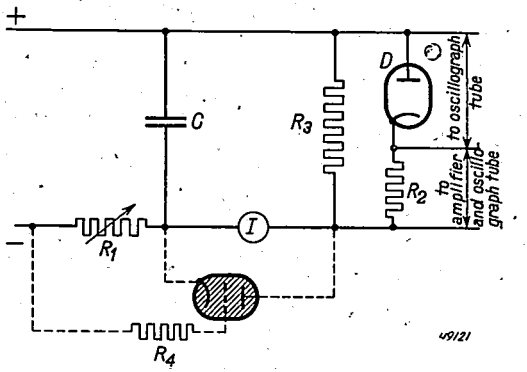


Fig. 2. Diagram of the electrical connections of the apparatus for the oscillographic measurement of the i,V-characteristic.

This circuit operates as follows: if C is charged, a potential difference is generated across R_1 which is put on the grid of the triode, thereby suppressing a discharge. At the end of the charging process this potential difference falls away and a discharge can take place. Hereafter the process is repeated. The resistance R_4 (a few thousand ohms) prevents overloading of the grid during the discharge. The values of R_1 and C determine the frequency of discharge.

If $C = 0{\cdot}25$ μF, R_1 must have values of 10^4 to 10^6 ohms in order that the pulse-recurrence frequency may be varied from about 10^2 to 1 per second.

In order to get reproducible results the load on the cathode under investigation has to be kept as low as possible. Therefore the value of C, the D.C. voltage and the pulse-recurrence frequency were reduced until the emission current remained constant during the measurements. This is shown by a steady i,V-characteristic of the diode even when exposed to the pulses for some hours. In the series of measurements described here we used a condenser of 0·25 μF, a D.C. voltage of 400-500 volts and a pulse-recurrence frequency of 25-50 c/sec.

Moreover, a resistance (R_3) of 300 ohms was placed parallel to the diode under investigation to reduce the load on small or feebly emitting cathodes. Therefore the "RC-time" was maximally about 10^{-4} sec. Under these circumstances the maximum average current through a diode was about 5 mA.

The voltages generated during a flash across the diode D are conducted

to the horizontal deflection plates of an oscillograph tube. This tube is of the type DB 16-1, working at an anode voltage of 2000 V. By using this rather low voltage the sensitivity of the tube was increased.

The voltage across R_2, which is only a few volts because R_2 has a small resistance relative to the diode, is directly proportional to the current through D. It is amplified in an adjustable resistance amplifier, transmitting frequencies between 10^3 and 10^5 c/sec without distortion at an output voltage of maximum 300 volts, and having a capacity of 25 cm between one of the plates and the earth. The output is then conducted to the vertical deflection plates of the oscillograph tube.

The oscillograph was given an excentric zero point by means of constant voltages applied to the plates, in order that the whole screen could be utilized. The electron beam was switched on only during the flashes. Vertical and horizontal scales were drawn on the screen. The voltage scale (0-250 V) was calibrated by means of an A.C. source of known peak value. The current scale was calibrated by replacing the diode by a resistance of known value (accuracy $\pm$ 1 %) and adjusting the amplifier. The really straight line shown on the screen in this case showed that no distortion took place in the entire apparatus.

The accuracy of the readings on the scales was about 1% at the voltages and currents measured in these experiments, but small deviations of the zero point and inaccuracy in the drawing of the scales increased the possible error to $\pm$ 2 %.

4. *Determination of the distance between anode and surface of the coating*

The accurate determination of the distance between the surface of the coating and the anode was of great importance. For this purpose a comparator was used with an accuracy of $\pm 2\mu$. With this instrument the distance between L and M (see fig. 1) was determined. First of all this distance was measured when the anode touched the cathode surface after temperature equilibrium has been established (usually after 15 minutes). Apparently the thermal contact between cathode and anode was only very slight, probably because of the insulating properties of the oxide coating, since the temperature did not alter perceptibly when contact was made. Though by means of the comparator it was not possible to see whether anode and coating really touched each other, this could easily be detected by a very marked discontinuity in the change of the shape of the characteristic, when the distance was gradually diminished. As shown in *fig. 3* the characteristic changed very suddenly from curve 2 to curve 3.

At the moment the curve changed, another phenomenon occurred. By means of the microscope of the comparator small sparks could be seen, three or four at a time, moving rapidly along the contact layer. This effect was

reproducible and did not harm the emission of the cathode. We could not detect any difference between two characteristics, one before and one after contact had been made. This is true for activated cathodes only, as less well activated cathodes were activated rapidly under these conditions.

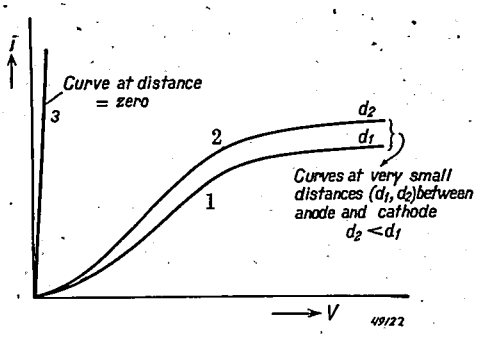


Fig. 3. Schematically drawn characteristics at small and zero coating-anode distances.

The surfaces of coating and anode had to be perfectly plane and parallel to each other. If this was not the case the tubes were discarded. The selection was effected by means of the "sparking" phenomenon described above. Only if the sparks originated from the whole surface of the coating were the cathodes used for further experiments. Small unevennesses were removed by dropping the anode upon the cathode a number of times from a height of about 0·5 cm.

By subtracting the distance LM at any position of the anode from the "zero point" distance, the distances between cathode and anode were determined. After a complete set of characteristics at various distances had been measured, the "zero point" distance was checked. If the agreement was not within 2 μ the whole set of observations was discarded.

5. *Results of the measurements*

As an example we shall now give some of the results obtained with tube V 41 *). The BaO-SrO coating was prepared by decomposition of a sprayed coating of $BaCO_3$-$SrCO_3$ of 25 μ thickness. After conversion this cathode was very well activated by drawing emission, and aged for 24 hours by drawing 12 mA/cm^2 emission at 800 °C.

At a temperature of 1078 °K we observed the characteristics of the diode at cathode-anode distances of 48, 145 and 250 μ (*fig. 4*).

*) Some other tubes of the same construction have been investigated in the same way as V 41, the results being qualitatively in accordance with each other.

From fig. 4 we obtained the data for *fig.* 5, in which figure the graphical extrapolation to the distance $d = 0$ was made.

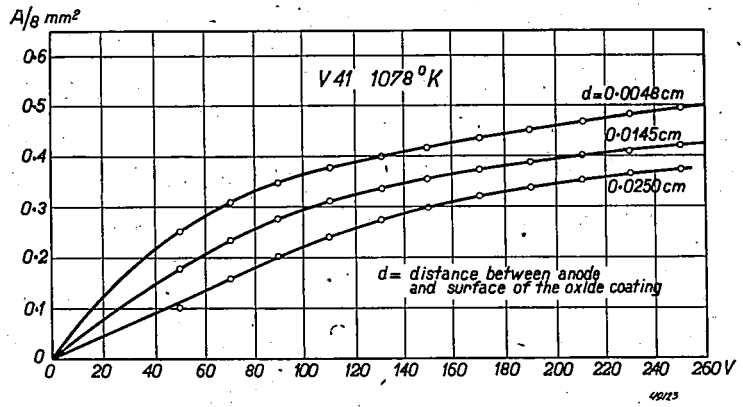


Fig. 4. i,V-characteristics in a diode for various cathode-anode distances at a cathode temperature of 1078 °K. Tube V 41 with oxide-coated cathode; oxide is equimolecular BaO-SrO of 25 μ thickness.

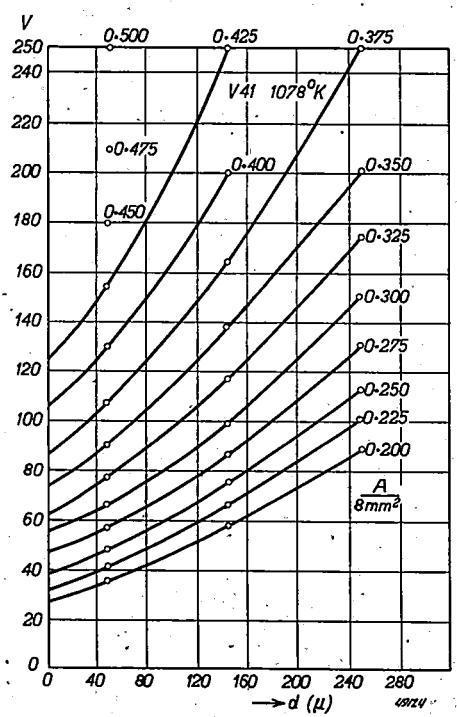


Fig. 5. Potential distribution in the vacuum gap at a cathode temperature of 1078 °K, for various current densities. Tube V 41.

Such an extrapolation might be very inaccurate. However, it will be shown in the next section that in our case the extrapolation could be effected without much inaccuracy.

The points of intersection of the various curves with the voltage axis are plotted in *fig. 6* as a function of the current density, thus giving the current-voltage characteristic of the oxide coating.

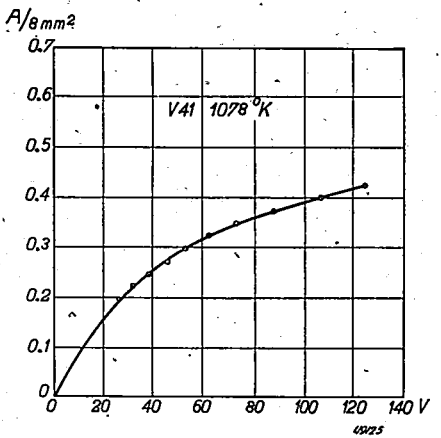


Fig. 6. i,V-characteristic of an oxide coating at a cathode temperature of 1078 °K. Tube V 41.

With the same tube (V 41) we carried out more elaborate experiments involving other distances, *viz.* 5, 12, 28, 43 and 49 μ at a cathode tempera-

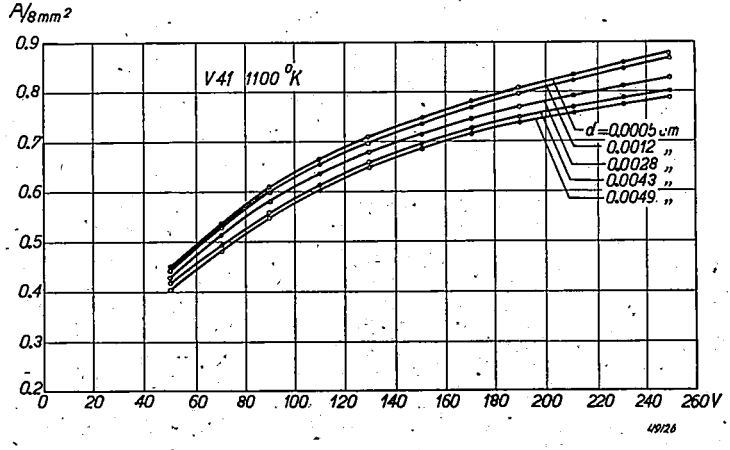


Fig. 7. Diode i,V-characteristic for various cathode-anode distances at a cathode temperature of 1100 °K. Tube V 41.

ture of 1100 °K instead of 1078 °K. In *fig. 7* the i,V-characteristics for the diode are plotted. The extrapolation to zero distance is shown in *fig. 8*, while the i,V-characteristic of the whole oxide coating is given in *fig. 9*.

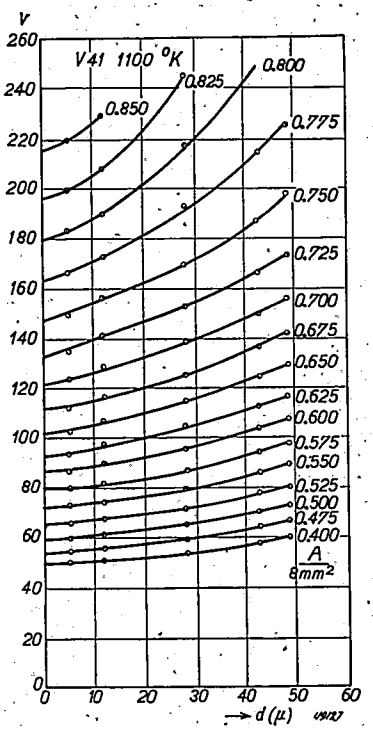


Fig. 8. Potential distribution in the vacuum gap at a cathode temperature of 1100 °K. Tube V 41.

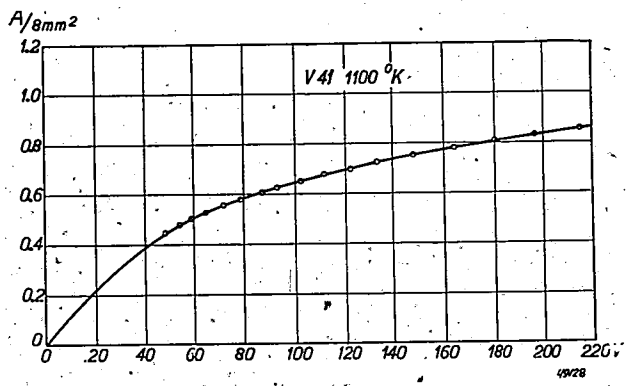


Fig. 9. i,V-characteristic of an oxide coating at 1100 °K. Tube V 41.

The graphs in *figs 10, 11* and *12* represent the results of similar experiments carried out at a cathode temperature of 976 °K.

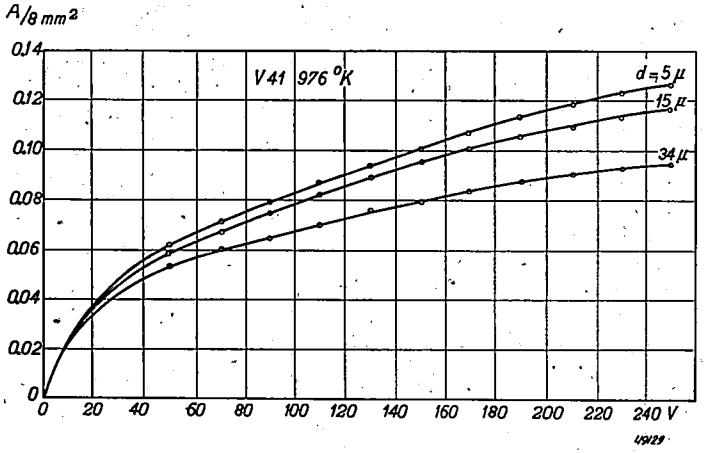


Fig. 10. Diode i,V-characteristics for various cathode-anode distances at a cathode temperature of 976 °K. Tube V 41.

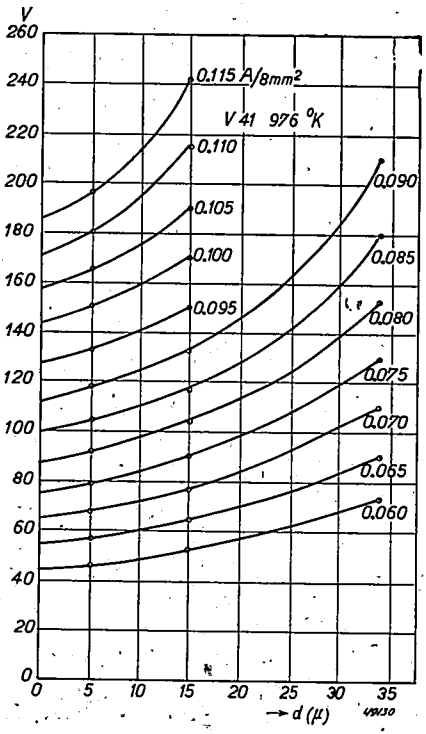


Fig. 11. Potential distribution in the vacuum gap at a cathode temperature of 976 °K. Tube V 41.

6. *Discussion of results*

Examination of the figs 6, 9 and 12 reveals the hitherto unknown fact that, at the highest current densities observed, fairly high voltages are established across the whole oxide coating. For example, at 1100 °K and a current density of 0·85 A/8 mm² (that is 10·6 A/cm²) the potential difference across the oxide coating is 215 volts, a value far exceeding the potential in the vacuum gap.

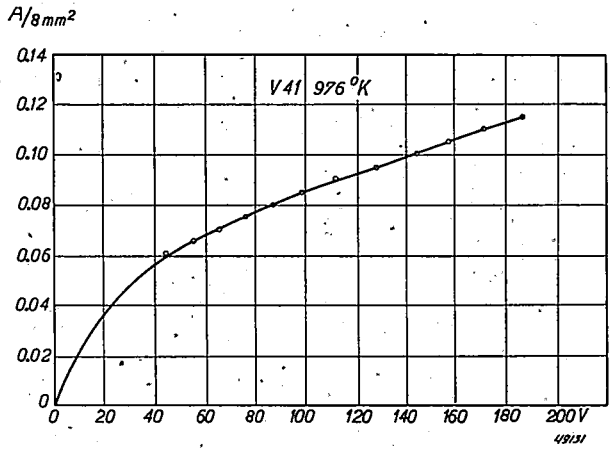


Fig. 12. i,V-characteristic of an oxide coating at a cathode temperature of 976 °K. Tube V 41.

A discussion of the validity of the extrapolation method with which we obtained this remarkable result falls into two parts:

a) a discussion as to whether the curves in figs 5, 8 and 11 really represent the potential distribution in the vacuum gap;

b) a discussion on the reliability of the extrapolation itself.

With regard to a) we note that if the logarithm of the potential of a point in the vacuum gap with respect to the surface of the coating is plotted as a function of the logarithm of the distance of this point to the surface of the cathode, the current density being a parameter, we find straight lines, mostly having a slope of 4/3 as is required by the Langmuir formula for a space-charge-limited thermionic current:

$$I^{2/3} = \frac{(2{\cdot}334.10^{-6})^{2/3} \cdot O^{2/3}}{d^{4/3}} V$$

(where O denotes the surface of the cathode and d the distance from the cathode).

As an example the data for 1100 °K are given in *fig. 13*.

On the other hand, if the logarithm of the current density is plotted as a function of the logarithm of the potential of a point in the vacuum gap with respect to the surface of the cathode, the distance now being a parameter, we do not find a relation of the form $I^{2/3} = aV$ as is required by the Langmuir formula (*fig. 14*).

Therefore it is not possible to conclude that these curves really represent the potential distribution in the vacuum gap, from the shape and the relative distance of the curves in figs 5, 8 and 11 only.

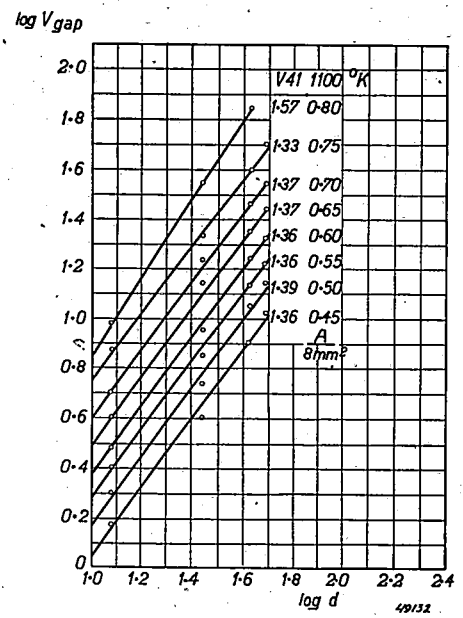


Fig. 13. Potential distribution in the vacuum gap with respect to the surface of the coating, plotted logarithmically, the current density being a parameter. The figures in the left-hand column indicate the slope of each line (V 41, 1100 °K).

In a further discussion of a) we first assume that no reflected or secondary electrons are emitted from the anode.

In this case it will be clear that, if the field in the gap is homogeneous, for a fixed current density a point at a certain distance (*e.g.* 48 μ) from the surface of the oxide layer must always have the same potential with respect to the core, irrespective of the anode being situated at this point or not. Thus in a homogeneous field the potential with respect to the core in a certain point can be measured by placing the anode at this point.

It is easy to see that there are two causes making this field inhomogeneous. In the first place we have the possibility that at larger distances the electron beam will diverge. In this case the potential for a

certain current density will increase less with increasing distance than would have been the case in a homogeneous field. Secondly, according to our results, at high current densities a potential difference of 100-200 V

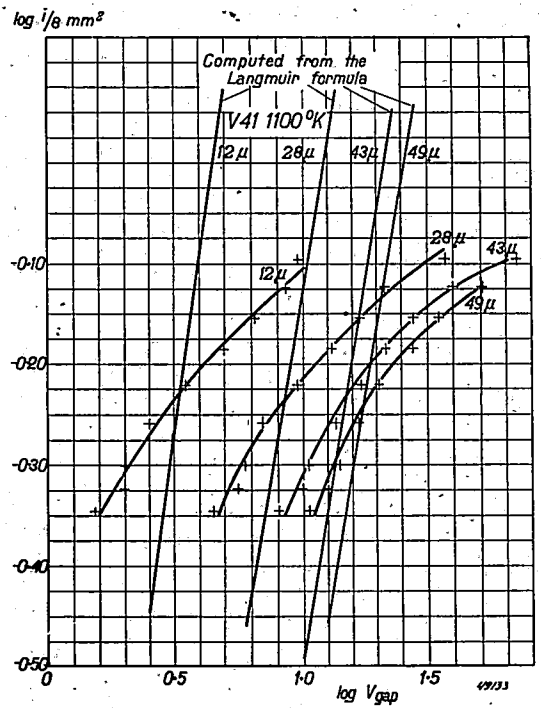


Fig. 14. Comparison of the Langmuir relation between log i and log V at the given dimensions and the observed relation (the distance being a parameter).

can be established across the oxide layer, by which a stray field at its circumference can come into existence, which will bend the electrons from the circumference of the beam in the direction of the centre. This is schematically drawn in *fig. 15*.

The magnitude of these two effects cannot be easily estimated; we only can say that they will counteract each other.

Although it is therefore uncertain in how far the curves in figs 5, 8 and 11 represent the potential distribution in the vacuum gap, we may in any case assume that their effect will decrease with decreasing distance, in view of the nature of the two possibilities of inhomogeneity. With decreasing distance the observed curves will therefore approach the true potential-distribution curves more and more closely, not only with regard to the shape but also to the position.

Therefore in discussing b) we may observe that there is no reason to fear that the high potentials of the surface of the layer, which were obtained by extrapolation of the V,d-curve to zero distance, assuming it to remain concave to the V-axis, are only apparent, because in that case there would be a sharp decrease of the potential for distances smaller than the smallest distance used. For these very small distances the potential curve would have to be wholly or partly convex with regard to the V-axis. This however is impossible, since the Poisson equation

$$\Delta V = 4\pi\varrho$$

would then require a positive space charge for these small distances.

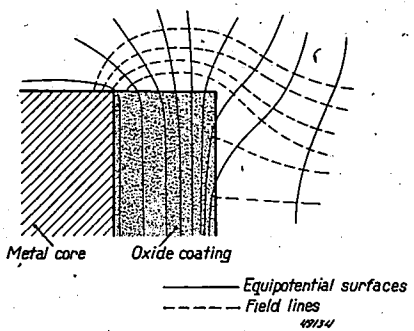


Fig. 15. The stray field at the circumference of the oxide coating (schematically).

On the contrary, the negative space charge at smaller distances must be more pronounced, as the velocities of the electrons are smaller in this region than at larger distances.

The curves can at small distances only be more or at least equally concave. We see, therefore, that whether the field is homogeneous or not the extrapolation to zero distance is probably justified.

We will now discuss the case that reflected or secondary electrons are present which are emitted from the anode.

As a possible consequence of the presence of high potential differences across the layer, the electrons may reach the anode with velocities totally different from those we would expect without these high potential differences.

The electrons that have been emitted from the outer grains of the oxide layer will attain only a low velocity. As a result of the granular structure of the layer, however, a larger or smaller part of the electrons can also be emitted from grains lying in the inferior of the coating, which may have a much lower potential. These electrons can attain a higher velocity. It is possible, therefore, that the velocity distribution can extend over some tens of electron volts. This velocity distribution which possibly may be deduced

from our measurements, may have consequences if we assume that the electrons reaching the anode will be partly reflected or cause secondary emission. In this case one could think, as may be clear from the following consideration, that the real potential-distribution curve between cathode and anode lies lower than the curves that can be deduced from our experiments (figs 5, 8 and 11).

The electrons that have been emitted from grains lying in the interior of the coating will, after reflection against the anode, have a greater velocity than the velocity corresponding to the potential difference between anode and the surface layer of the cathode. Therefore, they may reach the surface of the cathode.

If the current density is constant, as is the case in our measurements, the countercurrent formed by the reflected electrons has to be compensated by an extra current originating from the cathode. This is schematically drawn in *fig. 16.*

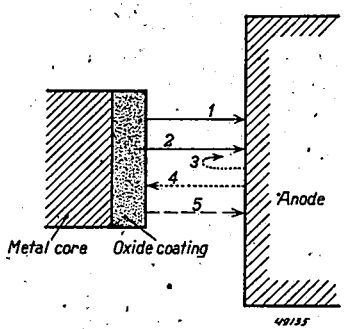


Fig. 16. Components of the current in the vacuum gap:
1. primary electrons emitted from the outer grains,
2. primary electrons emitted from grains lying in the interior,
3. secondary electrons emitted from the anode because of 1 and 2,
4. reflected electrons from current 2,
5. primary electrons emitted to compensate for the reflected electrons of 4.

When the distance is reduced, the potential differences between cathode and anode, which according to our measurements are already small, decrease even further. Thus increasingly more elastically and non-elastically reflected electrons, and at last secondary electrons from the anode will reach the cathode *), and therefore require an ever increasing extra current.

As the current density in the anode circuit remains constant, we find an ever increasing space charge in the vacuum gap as the distance is reduced, necessitating a higher anode potential. If we assume secondary emission of the anode, the curves in figs 5, 8 and 11, in contradiction with the two

*) or at least cover an important part of the distance anode-cathode.

effects already dealt with, will at smaller distances deviate ever more from the true potential distribution in such a way that at smaller distances we find a potential that is too high. The high potentials obtained by our method of extrapolation could be explained in this manner.

However, this whole argument only has any sense if we accept the high potentials of the surface of the oxide coating obtained by this method and the velocity distribution of the electrons, which possibly can be deduced from it. Therefore, if the secondary emission of the anode actually would have any importance the high potentials could only be partly explained in this way.

It can be estimated by a rather lengthy argument which will be omitted here that, with the known values of the secondary-emission coefficient of Ni, Ba and BaO on Ni and of the velocity distribution of the secondary electrons in dependence on the velocity of the primary electrons [4]), [5]), [6]), [7]), [8]), this effect will have hardly any importance.

Therefore we can say that the current-voltage curves given by us will not be far from the truth.

Thus it is established that the current-voltage curve of the whole well-activated coating under true working conditions bends towards the voltage axis in this way that at higher current densities unexpectedly high potentials are established across the whole oxide coating.

It is uncertain whether this is also the case for an oxide coating situated between two metal contacts at high current densities. Investigations on this subject are in progress. Experiments in order to gain an insight in the velocity distribution of the emitted electrons under these conditions are also being carried out.

Acknowledgement

The authors wish to express their sincerest thanks to Dr H. C. Hamaker for his valuable criticisms and continual interest.

Eindhoven, December 1946

REFERENCES

[1]) E. A. Coomes, J. appl. Phys. **17**, 647-654, 1946.
[2]) Tj. Douma and P. Zijlstra, Philips Transmitting News **5**, 89-96, 1938; *ibid.* **6**, 18-24, 1939.
[3]) W. Heinze, *e.g.*, Ann. Phys., Lpz. **16**, 41-76, 1933.
[4]) H. Bruining, J. H. de Boer and W. G. Burgers, Physica, 's-Grav. **4**, 267-275, 1937.
[5]) H. E. Farnsworth, Phys. Rev. **25**, 41-57, 1925.
[6]) H. Bruining and J. H. de Boer, Physica, 's-Grav. **5**, 913-917, 1938.
[7]) R. Warnecke, J. Phys. Radium **7**, 270-280, 1936.
[8]) H. Bruining and J. H. de Boer, Physica, 's-Grav. **5**, 17-30, 1938.

R 47 Philips Res. Rep. 2, 205-227, 1947

THE REACTION BETWEEN CARBON AND OXYGEN IN LIQUID IRON

by J. D. FAST

541.123:621.791.75:669.18

Summary

The main subject of this paper is the reaction

$$[FeO] + [C] \rightleftarrows [Fe] + CO\,, \qquad (1)$$

brackets indicating the components of a homogeneous liquid phase. Assuming a mixture of liquid Fe and liquid FeO to possess a Gibbs entropy of mixing and a Van Laar heat of mixing, the activity of FeO in liquid Fe is computed and compared with observed solubilities. It is given by the equation

$$a_{FeO} = x_{FeO}\, e^{9210(1-x_{FeO})^2/T},$$

pure liquid FeO being the standard of reference. Next the activity of C in Fe is derived by comparing the composition of CO-CO_2 gas phases in equilibrium with liquid iron and with graphite, using spectroscopic and thermal data to calculate the latter equilibrium. For both of these equilibria we may write

$$\log K = \log \frac{p^2_{CO}}{p_{CO_2} \cdot a_C} = -\frac{8435}{T} + 8{\cdot}835\,,$$

if we put $a_C = 1$ for the equilibrium with graphite (pure solid graphite being the standard of reference) and

$$a_C = x_C\, e^{1800/T-1{\cdot}5}$$

for the equilibrium with liquid iron, containing carbon up to concentrations of 1·0 wt % C (mole fraction $x_C = 0{\cdot}045$), and independent of the concentration of FeO in the liquid phase.

The quotient $(p_{CO} \cdot a_{Fe})/(a_{FeO} \cdot a_C)$ computed from the activities is a constant only if the C-concentration is less than 0·1 wt%, and this value is adopted as the correct constant of reaction (1). This is confirmed by independent arguments, leading to the formula

$$\log K = \log \frac{p_{CO} \cdot a_{Fe}}{a_{FeO} \cdot a_C} = -\frac{6750}{T} + 7{\cdot}22\,,$$

for the reaction constant of (1) at temperatures in the neighbourhood of the melting point of iron.

The equation for a_{FeO} loses its validity for carbon concentrations above 0·1 wt%, the activity of FeO being lowered by the presence of C.

1. *Introduction*

The reaction between carbon and oxygen, both dissolved in liquid iron, is of fundamental importance to steel making and arc-welding practice. In this paper we will deduce the chemical constants of this reaction from a consideration of the equilibrium between a) a liquid iron phase in which small quantities of oxygen and carbon are dissolved, and b) a gas phase consisting of a mixture of carbon monoxide and carbon dioxide.

Since oxygen is only very slightly soluble in liquid iron, it will make

a negligibly small difference in the molar concentrations whether we consider the oxygen dissolved as O or as FeO; we will assume that it is dissolved in the form of the ferrous oxide FeO.

Between the two phases mentioned above the following reactions can take place:

$$[FeO] + [C] \rightleftarrows [Fe] + CO, \tag{1}$$

$$2\,[FeO] + [C] \rightleftarrows 2\,[Fe] + CO_2, \tag{2}$$

$$[FeO] + CO \rightleftarrows [Fe] + CO_2, \tag{3}$$

$$[C] + CO_2 \rightleftarrows 2\,CO, \tag{4}$$

where the components of the liquid phase have been indicated by brackets. Since (3) is equal to the difference of (2) and (1), and since (4) is equal to twice equation (1) minus (2), two of these four equations will suffice to specify the chemical processes that may occur. As, moreover, under practical conditions the gas phase consists chiefly of CO, the concentration of CO_2 being seldom more than a few per cent, reaction (1) will be by far the most important one, and on this reaction we will mainly focus our attention.

2. *Thermodynamical formulae*

In this section we will briefly recapitulate the various thermodynamic principles and equations that will subsequently be used to deduce the constants of the reaction under consideration.

If in a chemical system in which thermodynamic equilibrium has been established, we visualize a further change to take place at constant pressure and temperature and under conditions of lasting equilibrium, then the corresponding change in the free enthalpy *) must be zero, $\Delta G = 0$. Denoting the chemical potentials (or the partial free enthalpies) of the separate substances, after Gibbs, by the symbol μ the equilibrium condition will be

$$\Delta G = \Sigma\, \nu_i\, \mu_i = 0, \tag{5}$$

where the ν_i's are equal to the coefficients with which each substance enters in the reaction equation. In the particular case of reaction (1) the ν_i's are all unity, so that we must have

$$\Delta G = \mu_{Fe} + \mu_{CO} - \mu_{FeO} - \mu_C = 0. \tag{6}$$

If, following Lewis, we introduce the activity a defined by

$$\mu = \mu^0 + RT \ln a, \tag{7}$$

*) The free enthalpy, $G = U - TS + pV$, is also frequently called the thermodynamic potential, the Gibbs free energy, or simply the free energy; in American literature it is usually indicated by the symbol F.

where μ^0 is the chemical potential of the substance under consideration in a certain standard state of reference (state of unit activity), equation (5) will be transformed into

$$\Delta G^0 = \Sigma \nu_i \mu_i^0 = -\Sigma RT \ln a_i^{\nu_i} \quad \text{or} \tag{8}$$

$$\Delta G^0 = -RT \ln \Pi a_i^{\nu_i} . \tag{8a}$$

It will be seen from this expression that the use of activities instead of chemical potentials is equivalent to the introduction of a reaction constant

$$K = \Pi a_i^{\nu_i} \tag{9}$$

which in the case of reaction (1) will take the particular value

$$K_1 = a_{Fe}^1 a_{CO}^1 a_{FeO}^{-1} a_C^{-1} . \tag{10}$$

If, according to accepted principles, we define a gas in the ideal state and at a pressure of 1 atmosphere to be in a state of unit activity then, as long as the ideal-gas laws apply, the activity of the gas will be equal to its pressure, and equation (10) becomes

$$K_1 = \frac{p_{CO} \cdot a_{Fe}}{a_{FeO} \cdot a_C} . \tag{11}$$

To compute the reaction constant by means of this expression we require to know the value of the activities of Fe, FeO and C in liquid steel. On the base of certain simplifying assumptions the activities of Fe and FeO will be deduced from experimental data in section 3, and that of C will be deduced in section 4; in section 5 the results will be combined for the calculation of K_1. Throughout these calculations unit activity will be assigned to liquid Fe, liquid FeO and graphite at a pressure of 1 atmosphere.

Rewriting (8a) as

$$\Delta G^0 = -RT \ln K \tag{8b}$$

and bearing in mind that

$$\Delta G^0 = \Delta H^0 - T\Delta S^0 , \tag{12}$$

it will be seen that the value of K_1 can also be derived from the reaction enthalpy ΔH and the reaction entropy ΔS, if the reaction is considered as taking place between the substances in their states of unit activity; ΔH^0 and ΔS^0 are the changes in enthalpy and entropy, if at a constant value of the temperature one mole liquid FeO and one mole graphite are transformed into one mole liquid Fe and one mole CO of atmospheric pressure.

In section 5 of this paper K_1 will be computed from equations (8b) and (12) and the result will be compared with the values derived from (11).

As we pointed out on page 206 the four reactions (1) to (4) are not inde-

pendent, and since the free enthalpy is an additive quantity we will have

$$\Delta G_3^0 = \Delta G_2^0 - \Delta G_1^0 , \tag{13}$$

$$\Delta G_4^0 = 2\,\Delta G_1^0 - \Delta G_2^0 , \tag{14}$$

or, by (8b)

$$K_3 = K_2/K_1 , \tag{15}$$

$$K_4 = K_1^2/K_2 . \tag{16}$$

In sections 4 and 5, K_4 will be computed as well as K_1 and, if required, K_2 and K_3 may be computed from these two.

3. *The activity of ferrous oxide, dissolved in liquid iron*

3a. *Theoretical postulates*

That part of the iron-oxygen phase diagram which has a bearing on our present problem has been reproduced in *fig. 1* [1]). The position of the line

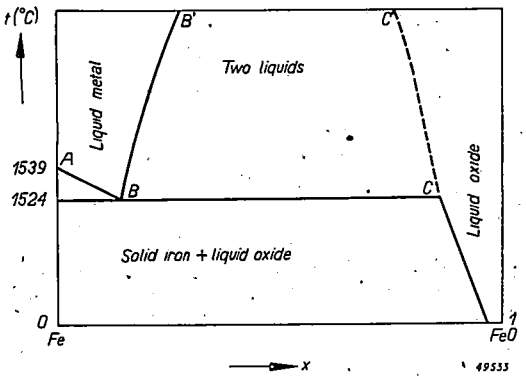


Fig. 1. Schematic phase diagram of the system Fe-FeO in the neighbourhood of the melting point of iron.

BB' which represents the solubility of FeO in Fe has been determined with fair accuracy, but the position of the line CC' giving the solubility of Fe in FeO is much less certain. *) It is evident, however, that both solubilities are small and that they both increase with increasing temperature. The solubility of oxygen in solid iron is so slight that it has been entirely neglected.

As no data are available concerning the changes in entropy and enthalpy upon mixing liquid Fe with liquid FeO, we will investigate this mixture

*) Darken and Gurry [2]) give the coordinates of point C as 1524 °C and 22·60 % O, the composition of the liquid then being $Fe_{0.98}O$ or $FeO_{1.02}$; hence, according to these investigators, FeO of purely stoichiometric composition does exist in the liquid no more than in the solid state. For the purpose of this section the point is, however, of no consequence, since we will only be considering phases represented by points at the left of the line BB', that is, with very small contents of oxygen.

on the base of two simple suppositions which are frequently made under similar circumstances when the mixing of two solids or two fluids is being considered [3]). We will assume that the heat to be supplied in mixing x moles FeO with $(1-x)$ moles Fe, at constant pressure and temperature, is a so-called Van Laar heat (enthalpy) of mixing, given by

$$\Delta H = Cx(1-x), \tag{17}$$

where the constant C is to be determined from experiment, and that the changes in entropy involved in the mixing process are represented by the terms of the Gibbs paradox *)

$$\Delta S = -R\,[x \ln x + (1-x) \ln (1-x)]\,. \tag{18}$$

The free enthalpy of mixing $\Delta G = \Delta H - T\Delta S$ will then be

$$\Delta G = Cx(1-x) + RT\,[x \ln x + (1-x) \ln (1-x)]\,. \tag{19}$$

In *fig. 2b* the parabola represents ΔH as a function of x, and the other curves represent ΔG for different values of the ratio RT/C, the ΔH-curve coinciding, of course, with the ΔG-curve for $T = 0$. *Fig. 2a* shows $-\Delta S$ as a function of x.

For values of T between zero and $C/2R$ the ΔG-curves show two minima which gradually approach one another with increasing temperatures; finally, for $RT/C = \frac{1}{2}$, they merge into a single minimum at the centre. When equilibrium has been established, the free enthalpy must be a minimum and this requires that mixtures of a composition corresponding to a point at the left of point P or at the right of point Q will be stable, where PQ is the common tangent of both branches of the ΔG-curve (see fig. 2). On the other hand, mixtures corresponding to points between P and Q will be unstable and will separate into two different phases of compositions given by the points of contact, P and Q, coinciding with the two minima in the special case under consideration.

Hence the position of the curves BB' and CC' in fig. 1 has to correspond with the position of points P and Q in fig. 2. Owing to the symmetry of our basic formulae, (17) and (18), with respect to x and $(1-x)$, points P and Q will lie symmetrically with respect to $x = \frac{1}{2}$ and, as we now see, this implies that the solubility of FeO in Fe is equal to the solubility of Fe in FeO.

The exact position of the tangent points P and Q may be deduced by putting the first derivative of (19) equal to zero:

$$\frac{\mathrm{d}(\Delta G)}{\mathrm{d}x} = C(1-2x) + RT\,\{\ln x - \ln (1-x)\} = 0\,, \tag{20}$$

*) See appendix at the end of the article.

which leads to

$$T = \frac{C}{R} \frac{1-2x}{\ln \frac{1-x}{x}}, \tag{21}$$

where x now stands for the solubility. Inserting the numerical value of the gas constant (in cal/degree) and introducing 10-base logarithms we obtain

$$T = \frac{C}{4{\cdot}575} \frac{1-2x}{\log \frac{1-x}{x}}. \tag{22}$$

The ΔG-curves in fig. 2 having two minima will also have two points of inflexion, that is, points where the second derivative disappears. With increasing temperatures these points of inflexion will move towards each other finally becoming coincident at the centre, that is, at $x = \frac{1}{2}$. Differentiating (20) a second time we obtain

$$\frac{d^2(\Delta G)}{dx^2} = -2\,C + RT\left(\frac{1}{x} + \frac{1}{1-x}\right),$$

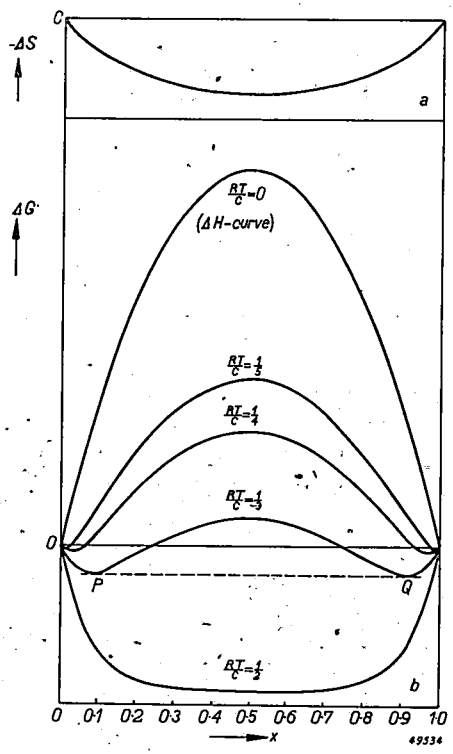


Fig. 2. Negative entropy of mixing $-\Delta S$, enthalpy of mixing ΔH and free enthalpy of mixing ΔG as functions of the composition if ΔH and ΔS are given by (17) and (18).

and, by putting $x = \frac{1}{2}$ and making this expression equal to zero, we obtain the critical temperature T_k as

$$T_k = C/2R. \tag{23}$$

At this temperature the miscibility curve of fig. 1 (BB', CC') will reach its maximum and above this temperature both components, Fe and FeO, will be miscible in all proportions.

Our basic assumption with regard to the heat evolved in mixing liquid Fe and liquid FeO has been expressed in (17) in terms of the molar fractions x and $(1-x)$. Instead of this so-called integral heat of mixing it is frequently convenient to consider the differential heats of mixing Δh_{FeO} and Δh_{Fe}, where Δh_{FeO}, for instance, represents the heat to be supplied if at constant values of temperature and pressure one mole of FeO is dissolved in a volume of the liquid so large that the molar concentrations do not change appreciably during the process. Obviously we have

$$\Delta h_{FeO} = h_{FeO} - h^0_{FeO}, \tag{24}$$

h_{FeO} being the partial molar enthalpy of FeO in the solution and h^0_{FeO} the molar enthalpy of pure liquid FeO.

Now it may be remembered that between the integral and the differential quantities the following relations subsist:

$$\Delta h_{FeO} = \Delta H + (1-x)\frac{\partial(\Delta H)}{\partial x}, \tag{25}$$

$$\Delta h_{Fe} = \Delta H - x\frac{\partial(\Delta H)}{\partial x}, \tag{26}$$

which upon introducing the value of ΔH as given by (17) yield

$$\Delta h_{FeO} = C(1-x)^2, \tag{27}$$

$$\Delta h_{Fe} = Cx^2. \tag{28}$$

The same equations will hold in the case of the entropy, *viz.*

$$\Delta s_{FeO} = \Delta S + (1-x)\frac{\partial(\Delta S)}{\partial x}, \tag{29}$$

$$\Delta s_{Fe} = \Delta S - x\frac{\partial(\Delta S)}{\partial x}. \tag{30}$$

which with (18) combine to

$$\Delta s_{FeO} = -R \ln x, \tag{31}$$

$$\Delta s_{Fe} = -R \ln (1-x). \tag{32}$$

Finally, the changes in the chemical potential accompanying the differential solution of FeO and Fe are deducible from

$$\Delta\mu_{FeO} = \Delta G + (1-x)\frac{\partial(\Delta G)}{\partial x}, \tag{33}$$

$$\Delta\mu_{Fe} = \Delta G - x\frac{\partial(\Delta G)}{\partial x}, \tag{34}$$

and, inserting (19) in these expressions, we obtain

$$\Delta\mu_{FeO} = C(1-x)^2 + RT\ln x, \tag{35}$$

$$\Delta\mu_{Fe} = Cx^2 + RT\ln(1-x), \tag{36}$$

a result that might also have been obtained by combining, respectively, (27) with (31) and (28) with (32), since

$$\Delta\mu = \Delta h - T\Delta s. \tag{37}$$

By making use of the relation $\Delta\mu = \mu-\mu^0$, we obtain from (35), (36) and (7) the following expressions for the activities of FeO and Fe in their liquid mixtures:

$$\ln a_{FeO} = \frac{C(1-x)^2}{RT} + \ln x, \tag{38}$$

$$\ln a_{Fe} = \frac{Cx^2}{RT} + \ln(1-x). \tag{39}$$

These are well-known formulae, which hold for all binary mixtures that possess an ideal (Gibbs) entropy of mixing and a Van Laar heat of mixing (formula (17)).

The states of unit activity to which equations (38) and (39) refer are those of pure liquid Fe and pure liquid FeO; as standard for FeO we might, however, as well have chosen a liquid of a slightly different composition, for instance $Fe_{0.95}O$, since in the region under consideration (at the left of the line BB' in fig. 1) this will entail only negligible changes in the values of x.

3b. *Comparison with experimental data*

If the basic equations (17) and (18) from which we started in the previous section actually apply to liquid iron-ferrous oxide mixtures we must expect to obtain a straight line through the origin, when plotting $(1-2x)/\log\frac{1-x}{x}$ against the temperature T, where x now stands for the solubility of FeO in Fe (equation (22)).

The most accurate determinations of the solubility of FeO (or O) in molten Fe have been made by Chipman and Fetters[4]), and as many

of our readers will not have direct access to the journal in which their data have been published, a full record of their observations has been reproduced in table I; the molar fraction x and the value of the function $(1-2x)/\log\frac{1-x}{x}$ computed from these observations have been added in the third and the fourth columns.

TABLE I

Solubility of oxygen in liquid iron according to the observations of Chipman and Fetters[4]).

temp. °K	wt % O	mol. fr. FeO (x)	$\frac{1-2x}{\log\frac{1-x}{x}}$	temp. °K	wt % O	mol. fr. FeO (x)	$\frac{1-2x}{\log\frac{1-x}{x}}$
1798	0·184	0·0064	0·451	1892	0·241	0·0084	0·475
1818	0·183	0·0064	0·450	1893	0·232	0·0081	0·471
1845	0·186	0·0065	0·452	1894	0·241	0·0084	0·475
1848	0·203	0·0071	0·459	1896	0·240	0·0084	0·474
1850	0·192	0·0067	0·454	1896	0·223	0·0078	0·468
1851	0·181	0·0063	0·449	1903	0·238	0·0083	0·473
1863	0·216	0·0075[5]	0·465	1916	0·264	0·0092	0·483
1865	0·223	0·0078	0·468	1927	0·245	0·0085[5]	0·476
1866	0·229	0·0080	0·470	1934	0·262	0·0091[5]	0·482
1873	0·231	0·0081	0·471	1943	0·267	0·0093	0·484
1873	0·211	0·0074	0·463	1953	0·277	0·0097	0·488
1873	0·215	0·0075	0·464	1974	0·308	0·0107[5]	0·498
1888	0·220	0·0077	0·467	1988	0·300	0·0105	0·496
1890	0·219	0·0076[5]	0·466	2083	0·384	0·0134	0·521

Because these observations are of a rather difficult nature, a high accuracy cannot be expected but, as will be seen from *fig. 3*, a plot of $(1-2x)/\log\frac{1-x}{x}$ against T can satisfactorily be represented by a straight line through the origin, as predicted by theory; numerically this relation is expressed by

$$T = 4000\ (1-2x)/\log\frac{1-x}{x}, \qquad (40)$$

from which it follows that the constant C in equations (17), (22) and (38) has the value

$$C = 4000 \times 4{\cdot}575 = 18300 \text{ cal}, \qquad (41)$$

and that the critical temperature is

$$T_k = C/2R = \text{appr. } 4600 \text{ °K},$$

a value of little interest, since a) the boiling point of iron must be

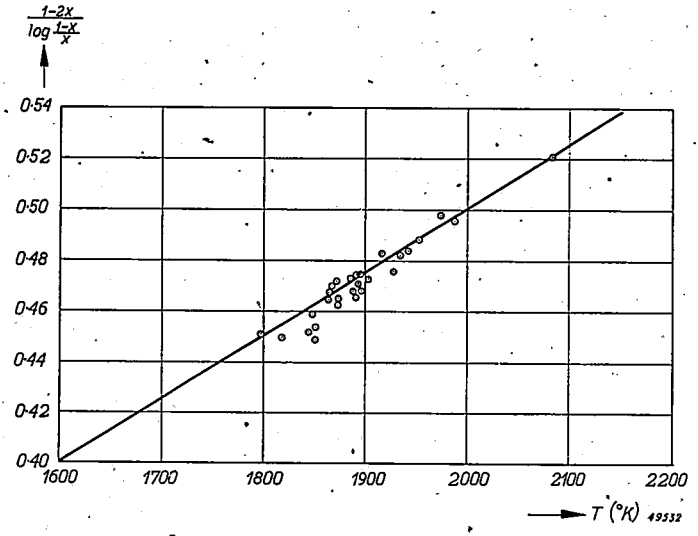


Fig. 3. Plot of $(1-2x)/\log\frac{1-x}{x}$ *versus* absolute temperature. The continuation of the straight line passes through the origin.

somewhere in the neighbourhood of 2700 to 2800 °K, according to most of the data now available, and b) the deviation from Van Laar-Gibbs behaviour may be considerable at high concentrations.

To conclude this section the solubilities of FeO as computed from equation (40) for temperatures between the melting point and the boiling point of iron have been entered in table II.

TABLE II

Solubility of oxygen in liquid iron as computed from equation (40).

T (°K)	$\frac{1-2x}{\log\frac{1-x}{x}}$	x	wt % O
1800	0·450	0·0063[5]	0·182
1850	0·462[5]	0·0073[5]	0·210
1900	0·475	0·0084[5]	0·241
1950	0·487[5]	0·0096[5]	0·275
2000	0·500	0·0109[5]	0·312
2100	0·525	0·0139	0·396
2200	0·550	0·0173	0·493
2300	0·575	0·0211	0·603
2400	0·600	0·0255	0·727
2500	0·625	0·0305	0·866
2600	0·650	0·0360	1·021
2700	0·675	0·0421	1·192

4. *The activity of carbon in liquid iron*

From a comparison of the reaction of Boudouard,

$$C + CO_2 \rightleftarrows 2\,CO\,, \tag{42}$$

where C represents graphite, and the reaction

$$[C] + CO_2 \rightleftarrows 2\,CO, \tag{4}$$

where [C] stands for carbon dissolved in liquid iron, it is evident that the activity of carbon in iron with reference to graphite is given by

$$a_C = \frac{(p^2_{CO}/p_{CO_2}) \text{ in equilibrium with [C] in Fe}}{(p^2_{CO}/p_{CO_2}) \text{ in equilibrium with graphite}}. \tag{43}$$

For the melting temperature of iron the numerator of this expression follows at once from the equilibrium pressures of CO and CO_2 in contact with liquid iron as observed by Marshall and Chipman (see page 217). The denominator of (43) cannot, however, be computed in such a simple manner, since the Boudouard reaction has only been studied at much lower temperatures from which a reliable extrapolation to higher temperatures cannot be carried out, owing to the rather exceptional temperature variation of the reaction entropy and the reaction enthalpy. Moreover, the results of several authors are not concordant and as a consequence we have preferred to make an *a priori* calculation of the reaction constant of the Boudouard equilibrium using calorimetric and spectroscopic data.

For the three substances entering in equation (42) the standard values of the enthalpy of formation, ΔH^0, in cal/mole and of the entropy; s^0, in cal/(degree . mole) at 25 °C and 1 atmosphere have been compiled in table III.

TABLE III

Molar enthalpy of formation, ΔH^0, and molar entropy, s^0, of CO_2, CO, and C at 25 °C and 1 atm.

	$\Delta H^0_{298\cdot1}$ cal/mole	$s^0_{298\cdot1}$ cal/(degree . mole)
CO_2	−94030	51·07
CO	−26394	47·32
C (graphite)	0	1·40

From these data the reaction enthalpy and the reaction entropy of (42) may be computed as

$$\Delta H^0_{298\cdot1} = -2 \times 26\,394 + 94\,030 = 41\,242 \text{ cal}$$

and

$$\Delta S^0_{298\cdot1} = 2 \times 47{\cdot}32 - 51{\cdot}07 - 1{\cdot}40 = 42{\cdot}17 \text{ cal/degree}.$$

These values hold only at room temperature, but the corresponding data for higher temperatures may be obtained from them by applying a correction which may be computed from the molar enthalpies, $(h^0—h_0^0)$, and the molar entropies, $(s^0—s_0^0)$, calculated from spectroscopic data for CO [5]) and for CO_2 [6]) and from the specific heats for C [7]); these have been collected in table IV.

TABLE IV

Molar enthalpy, h^0, and molar entropy, s^0, for CO_2, CO, and C (graphite) with reference to absolute zero.

temp. °K	CO_2		CO		C	
	$h^0—h_0^0$ cal/mole	$s^0—s_0^0$ cal/(degree. mole)	$h^0—h_0^0$ cal/mole	$s^0—s_0^0$ cal/(degree. mole)	$h_0—h_0^0$ cal/mole	$s^0—s_0^0$ cal/(degree. mole)
298·1	2237	51·07	2072	47·32	258	1·40
1200	12886	66·79	8872	57·59	4135	6·84
1400	15624	68·90	10530	58·87	5239	7·69
1600	18430	70·75	12210	59·98	6378	8·45
1800	21275	72.44	13915	61·00	7551	9·14

Using these data we obtain the reaction enthalpy at higher temperatures as

$$\Delta H_T^0 = \Delta H_{298\cdot1}^0 + 2(h_T^0 - h_{298\cdot1}^0)_{CO} - (h_T^0 - h_{298\cdot1}^0)_{CO_2} - (h_T^0 - h_{298\cdot1}^0)_C, \qquad (44)$$

giving for instance in the particular case $T = 1200$ °K:

$$\Delta H_{1200}^0 = 41\,242 + 2 \times 6\,800 - 10\,649 - 3\,877 = 40\,316 \text{ cal.}$$

Similar equations hold for other temperatures and for the entropy.

The values of ΔH^0 and ΔS^0 computed in this manner have been entered in columns 2 and 3 of table V. The reaction constant is given by the equations (8b) and (12) as

$$\log K = -\frac{\Delta H^0}{4{\cdot}575\,T} + \frac{\Delta S^0}{4{\cdot}575} \qquad (45)$$

and has been entered in columns 4 and 5 ($\log K$ and K).

TABLE V

Reaction enthalpy, reaction entropy and reaction constant of the Boudouard reaction $C + CO_2 \rightleftarrows 2\,CO$.

temp. ° K	ΔH^0 cal	ΔS^0 cal/degree	$\log K$	K
1200	40316	41·55	1·739	54·8
1400	39790	41·15	2·783	601
1600	39205	40·76	3·553	3570
1800	38597	40·42	4·148	14100

By using (45) it will easily be verified that in the neighbourhood of 1800 °K we have

$$\log K_{42} = -8435/T + 8{\cdot}835\,, \qquad (46)$$

and, with the aid of this expression, the denominator of (43) can now be computed in the range of temperatures in which we are interested.

To calculate the numerator of (43) we will make use of the data of Marshall and Chipman [8]) who determined the composition of the gas phase, in equilibrium with liquid iron in which varying quantities of O and C were dissolved. For a pressure of 1 atmosphere and for carbon concentrations varying from 0·02 to 1·00 wt%, their data are reproduced in table VI.

TABLE VI

Composition of gas phase (CO, CO_2) and liquid phase (Fe, FeO and C) in equilibrium at 1813 °K and 1 atmosphere according to the experiments of Marshall and Chipman [8]).

gas phase	metallic phase		gas phase	metallic phase	
% CO_2	wt % C	wt % O	% CO_2	wt % C	wt % O
10·0	0·02	0·090	1·15	0·20	0·011
4·3	0·05	0·038	0·46	0·50	0·0052
2·25	0·10	0·020	0·23	1·00	0·0035

By combining the various data and formulae given above we have finally calculated:

a) the mole fractions of C, FeO and Fe, from table VI,

b) the activity of C, which in virtue of (43) and (46) is

$$a_C = \frac{(p^2_{CO}/p_{CO_2}) \text{ from table VI}}{15\,200}, \qquad (47)$$

c) the activity of FeO, which by (38) and (41) is

$$\log a_{FeO} = 2{\cdot}206\ (1-x)^2 + \log x\,, \qquad (48)$$

and d) the activity coefficients $k = a/x$, for C and FeO. All these data have been compiled in table VII; the activity of Fe has been omitted, since it can simply be put equal to the molar fraction x_{Fe}.

TABLE VII

Calculated values of molar fractions, activities and activity coefficients of C and FeO in liquid iron at 1813 °K.

x_C	a_C	k_C	x_{FeO}	a_{FeO}	k_{FeO}	x_{Fe}
0·00093	0·00053	0·57	0·00315	0·491	156	0·996
0·0023	0·0014	0·61	0·00133	0·211	159	0·996
0·0046	0·0028	0·61	0·00070	0·112	160	0·995
0·0093	0·0056	0·60	0·00038	0·061	160	0·990
0·023	0·014	0·61	0·00018	0·029	160	0·977
0·045	0·028	0·62	0·00012	0·019	160	0·955

5. *The equilibrium constant of the reaction* [FeO] + [C] $\rightleftarrows$ [Fe] + CO

The equilibrium constant for the reaction (1), computed by means of equation (11) from the data compiled in tables VI and VII, has been entered in table VIII. As we see from this table, K is not a constant but decreases with increasing carbon concentration; this is presumably due to a lowering of the activities of FeO by the presence of C; for the activities of FeO as derived in section 3 refer to carbon-free solutions, whereas the activities of C, computed from the experiments of Marshall and Chipman, refer to solutions containing both FeO and C.

TABLE VIII

Equilibrium "constant" of the reaction [FeO] + [C] $\rightleftarrows$ [Fe] + CO at 1813 °K, as calculated from the activities.

p_{CO}	a_{Fe}	a_{FeO}	a_C	K
0·900	0·996	0·491	0·00053	3450
0·957	0·996	0·211	0·0014	3230
0·9775	0·995	0·112	0·0028	3100
0·9885	0·990	0·061	0·0056	2860
0·9954	0·977	0·029	0·014	2400
0·9977	0·955	0·019	0·028	1790

In the region here considered, the activity coefficient of carbon is very nearly constant and equal to 0·61 as will be seen from table VII; assuming that this is the correct value and that the first value entered in table VII ($k_C = 0·57$) is due to experimental errors *), the value for the reaction constant in the top line of table VIII must be corrected to $K = 3250$. With this correction the reaction constant becomes practically constant and equal to 3240 for the two lowest carbon concentrations (0·02 and 0·05%) and this will as yet be accepted as the correct value.

To check this conclusion we will proceed to calculate K by different arguments this time starting from the reaction **)

$$FeO_s + CO \rightleftarrows Fe_s + CO_2, \qquad (49)$$

which has been the subject of many thorough investigations.

These investigations brought to light that solid FeO of purely stoichiometric composition does not exist [9]), whence it is commonly preferred to indicate the solid FeO-phase as the wüstite phase. At 1000 °C this wüstite phase is existent from 1·048 O-atoms per Fe-atom, where wüstite is in equilibrium with Fe, to 1·140 O-atoms per Fe-atom, where wüstite is in

*) This will mean that a solution containing 0·02% C and 0·09% O will be in equilibrium with a gas phase containing 9·4% CO_2 instead of 10% as observed by Marshall and Chipman.

**) In the equations that now follow the suffix s indicates the solid state and the suffix l the liquid state.

equilibrium with magnetite Fe_3O_4[10]) (see *fig. 4*). Wüstite has a crystal lattice of the sodium-chloride type, with all sites for oxygen atoms fully occupied, but a number of the iron sites unoccupied. In equilibrium with metallic iron the number of unoccupied sites is nearly independent of the temperature, the composition of the phase being $Fe_{0.95}O$ (table IX); to be more precise, equation (49) should therefore be rewritten as

$$Fe_{0.95}O_s + CO \gtrless 0.95\,Fe_s + CO_2 . \qquad (50)$$

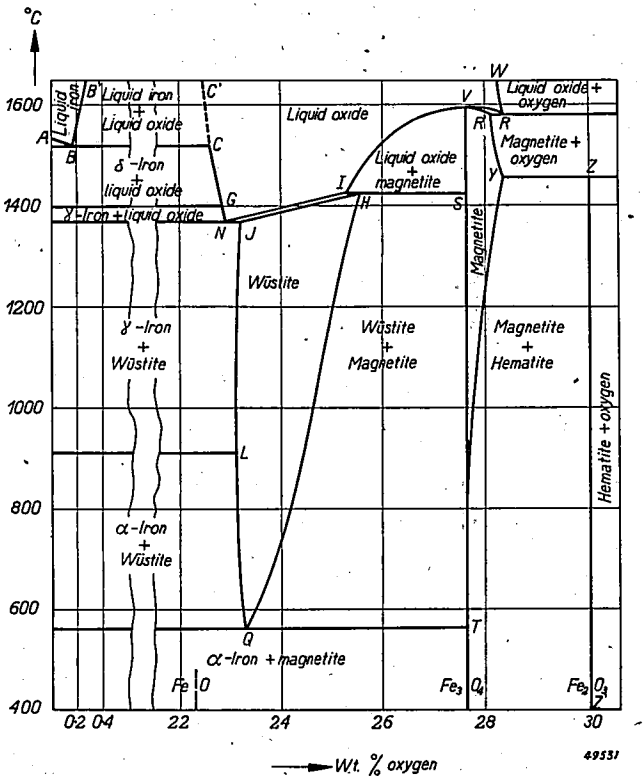


Fig. 4. The temperature-composition phase-diagram of the iron-oxygen system at a total pressure of one atmosphere ($CO_2 + CO$) or O_2. The following values for certain fixed points are recommended by Darken and Gurry[2]):

point	°C	% O	p_{CO_2}/p_{CO}	point	°C	% O	p_{O_2} (atm)
A	1535			S	1424	27·64	
B	1524	0·16	0·209	V	1597	27·64	0·0575
C	1524	22·60	0·209	R′	1583	28·07	1
G	1400	22·84	0·263	R	1583	28·30	1
N	1371	22·91	0·282	Y	1457	28·36	1
J	1371	23·16	0·282	Z	1457	30·04	1
Q	560	23·26	1·05	Z′		30·06	

When in the region from 800 to 1350 °C the logarithm of the reaction constant, $K_{50} = p_{CO_2}/p_{CO}$, of this reaction, as deduced from experiment, is plotted against the reciprocal of the absolute temperature, a straight line results as shown in *fig. 5*; this line is represented by the equation

$$\log K_{50} = 850/T - 1{\cdot}068 \quad (51)$$

In table IX the experimental values of $\log K_{50}$ [10]) have been compared with those computed from this expression.

TABLE IX

Experimental and calculated values of the equilibrium constant of the reaction $Fe_{0{\cdot}95}O + CO \rightleftarrows 0{\cdot}95\,Fe + CO_2$.

temp. °C	K (exper.)	$\log K + 1$ (exper.)	$\log K + 1$ (calcul.)	wt % O in "$Fe_{0{\cdot}95}O$"
600	0·895	0·952	—	23·22
700	0·667	0·824	—	23·15
800	0·532	0·726	0·724	23·11
900	0·460	0·663	0·657	23·10
1000	0·396	0·598	0·600	23·09
1100	0·355	0·550	0·551	23·10
1200	0·322	0·508	0·509	23·11
1300	0·297	0·473	0·472	23·13
1350	0·285	0·455	0·456	23·15

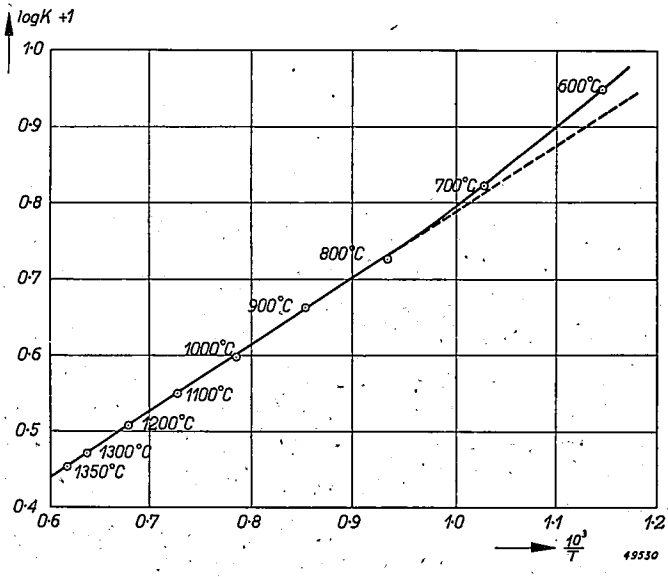


Fig. 5. Plot of the logarithm of the equilibrium constant of the reaction $Fe_{0{\cdot}95}O + CO \rightleftarrows 0{\cdot}95\,Fe + CO_2$ against the reciprocal of absolute temperature.

To change over from reaction (50) to the reaction

$$Fe_{0{\cdot}95}O_l + CO \rightleftarrows 0{\cdot}95\ Fe_l + CO_2, \qquad (52)$$

where wüstite and iron are now both in the liquid state, we require to know the thermal effects taking place during the processes of fusion.

Now, according to various investigators, the heat of fusion of iron [11]) is appr. 3600 cal/mole, the melting point 1812 °K [12]), and hence the entropy of fusion is 1·99 cal/(degree.mole). Likewise we have, according to the observations of Darken and Gurry [2]), for wüstite of the compositon $Fe_{0{\cdot}95}O$ a heat of fusion of 7230 cal/mole, and an entropy of fusion of 4·40 cal/(degree . mole).

Hence the transition from reaction (50) to reaction (52) will be accompanied by a change in the reaction enthalpy of

$$-7230 + 0{\cdot}95 \times 3600 = -3810 \text{ cal},$$

and by a change in the reaction entropy of

$$-4{\cdot}40 + 0{\cdot}95 \times 1{\cdot}99 = -2{\cdot}51 \text{ cal/degree},$$

so that from equations (45) and (51) we obtain for the reaction (52)

$$\log K_{52} = 1685/T - 1{\cdot}615\,. \qquad (53)$$

Finally, from reaction (52) we may derive the reaction constant for the reaction between liquid wüstite and solid graphite by means of the reaction of Boudouard

$$Fe_{0{\cdot}95}O_l + CO \rightleftarrows 0{\cdot}95\ Fe_l + CO_2 \qquad (52)$$

$$+ \quad C_s + CO_2 \rightleftarrows 2\ CO \qquad (42)$$

$$Fe_{0{\cdot}95}O_l + C_s \rightleftarrows 0{\cdot}95\ Fe_l + CO \qquad (54)$$

We infer from these equations that in the neighbourhood of the melting point of iron the reaction constant of (54) will be given by

$$\log K_{54} = \log K_{52} + \log K_{42}, \qquad (55)$$

or, in virtue of (53) and (46), by

$$\log K_{54} = -6750/T + 7{\cdot}22\,, \qquad (56)$$

which for $T = 1813$ °K yields the value $K = 3140$ in satisfactory agreement with the value 3240 deduced from the activities. This provides supplementary confirmation of our assumption that at 1813 °K the activity coefficient for carbon dissolved in liquid iron is 0·61 at low concentrations, and that a liquid phase containing 0·02% C and 0·09% O will at 1813 °K be in equilibrium with a gas phase containing 9·4% CO_2 instead of the 10%, given in table VI.

Perhaps it is not superfluous to note that the reaction constants for reactions between the pure substances Fe, "FeO" and C have the same values as when the substances are mixed in a liquid phase, since the pure substances were chosen as our standard states of reference.

6. *The activities of carbon and oxygen simultaneously dissolved in liquid iron*

We are now able to compute the activities of "FeO" in liquid Fe in the presence of C and to compare the resulting data with the activities in a carbon-free liquid calculated in section 5. In virtue of (52) the activity of "FeO" in the presence of carbon will be given by

$$a_{\text{"FeO"}} = \frac{a_{\text{Fe}} \cdot p_{\text{CO}_2}}{K_{52} \cdot p_{\text{CO}}}, \tag{57}$$

where K_{52} as computed from (53) assumes the value $K = 0{\cdot}206$ at 1813 °K *). From this equation and equation (38), and using the experimental data of Marshall and Chipman, corrected as indicated above, we arrived at the data presented in table X.

TABLE X

Activity of "FeO" in liquid iron containing carbon and in carbon-free liquid iron at 1813 °K.

wt% O	solutions containing carbon				carbon-free solutions	
	gas phase % CO_2	wt % C	$a_{\text{"FeO"}}$ from (57)	$k_{\text{"FeO"}}$	$a_{\text{"FeO"}}$ from (38)	$k_{\text{"FeO"}}$
0·090	9·40	0·02	0·502	159	0·491	156
0·038	4·30	0·05	0·217	163	0·211	159
0·020	2·25	0·10	0·111	159	0·112	160
0·011	1·15	0·20	0·056	145	0·061	160
0·0052	0·46	0·50	0·022	120	0·029	160
0·0035	0·23	1·00	0·011	90	0·019	160

Up to carbon concentrations of 0·1% the agreement of the two sets of "FeO" activities, computed in independent ways, is remarkably good; this is, of course, the same agreement which has previously been expressed in the close equality of the reaction constants K_{54}, noted on page 221. It will be seen, that carbon concentrations above 0·1% produce a very marked decrease in the activity coefficient of "FeO".

According to formula (38) and table X the activities of "FeO" in liquid solutions containing less than 0·1% C can satisfactorily be approximated by the expression

$$a_{\text{"FeO"}} = x e^{9210(1-x)^2/T}, \tag{58}$$

*) Strictly we should write $a_{\text{Fe}}^{0{\cdot}95}$ in (57), but the difference is of no practical consequence and hence we will stick to the simpler notation.

where x is the molar fraction of "FeO" and e is the base of the naperian logarithms.

If the concentration of FeO is very small, x may be neglected in the exponent so that we obtain

$$a_{\text{"FeO"}} = xe^{9210/T}, \tag{59}$$

which, for $T = 1813$ °K, yields

$$a_{\text{"FeO"}} = 160\, x_{\text{"FeO"}}. \tag{60}$$

At relatively low temperatures the solubility of oxygen in liquid iron is so small that (59) and (60) will approximately hold even for saturated solutions. Moreover the activity of "FeO" in the liquid will differ only slightly from unity in this case and consequently the solubility x_{sol} at 1813 °K will about equal 1/160. This may be compared with the value in table I. It follows that for non-saturated solutions at relatively low temperatures we may write

$$a_{\text{"FeO"}} \backsim x/x_{sol} \tag{61}$$

where x_{sol} is the solubility of "FeO" expressed as molar fraction (tables I and II).

At 1813 °K the activity of C is satisfactorily approximated by

$$a_C = 0{\cdot}61\, x_C \tag{62}$$

up to concentration of 1·0% C and independent of the oxygen concentration.

Detailed information concerning the structure of the liquid solutions is lacking but we know that in an electric field carbon moves in solid iron as a positive ion [13]) whereas oxygen moves in solid zirconium as a negative ion [14]). These observations suggest that carbon and oxygen may be dissolved in liquid iron in the form of positive and negative ions and this would entail mutual lowering of the activity coefficients of both components. That such a decrease in the activity coefficient has not been observed in the case of carbon is presumably a consequence of the very low oxygen concentrations which can be realized in practice owing to the low solubility of FeO in Fe.

7. *Data and formulae of technical importance*

Recapitulating the results of the foregoing sections we see that for the technically important reactions,

$$[\text{FeO}] + [\text{C}] \rightleftarrows [\text{Fe}] + \text{CO} \tag{1}$$

and

$$[\text{C}] + \text{CO}_2 \rightleftarrows 2\,\text{CO}, \tag{4}$$

the reaction constants are given by *)

$$\log K_1 = \log \frac{p_{CO} \cdot a_{Fe}}{a_{FeO} \cdot a_C} = -\frac{6750}{T} + 7{\cdot}22 \qquad (56)$$

and

$$\log K_4 = \log \frac{p_{CO}^2}{p_{CO_2} \cdot a_C} = -\frac{8435}{T} + 8{\cdot}835\,. \qquad (46)$$

In these equations we may put

$$p_{CO} = \text{CO-pressure in atmospheres},$$

$$a_{Fe} = 1\,,$$

$$a_{FeO} = x_{FeO}\, e^{9210\,(1-x)^2/T} \backsimeq \frac{x}{x_{sol}}. \qquad (58)$$

This last expression only holds for C-concentrations below 0·1 wt%, the deviations at higher concentrations being indicated in table X. Equations (56), (46) and (58) may be used over the entire range of temperatures of interest to the steelmaker (1800-2000 °K); the maximum error in K_1 or K_4 may be estimated at about 5%.

For the activity of carbon we have likewise

$$a_C = 0{\cdot}61\, x_C, \qquad (62)$$

an equation which has, so far, only been deduced for a temperature of 1813 °K. By the following arguments it is possible, however, to arrive at a fairly reliable estimate of the activity at different temperatures.

According to Basset [15]) the triple point of C is located at 4000 °K. Moreover, all but a few of the elements possess entropies of fusion lying between R and $2\,R$ (R = gas constant), so that it is reasonable to assume for C a value of $1{\cdot}5\,R$. The heat of fusion will then be $4000 \times 1{\cdot}5\,R = 6000\,R$.

In addition we will suppose that liquid carbon and liquid iron possess a Gibbs entropy of mixing and a Van Laar heat of mixing. The differential entropy of solution and the differential heat of solution for *liquid* carbon in liquid iron will then be represented by equations (27) and (31) which for the solution of *solid* carbon (graphite) in liquid iron will be transformed into

$$\Delta h_C = \quad C(1-x)^2 + 6000\,R \qquad (63)$$

and

$$\Delta s_C = -R \ln x + 1{\cdot}5\,R\,. \qquad (64)$$

*) The use of the concept of activity and the choice of definite states of unit activity implies that the constants of the reactions between the pure substances have the same values as for those between the substances dissolved in liquid iron (compare sections 2 and 5).

For the change in chemical potential accompanying the differential solution of graphite in liquid iron we have by (37)

$$\Delta\mu = C(1-x)^2 + 6000\,R + RT\ln x - 1{\cdot}5\,RT, \tag{65}$$

which by (7) may be rewritten as

$$RT\ln a = C(1-x)^2 + 6000\,R + RT\ln x - 1{\cdot}5\,RT\,. \tag{66}$$

In the case under consideration x is always very small and the term $(1-x)^2$ may be replaced by unity. And, since for $T = 1813$ °K (66) must reduce to (62), we find

$$C = -8400 \text{ cal/mole}\,, \tag{67}$$

by which (66) transforms into

$$\ln a_C = 3600/RT + \ln x_C - 1{\cdot}5 \tag{68}$$

or

$$a_C = x_C\, e^{1800/T - 1{\cdot}5}\,. \tag{68a}$$

For 2000 °K we have from this equation

$$a_C = 0{\cdot}55\,x_C\,. \tag{69}$$

We started these arguments on the assumption that the entropy of fusion of carbon was $1{\cdot}5\,R$; had we adopted the values R or $2\,R$ instead, we should have arrived at, respectively, $a_C = 0{\cdot}58\,x_C$ and $a_C = 0{\cdot}52\,x_C$, which indicates that the maximum error committed in computing the activity of C from (68) is probably not much greater than 5%.

Appendix concerning equations (17) and (18)

From an atomistic point of view acceptance of equations (17) and (18) involves at least three different assumptions. Firstly it is supposed that the heat of mixing is so small that the distribution of the molecules in the liquid is still purely a random one; secondly the internal energy is supposed to be the sum of the energies of interaction of pairs of molecules; and thirdly it is supposed that the molecules of the two substances that are mixed together differ only slightly in size and in shape.

If at a certain moment we consider the configuration of the liquid as fixed. the probability of finding a molecule of the first kind at a certain point will be $(1-x)$, and the probability of finding a molecule of the second kind will be x. Likewise, the probability of finding in two neighbouring positions a combination 1—1 will be $(1-x)^2$, the probability of finding the combination 2—2 will be x^2, and the probability of finding either 1—2 or 2—1 will be $2x(1-x)$.

Let the average number of neighbours for each molecule be z, then there will be $Nz/2$ bonds per mole, of which

$$\frac{Nz}{2}(1-x)^2 \text{ are 1—1 bonds,}$$

$$\frac{Nz}{2}x^2 \text{ are 2—2 bonds, and}$$

$$Nz\,x(1-x) \text{ are 1—2 bonds.}$$

With the aid of these expressions it is now easy to compute the enthalpy of mixing. Let us conceive the liquid as being condensed from a state of reference in which all the molecules are so far apart that they no longer exert forces on each other (an ideal-gas state), and let

the energy of interaction of neighbouring pairs of molecules 1—1, 2—2 and 1—2, in the liquid be, respectively, w_{11}, w_{22} and w_{12}, then the energy of the mixture will be *)

$$U_{12} = \frac{Nzw_{11}}{2}(1-x)^2 + \frac{Nzw_{22}}{2}x^2 + Nzw_{12}x(1-x), \tag{70}$$

and the energy set free in the separate formation of $(1-x)$ moles of liquid 1 and x moles of liquid 2 will be

$$(1-x)U_1 + xU_2 = \frac{Nzw_{11}}{2}(1-x) + \frac{Nzw_{22}}{2}x. \tag{71}$$

If it be assumed that the mixing process is additive with respect to volume, the enthalpy of mixing will be equal to the energy of mixing which is given by the difference between (70) and (71):

$$\Delta H = x(1-x)\,Nz\left\{w_{12} - \frac{w_{11} + w_{22}}{2}\right\}. \tag{72}$$

Thus the constant C in equation (17) is now seen to be equal to

$$C = Nz\left\{w_{12} - \frac{w_{11} + w_{22}}{2}\right\}. \tag{73}$$

Equation (18), too, can be deduced in a simple manner from statistical arguments. For $N(1-x)$ molecules of type 1 and Nx molecules of type 2 can be distributed over N different positions in

$$g = \frac{N!}{(Nx)!\,[N(1-x)]!} \tag{74}$$

different ways. By inserting this in the statistical definition of the entropy

$$\Delta S = k \ln g \tag{75}$$

(where k is the Boltzmann constant) and making use of Stirling's approximation

$$\ln N! = N \ln N - N, \tag{76}$$

we obtain at once for the entropy of mixing

$$\Delta S = k\,[-Nx \ln x - N(1-x)\ln(1-x)], \tag{77}$$

a formula identical with (18) since

$$kN = R. \tag{78}$$

This derivation presupposes that the contributions to the entropy from other sources (for instance those resulting from thermal agitation) do not undergo changes in the mixing process, so that the entropy of mixing is purely a consequence of the different kinds of molecules being distributed at random in the mixture.

Furthermore, it is a necessary condition that the heat of mixing does not differ much from zero, since otherwise the 1—2 bond or the other bonds will have preference and the distribution of the molecules in the mixture will no longer be a random one.

We are now in a position to specify with greater precision in how far the three assumptions mentioned at the beginning are satisfied. The term within the brackets in (73) measures the energetic advantage of 1—2 bonds over the 1—1 and 2—2 bonds. This amount of energy is according to this equation C/z for N bonds, and, putting z at about 10, this will be of the order of 2 kcal (see (41)). The thermal energy, on the other hand, per N bonds will be of the order of RT, that is *circa* 4 kcal for the range of temperatures in which we are interested. Consequently the first supposition will be satisfied.

The second condition also will probably be fulfilled.

The third condition that the different molecules should approximately be the same in size and shape will not be satisfied very closely in the case under consideration, and consequently deviations in enthalpy and entropy would be expected at concentrations higher than those that can be attained.

*) For the sake of simplicity only the interaction between direct neighbours has been taken into account; the more exact derivation of U does not alter the result fundamentaally.

Eindhoven, January 1947

REFERENCES

1) Compare *e.g.* R. Durrer, Die Metallurgie des Eisens, Berlin 1943.
2) L. S. Darken & R. W. Gurry, J. Am. Chem. Soc. **68**, 798-816, 1946.
3) Compare J. J. van Laar, Die Thermodynamik einheitlicher Stoffe und binärer Gemische, Groningen 1936;
J. H. Hildebrand, Solubility of non-electrolytes, New York 1936;
C. Wagner, Handbuch der Metallphysik **I**, 2, Leipzig 1940.
4) J. Chipman & K. L. Fetters, Trans. Amer. Soc. Metals **29**, 953-967, 1941.
See for the earlier determinations:
C. H. Herty, Min. Metallurg. Investigations, Bull. Nr 34, 1927;
F. Körber & W. Oelsen, Mitt. K.W.I. Eisenforsch. **14**, 181-204, 1932.
5) J. O. Clayton & W. F. Giauque, J. Am. Chem. Soc. **54**, 2610-2626, 1932;
H. L. Johnston & C. O. Davis, J. Am. Chem. Soc. **56**, 271-276, 1934.
6) A. R. Gordon & C. Barnes, J. Phys. Chem. **36**, 1143-1151, 1932;
L. S. Kassel, J. Am. Chem. Soc. **56**, 1838-1842, 1934.
7) L. Terebesi, Helvetica Chim. Acta **17**, 804-819, 1934.
8) S. Marshall & J. Chipman, Trans. Amer. Soc. Metals **30**, 695-746, 1942.
9) E. R. Jette & F. Foote, J. Chem. Phys. **1**, 29-36, 1933.
10) L. S. Darken & R. W. Gurry, J. Am. Chem. Soc. **67**, 1398-1412, 1945;
R. Schenck, Th. Dingmann, P. H. Kirscht & H. Wesselkock, Z. anorg. allg. Chem. **182**, 97-117, 1929.
11) See H. E. Cleaves & J. G. Thompson, The metal iron, New York 1935.
12) Wm. F. Roeser & H. T. Wensel, J. Res. Bur. Stand. **26**, 273-287, 1941.
H. E. Cleaves & J. M. Hiegel, J. Res. Bur. Stand. **28**, 643-667, 1942.
13) W. Seith & O. Kubaschewski, Z. Elektrochem. **41**, 551-558, 1935.
14) J. H. de Boer & J. D. Fast, Recueil Trav. Chim. Pays-Bas **59**, 161-167, 1940.
15) J. Basset, J. de Physique **10**, 217-228, 1939.

Philips Res. Rep. **2**, 228-240, 1947

CALCULATION OF THE INPUT IMPEDANCE OF A SPECIAL ANTENNA

by C. J. BOUWKAMP

621.396.671

Summary

A calculation is given of the input impedance of antennae consisting of a vertical wire fed against a system of two or four equal horizontal wires. The latter are placed end to end and symmetrically around the base of the antenna proper. The investigation is based on the assumption of sinusoidal current distribution. For a quarter-wave antenna the radiation resistance is found to be approximately 20 ohms, both for two-wire and for four-wire systems.

1. *Introduction*

As is well known, the radiation resistance of a centre-fed half-wave antenna of small circular cross-section is approximately 70 ohms, provided there are no obstacles in its nearby zone. Similarly, a base-fed quarter-wave antenna, when erected vertically upon a perfectly conducting plane earth, has a radiation resistance of about $\frac{1}{2} \times 70 = 35$ ohms. Although in practice the earth does not act as a perfect mirror, such ideal behaviour can be approached by applying a suitable system of grounded wires. In order to be fully effective the ground net should have linear dimensions of the order of some wavelengths. Only in this case are the driving properties of the antenna quite independent of variations in ground structure.

Similar problems are encountered in other antenna design. For instance, in order to avoid disturbances due to "antenna currents" induced in the feeder system, one might feed the quarter-wave antenna against a large circular metal plate (*fig. 1a*). Though, obviously, a circular plate of diameter $\lambda/2$ will not be sufficient to act as a perfectly conducting plane earth, it was not expected that the radiation resistance of the antenna system of fig. 1*a* would differ appreciably from 35 ohms. Measurements [1]), however, yielded values of the order of 15 to 20 ohms; thus the radiation resistance is halved once more.

The question arose whether some theory could predict such low resistance values. A possible way of approach might be provided by Schelkunoff's general antenna theory [2]), by considering the plate and antenna as parts of conical surfaces (of angles $\pi/2$, 0, respectively). We have followed a different way, however. Presumably one may replace the full plate by a suitable number of thin cylindrical wires, placed symmetrically in the plane of the former circular disc, and pointing radially outwards from the feeder ends.

When six such $\lambda/4$ wires were present the radiation resistance was again of the order of 20 ohms [1]).

This led us to the following theoretical investigation. The simplest examples of antenna systems that may show the effect mentioned are clearly those containing two or four horizontal wires of dimensions equal to those of the antenna proper. Though a rigorous theory should be able to *calculate* the current distribution along the wires, we shall here *assume* it. As follows from other antenna investigations, both practical and theoretical, the assumption of a sinusoidal current distribution can certainly be justified for very thin wires. That means, if l denotes the length and a the radius of the cross-section of the wires, then the constant

$$\Omega = 2 \ln \left(\frac{2l}{a} \right)$$

should be large ($\Omega \geqslant 10$, for instance; thus $l/a \geqslant 75$).

As usual we confine ourselves to harmonic vibrations of angular frequency ω. The time factor $\exp(j\omega t)$ will be omitted farther on.

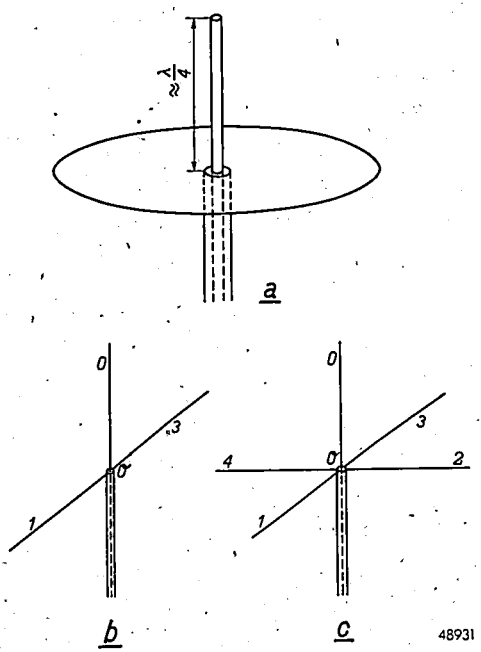


Fig. 1. Special antenna systems; quarter-wave antenna fed against
a) a circular plate,
b) two horizontal wires,
c) four horizontal wires.

In the following Z_0 denotes the input impedance of an antenna fed against an *infinitely large*, perfectly conducting, plate. Z_2 and Z_4 denote the impedances of the same antenna when fed against a system of two or four horizontal wires respectively. In section 2 a remarkable relationship between these three impedance functions is established.

2. *A relation between* Z_0, Z_2, Z_4

According to our assumption concerning the current distribution we may write for the current in any wire

$$I(s) = A\, f(s) = A \sin k(l - s)\,.$$

In this expression k $(= 2\pi/\lambda)$ denotes the wave number; A is an amplitude factor, and s is the distance along the wire measured from the centre Q of the antenna system.

The horizontal wires will be numbered 1, 3 (case a) or 1, 2, 3, 4 (case b) whilst the vertical wire is distinguished from the horizontal ones by the suffix 0.

Hence, for the currents in the two cases,

a) two-wire system:

$$\begin{aligned} I_0 &= 2A \sin k(l - s_0)\,, \\ I_n &= -A \sin k(l - s_n)\,, \qquad (n = 1,\ 3), \end{aligned}$$

b) four-wire system:

$$\begin{aligned} I_0 &= 4A \sin k(l - s_0)\,, \\ I_n &= -A \sin k(l - s_n)\,, \qquad (n = 1, 2, 3, 4). \end{aligned}$$

The currents are supposed to flow in the axes of the corresponding wires. Next the field due to these assumed currents has to be calculated, for instance by the vector-potential method. Especially required is the tangential component $E_t(s)$ of the electric field at the surface of the wires, since the complex power W can be obtained by integrating $-I(s)\,E_t(s)$ along the wires.

Let E_{mn} denote the component of the electric field parallel to, and at the surface of, the nth wire, that is due to a current $f(s)$ flowing in the mth wire. Owing to the symmetry of the systems (*figs 1b,c*) there are only three different types of functions E_{mn}. Firstly, E_{00} describes the coupling of any current with its own field. Secondly, E_{01} describes the coupling of two mutually perpendicular currents, and lastly E_{13} is characteristic for the coupling of currents in the same line.

If then E_n denotes the total electric field parallel to s_n, we readily obtain

a)
$$\begin{aligned} E_0 &= 2A\{E_{00} - E_{01}\}, \\ E_n &= A\{2E_{01} - E_{00} - E_{13}\}, \qquad (n = 1, 3), \end{aligned}$$

b)
$$\begin{aligned} E_0 &= 4A\{E_{00} - E_{01}\}, \\ E_n &= A\{2E_{01} - E_{00} - E_{13}\}, \qquad (n = 1, 2, 3, 4). \end{aligned}$$

As already mentioned, the input impedance Z can be calculated from

$$Z A_0^2 = W = -\sum_n \int_0^l I_n E_n \, ds_n ,$$

where A_0 is the current in the vertical wire, and at its base:

a) $$A_0 = 2 A \sin kl ,$$

b) $$A_0 = 4 A \sin kl .$$

Upon introducing the definitions

$$U_1 = -\int_0^l \sin k(l-s) E_{00}(s) \, ds , \tag{1}$$

$$U_2 = -\int_0^l \sin k(l-s) E_{13}(s) \, ds , \tag{2}$$

$$U_3 = -\int_0^l \sin k(l-s) E_{01}(s) \, ds , \tag{3}$$

we obtain for the input impedances Z_2, Z_4:

$$Z_2 \sin^2 kl = \tfrac{3}{2} U_1 + \tfrac{1}{2} U_2 - 2 U_3 , \tag{4}$$

$$Z_4 \sin^2 kl = \tfrac{5}{4} U_1 + \tfrac{1}{2} U_2 - \tfrac{3}{2} U_3 . \tag{5}$$

At the same time it is easily verified that Z_0 is expressible in terms of U_1, U_2, namely

$$Z_0 \sin^2 kl = U_1 - U_2 . \tag{6}$$

If the U-functions are eliminated from (4), (5), (6), the following remarkable relation is obtained:

$$Z_0 + 6 Z_2 = 8 Z_4 . \tag{7}$$

This relation holds for any frequency whenever the wires are identical. It is clearly due to properties of symmetry. Analogous identities may be found for multi-wire systems of more complicated character.

3. *The impedances in terms of the vector potentials*

In order to obtain numerical values for the impedances we have to evaluate the functions U_1, U_2, U_3. The functions U_1, U_2 do not give much difficulty as they can be expressed in terms of known functions. U_3, on the other hand, is more troublesome and requires some graphical integration.

Let us take a very thin wire along the z-axis of length l and diameter $2a$, the ends of the wire lying at $z = l$ and $z = 0$. The currents flowing in the axis are distributed according to

$$f(z) = \sin k(l-z) .$$

Let $E_z(z, r)$ be the z-component of the electric field, as a function of the cylindrical coordinates (z, r) around the wire. According to the definitions in section 2 we have evidently $(0 \leqslant s \leqslant l)$:

$$\begin{aligned} E_{00}(s) &= \quad E_z(s, a)\,, \\ E_{13}(s) &= -E_z(-s, a)\,. \end{aligned} \tag{8}$$

E_z can be derived from the vector potential which has only one component Π, *viz.*:

$$jk\,E_z = k^2\,\Pi + \frac{\partial^2 \Pi}{\partial z^2}, \tag{9}$$

$$\Pi(z, r) = \frac{1}{c}\int_0^l \sin k\,(l-\zeta)\,\frac{e^{-jk\sqrt{(z-\zeta)^2+r^2}}}{\sqrt{(z-\zeta)^2+r^2}}\,\mathrm{d}\zeta\,. \tag{10}$$

The integral (10) can be solved explicitly in terms of the well-known sine and cosine integrals, Si and Ci respectively. Upon introducing their combination

$$N(u) = Ci(u) - j\,Si(u) \tag{11}$$

one obtains

$$2\,cj\Pi(z, r) = e^{-jkl+jkz}\{N(u_1) - N(u_3)\} + e^{jkl-jkz}\{N(u_2) - N(u_4)\}, \tag{12}$$

where the arguments of the N-function are abbreviations for

$$\begin{aligned} u_1 &= k\sqrt{(z-l)^2+r^2} + kz - kl\,, \\ u_2 &= k\sqrt{(z-l)^2+r^2} - kz + kl\,, \\ u_3 &= k\sqrt{z^2+r^2} + kz\,, \\ u_4 &= k\sqrt{z^2+r^2} - kz\,. \end{aligned}$$

Let us first show how the function U_1 can be determined. From equations (1), (8), (9), we obtain

$$-jk\,U_1 = \int_0^l \left(k^2\Pi + \frac{\partial^2 \Pi}{\partial \zeta^2}\right) \sin k(l-\zeta)\,\mathrm{d}\zeta\,,$$

if it is understood that $\Pi = \Pi\,(\zeta, a)$, as given by (12). Integrating twice by parts with respect to the term $\partial^2 \Pi/\partial\zeta^2$ leads to $(z = a)$

$$-jk\,U_1 = \left\{f(\zeta)\,\frac{\partial \Pi}{\partial \zeta} - \Pi f'(\zeta)\right\}_{s=0}^{s=l},$$

and hence, when a prime in Π' denotes differentiation with respect to the first argument (z),

$$U_1 = j\Pi(l,a) - j\cos kl\;\Pi(0, a) - j\,(\sin kl/k)\,\Pi'\,(0, a)\,. \tag{13}$$

For the second U-function a similar expression holds. From

$$jk\,U_2 = \int_{-l}^{0} \left(k^2 \Pi + \frac{\partial^2 \Pi}{\partial \zeta^2}\right) \sin k(l+\zeta)\,\mathrm{d}\zeta$$

it follows that

$$U_2 = -j\,\Pi(-l, a) + j\cos kl\;\Pi(0, a) - j\,\frac{\sin kl}{k}\,\Pi'(0, a) \qquad (14)$$

Thus we have expressed U_1, U_2, in terms of the surface values of the vector potential and its derivative.

The remaining function U_3 is less tractable. E_{01}, as defined in section 2, is roughly the radial component $E_r(0, r)$ of the electric field of the current considered above. There is, however, no sense in substituting in (3) $E_{01}(s) = E_r(0, s)$ for all values of s, because then the integral (3) would diverge at $s = 0$.

Therefore the following compromise may be made: (*i*) the end of the horizontal wire does not lie at $r = 0$ but at $r = a$, (*ii*) for $a \leqslant s \leqslant l$ we take $E_{01}(s) = E_r(0, s)$. It then follows from (3) that

$$U_3 = -\int_a^l E_r(0, s) \sin k(l-s)\,\mathrm{d}s\,.$$

As the radial field E_r can be obtained from

$$jkE_r = \frac{\partial^2 \Pi}{\partial z \partial r},$$

we have thus ($z = 0$)

$$-jk\,U_3 = \int_a^l \frac{\partial^2 \Pi}{\partial z \partial s} \sin k(l-s)\,\mathrm{d}s\,.$$

Furthermore, a partial integration leads to

$$U_3 = -\frac{j \sin kl}{k}\,\Pi'(0, a) + j\int_a^l \Pi'(0, s) \cos k(l-s)\,\mathrm{d}s\,. \qquad (15)$$

A mere glance at (13), (14), (15) shows that the coefficients of $\Pi'(0, a)$ are mutually equal. Moreover, in the right-hand side of any of the expressions (4), (5), (6) the sum of the numerical coefficients of the U's vanishes. Therefore, the final expressions for the impedances will not contain terms with $\Pi'(0, a)$. This was to be expected, as the impedances of thin-wired antennae are generally of the order of $\ln a$, whilst, on the contrary, $\Pi'(0, a)$ becomes large of the order of l/a for $a \to 0$.

Introducing the values of the U-functions into (6), (4), (5) we obtain, respectively,

$$Z_0 = \frac{j}{\sin^2 kl}\{\Pi(l, a) + \Pi(-l, a) - 2\cos kl\ \Pi(0, a)\} \tag{16}$$

$$Z_2 = \frac{j}{\sin^2 kl}\{\tfrac{3}{2}\Pi(l, a) - \tfrac{1}{2}\Pi(-l, a) - \cos kl\ \Pi(0,a) - 2P\}, \tag{17}$$

$$Z_4 = \frac{j}{\sin^2 kl}\{\tfrac{5}{4}\Pi(l, a) - \tfrac{1}{4}\Pi(-l, a) - \cos kl\ \Pi(0, a) - \tfrac{3}{2}P\}, \tag{18}$$

where P is an abbreviation for

$$P = \int_a^l \Pi'(0, s)\cos k(l-s)\,\mathrm{d}s\,. \tag{19}$$

4. *Further transformation of the impedance expressions*

It is convenient to measure the antenna dimensions electrically, thus in radian units. This can be achieved by the introduction of dimensionless quantities:

$$x = kl = \frac{2\pi l}{\lambda},$$

$$y = ka = \frac{2\pi a}{\lambda},$$

$$\Omega = 2\ln\left(\frac{2l}{a}\right) = 2\ln\left(\frac{2x}{y}\right).$$

Then from (12)

$$2\,cj\Pi(l, a) = 2\,N(y) - N(\sqrt{x^2+y^2} - x) - N(\sqrt{x^2+y^2} + x), \tag{20}$$

$$2\,cj\Pi(-l, a) = e^{-2jx}[N(\sqrt{4x^2+y^2} - 2x) - N(\sqrt{x^2+y^2} - x)] + e^{2jx}[N(\sqrt{4x^2+y^2} + 2x) - N(\sqrt{x^2+y^2} + x), \tag{21}$$

$$2\,cj\Pi(0, a) = e^{-jx}[N(\sqrt{x^2+y^2} - x) - N(y)] + e^{jx}[N(\sqrt{x^2+y^2} + x) - N(y)]\,. \tag{22}$$

Further, as the function $N(u)$ becomes infinitely large for $u \to 0$ as $\ln u$, we define (γ is Euler's constant $= 0{\cdot}5772\ldots$)

$$E(u) = \gamma + \ln u - N(u) = C(u) + jS(u)\,,$$

where the notations $S(u)$, $C(u)$ are the same as adopted in a previous paper [3]):

$$S(u) = \int_0^u \frac{\sin t}{t}\,\mathrm{d}t, \quad C(u) = \int_0^u \frac{1-\cos t}{t}\,\mathrm{d}t\,.$$

Now let y tend to zero in the formulae (20), (21), (22). Furthermore, let only terms of the order of $\ln y$ be taken into account, small terms of the order y being disregarded. Then one will get

$$2\,cj\Pi(l, a) \approx E(2x), \tag{23}$$

$$2\,cj\Pi(-l, a) \approx j \ln 4 \sin 2x + e^{2jx}\{E(2x) - E(4x)\}, \tag{24}$$

$$2\,cj\Pi(0, a) \approx j\Omega \sin x - e^{jx} E(2x). \tag{25}$$

The integral P of (19) has to be transformed in a similar way. Upon introducing $ks = \xi$ in (12), we obtain, after performing the differentiation in $\Pi'(0, s)$:

$$2\,cj\Pi'(0, s) = 2\,jk \sin x \frac{e^{-j\xi}}{\xi} + jk\,e^{-jx}\{N(\sqrt{x^2+\xi^2} - x) - N(\xi)\} -$$

$$- jk\,e^{jx}\{N(\sqrt{x^2+\xi^2} + x) - N(\xi)\},$$

and consequently, from (19),

$$2\,cP = 2 \sin x \int_y^x \cos(x-\xi)\left[\frac{e^{-j\xi}}{\xi} + j\,N(\xi)\right] d\xi +$$

$$+ \int_y^x \cos(x-\xi)\,[e^{-jx} N(\sqrt{x^2+\xi^2} - x) - e^{jx} N(\sqrt{x^2+\xi^2} + x)]\, d\xi.$$

The first integral on the right-hand side can be solved explicitly, with the result:

$$j \sin(x-y)\, N(y) + e^{jx}\{N(2x) - N(2y)\}.$$

The second integral can be transformed by a partial integration, leading to:

$$\sin(x-y)\,[e^{-jx} N(\sqrt{x^2+y^2} - x) - e^{jx} N(\sqrt{x^2+y^2} + x)] +$$

$$+ 2x \int_y^x \frac{\sin(x-\xi)}{\xi} \frac{e^{-j\sqrt{x^2+\xi^2}}}{\sqrt{x^2+\xi^2}}\, d\xi.$$

Probably the remaining integral cannot be expressed in tabulated functions. Thus we have to use some graphical method. As the integrand becomes large at the lower limit $\xi = y$, it is convenient to remove first this singularity. To that end we define

$$F(x) = \int_y^x \left\{\sin x\, e^{-jx} - \sin(x-\xi) \frac{x e^{-j\sqrt{x^2+\xi^2}}}{\sqrt{x^2+\xi^2}}\right\} \frac{d\xi}{\xi}. \tag{26}$$

In terms of this function the remaining integral can be expressed as

$$2x \int_y^x \frac{\sin(x-\xi)}{\xi} \frac{e^{-j\sqrt{x^2+\xi^2}}}{\sqrt{x^2+\xi^2}}\, d\xi \approx 2 \sin x\, e^{-jx} \ln\left(\frac{x}{y}\right) - 2\,F(x).$$

We then readily obtain the final expression for the integral P, *viz.*:

$$2\,c\,P \approx \tfrac{1}{2}\,\Omega \sin 2x - 2\,F(x) - \ln 4 \sin 2x - \sin x\, e^{jx}\,E(2x)\,, \qquad (27)$$

and next the formulae for the impedances. In order to obtain the latter in practical units (ohms) we have to take $c = \frac{1}{3\,0}$. Substitution of (23), (24), (25), (27), in (16), (17), (18), gives

$$\tfrac{1}{1\,5}\,Z_0 \sin^2 x = -j\Omega^* \sin 2x + 2\,[E(2x) + e^{2jx}\{E(2x) - \tfrac{1}{2}\,E(4x)\}], \qquad (28)$$

$$\tfrac{1}{1\,5}\,Z_2 \sin^2 x = -\tfrac{3}{2}\,j\Omega^* \sin 2x + E(2x) + e^{2jx}\{E(2x) + \tfrac{1}{2}E(4x)\} + 4jF(x)\,, \qquad (29)$$

$$\tfrac{1}{1\,5}\,Z_4 \sin^2 x = -\tfrac{5}{4}\,j\Omega^* \sin 2x + E(2x) + e^{2jx}\{E(2x) + \tfrac{1}{4}E(4x)\} + 3jF(x)\,, \qquad (30)$$

where Ω^* denotes the new characteristic constant

$$\Omega^* = \Omega - \ln 4 = 2 \ln (l/a)\,.$$

We may notice here that our theoretical impedance formulae have only sense in the range $0 \leqslant x \leqslant 2$ (they are at least invalid in the neighbourhood of anti-resonance). Moreover, the wires must be thin ($\Omega \geqslant 10$). If these conditions are not fulfilled the "classical" method, based upon the *assumption* of sinusoidal currents, is known to be insufficient.

5. *Evaluation of the definite integral for* $F(x)$

Upon introducing a new variable of integration $\xi = xt$ in (26) one will get

$$F(x) = e^{-jx}\{H(x) + G(x)\}\,,$$

where

$$H(x) = \int_0^1 \frac{\sin x - \sin (x - xt)}{t}\,\mathrm{d}t = C(x) \sin x + S(x) \cos x\,,$$

$$G(x) = \int_0^1 \sin (x - xt)\left\{1 - \frac{e^{jx(1-\sqrt{1+t^2})}}{\sqrt{1+t^2}}\right\}\frac{\mathrm{d}t}{t}. \qquad (31)$$

Thus

$$F(x) = e^{-jx}\{C(x) \sin x + S(x) \cos x + G(x)\}. \qquad (32)$$

The definite integral (31) can be easily evaluated graphically because (*i*) the limits of integration are independent of x, (*ii*) the integrand vanishes at the limits, and (*iii*) for $x \approx \pi/2$ (resonance) the integrand is small throughout and is only slightly dependent on x.

It appears that the real as well as the imaginary part of $G(x)$ is practically a linear function of x, such that within one per cent the following approximation holds:

$$G(x) \approx (0{\cdot}0712 + 0{\cdot}1530\,j)\,x - (0{\cdot}0012 + 0{\cdot}0905\,j), (1 \leqslant x \leqslant 2). \qquad (33)$$

More accurate values of $G(x)$ and values of $F(x)$ are shown in table I.

TABLE I

x	Re $G(x)$	Im $G(x)$	Re $F(x)$	Im $F(x)$
1·00	0·0697	0·0654	0·4779	—0·6233
1·05	0·0733	0·0717	0·4568	—0·6523
1·10	0·0768	0·0783	0·4326	—0·6773
1·15	0·0805	0·0850	0·4055	—0·6980
1·20	0·0840	0·0919	0·3762	—0·7138
1·25	0·0876	0·0990	0·3451	—0·7247
1·30	0·0912	0·1063	0·3131	—0·7303
1·35	0·0948	0·1137	0·2805	—0·7305
1·40	0·0984	0·1212	0·2481	—0·7253
1·45	0·1019	0·1290	0·2166	—0·7146
1·50	0·1056	0·1368	0·1867	—0·6989
1·55	0·1092	0·1448	0·1589	—0·6778
1.60	0·1127	0·1529	0·1351	—0·6516
1·65	0·1164	0·1610	0·1122	—0·6209
1·70	0·1199	0·1692	0·0945	—0·5859
1·75	0·1235	0·1774	0·0813	—0·5468
1·80	0·1271	0·1857	0·0730	—0·5044
1·85	0·1308	0·1940	0·0701	—0·4593
1·90	0·1344	0·2023	0·0730	—0·4119
1·95	0·1380	0·2107	0·0822	—0·3629
2·00	0·1416	0·2190	0·0976	—0·3131

6. *Tables for the impedance functions. Example*

Table II shows numerical values of the three auxiliary functions F_0, F_2, F_4, as defined by

$$F_0(x) = 30\,[E(2x) + e^{2jx}\{E(2x) - \tfrac{1}{2}\,E(4x)\}]\,, \tag{34}$$

$$F_2(x) = 15\,[E(2x) + e^{2jx}\{E(2x) + \tfrac{1}{2}\,E(4x)\} + 4\,j\,F(x)]\,, \tag{35}$$

$$F_4(x) = 15\,[E(2x) + e^{2jx}\{E(2x) + \tfrac{1}{4}\,E(4x)\} + 3\,j\,F(x)]\,. \tag{36}$$

They describe the parts of the impedances not depending upon the wire diameter, *viz.*:

$$Z_0 = -60\,j \ln\left(\frac{l}{a}\right)\cot x + \frac{F_0(x)}{\sin^2 x}, \tag{37}$$

$$Z_2 = -90\,j \ln\left(\frac{l}{a}\right)\cot x + \frac{F_2(x)}{\sin^2 x}, \tag{38}$$

$$Z_4 = -75\,j \ln\left(\frac{l}{a}\right)\cot x + \frac{F_4(x)}{\sin^2 x}. \tag{39}$$

TABLE II

x	$F_0(x)$	$F_2(x)$	$F_4(x)$
1·00	8·166 + 33·507 j	4·365 + 63·158 j	4·294 + 51·557 j
1·05	9·720 + 33·054 j	5·232 + 59·164 j	5·139 + 48·504 j
1·10	11·451 + 32·409 j	6·210 + 54·612 j	6·089 + 45·010 j
1·15	13·365 + 31·605 j	7·313 + 49·548 j	7·155 + 41·112 j
1·20	15·462 + 30·657 j	8·528 + 44·058 j	8·329 + 36·876 j
1·25	17·745 + 29·583 j	9·881 + 38·210 j	9·629 + 32·355 j
1·30	20·448 + 28·263 j	11·243 + 32·169 j	10·988 + 27·660 j
1·35	22·866 + 27·144 j	12·969 + 25·802 j	12·585 + 22·744 j
1·40	25·695 + 25·833 j	14·723 + 19·421 j	14·254 + 17·794 j
1·45	28·686 + 24·489 j	16·602 + 13·044 j	16·037 + 12·844 j
1·50	31·848 + 23·142 j	18·651 + 6·774 j	17·969 + 7·973 j
1·55	35·148 + 21·816 j	20·824 + 0·702 j	20·012 + 3·254 j
1·60	38·589 + 20·529 j	23·130 − 5·016 j	22·171 − 1·196 j
1·65	42·150 + 19·317 j	25·584 − 10·502 j	24·457 − 5·462 j
1·70	45·804 + 18·201 j	28·167 − 15·437 j	26·851 − 9·303 j
1·75	49·551 + 17·202 j	30·872 − 19·827 j	29·348 − 12·720 j
1·80	53·355 + 16·343 j	33·704 − 23·592 j	31·947 − 15·651 j
1·85	57·207 + 15·645 j	36·668 − 26·648 j	34·652 − 18·030 j
1·90	61·074 + 15·129 j	39·740 − 28·947 j	37·439 − 19·819 j
1·95	64·935 + 14·805 j	42·917 − 30·404 j	40·304 − 20·952 j
2·00	68·763 + 14·691 j	46·193 − 30·999 j	43·240 − 21·413 j

The real parts of these F_0, F_2, F_4 indicate twice the amounts of energy radiated into free space per second with reference to unit current amplitude. The physical input resistances (for unit input current) are obtained when the real parts of the F-functions are divided by $\sin^2 x$. It is to be noticed that $F_0(x) + 6\,F_2(x) = 8\,F_4(x)$, in consequence of (7).

Figs 2 and *3* show the curves for real and imaginary parts of the three different input impedances for $l/a = 75$. Fig. 2 clearly shows the decrease in radiation resistance with respect to R_0. Moreover the amount is of the order of 20 ohms if the antennae are tuned in their first resonance ($x \approx \pi/2$). The resistance values seem rather independent of the number of horizontal wires applied. This is obviously not true for the reactive parts; consequently the tuning conditions, *e.g.* the resonance shortening, may be slightly different for different numbers of horizontal wires.

An antenna is resonant if its dimensions are chosen such that, for the applied frequency, the imaginary part of Z vanishes. The resonance shortening in per cent for the antennae under consideration is

$$p = -\frac{\Delta l}{l} \times 100 = 100\left(1 - \frac{2}{\pi} x_{\text{res}}\right). \tag{40}$$

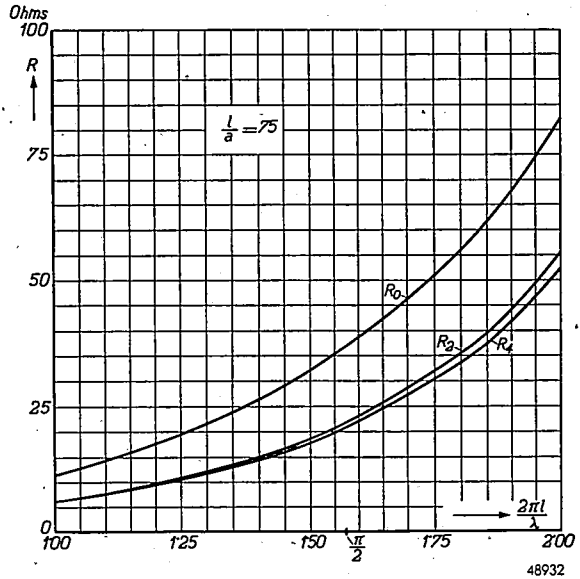


Fig. 2. Radiation-resistance curves:
R_0: antenna upon perfectly conducting plane earth (that of fig. 1*a* with infinitely large plate), R_2: antenna of fig. 1*b*, R_4: antenna of fig. 1*c*.
Antenna and horizontal wires are equal in length and in diameter.

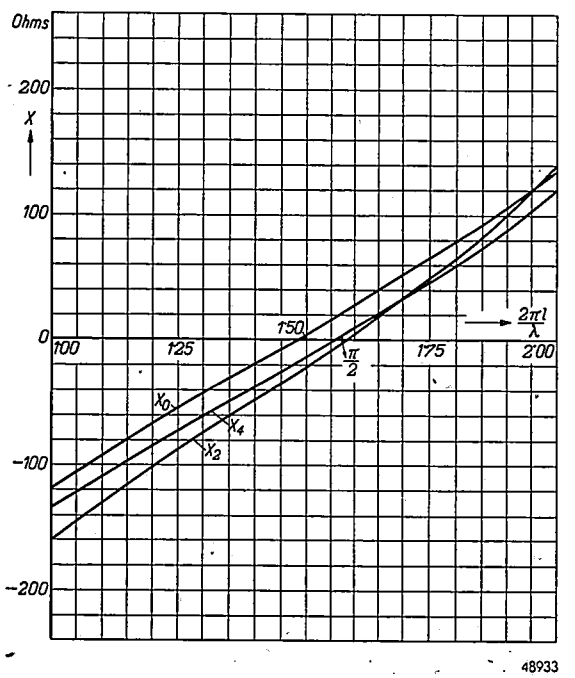


Fig. 3. Radiation-reactance curves. For notations see fig. 2 with X substituted for R.

We obtained for the three different cases, as far as the main terms are concerned, respectively,

$$p_0 = \frac{10}{3\pi} \frac{\operatorname{Im} F_0(\pi/2)}{\ln (l/a)} = \frac{22{\cdot}6}{\ln (l/a)}, \tag{41}$$

$$p_2 = \frac{20}{9\pi} \frac{\operatorname{Im} F_2(\pi/2)}{\ln (l/a)} = -\frac{1{\cdot}9}{\ln (l/a)}, \tag{42}$$

$$p_4 = \frac{8}{3\pi} \frac{\operatorname{Im} F_4(\pi/2)}{\ln (l/a)} = \frac{1{\cdot}8}{\ln (l/a)}. \tag{43}$$

It is interesting to note that in case of two horizontal wires the antenna has to be *lengthened* in order to be resonant. Further, the corrections for two and for four horizontal wires are in opposite directions, and approximately equal to one-tenth of that for a free $\lambda/2$ dipole. As the latter correction is 5%, approximately, the corrections p_2, p_4 are only of the order of 0·5 per cent and therefore almost negligible ($\Omega \geqslant 10$).

For infinitely thin wires resonance occurs at $x = \pi/2$. The value $Z(\pi/2)$ is independent of (l/a):

$$Z_0(\pi/2) = 36{\cdot}6 + 21{\cdot}3\,j,$$
$$Z_2(\pi/2) = 21{\cdot}7 - 1{\cdot}7\,j,$$
$$Z_4(\pi/2) = 20{\cdot}9 + 1{\cdot}4\,j.$$

Eindhoven, November 1946

REFERENCES

1) J. K. Schouten, N.S.F. Laboratory, Hilversum. Private communication.
2) S. A. Schelkunoff, Proc. Inst. Radio Engrs **29**, 493-521, 1941.
3) C. J. Bouwkamp, Philips Res. Rep. **1**, 65-76, 1945/46.

VOL. 2 NO. 4 AUGUST 1947

Philips Research Reports

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

R 49 *Philips Res. Rep. 2, 241-259, 1947*

ON A NON-LINEAR NOISE PROBLEM

by F. L. H. M. STUMPERS 621.396.822

Summary

A rectangular noise spectrum is applied to a valve with a non-linear current-voltage characteristic. The energy frequency spectrum is computed. It is shown that the partial spectra resulting around multifolds of the original central frequency have different forms. They are distinguished by their order. If the characteristic is given in the form of a polynomial or of a power series, a formula is obtained from which all partial spectra can be computed directly. Finally, the presence of one or more carriers is taken into account.

1. *Characteristic of degree p. One term*

We shall consider the following situation: Noise from a normal source is passed through a filter with a rectangular amplitude-frequency characteristic. The output is applied to a valve with a non-linear current-voltage curve. What is the resulting spectrum? This subject was tackled by Fränz [1]). Later this author withdrew part of his paper, and indicated another way for solving the problem without, however, giving explicit results [2]). Here we will show the results in a simple form, by applying the methods of Fränz's first paper correctly. We shall proof that the same result is reached by the second method.

For our computation the noise at the entrance of the filter be registered during a certain time, which we take from $-\pi$ to π. Of this function we make a Fourier spectrum:

$$V(t) = \sum_0^\infty{}^n (a_n \cos nt - b_n \sin nt). \qquad (1)$$

If we make the registration a large number of times, a_n as well as b_n will have a normal or gaussian probability distribution, with the same standard deviation for all n:

$$W(a_n)\, da_n = (\pi C)^{-1/2}\, e^{-a_n^2/C}\, da_n . \qquad (2)$$

This rule holds for radio spectra, as we will consider here, but fails at much higher frequencies. The distributions of the variables a and b are mutually

independent. Whatever value of a_n is found in a certain registration, it does not tell anything about b_n or about any other coefficient. For our purpose it is sometimes useful to introduce polar coordinates. The probability distributions of moduli and phase angles are then:

$$W(r_n)\,\mathrm{d}r_n = \frac{2\,r_n}{C}\,e^{-r_n^2/C}\,\mathrm{d}r_n, \tag{3}$$

$$W(\vartheta_n)\,\mathrm{d}\vartheta_n = \frac{1}{2\pi}\,\mathrm{d}\vartheta_n\,. \tag{4}$$

All phase angles have equal probability.

We will regard such spectra as result from a filter that only passes frequencies between N_1 and N_2. The ratios between amplitudes in the transmitted part of the spectrum are not affected. As for our computation the phase characteristic is not important, such a filter can be approximated. Now we shall apply this spectrum to a device with the current-voltage relation:

$$I = V^p \qquad (p \text{ integer}). \tag{5}$$

Our input signal is:

$$V = \sum_{N_1}^{N_2} r_n \cos\,(nt + \vartheta_n). \tag{6}$$

If $p = 2$:

$$I = V^2 = \tfrac{1}{2}\sum_{N_1}^{N_2}\sum_{N_1}^{N_2} r_n\,r_m \cos\{(n+m)\,t + \vartheta_n + \vartheta_m\} + \\ + \tfrac{1}{2}\sum_{N_1}^{N_2}\sum_{N_1}^{N_2} r_n\,r_m \cos\{(n-m)\,t + \vartheta_n - \vartheta_m\}\,. \tag{7}$$

In the last term the frequency will be negative a number of times. It does not make sense, however, to distinguish between positive and negative frequencies. Therefore, the last term in (7) can better be written:

$$\tfrac{1}{2}\sum_{N_1}^{N_2} r_n^2 + \sum_{N_1}^{N_2}\sum_{N_1}^{n-1} r_n\,r_m \cos\{(n-m)\,t + \vartheta_n - \vartheta_m\}. \tag{7a}$$

In our following formulae negative frequencies will sometimes occur, as otherwise the regularity of the formulae would be disturbed. The reader should keep in mind, however, that the significance is stated better by rewriting as above.

If $p = 3$ $(I = V^3)$, the resulting spectrum is:

$$I = V^3 = \tfrac{1}{4}\sum\sum\sum r_n\,r_m\,r_s \cos\{(n+m+s)\,t + \vartheta_n + \vartheta_m + \vartheta_s\} + \\ + \tfrac{1}{4}\sum\sum\sum r_n\,r_m\,r_s \cos\{(n+m-s)\,t + \vartheta_n + \vartheta_m - \vartheta_s\}. \tag{8}$$

To give the reader a good impression of the construction of spectra of higher degree, we shall give one more spectrum:

$$I = V^4 = \tfrac{1}{8}\Sigma\Sigma\Sigma\Sigma r_n r_m r_s r_q \cos\{(n+m+s+q)t+\vartheta_n+\vartheta_m+\vartheta_s+\vartheta_q\}+$$
$$+\tfrac{1}{2}\Sigma\Sigma\Sigma\Sigma r_n r_m r_s r_q \cos\{(n+m+s-q)t+\vartheta_n+\vartheta_m+\vartheta_s-\vartheta_q\}+$$
$$+\tfrac{3}{8}\Sigma\Sigma\Sigma\Sigma r_n r_m r_s r_q \cos\{(n+m-s-q)t+\vartheta_n+\vartheta_m-\vartheta_s-\vartheta_q\}. \quad (9)$$

If the original spectrum has a central frequency ω, a characteristic of degree k obviously gives new spectra with central frequencies $k\omega$, $(k-2)\omega$, $(k-4)\omega$, *etc.* To distinguish between these spectra we shall call k, $k-2$, $k-4$, *etc.*, the order of the spectrum. The cofactors of the development for a characteristic of degree k are:

$$A_{k,k} = 2^{1-k}\binom{k}{0};\ A_{k,k-2} = 2^{1-k}\binom{k}{1};\ A_{k,k-4} = 2^{1-k}\binom{k}{2}, \text{etc.} \quad (10)$$

The last term, however, is $A_{k,1} = 2^{1-k}\binom{k}{\frac{1}{2}k-\frac{1}{2}}$ for k odd, and $A_{k,0} = 2^{-k}\binom{k}{\frac{1}{2}k}$ for k even. The first suffix of A indicates the degree of the characteristic, and the second the order of the spectrum. The lower number in the binomial coefficient indicates the number of minus signs in the expression for the frequency.

The original sum V contains $(N_2-N_1+1) = N$ terms. Each of the sums in the spectrum of degree p therefore has N^p terms. To find how the energy is distributed over the spectrum, we have to answer two questions: How many terms out of the N^p result in the same frequency? How many terms belong to a group with the same phase angle?

Beginning with the first question, this one is identical with the following: In how many ways can we write a certain number as the sum of p others, provided we use only numbers between N_1 and N_2. If the same numbers occur in a different sequence, this has to be counted as a distinct way. It is clear that the distribution of the number of ways does not change if we make the first choice from N consecutive numbers, the second from N other consecutive numbers, *etc.* The only effect here is a constant shift of the distribution as a whole. We can also take the numbers from $-N_2$ to $-N_1$. Thus the numbers of terms, resulting in the same frequency, are distributed in the same way in spectra of different order, but of equal degree. When we take our numbers from 0 to $N-1$ each time, our question is obviously equivalent to: What is the cofactor of z^n in the development of:

$$(1+z+z^2+\ldots+z^{N-1})^p = (1-z^N)^p\,(1-z)^{-p} =$$
$$= \left\{1 - p\,z^N + \binom{p}{2}z^{2N} - \binom{p}{3}z^{3N} + \ldots + (-1)^p\binom{p}{p}z^{pN}\right\}$$
$$\left\{1 + pz + \binom{p+1}{2}z^2 + \binom{p+2}{3}z^3 + \ldots\right\}.$$

This cofactor is:

$$\binom{p+n-1}{n}-\binom{p}{1}\binom{p+n-1-N}{n-N}+\binom{p}{2}\binom{p+n-1-2N}{n-2N}+\ldots=$$
$$=\binom{p+n-1}{p-1}-\binom{p}{1}\binom{p+n-1-N}{p-1}+\ldots.$$

As soon as the upper term in the binominal coefficient is smaller than the lower one, the coefficients are zero. The development therefore stops automatically.

If $n = aN >> 1$, a good approximation is given by:

$$\frac{(aN)^{p-1}}{(p-1)!}-\binom{p}{1}\frac{\{(a-1)N\}^{p-1}}{(p-1)!}+\ldots=$$
$$=\frac{N^{p-1}}{(p-1)!}\left\{a^{p-1}-\binom{p}{1}(a-1)^{p-1}+\binom{p}{2}(a-2)^{p-1}+\ldots\right\}.$$

The summation has to be continued until $(-1)^k\binom{p}{k}(a-k)^{p-1}$, where $a-k \geqq 0 > a-k-1$.

As the maximum occurs at $n = \frac{1}{2}Np$, it is useful to make a change of coordinates, and to put $x = a-p/2$. The number of terms resulting in the same sum $N(p/2 \pm x)$ is then given by $N^{p-1}f_p(x)$, where

$$f_p(x)=\frac{1}{(p-1)!}\left\{\left(-x+\frac{p}{2}\right)^{p-1}-\binom{p}{1}\left(-x+\frac{p}{2}-1\right)^{p-1}+\right.$$
$$\left.+\binom{p}{2}\left(-x+\frac{p}{2}-2\right)^{p-1}+\ldots\right\}. \qquad (11)$$

The development is valid for $x \geqq 0$, and has to be continued until $-x + p/2 - k \geqq 0 > -x + p/2 - k - 1$. For $x < 0$ the function is defined by $f_p(x) = f_p(|x|)$.

The total number of terms giving a sum between $N(p/2 + x_1)$ and $N(p/2 + x_2)$ is $N^{p-1}\{f_p(y_1) + f_p(y_2) + \ldots + f_p(y_s)\}$, where $y_1, y_2 \ldots y_s$ are the s integers between those limits. If s is large this number is approximated by

$$N^p\int_{x_1}^{x_2} f_p(x)\,dx\,.$$

The total number of terms is N^p. Therefore,

$$\int_{-\infty}^{+\infty} f_p(x)\,dx = 1\,. \qquad (11a)$$

This is valid for both the precise and the approximate expression.

Another way of deriving $f_p(x)$ is to start with a continuum, extending from $-\frac{1}{2}$ to $\frac{1}{2}$, and to remark that every combination in the pth-degree spectrum is made up by a combination of the $(p-1)$th-degree spectrum and the original spectrum:

$$f_p(x) = \int_{-\infty}^{+\infty} f_{p-1}(x-y)\, f_1(y)\, dy = \int_{-\infty}^{+\infty} f_1(x-y)\, f_{p-1}(y)\, dy = \int_{x-1/2}^{x+1/2} f_{p-1}(y)\, dy . \quad (12)$$

Or: $f_p(x)$ is the p-fold convolution of the rectangular distribution with itself. Fränz was the first to introduce these functions in connection with noise problems, and he calls $f_p(x)$ the normal spectrum of degree p. With the help of the Laplace transform, formula (11) can be derived directly from (12) and the condition

$$f_1(x) = \begin{cases} 0, & x < -\frac{1}{2}, \\ 1, & -\frac{1}{2} < x < \frac{1}{2}, \\ 0, & x < \frac{1}{2}. \end{cases}$$

This transform leads also to the result

$$f_p(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-i\omega x} \left(\frac{\sin \frac{\omega}{2}}{\frac{\omega}{2}} \right)^p d\omega . \quad (13)$$

In the same way one can proof that for higher p the function $f_p(x)$ approximates a normal distribution [7]).

It might be interesting, from a historical point of view, to note that the question discussed here is identical with the problem of De Moivre: Given n dice, each with f sides, numerated 1 to f; let these dice be thrown at random; one is then asked to calculate the number of ways in which the sum can equal a given number p. The formula containing the solution of this and analogous problems was published without proof by De Moivre in his "Mensura sortis" (1711), and a simple proof was given in his "Miscellanea analytica" (1730). The limits for n large were considered by Laplace ("Théorie analytique des probabilités", 1812) who also proved that the normal distribution is approximated for higher p. These historical notes can be found *i.a.* in Czuber's historical review [3]). Later on the problem was reconsidered many times without acknowledgment of earlier treatment. In 1896 Maurer [4]) came to the problem in a discussion on averages, and in the "Boltzmann Festschrift" of 1904, Sommerfeld [5]) gave a geometrical treatment. In 1940 Fränz comes to it from the noise side, and in 1944 Silberstein [6]) starts anew. In the discussion following the last paper reference to some earlier papers was given, but not to those mentioned here.

Now we come to the second question, which is closely associated with the energy spectrum. If the current is given by $I = V^p$, we ask for the energy produced by this current in a resistance of 1 ohm, and for the distribution of this energy over the spectrum. The energy is defined by $\frac{1}{2}\overline{I^2}$, where the average has to be taken over a full period ($-\pi$ to π). First we shall con-

sider the spectrum of degree p and order p. For the moment we shall leave the cofactor $A_{p,p}$ out of account. We shall correct for this later on.

One of the terms in this spectrum is

$$r_{n_1}^{k_1} r_{n_2}^{k_2} \dots r_{n_m}^{k_m} \times$$
$$\times \cos \{(k_1 n_1 + k_2 n_2 + \dots + k_m n_m)\, t + k_1 \vartheta_{n_1} + k_2 \vartheta_{n_2} + \dots k_m \vartheta_{n_m}\}$$

with $\quad k_1 + k_2 + k_3 + \dots + k_m = p, \text{ and } k_1 n_1 + k_2 n_2 + \dots + k_m n_m - \tfrac{1}{2} p\,(N_1 + N_2) = xN.$

This term only can give a contribution to the mean square with terms having the same phase angle. There is a group of

$$\frac{p!}{k_1!\, k_2! \dots k_m!}$$

terms with this phase angle, and its contribution to the energy is therefore

$$\tfrac{1}{2} \left\{ \frac{p!}{k_1!\, k_2! \dots k_m!} \right\}^2 \overline{r_{n_1}^{2k_1}}\, \overline{r_{n_2}^{2k_2}}\, \overline{r_{n_3}^{2k_3}} \dots \overline{r_{n_m}^{2k_m}}$$

In accordance with formula (2) we get:

$$\overline{r^{2k}} = \int_0^\infty \frac{2}{C} e^{-r^2/C}\, r^{2k+1}\, \mathrm{d}r = k!\, C^k. \tag{14}$$

The energy of the group is

$$\tfrac{1}{2} \frac{(p!)^2\, C^p}{k_1!\, k_2! \dots k_m!}.$$

Dividing by the number of terms, we get the energy per term, independent of the structure of the group: $\frac{1}{2} p!\, C^p$. The distribution of the energy over the spectrum is thus conform to the distribution of the number of terms with the same frequency over the spectrum. The energy spectrum is thus given by:

$$E_{p,p}(x) = \tfrac{1}{2} p!\, C^p\, N^{p-1} f_p(x). \tag{15}$$

The energy for x between the values x_1 and x_2 is computed from:

$$N \int_{x_1}^{x_2} E_{p,p}(x)\, \mathrm{d}x = \tfrac{1}{2} (NC)^p\, p! \int_{x_1}^{x_2} f_p(x)\, \mathrm{d}x \tag{16}$$

(here $(x_1 - x_2)N$ is still a large number).

The problem is a bit more complicated if we come to the spectrum of degree p and order $p-2$. Here we have $p-1$ positive signs and one negative sign. In the following we have to make an exception for the spectrum of order 0, to which we will refer later on. Again we omit $A_{p,p-2}$.

Now in the spectrum of order $p-2$, one of the terms is

$$r_{n_1}^{k_1} r_{n_2}^{k_2} \dots r_{n_{m-1}}^{k_{m-1}} r_{n_m} \times$$
$$\cos\{(k_1 n_1 + k_2 n_2 + \dots + k_{m-1} n_{m-1} - n_m) t + k_1 \vartheta_1 + \dots + k_{m-1} \vartheta_{n_{m-1}} - \vartheta_{n_m}\}.$$

If the number n_m is not found in $n_1, n_2, \dots n_{m-1}$, the problem is nearly the same as before. The number of terms belonging to a group is

$$\frac{(p-1)!}{k_1!\, k_2! \dots k_{m-1}!},$$

and the energy of the group is

$$\tfrac{1}{2} \frac{(p-1)!^2\, C^p}{k_1!\, k_2! \dots k_{m-1}!},$$

thus giving per term $\frac{1}{2}\,(p-1)!\, C^p$.

However, if the number n_m is already contained in the numbers $n_1 \dots, n_{m-1}$, the situation is different. We take $n_1 = n_m$. Now the term belongs to a group A with all terms

$$r_{n_1}^{k_1-1} r_{n_2}^{k_2} \dots r_{n_{m-1}}^{k_{m-1}} r_s^2 \times$$
$$\cos[\{(k_1-1) n_1 + k_2 n_2 + \dots + k_{m-1} n_{m-1}\} t + (k_1-1)\vartheta_1 + k_2 \vartheta_2 + \dots k_{m-1} \vartheta_{m-1}],$$

where s may be any number subject to $N_1 \leqq s \leqq N_2$.
The number of terms in the group is

$$\frac{(p-1)!}{(k_1-1)!\, k_2! \dots k_{m-1}!} \left\{(N-m+1) + \frac{1}{k_1} + \frac{1}{k_2+1} + \dots + \frac{1}{k_{m-1}+1}\right\},$$

where $N-m+1$ refers to s not equal to $n_1 \dots, n_{m-1}$; $1/k_1$ refers to $s = n_1$, *etc.*

Apart from r_s^2 the term occurs in the spectrum of degree $p-2$ and order $p-2$, and there it belongs to a group of

$$\frac{(p-2)!}{(k_1-1)!\;\, k_2! \dots\; k_{m-1}!}$$

terms with the same phase angle.

The energy of all terms belonging to the group A is the average of

$$\tfrac{1}{2}\left\{\frac{(p-1)!}{(k_1-1)!\, k_2! \dots k_{m-1}!}\right\}^2 \left\{\Sigma^s\, r_{n_1}^{k_1-1} r_{n_2}^{k_2} \dots r_{n_{m-1}}^{k_{m-1}} r_s^2 + \right.$$
$$\left. + \frac{1}{k_1} r_{n_1}^{k_1+1} r_{n_2}^{k_2} \dots r_{n_{m-1}}^{k_{m-1}} + \dots + \frac{1}{k_{m-1}+1} r_{n_1}^{k_1-1} r_{n_2}^{k_2} \dots r_{n_{m-1}}^{k_{m-1}+2}\right\}^2,$$

where the summation has to be taken over all s, subject to $N_1 \leqq s < N_2$, apart from $s = n_1, s = n_2, \dots, s = n_{m-1}$.

By repeated application of formula (14) we find the energy to be

$$\tfrac{1}{2}\frac{(p-1)!^2\quad C^p}{(k_1-1)!\,k_2!\ldots k_{m-1}!}\left\{N^2+N-m+1+\frac{1}{k_1}+\frac{1}{k_2+1}+\ldots+\frac{1}{k_{m-1}+1}\right\}.$$

This we can interprete as an energy $\frac{1}{2}(p-1)!\,C^p$ for every term belonging to the spectrum of degree p, and an extra amount $\frac{1}{2}(p-1)\;(p-1)!\,N^2C^p$ for every term in the spectrum of degree $p-2$. The total energy in the spectrum of degree p and order $p-2$ is thus distributed as follows:

$$E_{p,p-2}(x) = \tfrac{1}{2}(p-1)!\,C^p\,N^{p-1}\left\{f_p(x)+(p-1)\,f_{p-2}(x)\right\}, \qquad (17)$$

and the energy between x_1 and x_2 is $N\int_{x_1}^{x_2} E_{p,p-2}(x)\,dx$.

In the same way we can make the computation for the spectrum of order $p-4$ and degree p. Its energy distribution is given by:

$$E_{p,p-4}(x)=\tfrac{1}{2}(p-2)!\,2!\,N^{p-1}C^p\left\{f_p(x)+\binom{2}{1}\binom{p-2}{1}f_{p-2}(x)+\binom{p-2}{2}f_{p-4}(x)\right\}. \qquad (18)$$

And, analogously, the energy distribution for the spectrum of order $p-6$ and degree p is:

$$E_{p,p-6}(x) = \tfrac{1}{2}(p-3)!\,3!\,N^{p-1}C^p\times$$
$$\times\left\{f_p(x)+\binom{3}{1}\binom{p-3}{1}f_{p-2}(x)+\binom{3}{2}\binom{p-3}{2}f_{p-4}(x)+\binom{p-3}{3}f_{p-6}(x)\right\}. \qquad (19)$$

For odd powers this treatment is quite general. For even powers we have still to consider the spectrum of order 0. In this spectrum, apart from the direct-current term, all terms occur twice; once with a positive frequency and once with a negative frequency of the same absolute value. As we have already pointed out, it makes no sense to distinguish between positive and negative frequencies; therefore, these terms have to be taken together before squaring. Thus the energy derived in formulae (16) to (19) has to be multiplied by 4 for the positive frequencies and by 2 for the direct current.

The above formulae can also be derived by a generalization from examples in which only few frequencies occur.

We have not yet defined $f_0(x)$. For the discontinuous spectrum $f_0(0) = N$, and $f_0(x) = 0$, if $x \neq 0$. If we use integrals (or make the spectrum continuous) $f_0(x) = \delta(x)$; $\delta(x) = 0$, if $x \neq 0$; $\int_{-\infty}^{+\infty}\delta(x)\,dx = 1$. This definition is indicated by the recurrence relation. Obviously the direct-current term holds a special position, as it is not zero in the average.

We still have to correct for the omitted cofactors $A_{p,p}$, *etc.* As an example we will compute all spectra of degree 6. The cofactors are

$$A_{6,6} = 2^{-5};\ A_{6,4} = 2^{-5}\binom{6}{1};\ A_{6,2} = 2^{-5}\binom{6}{2};\ A_{6,0} = 2^{-6}\binom{6}{3}.$$

The energy spectra are now given by

$$E_{6,6}(x) = 2^{-11} \cdot 6!\, N^5\, C^6\, f_6(x-6\omega)\,,$$
$$E_{6,4}(x) = 2^{-11} \cdot 6 \cdot 6!\, N^5\, C^6 \left\{ f_6(x-4\omega) + 5\, f_4(x-4\omega) \right\},$$
$$E_{6,2}(x) = 2^{-11} \binom{6}{2} 6!\, N^5\, C^6 \left\{ f_6(x-2\omega) + 8\, f_4(x-2\omega) + 6\, f_2(x-2\omega) \right\},$$
$$E_{6,0}(x) = 2^{-12} \binom{6}{3} 6!\, N^5\, C^6 \left\{ 2\, f_6(x) + 18\, f_4(x) + 18\, f_2(x) + f_0(x) \right\},$$
$$N\omega = \frac{N_1 + N_2}{2}.$$

The last term has been corrected for zero order, and is only valid for $x \geqq 0$. From formula (12) the numerical values can easily be deduced. The form of the spectra is seen in *fig. 1*. It is clear, that in the spectra of lower

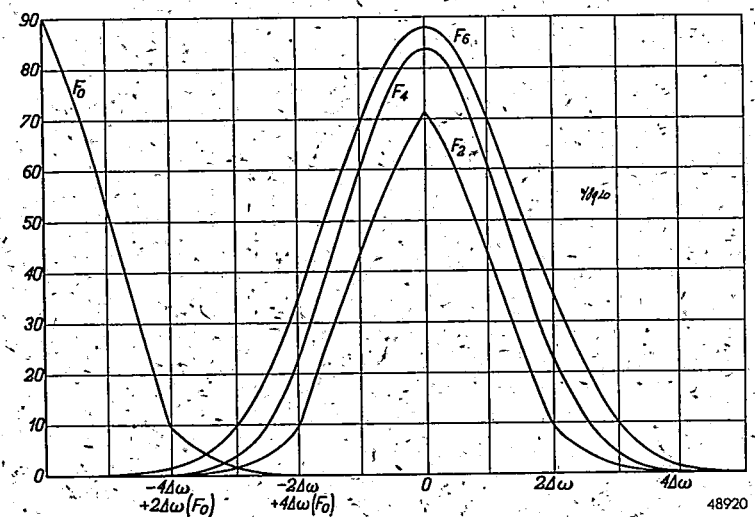


Fig. 1. Form of the energy spectra for a characteristic $I = V^6$. F_4 is the spectrum of order 4, *etc.* $\Delta\omega$ is the bandwidth of the original noise spectrum.

order the "normal spectra" of the corresponding degree have a very marked influence. In *fig. 2* the spectra are drawn on scale (logarithmic).

Spectra of different order do not influence each other as they do not give the same combination of phase angles. So, even if the original spectrum is so broad as to cause overlapping of higher-order spectra, the energy can be computed by simple addition.

Herewith we finish the computation of the output-energy spectrum of a non-linear device with a characteristic $I = V^p$.

2. *Polynomial or power-series characteristic*

We will now consider a polynomial characteristic:

$$I = a_0 + a_1 V + a_2 V^2 + \ldots a_p V^p.$$

Terms of even and odd degree do not influence each other, as they can never give spectra of the same order. There can, however, be combination

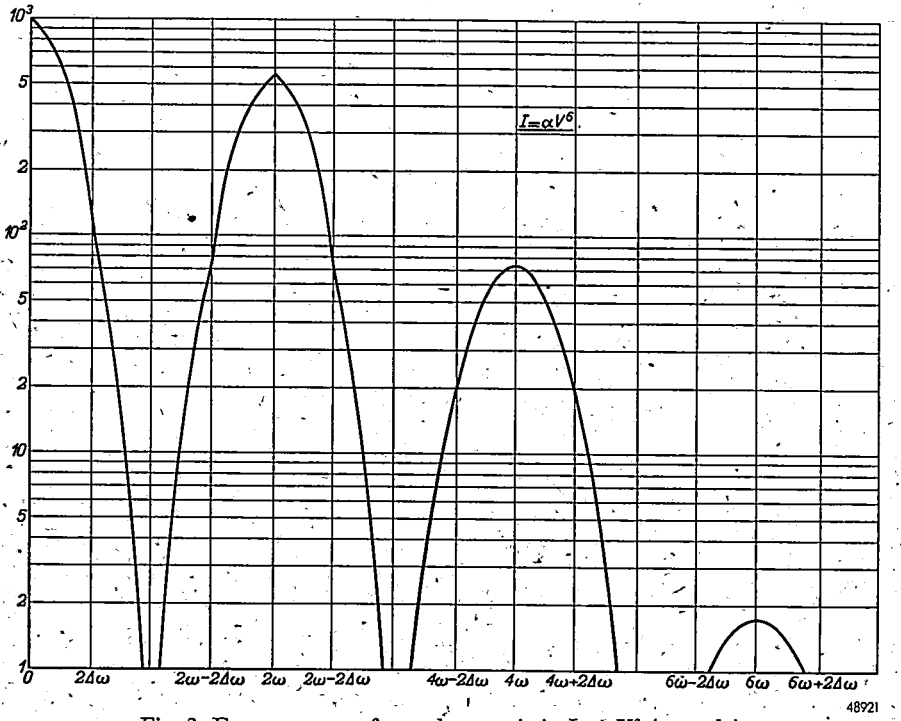


Fig. 2. Energy spectra for a characteristic $I = V^6$ (on scale).

terms in the energy if the degrees differ an even amount. Consider the spectrum of order $p-2$. Here only the terms

$$I = a_{p-2} V^{p-2} + a_p V^p$$

have to be taken into account. With each term in the spectrum of degree $p-2$ and order $p-2$, there correspond about $(p-1)\,N$ terms in the spectrum of degree p and order $p-2$. The spectrum of order $p-2$ contains terms of the form:

$$2^{3-p} \cdot a_{p-2} \Sigma\Sigma \ldots \Sigma r_{n_1}^{k_1} r_{n_2}^{k_2} \ldots r_{nl}^{kl} \cos\{(k_1 n_1 + k_2 n_2 + \ldots k_l n_l)t + \vartheta_{n_1} + \ldots \vartheta_{nl}\} +$$
$$+ 2^{1-p} \cdot a_p \cdot p\, \Sigma\Sigma \ldots \Sigma r_{n_1}^{k_1} r_{n_2}^{k_2} \ldots r_{nl}^{kl} r_s^2 \cos\{(k_1 n_1 + k_2 n_2 + \ldots k_l n_l)t + \vartheta_{n_1} + \ldots \vartheta_{nl}\},$$

where $k_1 + k_2 + \ldots + k_l = p-2$, and s can have all values subject to $N_1 \leqq s \leqq N_2$.

In both terms we get the same phase angle, and therefore their product appears in the energy spectrum. Using nearly the same reasoning as before for the characteristic $I = V^p$ and the spectrum of order $p-2$ we can easily prove that (leaving cofactors out of account) to every term in the first sum corresponds an energy contribution in the product of $N(p-1)!\, C^{p-1}$. This is independent of the structure of the term.

If $\beta_{p-2} = 2^{3-p} \cdot \alpha_{p-2}$ and $\beta_{p-2} = 2^{1-p} \cdot p \cdot \alpha_p$ the mixed term in the energy is: $\beta_p\, \beta_{p-2}\, (p-1)!\, C^{p-1}\, N^{p-2}\, f_{p-2}(x)$.

The results for the terms containing α_p^2 and α_{p-2}^2 can be taken from the computation for a one-term characteristic. In this way we get the total energy distribution of the spectrum of order $p-2$:

$$E_{p-2}(x) = \tfrac{1}{2}(p-2)!\, C^{p-2}\, N^{p-3} [\beta_{p-2}^2 f_{p-2}(x) + 2\beta_p\, \beta_{p-2}\, NC\,(p-1) f_{p-2}(x) + \\ + \beta_p^2\,(p-1)\, N^2 C^2 \{f_p(x) + (p-3) f_{p-2}(x)\}] \ldots \qquad (20)$$

As another example we take the spectrum of order $p-6$. In the same way as above new coefficients are introduced:

$$\beta_{p-6} = 2^{7-p}\alpha_{p-6};\ \beta_{p-4} = 2^{5-p}\binom{p-4}{1}\alpha_{p-4};\ \beta_{p-2} = 2^{3-p}\binom{p-2}{2}\alpha_{p-2};\ \beta_p = 2^{1-p}\binom{p}{3}\alpha_p .$$

The energy can then be given in the form:

$$E_{p-6}(x) = \tfrac{1}{2}\{\beta_{p-6}^2\, C_{p-6,p-6} + 2\beta_{p-6}\beta_{p-4}\, C_{p-6,p-4} + 2\beta_{p-6}\,\beta_{p-2}\, C_{p-6,p-2} + \ldots + \beta_p^2 C_{p,p}\},$$

where

$$C_{p-2s,\,p-2t} = (3-s)!\,(p-t-3)!\, N^{-1}\,(NC)^{p-t-s} \sum_0^3 k \binom{3-t}{3-s-k}\binom{p-3-s}{k} f_{p-2s-2k}(x);$$

k, s and t are integers subject to the conditions:

$$0 \leqq s \leqq 3,\ 0 \leqq t \leqq s,\ 0 \leqq k \leqq 3-s .$$

Of course, a correction has to be made if $p = 6$.

The reader will have little difficulty in generalizing this formula for spectra of other orders.

We can now apply our formulae to characteristics of the general form:

$$I = \sum_0^\infty p \frac{\alpha_p}{p!}\, V^p , \qquad (21)$$

and compute the energy spectra.

First we introduce functions $g_n(y)$ related to the original characteristic by:

$$g_n(y) = \sum_0^\infty s\, a_{n+2s} y^s / s! .$$

Consequently $g_{n+2}(y) = g_n'(y)$.

The computation of the energy spectrum of order zero gives the result:

$$E_0(x) = g_0^2(y)\, f_0(x) + 2y^2 g_2^2(y)\, f_2(x) + 2y^4 \frac{g_4^2(y)}{2!\,2!} f_4(x) + \ldots, \quad (22)$$

with $y = \frac{1}{4}\, NC$. This is valid for $x \geqq 0$; the correction necessary for zero order has been made.

The energy spectrum of order n is $(n \neq 0)$;

$$E_n(x) = 2y^n \sum_0^\infty k \frac{y^{2k} g_{n+2k}^2(y)}{(n+k)!\, k!} f_{n+2k}\left(x - \frac{N_0}{N} n\right). \quad (23)$$

N_0 is the original central frequency $\frac{1}{2}\,(N_1 + N_2)$.

In the special case of an exponential characteristic all coefficients a_n are 1 and the functions $g_n(y)$ are equal to e^y. For an exponential characteristic the total energy of order zero is computed with the help of equation (11a)

$$E_0 = e^{\frac{1}{2}NC}\, I_0\left(\frac{NC}{2}\right). \quad (24)$$

The total energy of the spectrum of the nth order is:

$$E_n = 2e^{\frac{1}{2}NC}\, I_n\left(\frac{NC}{2}\right) \quad (25)$$

Here

$$I_n(z) = i^{-n} J_n(iz), \quad (I_n(z) \text{ not be to confused with the current } I),$$

and $J_n(iz)$ is the Bessel function of order n and argument iz. If the characteristic is given in the form:

$$\frac{I}{I_0} = e^{\frac{V}{V_0}},$$

and if the effective noise voltage is v_n, we get instead of formula (25)

$$E_n = 2\, e^{\frac{V_n^2}{V_0^2}}\, I_n\left(\frac{V_n^2}{V_0^2}\right). \quad (25a)$$

With our formulae we can also compute what part of the energy is contained in a certain frequency band. In *fig. 3* one can see in what band 50 or 90% of the total low-frequency energy is present for $I = V^p$ and $p = 2,\ 4,\ 6,\ 8$.

Our computation involved only one integral (14). The proof of the formulae only needs a careful counting of the cases in which the same phase angle is present. Moreover the formulae are valid without any approximation for any number N and not only for $N = \infty$. One has of course to replace

the approximate formula $f_p(x)$ (valid for large N) by the original formula of De Moivre. If one considers a few examples for very low values of N, the general trend in the structure of the energy spectrum soon comes forward. The general proof, however, is rather cumbersome, though only elementary operations are included.

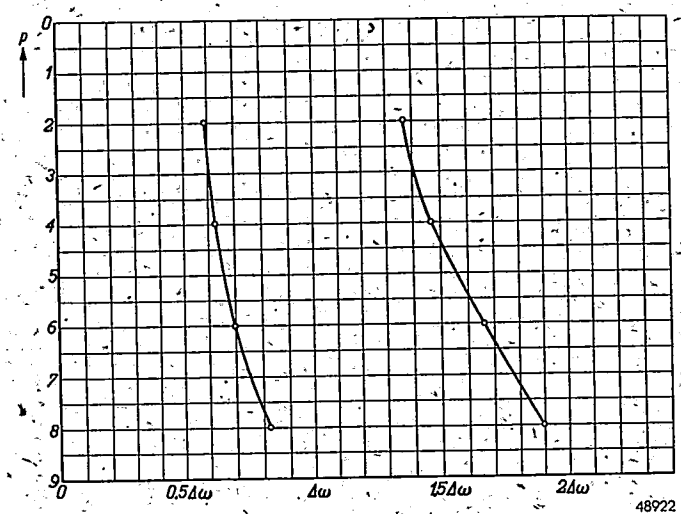


Fig. 3. The band containing 50% of the total low-frequency energy for $I = V^p$. Also the band containing 90% of this energy. Bandwidth of the original noise spectrum: $\Delta\omega$.

For this reason we will give a straight-forward proof of the general formula (23) in the way Fränz indicated in his correction [2]).

The characteristic is:

$$I = \sum_0^\infty{}_p \frac{a_p}{p!} V^p \,. \tag{21}$$

A well-known application of the theory of functions gives:

$$\frac{V^p}{p!} = \frac{1}{2\pi} \oint \frac{e^{iVz}}{(iz)^{p+1}} \, dz \,. \tag{26}$$

The Fourier spectrum of the original noise as observed from $-\pi$ until π is:

$$V = \sum_{N_1}^{N_2}{}_n (a_n \cos nt - b_n \sin nt) \,. \tag{1}$$

The probability distribution of a_n and b_n is given by:

$$W(a_n)\, da_n = \frac{1}{\sqrt{\pi C}} e^{-a_n^2/C} \, da_n \,, \tag{2}$$

$$W(b_n)\,\mathrm{d}b_n = \frac{1}{\sqrt{\pi C}}\, e^{-b_n^2/C}\,\mathrm{d}b_n\,. \tag{2a}$$

We make a Fourier spectrum of I:

$$I = \mathrm{Re}\sum_0^\infty f_m\, e^{imt}, \tag{27}$$

where

$$f_m = \frac{1}{\pi}\int_0^{2\pi} I(t)\, e^{-imt}\,\mathrm{d}t \quad (m \neq 0)\,, \tag{28}$$

$$f_0 = \frac{1}{\pi}\int_0^{2\pi} I(t)\,\mathrm{d}t\,. \tag{28a}$$

The energy for the frequency n is given by $\frac{1}{2} f_n f_n^*$. For the direct-current term it is f_0^2.

As usual the * indicates the complex conjugated. The energies are functions of the amplitudes a_n and b_n. We can make an average over a great many observation periods 2π. Then

$$\overline{f_m f_m^*} = \int f_m f_m^*(a_{N_1}, b_{N_1}, \ldots a_{N_2}, b_{N_2}) \prod_l W(a_l,b_l)\mathrm{d}a_l\mathrm{d}b_l\,. \tag{29}$$

Thus formula (29) is a shorter way of writing the integral that results from the following procedure: Introduce the expression (1) in the integral (26), and the result of this operation in (21). Now replace $I(t)$ in (28) by the expression just obtained. f_m is written as an integral over t and z involving summations over n and p. In the same way f_m^* is written as an integral over t_2 and z_2 with summations over n and q. Finally, together with these results, (2) and (2a) are introduced in (29).

First of all we carry out the integration over a_n:

$$\frac{1}{\sqrt{\pi C}}\int_{-\infty}^{+\infty} \exp\left\{-\frac{a_n^2}{C} + i(a_n z_1 \cos nt_1 + a_n z_2 \cos nt_2)\right\}\mathrm{d}a_n = \\ = e^{-\frac{C}{4}(z_1\cos nt_1 + z_2 \cos nt_2)^2}\,. \tag{30}$$

In the same way the integration with respect to b_n gives:

$$\frac{1}{\sqrt{\pi C}}\int_{-\infty}^{+\infty} \exp\left\{-\frac{b_n^2}{C} - i(b_n z_1 \sin nt_1 + b_n z_2 \sin nt_2)\right\}\mathrm{d}b_n = \\ = e^{-\frac{C}{4}(z_1\sin nt_1 + z_2 \sin nt_2)^2}\,. \tag{31}$$

The product of all $2N$ integrations for the a's and b's is:

$$e^{-\frac{NC}{4}(z_1^2+z_2^2) - \frac{C}{2} z_1 z_2 \sum_{N_1}^{N_2} \cos m\,(t_1 - t_2)}.$$

Regarding the integrations over t_1 and t_2 we can see that the integrand is only a function of $(t_1 - t_2)$. As the integrand is periodic in both variables:

$$\frac{1}{\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} g(t_1-t_2)\,dt_1\,dt_2 = \frac{2}{\pi}\int_{-\pi}^{\pi} g(v)\,dv\,.$$

In our case this integral is:

$$-\frac{2}{\pi}\int_{-\pi}^{\pi}\exp\left\{-\frac{C}{2}z_1z_2\sum_{N_1}^{N_2}\cos mv - inv\right\}dv\,.$$

The sum in the exponent is:

$$\sum_{N_1}^{N_2}\cos mv = \frac{\sin\frac{Nv}{2}\,.\cos N_0 v}{\sin\frac{v}{2}}\,. \tag{32}$$

where $N_0 = \frac{1}{2}(N_1 + N_2)$ and $N = N_2 - N_1 + 1$.
Upon introduction of the new variable $Nv = \mu$, and for N increasing to infinity (while N_0/N and n/N remain finite), the integral approaches:

$$\frac{2}{\pi}\int_{-\infty}^{+\infty}\exp\left\{-\frac{NC}{2}z_1z_2\frac{2}{\mu}\sin\frac{\mu}{2}\cos\frac{N_0}{N}\mu - i\frac{n}{N}\mu\right\}d\mu =$$

$$\frac{2}{\pi}\sum_0^\infty{}_k\int_{-\infty}^{+\infty}\frac{(-1)^k}{k!}\left(\frac{NCz_1z_2}{4}\right)^k\left(\frac{\sin\mu/2}{\mu/2}\right)^k\sum_0^k{}_r\binom{k}{r}e^{i(k-2r)\frac{N_0}{N}\mu - i\frac{n}{N}\mu}\,d\mu.$$

Using equation (13) we can also write instead of this:

$$4\sum_0^\infty(-1)^k\left(\frac{NC}{4}z_1z_2\right)^k\sum_0^k{}_r\frac{1}{(k-r)!\,r!}f_k\left\{(2r-k)\frac{N_0}{N}+\frac{n}{N}\right\}.$$

We still have to perform the integrations over z_1 and z_2

$$\sum_0^\infty{}_p\sum_0^\infty{}_q\frac{a_p\,a_q}{4\pi^2}\oint\oint\frac{e^{-\frac{NC}{4}(z_1z_2+z_1z_2)}}{(iz_1)^{p+1}(iz_2)^{q+1}}(-z_1z_2)^k\,dz_1\,dz_2 =$$

$$= \left\{\Sigma\frac{a_p}{2\pi}\oint\frac{e^{-\frac{NC}{4}z_1z_2}}{(iz_1)^{p+1}}(iz_1)^k\,dz_1\right\}^2 =$$

$$= \begin{cases} = 0 & \text{, if } p-k \text{ is odd,} \\ = \left\{\sum_0^\infty{}_p\, a_p\left(\frac{NC}{4}\right)^{\frac{p-k}{2}}\frac{1}{\left(\frac{p-k}{2}\right)!}\right\}^2 & \text{, if } p-k \text{ is even.} \end{cases} \tag{33}$$

Summing all integrations and summations we get:

$$\overline{f_m f_m^*} = 4 \sum_0^\infty{}^k \left\{ \sum_0^\infty{}^p a_p \left(\frac{NC}{4}\right)^{\frac{p}{2}} \frac{1}{\left(\frac{p-k}{2}\right)!} \right\}^2 \sum_0^k{}_r \frac{1}{r!\,(k-r)!} f_k \left\{ (2r-k)\,\frac{N_0}{N} + \frac{m}{N} \right\}. \tag{34}$$

Now we collect all spectra of the same order b:

$$\begin{aligned} 2r-k &= -b, \quad & k &= 2r+b, & 0 \leqq r < \infty, \\ p-k &= 2l, & p &= 2r+2l+b, & 0 \leqq l < \infty. \end{aligned}$$

Introduction of these relations in (34) leads to the result:

$$\overline{f_m f_m^*} = 4 \sum_0^\infty{}^b \sum_0^\infty{}_r \left\{ \sum_0^\infty{}_l a_{b+2r+2l} \left(\frac{NC}{4}\right)^{r+l+\frac{b}{2}} \frac{1}{l!} \right\}^2 \frac{1}{(b+r)!\,r!} f_{2r+b} \left(-b\,\frac{N_0}{N} + \frac{m}{N} \right). \tag{35}$$

Be $g_{b+2r}(y) = \sum_0^\infty{}_l a_{b+2r+2l} \left(\frac{NC}{4}\right)^l \frac{1}{l!}$.

Then the energy for the frequency m in the spectrum of order b is:

$$E_b(m) = 2 \sum_0^\infty{}_r \frac{1}{(b+r)!\,r!} y^{b+2r}\, g_{b+2r}{}^2(y)\, f_{b+2r} \left(-b\,\frac{N_0}{N} + \frac{m}{N} \right). \tag{23a}$$

Only for the zero-order spectrum have we to correct the direct-current term, because we have to use the expression (28a) instead of (28) and f_0^2 instead of $\frac{1}{2} f_n f_n^*$. This leads directly to (22).

Noise spectra in presence of carrier waves

The problems resulting from the addition of one or more carrier waves can also be solved in both ways. We will only use the first method for a fourth-degree characteristic and one or two carriers. Original spectrum:

$$V = a \cos \omega_1 t + \Sigma\, r_n \cos (nt + \vartheta_n).$$

Output spectrum of the non-linear device (low-frequency terms only):

$$\begin{aligned} V^4 = {} & \tfrac{3}{8} a^4 + \tfrac{3}{2} a^2 \,\Sigma\Sigma\, r_n\, r_m \cos \left\{ (n-m)\, t + \vartheta_n - \vartheta_m \right\} + \\ & + \tfrac{3}{8} \,\Sigma\Sigma\Sigma\Sigma\, r_n\, r_m\, r_p\, r_q \cos \left\{ (n+m-p-q)\, t + \vartheta_n + \vartheta_m - \vartheta_p - \vartheta_q \right\} + \\ & + \tfrac{3}{2} a^3 \,\Sigma\, r_n \cos (nt - \omega_1 t + \vartheta_n) + \\ & + \tfrac{3}{2} a \,\Sigma\Sigma\Sigma\, r_n\, r_m\, r_p \cos \left\{ (n+m-p)\, t - \omega_1 t + \vartheta_n + \vartheta_m - \vartheta_p \right\} + \\ & + \tfrac{3}{4} a^2 \,\Sigma\Sigma\, r_n\, r_m \cos \left\{ (n+m)\, t - 2\omega_1 t + \vartheta_n + \vartheta_m \right\}. \end{aligned}$$

We have already met the first three terms; they are also present without the addition of a carrier. The energy computation for the fourth term is

analogous to the calculation of the energy spectrum of the first order for a third-degree polynomial characteristic. Keeping this in mind we obtain the following result. Energy spectrum:

$$E_0(x) = \tfrac{9}{64}\,(a^4 + 4a^2\,NC + 2\,N^2C^2)\,.\,f_0(x) + \tfrac{9}{2}\,N^2C^2\,(a^2 + NC)^2\,.\,f_2(x) + \\ + \tfrac{9}{8}\,N^4C^4\,.\,f_4(x) + \tfrac{9}{8}\,(a^2 + 2\,NC)^2\,a^2\,NC\,.\,f_1\,(x + \omega - \omega_1) + \\ + \tfrac{9}{4}\,a^2\,N^3\,C^3\,.\,f_3\,(x + \omega - \omega_1) + \tfrac{9}{16}\,a^4\,N^2C^2\,.\,f_2\,(x + 2\omega - 2\omega_1).$$

In this expression ω (central frequency of the input noise spectrum) and ω_1 are normalized in such a way that the bandwith of the input noise spectrum is the frequency unit. After introduction of effective voltages the total low-frequency energy can be written:

$$a^2/2 = V_s^2 \quad ; \quad NC/2 = V_n^2,$$

$$E_0 = \tfrac{9}{4}\,(V_s^4 + 4\,V_s^2\,V_n^2 + 2\,V_n^4)^2 + 36\,(V_s^2 + V_n^2)^2\,V_n^4 + 9\,V_n^8 + \\ + 18\,(V_s^2 + 2\,V_n^2)^2\,V_n^2\,V_s^2 + 36\,V_s^2\,V_n^6 + 9\,V_s^4\,V_n^4\,. \qquad (36)$$

The first term gives only direct-current energy. Leaving this out of account, the total noise plus signal energy bears the following ratio to the noise energy alone:

$$V_s/V_n = y\,,$$

$$\frac{E_{n+s}}{E_n} = \frac{36\,(1+y^2)^2 + 9 + 18\,(y^2+2)\,y^2 + 36\,y^2 + 9\,y^4}{36+9} = \\ = 1 + 4\,y^2 + 2{\cdot}6\,y^4 + 0{\cdot}4\,y^6\,. \qquad (37)$$

If two carriers are present:

$$V = a\cos\omega_1 t + b\cos\omega_2 t + \Sigma\, r_n \cos\,(nt + \vartheta_n),$$

and the characteristic is again $I = V^4$, the low-frequency terms in the spectrum are:

$$V^4 = \tfrac{3}{8}\,(a^4 + 4a^2b^2 + b^4) + \tfrac{3}{2}\,(a^2+b^2)\,\Sigma\Sigma\, r_n\, r_m \cos\{(n-m)\,t + \vartheta_n - \vartheta_m\} + \\ + \tfrac{3}{8}\,\Sigma\Sigma\Sigma\Sigma\, r_n\, r_m\, r_p\, r_q \cos\{(n+m-p-q)\,t + \vartheta_n + \vartheta_m - \vartheta_p - \vartheta_q\} + \\ + \tfrac{3}{4}\,a^2b^2\cos 2\,(\omega_2 - \omega_1)\,t + \\ + \tfrac{3}{2}\,ab\,(a^2+b^2)\cos\,(\omega_2 - \omega_1)\,t + \\ + 3\,ab\,\Sigma\Sigma\, r_n\, r_m \cos\{(\omega_2 - \omega_1)\,t + (n-m)\,t + \vartheta_n - \vartheta_m\} + \\ + \tfrac{3}{2}\,a\,(a^2 + 2\,b^2)\,\Sigma\, r_n \cos\{nt - \omega_1 t + \vartheta_n\} + \\ + \tfrac{3}{2}\,b\,(b^2 + 2a^2)\,\Sigma\, r_n \cos\{nt - \omega_2 t + \vartheta_n\} + \\ + \tfrac{3}{2}\,a\,\Sigma\Sigma\Sigma\, r_n\, r_m\, r_p \cos\{(n+m-p)\,t - \omega_1 t + \vartheta_n + \vartheta_m - \vartheta_p\} + \\ + \tfrac{3}{2}\,b\,\Sigma\Sigma\Sigma\, r_n\, r_m\, r_p \cos\{(n+m-p)\,t - \omega_2 t + \vartheta_n + \vartheta_m - \vartheta_p\} + \\ + \tfrac{3}{2}\,a^2\,b\,\Sigma\, r_n \cos\{nt + \omega_2 t - 2\omega_1 t + \vartheta_n\} + \\ + \tfrac{3}{2}\,ab^2\,\Sigma\, r_n \cos\{nt + \omega_1 t - 2\omega_2 t + \vartheta_n\} + \\ + \tfrac{3}{4}\,a^2\,\Sigma\Sigma\, r_n\, r_m \cos\{(n+m)\,t - 2\omega_1 t + \vartheta_n + \vartheta_m\} + \\ + \tfrac{3}{4}\,b^2\,\Sigma\Sigma\, r_n\, r_m \cos\{(n+m)\,t - 2\omega_2 t + \vartheta_n + \vartheta_m\} + \\ + \tfrac{3}{2}\,ab\,\Sigma\Sigma\, r_n\, r_m \cos\{(n+m)\,t - \omega_1 t - \omega_2 t + \vartheta_n + \vartheta_m\}\,.$$

Consequently the low-frequency energy spectrum can be built up by addition of the following energy spectra:

a) A noise spectrum beginning with the direct-current component and extending over twice the original bandwidth. This spectrum is computed in the same way as the zero-order spectrum for a polynomial characteristic of degree 4:

$$\tfrac{9}{16}\{(a^4 + 4\,a^2\,b^2 + b^4) + 2\,(a^2 + b^2)\,NC + N^2C^2\}^2 \cdot f_0(x) + \\ + \tfrac{9}{2}\,N^2C^2\,(a^2 + b^2 + NC)^2 \cdot f_2(x) + \tfrac{9}{8}\,N^4C^4 \cdot f_4(x).$$

Here we have considered continuous spectra. The first term in the expression is the direct-current energy.

b) A component with frequency $2\,\omega_1 - 2\,\omega_2$ and energy $\tfrac{9}{32}\,a^4b^4$.

c) A noise spectrum around the central frequency $\omega_1 - \omega_2$, built up in the same way as a zero-order spectrum for a polynomial characteristic of degree 2, but extending both sides from $\omega_1 - \omega_2$ (therefore the usual correction for zero order is not necessary here).

$$\tfrac{9}{8}\,a^2\,b^2\,(a^2 + b^2 + 2\,NC)^2 : f_0\,(x - \omega_1 + \omega_2) + \tfrac{9}{2}\,a^2\,b^2\,N^2C^2 \cdot f_2\,(x - \omega_1 + \omega_2).$$

If this formula gives energy at negative frequencies the amount of energy here should be added to the energy at the positive frequency of the same absolute value. This applies also to the following spectra.

d) A noise spectrum with central frequency $\omega - \omega_1$, to be computed in the same way as the first-order spectrum resulting from a polynomial characteristic of degree 3. As before ω is the central frequency of the original noise spectrum.

$$\tfrac{9}{8}\,NC\,a^2\,(a^2 + 2\,b^2 + 2\,NC)^2 \cdot f_1\,(x + \omega_1 - \omega) + \tfrac{9}{4}\,a^2N^3C^3 \cdot f\,(x + \omega_1 - \omega).$$

e) A noise spectrum with central frequency $\omega - \omega_2$. This follows directly from d) by interchanging a and b, ω_1 and ω_2.

$$\tfrac{9}{8}\,NC\,b^2\,(b^2 + 2\,a^2 + 2\,NC)^2 \cdot f_1\,(x + \omega_2 - \omega) + \tfrac{9}{4}\,b^2N^3C^3 \cdot f_3\,(x + \omega_2 - \omega).$$

f) Spectra with central frequencies $\omega + \omega_2 - 2\omega_1$ and $\omega + \omega_1 - 2\omega_2$ with the form of the original noise spectrum.

$$\tfrac{9}{8}\,a^4\,b^2\,NC \cdot f_1\,(x + 2\,\omega_1 - \omega_2 - \omega) + \tfrac{9}{8}\,a^2\,b^4\,NC \cdot f_1\,(x + 2\,\omega_2 - \omega_1 - \omega).$$

g) Spectra with central frequencies $2\omega - 2\omega_1$, $2\omega - \omega_1 - \omega_2$ and $2\omega - 2\omega_2$ with the form of a normal spectrum of degree 2.

$$\tfrac{9}{16}\,a^4\,N^2C^2 \cdot f_2\,(x + 2\,\omega_1 - 2\,\omega) + \tfrac{9}{16}\,b^4\,N^2C^2 \cdot f_2\,(x + 2\,\omega_2 - 2\,\omega) + \\ + \tfrac{9}{4}\,a^2\,b^2\,N^2C^2 \cdot f_2\,(x + \omega_1 + \omega_2 - 2\,\omega).$$

Again introducing effective voltages:

$$a^2/2 = V_s^2,\ b^2/2 = V_t^2,\ NC/2 = V_n^2,\ V_s/V_n = x,\ V_t/V_n = y,$$

and excluding the direct current, the ratio of the energy when the signal of frequency ω_1 is present to the energy when this signal is absent, is given by:

$$\frac{15+180(x^2+y^2)+117(x^4+y^4)+540x^2y^2+18(x^6+y^6)+234x^2y^2(x^2+y^2)+\frac{81}{2}x^4y^4+18x^2y^2(x^4+y^4)}{45+180y^2+117y^4+18y^6}$$

For y large, the result is $\approx 1+x^2$, as in the intermediate frequency spectrum. For a square-law characteristic the ratio is (computed in the same way):

$$[1+2(x^2+y^2)+2x^2y^2]/[1+2y^2].$$

For $y=6$ the results are

$1+1{\cdot}165\,x^2+0{\cdot}0612\,x^4+0{\cdot}00067\,x^6$, and $1+1{\cdot}014\,x^2$, respectively.

For $y=2$:

$$1+1{\cdot}910\,x^2+0{\cdot}449\,x^4+0{\cdot}0238\,x^6, \text{ and } 1+1{\cdot}111\,x^2.$$

If we compare these results with formula (37) we see that the higher-degree terms are much smaller than without the extra carrier. A large extra carrier has a linearizing effect.

For higher-degree spectra or more carrier waves the computation may require more time, but no essentially new problems are to be expected.

Therefore, we can finish the computation of the energy spectra resulting from the application of noise and carrier waves to a non-linear characteristic.

For literature on earlier treatments of this problem (mostly confined to square-law characteristics) we refer to Fränz's paper [1]).

Eindhoven, August 1946

REFERENCES

1) K. Fränz, Beiträge zur Berechnung des Verhältnisses von Signalspannung zu Rauschspannung an Ausgang von Empfängern, Elekt. Nach. Tech. **17**, 215-230, 1940.
2) K. Fränz, Beiträge zur Berechnung des Verhälnisses von Signalspannung zu Rauschspannung am Ausgang vom Empfängern, Elekt. Nach. Tech. **19**, 285-287, 1942.
3) E. Czuber, Die Entwicklung der Wahrscheinlichkeitstheorie und ihrer Anwendungen. Jahresber. Deutsche Math. Vereinigung **7**, II, 1-180, 1899.
4) L. Maurer, Über die Mittelwerthe der Funktionen einer reellen Variabelen. Math. Ann. **47**, 263-280, 1896.
5) A. Sommerfeld, Eine besonders anschauliche Ableitung des Gaussischen Fehlergesetzes, Boltzmann Festschrift (Ann. Phys.) Leipzig 1904, pp. 848-855.
6) L. Silberstein, The accumulation of chance effects and the gaussian frequency distribution, Phil. Mag. **35**, 395-404, 1944.
7) J. B. S. Haldane, Chance effects and the gaussian frequency distribution. Phil. Mag. **36**, 184-185, 1945.

R 50 Philips Res. Rep. 2, 260-280, 1947

HARDENING OF METALS BY INTERNAL OXIDATION

by J. L. MEIJERING and M. J. DRUYVESTEYN

PART II

621.785.613

Summary

As was described in Part I [18]), certain alloys of Ag, Cu and Ni with a few atomic % of a homogeneously dissolved metal having a sufficient affinity for oxygen can be dispersion-hardened by diffusing O into them. Two conditions must be satisfied. Firstly the oxide must be formed not as a surface layer, but dispersed in the interior of the alloy. In this connection the penetration of the reaction front and the oxide concentration was calculated. Secondly the dispersion must be very fine. The greater the affinity for O of the basis metal, the greater must be the affinity of the solute to produce oxide that conglomerates slowly enough. This was worked out in a tentative thermodynamical scheme.

In Part II diffusion coefficients of O in internally oxidized alloys of Ag and Cu are given. X-ray and electrical resistivity measurements support the view that the MgO and Al_2O_3 particles that harden silver are very small. The mechanical properties are not much affected by long annealings at high temperatures. Creep and recrystallization are slowed down considerably. A drawback is the intercrystalline brittleness of these materials, which is less serious when somewhat smaller hardnesses are aimed at. Single crystals are completely ductile. Some peculiar metallographic effects are explained.

7. *Numerical values of the coefficients of the diffusion of oxygen through internally oxidized alloys*

In Part I (section 2) a calculation was given of the penetration velocity of the boundary between the internally oxidized rim (*e.g.* Ag + MgO) and the unoxidized interior of the (dilute) alloy (Ag + Mg). The result was that the penetration is proportional to the square root of the time of heating in the O_2-containing atmosphere: $\xi = 2\sqrt{\alpha t}$. The constant α was a complicated function of c_2 (atomic fraction of Mg), c_1 (*idem* of O at the surface; this solubility is practically proportional to the square root of the O_2 pressure), D_2 (diffusion coefficient of Mg through Ag + Mg) and D_1 (*idem* of O through Ag + MgO).

In most practical cases (when D_2 is negligible and c_1 very small with respect to c_2), however, this function could be approximated by $\alpha = c_1 D_1/2c_2$, which is accurate enough for the purpose of this section. The equivalent equation $\xi^2 = 2\, c_1 D_1/c_2\, t$ can be obtained directly by integrating the differential equation $(c_1/\xi)\, D_1 \mathrm{d}t = c_2 \mathrm{d}\xi$, which expresses that the amount of O arriving at the reaction front along the concentration gradient c_1/ξ

(see *fig. 12*) during a short time dt converts an equal amount of Mg into MgO by pushing the front forward over the short distance dξ.

The internal oxidation of copper alloys also proceeds proportionally to $\sqrt{t}$, but the formation of a Cu_2O scale must be taken into account (section 3).

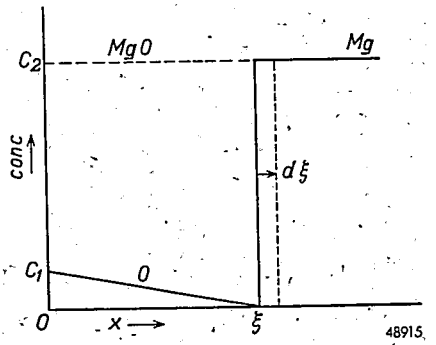


Fig. 12. Concentrations of O, Mg and MgO in Ag, as functions of x, the distance from the surface of the strip. Limiting case: only O diffuses, and $c_1 << c_2$. Schematically.

The theoretical proportionality of the depth of internal oxidation to the square root of the time *) and (for Ag alloys) to the square root of the oxygen pressure is confirmed experimentally. From the penetration of the oxidation boundary with time and the analysis of the alloy we get the product c_1D_1. When the valency of the solute metal is not 2, we must multiply c_2 by a factor which for Al is 3/2 and for Mn 1·81 (derived from the weight increment, see section 6).

The Ag alloys examined yielded more or less differing values of c_1D_1, which were distinctly higher than was to be expected from the work of Johnson and Larose [1]. The following special experiments with careful temperature control proved that these differences were real. In table III D_1 itself is given, and the required values of c_1 for 800 °C and 650 °C are taken from the measurements of the solubility of O in Ag by Steacie and Johnson [2]. Combining these measurements with those of Johnson and Larose [1] on the diffusion of oxygen through 0·3 mm-thick silver foil we find for D_1 at 754 °C: $4{\cdot}21.10^{-6}$ cm²/sec and at 608 °C: $1{\cdot}19.10^{-6}$ cm²/sec. Taking log D_1 as a linear function of $1/T$ the values at 800 °C and 650 °C in the table are found.

The acceleration of the diffusion of O by the presence of relatively coarse particles (*e.g.* CdO) can be due to the same mechanism which causes the high mobilities of H^+ and OH^- ions in water: an O atom arrives at one point of a CdO complex and another O atom starts off from another point of this same complex. When the particles (*e.g.* MgO) are so small that the

*) compare *fig. 16.*

TABLE III

Diffusion coefficients of O in oxidized Ag alloys *)

Weight %	At. %	D_1 800 °C (cm²/sec)	D_1 650 °C (cm²/sec)
0·23 Mg	1·01	$1{\cdot}25.10^{-5}$	$0{\cdot}52.10^{-5}$
0·046 Mg	0·20	$0{\cdot}9\ .10^{-5}$	
0·26 Al	1·03	$1{\cdot}7\ .10^{-5}$	
0·66 Mn	1·30	$1{\cdot}8\ .10^{-5}$	
0·79 Zn	1·30	$2{\cdot}1\ .10^{-5}$	$0{\cdot}79.10^{-5}$
0·96 Cd	0·92	$2{\cdot}2\ .10^{-5}$	
0·20 Cd	0·19	$1{\cdot}8\ .10^{-5}$	
Pure Ag, Johnson		$0{\cdot}58.10^{-5}$	$0{\cdot}18.10^{-5}$
Pure Ag, Larose			

Ag lattice is distorted, this can also accelerate the diffusion of oxygen. In this case it is possible that the solubility of O is changed too, so that the value of D_1 calculated with c_1 for unalloyed silver is not the true one.

For the activation energy of the diffusion of O in oxidized Ag alloys we find about 12 kcal against 15 kcal from Johnson and Larose's values for pure Ag.

Our results for the copper alloys are more homogeneous; as the accuracy of α **) is diminished by the formation of an external oxide scale, we are not convinced that the mutual differences are real, in contrast with the case of the silver alloys.

All values of D_1 in table IV are calculated with the help of Rhines and Matthewson's [4]) measurements of the solubility of oxygen in copper.

Smith's values are probably somewhat low, as the oxidation boundaries in the Si-copper were serrated by the formation of inner oxide films. As Ransley [5]) has measured the oxygen diffusion by annealing copper, containing Cu_2O inclusions, in a CO atmosphere and observing the displacement of the boundary of the Cu_2O-free copper, his values of c_1D_1 will be more easily affected by an insufficient solution velocity of Cu_2O than our values. Therefore we cannot say for certain that O is diffusing more rapidly through Cu containing oxide particles than through Cu that (apart of course from the homogeneously dissolved oxygen) is pure.

For the activation energy of the diffusion we find 50 kcal (Ransley: 46 kcal, Smith: 54 kcal).

*) *Note added in proof:* We have now been able to see the papers of Rhines *et. al.* which appeared during the war. They too find for Ag alloys [19]) (with Cd, Cu, In, Sn and Sb) a penetration that is a few times faster than is to be expected from [1]), but they suppose that Johnson and Larose's results are too low. For the Cu alloys [20]) the agreement is better.

**) Calculated with the formulae in section 3 for $\xi_0 = 0$.

TABLE IV

Diffusion coefficients of O in oxidized Cu alloys

Weight %	At. %	D_1 950 °C (cm²/sec)	D_1 800 °C (cm²/sec)
0·37 Be	2·55	$3{\cdot}7.10^{-6}$	$0{\cdot}2.10^{-6}$
0·64 Al	1·49	$3{\cdot}7.10^{-6}$	$0{\cdot}2.10^{-6}$
1·20 Ni	1·30	$4{\cdot}7.10^{-6}$	
2·26 Ni	2·44	$4{\cdot}0.10^{-6}$	
0·29 Zn	0·28	$3{\cdot}6.10^{-6}$	
0·97 Zn	0·94	$3{\cdot}2.10^{-6}$	
0·103 Si*)	0·23	$2{\cdot}9.10^{-6}$	$0{\cdot}13.10^{-6}$
Pure Cu, Ransley		$2{\cdot}1.10^{-6}$	$0{\cdot}15.10^{-6}$

The diffusion coefficient of O through nickel cannot be given, as c_1 is not known. For $c_1 D_1$ we find 10^{-9} cm²/sec at 1200 °C, (c_1 expressed in atomic fraction), in good accord with the value $1{\cdot}4.10^{-9}$ cm²/sec found by extrapolation from Smithells and Ransley's [6]) results at lower temperatures. These authors give values for D_1 which, however, must be far too low, as they substitute the atomic O fraction of the eutectic liquid for c_1.

8. *X-ray measurements*

For a number of silver and copper alloys the spacing of the face-centered lattice was measured before and after oxidation by Dr J. F. H. Custers in this laboratory**). The samples were thin rods etched into a point of about 0·3 mm diameter. CuKα, CuKβ and CoKα rays were used. The diameters of the highest-order rings were measured by Straumanis's and by Cohen's methods [7]). No diffraction lines of the oxide were observed. The silver alloys were oxidized at 800 °C in air, the copper alloys at 950 °C. The error of measurement in the spacings given, due to the X-ray measurements, is about 0·0002 Å.

From the table we see that the spacing of Mg-silver and Al-silver increases markedly by the oxidation to a value distinctly larger than that of pure Ag. The influence of the dissolved free oxygen in Ag is seen to be small. This result shows that the Mg and Al oxides are dissolved in the silver lattice, presumably in almost molecular dispersion, at least for a considerable part.

The silver alloys with Cd and Zn, which do not harden appreciably under these circumstances, give a spacing almost equal to that of pure Ag, indicating the formation of larger oxide particles.

*) Calculated from C. S. Smith [3]), supposing formation of SiO_2. An eventual formation of copper silicate would mean a higher value of D_1.

**) We thank Dr Custers for these precision measurements.

TABLE V

Spacings in Å units of annealed *) and oxidized alloys

Weight %	At. %	Annealed	Oxidized
Ag + 0·04 Be**)	0·5	4·0772	4·0768
0·36 Mg	1·58	4·0785	4·0837
0·26 Al	1·03	4·0762	4·0785
0·79 Zn	1·30	4·0768	4·0774
0·96 Cd	0·92	4·0774	4·0777
Ag pure		4·0774	4·0779
id. [8])		4·0772	
id. [9])		4·0779	
Cu + 0·20 Be	1·4	3·6039	3·6081
0·75 Mg	1·93	3·6152	3·6071
0·81 Al	1·89	3·6087	3·6070
Cu pure [8])		3·6078	

All Cu alloys have in the oxidized state a spacing that is almost equal to that of pure copper, although hardening takes place ***). This is in agreement with the copper-red colour of the oxidized alloys.

9. *Comparison with nitride hardening*

The picture given of hardening by oxidation is quite analogous to that mostly accepted for nitride hardening of iron (with Al and/or Cr) [10]). There are three remarkable differences:

1) the hardness attainable by nitriding is much higher,
2) the difference in heat of formation of AlN (or CrN) and Fe_4N is a good deal smaller than *e.g.* between BeO and Cu_2O,
3) the inner boundary of the "case" is far less sharp than the oxidation boundary even in Cd-silver.

The last point indicates that the reaction Al + N → AlN necessitates a considerable energy of activation, perhaps about 40 kcal. In the opposite sense of the reaction (necessary for conglomeration) this partly compensates the smaller relative stability of the nitride. Furthermore, this difference in stability between oxide and nitride will be less in molecular form (surrounded by metal) than in bulk, because it is plausible that the heat of solution †) in metals becomes less positive when we pass from

*) ZnAg and CdAg in N_2, rest *in vacuo*.
**) This Be concentration is rather uncertain.
***) Apparently with increasing conglomeration of the oxide the lattice constant (and the electrical resistivity, see section 12) approaches its final value more quickly than the hardness does. This is in accordance with the experience on the subject of overageing age-hardening alloys.
†) Compare part I, p. 99.

oxides to nitrides (and then to carbides), owing to the growing metallic nature of these compounds. The low temperatures used in nitriding will also be favourable for its effectiveness.

10. *Effect of concentration, oxygen pressure and oxidation temperature*

We will at first only consider alloys without inner oxide films.

The case of AlNi, where oxidation hardening is only marked at higher concentrations (see section 6) seems to be an exception. In BeCu 0·1 weight % Be already gave a hardness of 105 Vickers (against ± 50 without oxidation); 0·2 and 0·35 % Be gave practically the same hardness: ± 165 Vickers. This same type of hardness-concentration curve (first rather steep, then flatter) is found in the hardening silver alloys; it will partly be due to the increasing coagulation of the oxide (section 6).

We have not studied systematically the influence of the oxidation temperature on copper and nickel alloys, and it is clear that the oxygen pressure has no influence as long as a Cu_2O (resp. NiO) layer is formed.

In silver with 0·3 weight % Mg or Al the hardness produced at 800 °C in air was only slightly (5-10 Vickers) exceeded at 650 °C and/or in oxygen of 1 atm. Extremely low O_2 pressures gave a much smaller hardness: annealing 0·3 % MgAg for 6 hours at 850 °C in N_2 containing about 0·02 % O_2 (followed by $^1/_2$ hour 850 °C air to make sure that the oxidized rim was sufficiently thick) gave 100 Vickers, against 162 after $^1/_2$ hour 850 °C air alone. The penetration velocity in N_2 with 0·2 % O_2 confirmed that nitrogen itself cannot be reponsible for the rims; its solubility in solid silver is not measurable [2]).

In Ag with 0·15 % Be the hardness produced at 650 °C was 15 Vickers *lower* than at 800 °C, owing no doubt to insufficient solid solubility of Be, which already at 750 °C has decreased to 0·13 % [11]).

The different hardnesses of oxidized CdAg shown in table VI might be (partly) due to inequal softening (*e.g.* by crystal growth) during the three different heat treatments.

TABLE VI

Silver alloys. Vickers hardness ($2^1/_2$ kg load)

	0·96 Cd	0·79 Zn	0·58 Mg	1·14 Mg	1·03 Al
$6^1/_2$ h. 800 °C air	39	59	107*	70*	30
48 h. 800 °C air			170	150	
2 h. 800 °C O_2		75	215	125-165*	60
id. + $6^1/_2$ h. 800 °C air			220	170	55
7 h. 800 °C O_2				190	
40 h. 650 °C air	48	98	203	228	40-80
id. + $6^1/_2$ h. 800 °C air			199	212	40
20 h. 650 °C O_2	55	105	210	230	110

Pressure 1 atm. Concentrations in weight %.

Especially the influence of the temperature is considerable when the conditions for hardening are nearly fulfilled: ZnAg (compare section 6, p. 101).

The same is true when the penetration of oxygen is seriously opposed by *inner oxide films.*

The hardness values of 1·03 % AlAg in table VI are so low because the rims are too thin (0·02-0·03 mm after $6^1/_2$ h. 800° air, 0·1 mm after 20 h. 650° O_2). The hard and brittle nature of these rims is shown by the appearance of surface cracks on bending the (1 mm-thick) strips. The hardness values of MgAg marked with an asterisk are also too low; after 48 h. 800° air or 7 h. 800 °C O_2 the rims are thick enough (0·25 mm or more) for measuring the true hardness with $2^1/_2$ kg load. Apparently a concentration higher than 0·3 % Mg is only able to yield a greater hardness when the oxidation temperature is lowered and/or the oxygen pressure increased. With 2 % Mg at 650 °C in O_2, 260 Vickers was obtained.

In table VII the ratio of effective α (see part I, p. 86), calculated from the penetration of the oxidation, and the theoretical α, calculated according to section 7 with 0·23 % MgAg as basis, is given.

TABLE VII

$\alpha_{eff}/\alpha_{theor}$

	800 °C		650 °C	
	Air	O_2	Air	O_2
0·58 % Mg	0·1	0·7	>0·7	>0·6
1·14 % Mg	0·05	0·1	1·1	>1·2
1·03 % Al	0·005			0·05

In those cases where the oxidation of the strips was complete, only a minimum value could be derived. The formation of inner oxide films, and with it the factor $\alpha_{eff}/\alpha_{theor}$, can probably be influenced by several secondary circumstances; the three alloys concerned were in the form of rolled strips.

The accuracy of table VII is not so great that the factors of 1·14 % MgAg could not be less than 1, but even so the retarding effect of the many oxide films in MgAg oxidized at 650 °C (see *fig. 13*) looks surprisingly small. It is probable that the real factors are smaller, because at these higher oxide concentrations c_1D_1 (and therefore α_{theor}) will be still greater than in oxidized 0·23 % MgAg (comp. p. 262).

On the other hand the retarding effect of outflanked oxide films is lessened by the impoverishment in Mg behind them: as Mg^{++} ions will

diffuse much easier through MgO than O^{--} ions, the Mg atoms which reach the back of the films can be transformed into MgO. This must result in a fluctuation in the MgO concentration, being minimal just behind the films and maximal in front if them. As the depth of etching increases markedly with increasing MgO content the surface of an etched cross-section of a com-

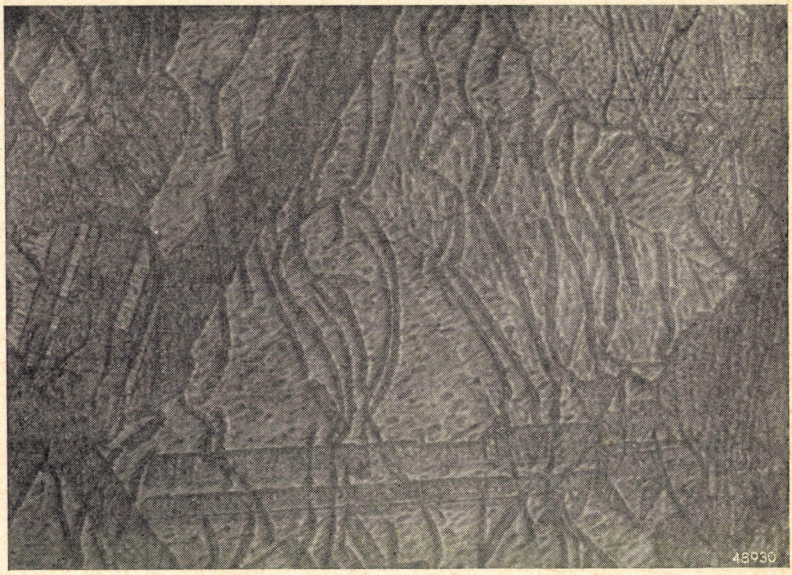


Fig. 13. Ag with 0·58 % Mg, 40 h. 650° air (+6½ h. 800° air); etched $NH_4OH + H_2O_2$. Magn. 600 ×. Inner oxide films, running roughly parallel to the surface of the strip. Arrow indicates direction of penetration of O.

pletely oxidized strip with many oxide films must have a relief as schematically represented in *fig. 14*. This is strikingly confirmed (we found the effect, before the interpretation given here, in all cases where 0·58 % and 1·14 % MgAg were oxidized completely or fairly deeply) by the following experiment: we look through a small 10 × lens at the specimen, holding it so that we see the light reflected on the bright surface of the holder

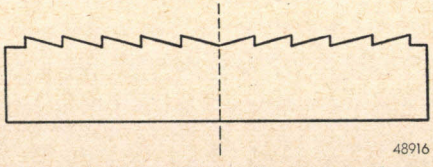


Fig. 14. Surface relief of etched cross-section of a completely oxidized strip with many inner oxide films parallel to the strip surface. Very schematical.

(in which the sectioned silver strip is clamped, polished and etched), the long side of the strip cross-section being perpendicular to the surface through the incoming and reflected rays. We now rotate the specimen some-10° about an axis passing through this long side. After a clock-wise rotation (viewed from the left-hand side) we see the half of the strip (when com-

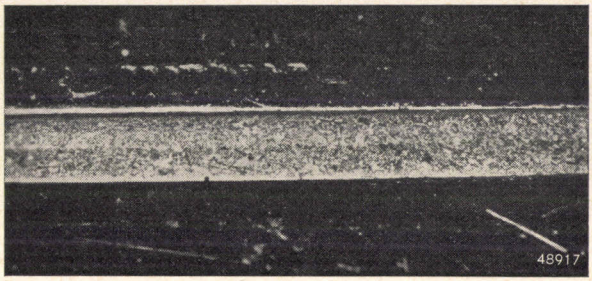


Fig. 15. Ag with 0·58 % Mg, 40 h. 650 °C air; etched $NH_4OH + H_2O_2$. Magn. 9 ×. Illumination oblique. Marked difference in reflection of the two halves of the cross-section suggesting a difference in inclination of the etched surface.

pletely oxidized) farthest from our eye darker than the nearer half. Rotation in the opposite direction shows the effect reversed. The "middle line" is beautifully sharp, straight and in its right place, showing that the retarding effect of the oxide films cannot be very great and is smoothed out statistically.

The effect cannot be due to one single "valley" over the whole cross-section, as the middle was not measurably deeper (< 1 μ) than the borders.

The pattern of the brittle fracture of these oxidized alloys with many inner oxide films shows that these are also followed by the fracture, besides the crystal boundaries. (In most cases these films and crystal boundaries run independently of each other).

11. *Influence of treatment before and after hardening of silver alloys*

Oxidizing a cold-*rolled* strip or the same alloy in the *cast* condition yielded practically the same hardness; only with Ag + 0·26 % Mg + 0·11 % Ni *) found we an appreciable difference (156 against 168 Vickers respectively, oxidized at 800° in air).

Pre-annealing in *hydrogen* (1 hour 800°) decreased the hardness after oxidation of 0·15 % BeAg at 650° from 111 to 96 Vickers; in view of the heterogeneous nature of this alloy at this temperature (see p. 265) this is

*) This addition is only partially in solid solution, and causes a grain refinement.

not strange. Pre-annealings in *nitrogen* had a greater softening effect, and it also occurred in more alloys, but as we did not realize at the time that the oxygen content of our gas (0·02 %) could not be ignored (see p. 265) we had not purified it.

No influence of the *rate of cooling* after oxidation was found: quenching in water, cooling in the air, or slow furnace-cooling during 15 hours to 100 °C gave practically the same hardness.

Very *long annealings in air* at 800 °C and 900 °C show a softening (see table VIII) which is probably for the greater part the result of conglomeration; crystal-growth is perhaps also a factor (comp. section 15).

TABLE VIII

Vickers hardness ($2^1/_2$ kg load)

	16 h. 800 °C	375 h. 800 °C	375 h. 800 °C + 117 h. 900 °C
0·1 % Mg	122	119	108
0·2 % Mg	153	149	124
0·4 % Mg *)	205	175	162
0·9 % Mg *)	181	189	171
0·5 % Al	160	161	145
0·7 % Mn	130	123	105
0·8 % Zn	68	58	52

Cast material. Concentrations in weight %.

Even after this long annealing a very clear "middle line" (part I, p. 92) was shown by the MnAg and 0·2 % MgAg strip cross-sections, illustrating the immobility of the oxide.

The expectation that removal of the free oxygen would accelerate the conglomeration (end of section 6) was fulfilled. After one hour *annealing in nitrogen* or a vacuum at 800 °C the fall in hardness was still small: ± 12 Vickers for 0·33 % MgAg and ± 5 Vickers for 0·26 % AlAg. It is noteworthy that, while oxidation of 0·79 % ZnAg in air at 800 °C (*e.g.* for 2 hours) gave only 60 Vickers, the greater hardness obtained by oxidizing at 650 °C (98) decreased only to 92 Vickers after annealing for 1 hour at 800 °C in N_2. This demonstrates clearly that the oxide conglomerates much faster in the earliest stages after its formation (when the oxidation boundary is just passing) than later on.

Ten hours annealing at 800 °C in N_2 (again with 0·02 % O_2) decreased

*) These two alloys were oxidized 2 minutes at 800 °C in oxygen beforehand.

the hardness of 0·33 % MgAg from 175 to 130 Vickers and that of 0·26% AlAg from 140 to 115 Vickers.

Annealing in *hydrogen* for 1 hour at 800 °C decreased the hardness of these two alloys to about 100 Vickers, also when an intermediate 10 hours annealing at 800 °C in N_2 was used to expel by far the greater part of the free oxygen.

The hardness of 0·79 % ZnAg (even when oxidized at 650 °C) fell back to 36 Vickers after a subsequent annealing of one hour at 800 °C in H_2, suggesting that at least the finer (and thus less stable) ZnO particles were reduced to Zn. These softened specimens were still quite brittle. This hydrogen embrittlement (well known in copper containing Cu_2O) is also shown by commercially pure silver containing only free oxygen, and is attributed to accumulations of H_2O *).

12. *Electrical resistance*

For a number of Ag and Cu alloys the electrical resistance was measured in the normal (annealed) and in the oxidized state. Drawn wires (diameter 0·5 till 1 mm) were used in most cases. Table IX gives the electrical resistivity in 10^{-6} Ω cm at 0 °C in the annealed and in the oxidized state.

The annealing generally took place in a high vacuum, the annealing temperature lying between 500 °C and 950 °C. The oxidation temperature was generally 800 °C for the silver alloys and 950 °C for the copper alloys. Just as in the case of the hardness, the oxidation temperature has an influence on the resistivity. Also other factors influence the resistance, *e.g.* a long time of annealing before or after oxidation. Thus the numbers given in table IX must be understood as only semi-quantitative.

As an example of the influence of different treatments we shall give the resistivity of some wires of Ag + 0·33 % Mg in 10^{-6} Ω cm. The drawn wire (diameter 0·75 mm) had a resistivity of 2·25. Annealing for one hour at 400 °C in vacuum gave a decrease to 2·14, annealing at 800 °C gave 2·05. Oxidation for 1 hour at 800 °C in air increased the resistivity to 2·32. When the drawn wire was directly oxidized (without annealing) at 800 °C we obtained 2·40. Oxidation for 40 hours at 650 °C in air gave 3·20. Heating for 2 hours in N_2 at 800 °C after oxidation gave a decrease in the resistance of 5 to 10 %. It is seen that the influence of these different treatments had an influence on the resistance of the same nature as it had on the hardness (see sections 10 and 11).

The results of table IX show that the resistance of the oxidized Cu alloys decreases to a value that is not very different from the value of pure

*) The detrimental effect of an H_2 annealing on oxidized silver and dilute Ag alloys has been studied more fully by Chaston [12]).

TABLE IX

Resistivity in 10^{-6} Ω cm at 0 °C of annealed (ϱ_a) and oxidized (ϱ_o) Ag and Cu alloys

	ϱ_a	ϱ_o	$\frac{\varrho_o - \varrho_a}{\varrho_a} 100$
Ag + 0·33 % Mg	2·05	2·32	+13%
Ag + 0·15 % Be	1·56	1·92	+23
Ag + 0·26 % Al	3·45	2·39	—31
Ag + 0·60 % Ti	2·39	1·94	—19
Ag + 0·79 % Zn	2·05	1·65	—20
pure Ag *)	1·75	1·57	—10
Cu + 0·75 % Mg	2·30	1·72	—25
Cu + 0·1 % Be	2·08	1·70	—18
Cu + 0·69 % Be	2·95	1·76	—40
Cu + 0·81 % Al	3·17	1·78	—44
Cu + 1·1 % Ti	8·45	1·78	—79
pure Cu	1·61	1·62	0

The % are weight %. Ag + 0·79% Zn was annealed in N_2 to cut down evaporation of Zn.

Cu **). For the silver alloys the results are quite different, only the non-hardening Zn alloy decreased to about the silver resistance. The other silver alloys which harden by oxidation show a resistance much higher than that of silver, for the bivalent Mg and Be even an increase of the resistance by oxidation occurs.

We will discuss the two main factors that have an influence on the resistance:

1) Internal strains cause generally an increase of the resistance as the lattice is distorted irregularly by these strains. This effect causes a decrease of the resistance by annealing. Tammann and Dreyer [13]) found that by annealing after 98 % rolling the resistance of pure Ag decreases by 5 %, the decrease of pure rolled Cu after annealing being only 1·6 %.

2) A small number of foreign ions, substituted on the normal lattice places, increases the resistance, this effect being for Cu and Ag alloys the larger the higher the valency of the alloying element [14]), as ions with a high electric charge will scatter the electron waves considerably. When precipitation occurs and larger particles are formed the resistance will decrease. It is uncertain whether the scattering of electron waves in an Ag lattice is highest on Mg ions or on MgO molecules.

To summarize our discussion we may state that the change of resistance by dissolving a small amount of a metal in, say, silver is due to two causes: the silver lattice is distorted and the electric field near a foreign ion or

*) Evidently not very pure.

) Compare note *) on p. 264.

atom differs from that near an Ag ion. The oxidation hardening causes a change in both factors. When hardening occurs the distortion is probably increased, the scattering of the electron waves near the foreign ions probably decreases, at least when larger particles are formed or when the scattering near the foreign ion is very large as is the case for trivalent and quadrivalent ions (Al and Ti). For Mg and Be silver alloys the distortion effect seems the largest, in the other cases the other effects probably overshadow this increase of the resistance.

13. *Visibility of old boundaries*

When the internal oxidation of a strip or wire (*e.g.* of an Ag or Cu alloy) is interrupted one or more times by cooling, the position of the successive intermediate oxidation boundaries is marked (after etching) by clearly visible lines (*fig. 16*), often of the same sharpness as the last boundary. The speed of cooling and reheating may vary between fairly wide limits without the effect being lost. Even quenching in water and afterwards putting the strip in the furnace again causes such a line. Experiments with wires of 0·75 mm diameter, cooled and heated rapidly by switching off and on the electric current passing through it, gave the same result.

One can separate the effects of cooling and heating when the lower temperature is not room temperature but *e.g.* 650 °C with silver alloys. Both changing the temperature rapidly from 800 °C to 650 °C or *vice versa* results in a visible line. With copper alloys these last experiments are less satisfactory because the contact with the copper oxide is often impeded by the cooling, while on the other hand it does not flake off (as is generally the case when we cool to room temperature) and may hamper the formation of a fresh layer.

An explanation of these curious lines may be offered by the difference between the temperature coefficient of diffusion of oxygen and that of metallic solutes in silver and copper. In section 7 we have seen that the activation energy for O through Ag is about 15 kcal; for Cd, In, Sn, Sb and Cu through Ag [15]) the mean value is 23 ± 1 kcal. For O through Cu the activation energy is 50 kcal (see section 7); Rhines and Mehl [16]) found for Al, Si and Zn through Cu about 40 kcal, for Be about 30 kcal.

Thus a temperature drop increases D_1/D_2 for Ag alloys and decreases it for Cu alloys. An inspection of the boundary conditions (see equation (2) in section 2) shows that a *sudden* drop of temperature during the oxidation of a Cu alloy (or a sudden rise with an Ag alloy) must result in a *backward* move of the oxidation boundary during a short time. Thus $d\xi/dt$ would be zero at two points close to each other, where c_{ox} would "theoretically" be infinite. Of course the assumptions underlying section 2 (especially that the diffusion does not depend on the — normally small — concentration

of the oxide formed) do not hold here, but it is clear that a narrow zone of relatively great oxide concentration will be formed, which can be revealed by etching. In the reverse cases (temperature rise with Cu and drop with Ag alloys) the reaction balance at the oxidation boundary is disturbed in the other direction and a zone poor in oxide will be formed, which is also revealed by etching.

The temperature variations will also change the solubility of the oxygen

Fig. 16. Ag with 0·23 % Mg, not far from short side of $2^1/_2$ mm strip. Annealed 3 times: 50 min + 150 min + 250 min 800 °C air. Etched $NH_4OH + H_2O_2$. Magn. 50 ×. Besides scratches 3 oxidation boundaries are seen. The last is oval, the second takes its turn farther on, and the first is seen to begin its bend only just inside the photographed area. In the region where all three run straight and parallel to the surface of the strip, the three mutual distances are equal, proving that the penetration of oxidation is proportional to the square root of (total) time (*i.e.* 50, 200 and 450 min respectively). The visibility of the first two boundaries is explained in section 13.

(c_1). This will result in a (small *)) effect in the same direction in Cu alloys, while in Ag alloys it will be in the opposite sense. Here c_1D_1/c_2D_2 varies very little with temperature, but this does not mean that the total effect in silver alloys is small, because the change in O concentration starts from the surface of the alloy, so that its effect at the oxidation boundary is less instantaneous than that of the change of D_1/D_2; we must expect a lag and a peak in the oxide concentration beside each other.

To get a, very rough, idea of the influence of the speed of the temperature change we have calculated how a Cu alloy must be cooled from 950 °C to make the oxidation boundary remain stationary. For the activation energies of D_1 and D_2 are taken 50 and 33·3 kcal; the complications arising from the external scale and the change of c_1 are ignored. Some points of the temperature-time curve are given in table X; the unit of time is the time during which the alloy is oxidized at 950° before being cooled.

From this we would conclude that the greater the change in temperature the less rapid it has to be to create a visible effect.

TABLE X

time	°C	time	°C
0	950	0·005	900
0·0001	940	0·06	850
0·0004	930	1	760
0·001	920	8	700

That the temperature changes act only via disturbances in the reaction balance is proved by the fact that intermediate boundaries are also shown (less sharp, but clearly visible) when the oxidation of a silver alloy is interrupted at constant temperature by replacing the oxidizing atmosphere by nitrogen for a while. Here, too, the process can be divided: after heating a strip of Ag with 0·3 % Mg 5 minutes in O_2, 30 minutes in air and 15 minutes in O_2 again, throughout at 800 °C, two intermediate boundaries were clearly revealed by etching.

Rhines **) has found peculiar regularly banded structures parallel to the oxidation boundary in oxidized Be-Cu alloys. We had not found these "Liesegang bands" at first, because they appear only after severe etching. The most beautiful ones we found in a 0·25 % BeCu single-crystal rod which after oxidation at 950 °C was rolled and annealed in

*) According to Rhines and Matthewson [4]) the solubility of O in Cu decreases from 950 °C to 700 °C only by 25%.

**) [17]); especially fig. 19 and 31.

nitrogen (see section 15). But often they were too vague to be visible at magnifications above, say, 200. Corresponding periodic variations in colour were often produced by tarnishing after etching. In many cases the bands could not be developed, especially when the oxidation was carried out in a furnace whose temperature was not automatically controlled by a relay but kept constant by hand. Specimens of the same alloy, after identical heat treatment in this furnace or in a relay-controlled Heraeus

Fig. 17. Round rod of Cu with 0·25 % Be, oxidized in air at 950 °C in relay-controlled furnace. Deep etch with HNO_3. Magn. 150 ×. Up to the oxidation boundary (inner circle) vague rings are visible, suggesting a Liesegang phenomenon, but caused by the temperature oscillations.

furnace, were polished and etched together in the same holder: only the last ones showed the bands. Thus it appeared that the regular temperature variations (total amplitude 12 °C; period 8 min 45 sec) were responsible.

To prove this we counted the bands. As they were often indistinct near the surface (this is also the case in Rhines's photographs 19 and 31) and the amount of alloy lost as scale introduced an uncertainty, they were oxidized twice; the number of bands between the intermediate and the final boundary was indeed found to be precisely the same as the number of relay periods (up to 50 in one experiment) during the second oxidation. Altering the speed of penetration by varying the Be content and the temperature (between 920 °C and 950 °C) gave the same result and the possi

bility of a mere coincidence was definitely discarded by comparing two Heraeus furnaces with periods of 6 min 30 sec and 8 min 45 sec respectively.

Thus our bands are no Liesegang bands at all; Rhines's bands could be explained if his furnace showed temperaure variations with a period of about 6 minutes. On the other hand they are no normal "intermediate boundaries" either: these appear much earlier upon etching (about at the same time as the crystal boundaries). The temperature variations are so small and slow (compare table X) that some amplifying process must play a part in rendering the bands visible. In fact heat treatments which should be favourable for coagulation of the oxide seem to favour the appearance of good bands, see *e.g.* the case of the single crystal mentioned above; the only bands we have found in silver were in an alloy of Ag with 0·5 % Al, heated (in a Heraeus furnace) in air for 375 hours at 800 °C plus 117 hours at 900 °C (see table VIII), while we could not find them in the same alloy heated only 185 hours at 800 °C. These bands in silver make it highly improbable that a periodic loosening of the contact between a Be-copper alloy and its scale plays an important part.

We cannot state definitely that the bands are not formed in other Cu alloys, as we have not sought them systematically and we could not even find them in all Be-Cu alloys; in some cases they were absent in flat strips but present in cylindrical rods of the same alloy. Rhines [17]) states that he has found "Liesegang bands" in alloys of copper with Be, Al and Si only (and not with some 30 other alloying elements), but in the Al-Cu and Si-Cu alloys they are of quite another type (irregular, not parallel to the oxidation boundary); these inner oxide films appear already before etching (after polishing), as we have described in section 6. They are not caused by temperature fluctuations, but in our opinion they cannot be regarded as a Liesegang phenomenon either, owing to their not being parallel to the diffusion front.

14. *Mechanical properties*

When internal oxidation results in a considerable hardening (thus not in the case of CdAg, ZnCu *etc.*), this is accompanied by a serious embrittlement in polycrystalline specimens. The elongation in tension of completely oxidized alloys is often only 1 or 2 %, especially when the hardness is 140 Vickers or more.

Silver with 0·15 weight % Be oxidized at 800 °C (hardness: 126 Vickers) gave 15% elongation (measured over 40 times the wire diameter). A very fine grain seems to increase the ductility: Ag with 0·26 % Mg + 0·11 % Ni gave 12 % elongation with ± 160 Vickers.

Oxidized 0·79 % ZnAg (at 650 °C) and 0·15 % BeAg with hardnesses of 100-125 Vickers were cold-rolled from 1 mm to 0·5 mm thickness

without macroscopically visible cracks, or with small cracks at the borders of the strips, while oxidized Ag with an equivalent amount of Mg or Al instantly developed visible intercrystalline cracks on rolling (or hammering), which, however, did not result in disintegration on further rolling for a long time.

The tensile strength of Ag and Cu hardened by oxidation seldom exceeds 35 kg/mm², which is strikingly low compared with *e.g.* cold-worked Cu of the same Vickers hardness.

Table XI presents the results of a 0·1 % BeCu wire partially oxidized when it had 0·75 mm diameter and then cold drawn to 0·10 mm diameter. It is compared with the unoxidized wire drawn to the same diameter, also after a subsequent half-hour annealing in N_2 at 400 °C or 800 °C. The tensile strength of the fully oxidized wire obtained by extrapolation is seen to decrease relatively little by the annealing. The increase by the cold-drawing of the oxidized wire (in the 0·75 mm stage the tensile strength was 35 kg/mm², with 9·5 % elongation) can therefore, in our opinion, only partly be a normal work-hardening. Perhaps the drawing process causes an adjustment at the crystal boundaries, thus increasing the abnormally low tensile strength.

TABLE XI

Tensile strength (σ_B) in kg/mm², and elongation (δ) measured over 50 × wire diameter, of 0·1 mm ⌀ 0·1 weight % Be copper.

Material / Subsequent treatment	Not oxidized		78% oxidized *)		Fully oxidized (extrapol.) **)
	σ_B	δ	σ_B	δ	σ_B
cold drawn	69·6	1·1%	67·0	0·3%	66
annealed ½ h. 400° N_2	40·1	19%	58·9	1·1%	64
id. ½ h. 800° N_2	34·8	16%	52·1	2·0%	57

The Brinell hardness at 600° was measured with a 10 mm Widia ball (load 250 kg during 10 seconds) on a 9 mm thick piece of 0·1 % Mg silver, unoxidized and oxidized 50 hours at 800° in air, which gave a rim some 4 mm thick. The hardnesses were 9 and 25 kg/mm² respectively.

*) 47 % of the radius, thus 22 % of the volume, not oxidized.

**) This extrapolation is rough for the two annealed conditions, because here the core of the 78% oxidized wire is much more ductile, and cannot have attained its full tensile strength at the moment of rupture. Thus the last two extrapolated values of σ_B will be too low, and its real decrease by annealing still less than in the table.

A great improvement in the creep properties by oxidation hardening was shown by simple bending tests on $^1/_2$ mm thick strips loaded during $^1/_2$ hour. For instance 1 hour 400 °C vacuum-annealed 0·1 % BeAg strips crept 10 times faster at 400 °C with 20 g load than the oxidized strips (800 °C air) at 530 °C with 50 g load.

These mechanical properties, together with good conductivity (as far as the Ag and Cu alloys are concerned) make these materials suitable for several uses, especially in the electronic industries.

15. *Single crystals*

Because of the intercrystalline character of the brittleness of the alloys hardened by oxidation we tried whether single crystals were ductile. This indeed proved to be the case. (Inner oxide films are not considered here.)

Rods of 3-3$^1/_2$ mm diameter were prepared *) by sucking molten Ag with 0·25 % Mg and Cu with 0·25 % Be respectively into thin-walled quartz-glass tubes. The glass was easily removed after quenching from 800° in water. Besides some less coarse-grained rods there were some which consisted of crystals between 10 and 60 mm length. Some silver-alloy rods were completely hardened by heating 27 hours at 800 °C in air, one copper-alloy rod (A) by heating 26 hours at 950 °C in air, having still 1·6 mm diameter after removal of the scale. Another copper rod (B) after 17$^1/_2$ hours was 2·15 mm in dia., having an unhardened core with a quarter of this diameter. They were all easily broken into their component crystals, but these themselves were quite ductile in bending **). Tensile tests gave 28 % and 21 % elongation respectively ***) for completely hardened MgAg and BeCu (against less than 1 % for a polycrystalline specimen of MgAg). The tensile strength was 31 and 28$^1/_2$ kg/mm^2 respectively.

Cold rolling to 0·5 mm was performed without the smallest crack. Vickers hardness after rolling was compared with that of rolled (0·5 mm) *un*oxidized polycrystalline MgAg and BeCu and not-rolled oxidized MgAg, after subsequent annealings in *nitrogen* at 300°, 400°, *etc.* to 800 °C. In table XII the results are also given of an MgAg and a BeCu specimen rolled (after oxidizing) from liquid-nitrogen temperature, but this silver specimen was polycrystalline (grain size about 1 mm^2).

The hardnesses of the oxidized single crystals are evidently lower than in polycrystalline form (see sections 1 and 11). The increase in hardness

*) These rods, like the other castings, were made in the laboratory foundry of Ir Meerkamp van Embden.

**) We found also that the brittleness of nitride-hardened iron + Al is less in single-crystal form.

***) Measured over 10 mm; the silver rods were machined to 2·5 mm dia. over the gauge length, while the copper one was tested without profile.

TABLE XII

Vickers hardness ($2^1/_2$ kg load)

Material \ Subsequent treatment	Oxid. and/or rolled	$+^1/_2$ h. 300 °C N_2	$+^1/_2$ h. 400 °C N_2	$+^1/_2$ h. 500 °C N_2	$+^1/_2$ h. 600 °C N_2	$+^1/_2$ h. 700 °C N_2	$+^1/_2$ h. 800 °C N_2
MgAg rolled *)	102	56	40	35			
oxidized **)	122	116	121	109	112	116	108
ox. + r. **)	146	138	143	133	123	125	122
ox. + r. −190° *)	176			194			168
BeCu rolled *)	168	163	68	59			
ox. + r. A **)	113	117	130	127	122	118	119
ox. + r. B **)	122	141	128	115	126	124	127
ox. + r. −190° **)	140			138			131

by rolling and the corresponding decrease by annealing at temperatures above the normal recrystallization temperatures of Ag and Cu are remarkably small, especially for BeCu. Metallographic examination after the last annealing (at 800 °C) of the rolled oxidized specimens showed that even in the MgAg no visible recrystallization had started; flow lines were still visible. Only in the BeCu crystal B had the unoxidized interior of course recrystallized completely, but *fig. 18* shows no signs that the sharp boundary is being crossed into the un-recrystallized exterior.

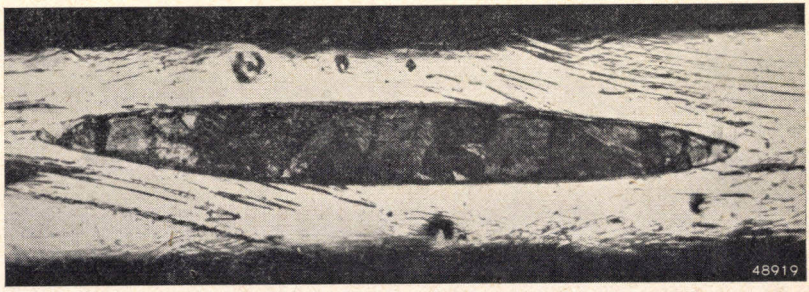


Fig. 18. Single crystal rod of Cu with 0·25 % Be, oxidized partially at 950 °C, cold rolled from 2·15 mm ⌀ to 0·5 mm thickness, and then annealed in N_2 subsequently at 300°, 400°, ... 800 °C, each time for $^1/_2$ hour. Etched $NH_4OH + H_2O_2$. Magn. 60×. Only the unoxidized interior recrystallized, flow lines in hardened exterior persisting.

*) polycrystalline **) single crystal

The lattice disturbances caused by cold work appear to play a minor rôle besides those around the oxide inclusions.

At the crystal boundaries the flow along the glide planes must adjust itself in some way to prevent rupture; the required flexibility can be provided by resolution of existing disturbances (with formation of new ones). This can take place easily in normal copper alloys and probably in age-hardened Cu with $2^1/_2$ % Be too, but in Cu + BeO or Ag + MgO it will be difficult, so that their intercrystalline brittleness is not so surprising.

In conclusion we wish to express our thanks to Mr P. Jongenburger for carrying out part of the experiments.

Eindhoven, July 1946

REFERENCES

1) F. M. G. Johnson and P. Larose, J. Am. Chem. Soc. **46**, 1377, 1924; **49**, 312, 1927.
2) E. W. R. Steacie and F. M. G. Johnson, Proc. Roy. Soc. **A 112**, 542, 1926.
3) C. S. Smith, Min. and Met. **13**, 481, 1932.
4) F. N. Rhines and C. H. Matthewson, Trans. A.I.M.M.E. **111**, 337, 1934.
5) C. E. Ransley, J. Inst. Met. (London) **65**, 147, 1939.
6) C. J. Smithells and C. E. Ransley, Proc. Roy. Soc. **A 155**, 195, 1936.
7) M. Straumanis and A. Ievins, Z. Phys. **98**, 461, 1936. M. U. Cohen, Z. Krist. **94**, 288, 1936.
8) E. A. Owen and E. L. Yates, Phil. Mag. (VII) **15**, 472, 1933.
9) E. R. Jette and F. Foote, J. Chem. Phys. **3**, 605, 1935.
10) M. S. Fisher and Z. Shaw, J. Iron-Steel Inst. **136**, 143, 1937.
11) H. A. Sloman, J. Inst. Met. (London) **54**, 161, 1934.
12) J. C. Chaston, J. Inst. Met. (London) **71**, 23, 1945.
13) G. Tammann and K. L. Dreyer, Ann. Phys. (Leipzig) **16**, 111, 1933.
14) J. O. Linde, Ann. Phys. (Leipzig) **10**, 52, 1931; **14**, 353, 1932; **15**, 219, 1932.
15) W. Seith and E. Peretti, Z. Elektrochem. **42**, 570, 1936.
16) F. N. Rhines and R. F. Mehl, Trans. A.I.M.M.E. **128**, 185, 1938.
17) F. N. Rhines, Met. Technol. **7**, T.P. 1162, 1940.
18) J. L. Meijering and M. J. Druyvesteyn, Philips Res. Rep. **2**, 81, 1947.
19) F. N. Rhines and A. H. Grobe, Met. Technol. 1942, T. P. 1439.
20) F. N. Rhines, W. A. Johnson and W. A. Anderson, Met. Technol. 1941, T. P. 1368.

R 51 *Philips Res. Rep.* **2**, 281-312, 1947

MINIMUM-COST CHOKES

by T. H. ODDIE and J. L. SALPETER

621.318.42-181

Summary

A design method is developed for chokes carrying A.C. only to enable the most economical dimensions to be found for any given electrical characteristics. Formulae and tables are given for rectangular types of chokes with and without limitations on the stacking height. It is also shown briefly how the method may be applied to chokes carrying D.C. with superimposed A.C.

LIST OF SYMBOLS

The numbers in parentheses denote the equations in which the symbols are defined or used for the first time.

a cm, length of the air gap (2.1.2)
b cm, height of the winding window, see figs 1 and 2 (2.2.3)
c cm, width of the winding window, see figs 1 and 2 (2.2.3)
d cm, width of the leg of the core inside the windings, see figs 1 and 2 (2.1.8)
e $= t/m$ (3.2.1)
f c/sec, frequency (2.1.4)
h cm, stacking height of the core, see figs 1 and 2 (2.1.8)
k_1, k_2 (6.4.1)
l cm, mean length of magnetic path in iron (2.1.1)
l_c cm, mean length of one turn of the windings (2.2.2)
m $= d/c$ (2.2.6)
n $= b/c$ (2.2.6)
p (5.2)
q (5.3)
r $= c/V^{1/3}$ (2.2.6)
s winding space factor (2.2.3)
t $= h/c$ (2.2.6)
u (2.3.9)
w (2.3.8)
x weight factor = ratio of the cost of the core material per unit volume to the cost of the windings per unit volume of conductor (6.1.2)
y $= V_c/2V_s$ (2.2.6)
z $y/(r^4mnt)^2$ (2.2.7)
A cm^2, area of cross-section of the core, including the spaces between the laminations (2.1.1)

A_c cm², total conducting area of cross-section of the windings (2.2.3)
B_m gauss, maximum flux density in the core material (2.1.3)
C (7.6)
D amp/cm², current density (2.4.1)
D_{max} upper limit for D (2.4.4)
E volt, root-mean-square reactive potential drop across the choke (2.1.4)
G (2.4.4)
H_m oersted, magnetizing force corresponding to B_m (2.3.5)
I amp, root-mean-square current (2.1.3)
K $= 1 + \mu aS/l$ (2.1.2)
L henry, inductance of the choke (2.1.1)
N number of turns (2.1.1)
R ohm, total resistance (2.2.8)
R_c ohm, resistance of the windings (2.2.2)
R_m ohm, resistance due to magnetic losses (2.2.1)
R_{max} maximum value of R that can be tolerated (6.1.3)
S stacking factor for the core (2.1.1)
V cm³, volume of the core, including the spaces between the laminations (2.1.8)
V_c cm³, volume of conductor (2.2.4)
V_{max} maximum value of V, corresponding to β_{max} (2.3.7)
V_{min} lower limit for V, corresponding to D_{max} (2.4.3)
α ratio of iron losses to total losses (6.2.5)
β $= (E/I)\mathrm{d}I/\mathrm{d}E$, linearity factor (2.3.1)
β_{max} maximum value of β that can be tolerated (2.3.7)
γ (6.4.1)
δ (6.2.6)
μ permeability of the core material (2.1.1)
ϱ watt/cm³, specific loss factor of the core material (2.2.1)
ϱ_c ohm. cm, resistivity of the winding material (2.2.2)
σ $= \varrho/B_m^2$, watt/cm³gauss² (5.1)

Unprimed symbols represent operating values.
Primed symbols represent starting values (2.3.10).

1. *Introduction*

The designer of a choke that is to be produced in reasonably large quantities is practically always faced with the task of obtaining the required electrical characteristics for as low a cost as possible. Even if he limits himself to one type of core, however, he is still faced with an infinite number of possible designs, of varying cost, but all giving the same electrical

results. And finally, since some workable design must usually be produced in a short time, a few trial calculations are made and the cheapest is chosen in the hope that it is somewhere near the most economical of all.

While the analogous problem for transformers has been treated by a number of writers [2] [4], there is little published work on the economical design of chokes, possibly because it is only recently that these have been needed in large quantities as ballasts for fluorescent lamps.

In the following sections the writers have endeavoured to treat the above problem in such a way that the cheapest design, for a rectangular choke, may be found with the minimum of searching, always assuming that the quantities to be produced are large enough to warrant the manufacture of special laminations of the required optimum shape. If this is not so, the methods developed may be extended fairly easily to cover a lamination of any particular shape.

Beginning with the assumption that the relative costs per unit volume for the windings of the choke compared with the core may be obtained (or are known), the problem of obtaining the minimum total cost becomes essentially one of varying the relative dimensions of the choke, and therefore the ratio of the volumes of windings and core, while keeping the electrical characteristics constant or at least below the limits required.

To do this the general equations relating the choke dimensions to its electrical properties are first brought, in section 2 below, to a form where the shape may be considered as a separate factor. The variations of this shape factor, and its optimum values, are considered in sections 3 and 4 for particular types of core. These optimum values are then used in sections 5 and 6 to obtain equations and tabulated data that yield the best solution for any particular case; examples of such solutions are included in the Appendix.

The treatment presented in this paper is meant for an alternating current of a sine-wave shape. In the case of a distorted waveform it will be necessary to find, in every particular case, the nearest possible equivalent sinusoidal current, *i.e.* a sinusoidal alternating current, giving the same copper and iron losses as the distorted one.

2. *Basic equations applicable to all rectangular cores*

2.1. *Current, voltage drop, and inductance*

The following fundamental relations [1]) apply to all chokes of rectangular shape.

The inductance, L henry, of the choke is given by

$$L = \frac{4\pi\ N^2\ \mu AS}{10^9\ lK} \qquad (2.1.1)$$

where

$$K = 1 + \frac{\mu a S}{l}, \tag{2.1.2}$$

and N = the number of turns,
μ gauss/oersted = the permeability of the core material,
A cm^2 = the area of cross-section of the core, including the spaces between the laminations,
l cm = the mean length of the magnetic path in iron,
a cm = the length of the air gap, and
S = the stacking factor for the core.

Also the root-mean-square current, I amp, is obtained from

$$0{\cdot}4\,\pi\sqrt{2}\,NI = \frac{B_m\, lK}{\mu}. \tag{2.1.3}$$

where B_m gauss = the maximum flux density in the core material itself, and since the root-mean-square reactive potential drop across the choke, E volt, is

$$E = 2\pi f L I, \tag{2.1.4}$$

equations (2.1.1) and (2.1.3) lead to

$$E = 4{\cdot}44 \times 10^{-8}\, N A\, B_m f S. \tag{2.1.5}$$

where f c/sec = the frequency considered.
Now (2.1.1) and (2.1.3) may also be solved for N and K, giving

$$N = \frac{\sqrt{2} \times 10^8\, L\, I\, l}{B_m\, V\, S}, \tag{2.1.6}$$

$$K = \frac{8\pi \times 10^7\, L\, I^2\, \mu}{B_m{}^2\, V\, S} \tag{2.1.7}$$

where

$$V = Al = dhl. \tag{2.1.8}$$

Here V cm^3 = the volume of the core, including the spaces between the laminations,
d cm = the width of the leg of the core inside the windings, see figs 1 and 2, and
h cm = the stacking height of the core, see figs 1 and 2.

2.2. *Losses*

From the definition of ϱ watt/cm^3, the specific loss factor of the core material, we have

$$I^2 R_m = V \varrho S \tag{2.2.1}$$

where R_m ohm = the resistance due to magnetic losses, while if the mean

length of one turn of the windings is l_c cm and the total conducting area of cross-section of the windings is A_c cm^2, the total length of the windings will be Nl_c and the cross-sectional area of each wire will be A_c/N, so that the resistance of the windings, R_c ohm, will be

$$R_c = \frac{\varrho_c \, l_c \, N^2}{A_c} = \frac{\varrho_c \, l_c \, N^2}{b \, c \, s} \tag{2.2.2}$$

where ϱ_c ohm.cm = the resistivity of the winding material, and, by definition,

$$A_c = b \, c \, s \tag{2.2.3}$$

where b cm = the height of the winding window, see figs 1 and 2,
c cm = the width of the winding window, see figs 1 and 2, and
s = the winding space factor.

Equation (2.1.6) may now be used to eliminate N from (2.2.2) giving

$$\frac{R_c}{L} = \frac{2 \times 10^{16} \, \varrho_c \, L \, I^2 \, l_c}{B_m^2 \, A^2 \, S^2 \, s \, b \, c}$$

and, since the total volume of conductor, V_c cm^3, will be

$$V_c = l_c \, A_c = l_c \, b \, s \, c, \tag{2.2.4}$$

$$\frac{R_c}{L} = \frac{2 \times 10^{16} \, \varrho_c \, L \, I^2 \, V_c}{B_m^2 \, S^2 \, (b \, c \, d \, h \, s)^2} \tag{2.2.5}$$

where equation (2.1.8) has been used to eliminate A.

In order to consider the merits of various shapes of core, it is better to replace the various dimensions by ratios. Let these ratios be represented as follows

$$\left.\begin{aligned} V_c &= 2 \, V \, s \, y, \\ c &= r \, V^{1/3}, \\ d &= m \, c, \\ b &= n \, c, \\ h &= t \, c. \end{aligned}\right\} \tag{2.2.6}$$

Then, putting in these ratios, combining equations (2.2.1) and (2.2.5), and substituting

$$z = \frac{y}{(r^4 m n t)^2} \tag{2.2.7}$$

we have for the total resistance, R ohm,

$$R = R_m + R_c = \frac{V \varrho s}{I^2} + \frac{4 \times 10^{16} \, \varrho_c \, L^2 \, I^2 \, z}{B_m^2 \, S^2 \, s \, V^{2/3}}. \tag{2.2.8}$$

The factor z involves ratios only and is therefore dimensionless and in a form suitable for consideration apart from the other parameters, and particularly apart from the volume of the core V. Values of z for various core shapes are dealt with in sections 3 and 4 below, where the lowest, and therefore the best, values are obtained.

2.3. *Linearity*

The linearity of a choke will be defined here by a linearity factor β, where

$$\beta = \frac{E}{I} \cdot \frac{\mathrm{d}I}{\mathrm{d}E}. \tag{2.3.1}$$

The quantity $\frac{\mathrm{d}I}{I}$ is obtained from equations (2.1.2) and (2.1.3) as

$$\frac{\mathrm{d}I}{I} = \left(1 - \frac{B_m}{\mu K} \cdot \frac{\mathrm{d}\mu}{\mathrm{d}B_m}\right) \frac{\mathrm{d}B_m}{B_m}. \tag{2.3.2}$$

But from (2.1.5)

$$\frac{\mathrm{d}B_m}{B_m} = \frac{\mathrm{d}E}{E}, \tag{2.3.3}$$

so that

$$\beta - 1 = -\frac{B_m}{\mu K} \cdot \frac{\mathrm{d}\mu}{\mathrm{d}B_m}. \tag{2.3.4}$$

If the magnetizing force corresponding to B_m is H_m oersted, since $B_m = \mu H_m$ we may substitute

$$\frac{\mathrm{d}\mu}{\mathrm{d}B_m} = \frac{\mu}{B_m}\left(1 - \mu \frac{\mathrm{d}H_m}{\mathrm{d}B_m}\right) \tag{2.3.5}$$

giving

$$\beta - 1 = \frac{1}{K}\left(\mu \frac{\mathrm{d}H_m}{\mathrm{d}B_m} - 1\right). \tag{2.3.6}$$

Equating the two values of K from (2.1.7) and (2.3.6) and putting the maximum value of V as $V_{\max}$ corresponding to the maximum value of β that can be tolerated, $\beta_{\max}$, gives

$$V_{\max} = \frac{8\pi \times 10^7 \, L \, I^2 (\beta_{\max} - 1)}{\left(B_m{}^2 \frac{\mathrm{d}H_m}{\mathrm{d}B_m} - B_m H_m\right) S} = \frac{w}{u} \text{ (say)} \tag{2.3.7}$$

where

$$w = \frac{8\pi \times 10^7 \, L \, I^2 \, (\beta_{\max} - 1)}{S}, \tag{2.3.8}$$

$$u = B_m{}^2 \frac{\mathrm{d}H_m}{\mathrm{d}B_m} - B_m \, H_m. \tag{2.3.9}$$

It may be seen that u is a function only of the properties of the particular magnetic material used (and of B_m).

In the case of ballast chokes for fluorescent lamps, limits for the value of β selected for design purposes may be found as follows.

For satisfactory starting of the lamp the starting potential drop E' and current I' may be specified (within limits) as well as the operating values. Taking unprimed symbols to represent operating values, and primed symbols for starting values, we have from (2.1.5) and (2.1.3)

$$\frac{E'}{E} = \frac{B_m'}{B_m},$$

$$\frac{I'}{I} = \frac{B_m' \; K' \; \mu}{B_m \; K \; \mu'},$$

so that

$$\frac{I'E}{IE'} = \frac{K' \, \mu}{K \, \mu'}. \tag{2.3.10}$$

But from (2.1.2) it is easily found that

$$\frac{K' \, \mu}{K \, \mu'} = 1 + \frac{\mu - \mu'}{K \mu'}. \tag{2.3.11}$$

Then from (2.3.10) and (2.3.11) applying (2.1.7)

$$K = \frac{(\mu - \mu') \; E'/E}{\mu' \; (I'/I - E'/E)} = \frac{8\pi \times 10^7 \; L \; I^2 \; \mu}{B_m{}^2 \; V \; S}, \tag{2.3.12}$$

and putting in the value of V from (2.3.7)

$$\frac{u}{B_m{}^2 \; (\beta - 1)} = \left(\frac{1}{\mu'} - \frac{1}{\mu}\right) \left(\frac{E'/E}{I'/I - E'/E}\right). \tag{2.3.13}$$

This relation enables β to be found when the starting and operating conditions are specified and when the form of the magnetization curve is known for the core material to be used. For a particular form of curve a more practical expression is derived in section 6.4 below.

2.4. *Current density*

When the current density in the choke windings is considered, a lower limit is placed on V in a manner resembling the effect of linearity dealt with above.

If D amp/cm^2 is the current density, it will be given by

$$D = \frac{IN}{bcs}. \tag{2.4.1}$$

Putting in N from (2.1.6), and using (2.2.6) and (2.2.7) gives

$$D = \frac{\sqrt{2} \times 10^8 \, L \, I^2 \, z^{1/2}}{B_m \, S \, s \, V^{4/3} \, y^{1/2}}. \tag{2.4.2}$$

Then if an upper limit D_{max} is set for the current density, V will have a lower limit V_{min} given by

$$V_{min} = \left(\frac{G^2 \, z}{B_m{}^2 \, y} \right)^{3/8}, \tag{2.4.3}$$

where

$$G = \frac{\sqrt{2} \times 10^8 \, L \, I^2}{S \, s \, D_{max}}. \tag{2.4.4}$$

These relations will not normally be needed, but there may occur special chokes where D_{max} must be taken into account.

3. *Optimum shapes for shell-type choke*

3.1. *Without special restrictions*

Taking first a shell-type choke as in *fig. 1*, it may be seen that the volumes of core material and conductor are given by

$$V = 2 \, h \, d \, (b + c + d), \tag{3.1.1}$$

$$V_c = 2 \, b \, c \, s \, (2c + d + h). \tag{3.1.2}$$

Substitution from equations (2.2.6) then gives

$$1 = 2 \, r^3 \, m \, t \, (1 + m + n), \tag{3.1.3}$$

$$y = r^3 \, n \, (2 + m + t). \tag{3.1.4}$$

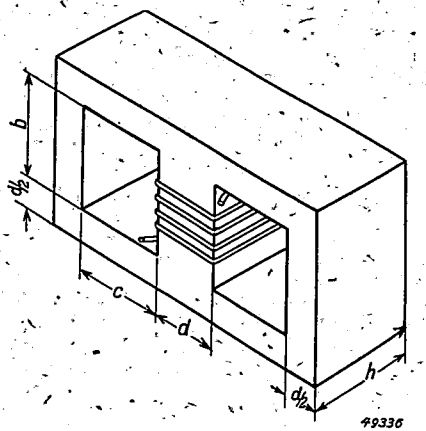


Fig. 1. Shell-type choke.

Considering now equation (2.2.8), the shape factor z may be made a minimum for any given value of y by a suitable choice of the ratios r, m, n, and t.

Even without any other restrictions on the shape, however, equations (3.1.3) and (3.1.4) must first be applied to eliminate two of the ratios say r and t. Thus

$$r^3 = \frac{2\,m\,y\,(1+m+n)-n}{2\,m\,n\,(1+m+n)\,(2+m)}, \tag{3.1.5}$$

$$t = \frac{n\,(2+m)}{2\,m\,y\,(1+m+n)-n}. \tag{3.1.6}$$

Further, if we consider y to be fixed, for z to be a minimum (r^4mnt) must be a maximum. With (3.1.5) and (3.1.6) we obtain

$$16\,(r^4\,m\,n\,t)^3 = \frac{2\,m\,n^2\,y\,(1+m+n)-n^3}{m\,(2+m)\,(1+m+n)^4}. \tag{3.1.7}$$

The ratios m and n must now be chosen to make this a maximum; this will occur when the partial derivatives of the right-hand side of (3.1.7) with respect to m and n are each equal to zero.

Carrying out these operations and solving each equation for y gives

$$y = \frac{n\,(1+6\,m+3\,m^2+n+m\,n)}{m^2\,(1+m+n)\,(7+4m+n)}, \tag{3.1.8}$$

$$y = \frac{n\,(3+3\,m-n)}{2\,m\,(1+m+n)\,(2+2\,m-n)}, \tag{3.1.9}$$

and solving these for n in terms of m gives

$$n\,(2+m) = 1-0{\cdot}5\,m^2 \pm (9+18\,m+12\,m^2+3\,m^3+0{\cdot}25\,m^4)^{1/2}. \tag{3.1.10}$$

Here only the positive root for n is of physical significance.

Now for selected values of m, n may be calculated from (3.1.10), then y from (3.1.8) or (3.1.9) and r and t from (3.1.5) and (3.1.6) respectively, and the optimum shape is determined for various values of y. Some results of such calculations are given in table I and show that the best possible z is about 56·2 at a value of y of about 1·71.

It will be shown later that the smaller values of y are also of practical interest in reducing the cost of any choke, but the higher values are of limited interest only, since the same result may nearly always be obtained with the use of less volume of conductor and the same volume of core. The only case where higher values of y may be needed is when the current density in the windings would exceed the upper limit discussed in section 2.4 above, unless a higher value of y were taken.

The data in table I are given in a more practical form in table III, and the latter should be used in calculations of particular chokes.

3.2. *With fixed ratio of stacking height to core width*

The ratio of stacking height to the core width is t/m, and table I shows that the values of this ratio for the shell-type choke without restriction are inconveniently large from the practical point of view, as winding is difficult on a very asymmetrical coilformer.

So consideration has been given to the case where t/m is restricted, say

$$t = e\,m\,. \tag{3.2.1}$$

Equations (3.1.3) and (3.1.4) still apply, and enable r, n, and t to be eliminated. Then differentiation with respect to the remaining variable m gives, by a procedure similar to that of section 3.1 above,

$$y = \frac{16 + 22\,m + 10\,e\,m + 7\,m^2 + 8\,e\,m^2 + e^2\,m^2}{2\,e\,m^2\,(12 + 12\,m + 2\,e\,m + m^2 + e\,m^2)}, \tag{3.2.2}$$

and also

$$n = \frac{2\,e\,m^2\,y\,(1 + m)}{2 + (1 + e)\,m - 2\,e\,m^2\,y}, \tag{3.2.3}$$

$$r^3 = \frac{2 + (1 + e)\,m - 2\,e\,m^2\,y}{2\,e\,m^2\,(1 + m)\{2 + (1 + e)\,m\}}. \tag{3.2.4}$$

From these equations the values in table IV have been calculated for the cases of $e = 1$, $e = 1{\cdot}5$, and $e = 2$. Here and in later tables only values of y lower than that giving the minimum z have been included, as the general features of the variations are similar to the case discussed in 3.1 above.

3.3. *With "wasteless" laminations*

If the laminations are of the so-called "wasteless" design of E and I shape, where the I is obtained from the punched out portions, we must have $m = 2$ and $n = 3$.

Then the other ratios follow from equations (3.1.3), (3.1.4), and (2.2.7), thus

$$t = \frac{4}{8y - 1}, \tag{3.3.1}$$

$$r^3 = \frac{8y - 1}{96}. \tag{3.3.2}$$

For those interested in this type of lamination, table II gives typical calculated values, together with the corresponding figures for z.

4. *Optimum shapes for core-type choke*

For the core-type choke with two windings on opposite sides as in *fig. 2*, the volumes of core material and conductor are now

$$V = 2\,h\,d\,(b + c + 2d), \tag{4.1}$$

$$V_c = 2\,b\,c\,s\,(c + d + h), \tag{4.2}$$

and putting in the ratios from (2.2.8),

$$1 = m\,t\,r^3\,(1 + 2\,m + n), \tag{4.3}$$

$$y = n\,r^3\,(1 + m + t). \tag{4.4}$$

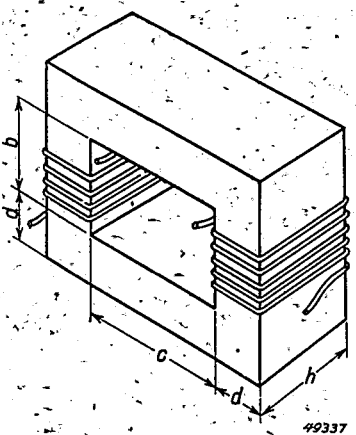


Fig. 2. Core-type choke.

By applying the transformation

$$\left.\begin{aligned} 2\,m &= m' \\ n &= n' \\ r &= 4^{1/3}\,r' \\ 2\,t &= t' \\ y &= 2y' \\ 2^{1/3}\,z &= z' \end{aligned}\right\} \tag{4.5}$$

where the unprimed letters now refer to the core-type choke and the primed ones to the shell-type, equations (4.3) and (4.4) reduce to the form of (3.1.3) and (3.1.4), and the results of the calculations for the shell-type choke may be converted directly for the core-type.

The new values of the various ratios are in tables V and VI for the cases without special restrictions and with restricted stacking heights respectively; these tables have been prepared by converting the figures of tables III and IV respectively.

5. *Minimum losses*

In the preceding sections it has been shown that the total resistance of the choke may be represented by equation (2.2.8), and that the factor z may be made a minimum by a suitable choice of the relative dimensions only.

To obtain the conditions for minimum losses when B_m or V is fixed, the loss factor ϱ must be expressed as a function of B_m. For common magnetic core materials we may write approximately

$$\varrho = \sigma B_m^2, \tag{5.1}$$

where σ watt/cm^{3}gauss^2 is a coefficient depending on the material used, but not upon B_m.

Also substituting, for brevity,

$$p = \frac{\sigma S}{I^2}, \tag{5.2}$$

$$q = \frac{4 \times 10^{16}\, \varrho_c'\, L^2\, I^2}{S^2\, s}, \tag{5.3}$$

(2.2.8) becomes

$$R = pV B_m^2 + \frac{q\, z}{V^{2/3} B_m^2}. \tag{5.4}$$

For fixed V and z then R is a minimum, R_0, when

$$\frac{\mathrm{d}R}{\mathrm{d}B_m} = 0, \qquad \text{that is}$$

$$B_m^2 = \left(\frac{q\, z}{p V^{2/3}}\right)^{1/2}, \tag{5.5}$$

making the iron and copper losses equal, and

$$R_0 = \left(\frac{p\, q\, z}{V^{2/3}}\right)^{1/2} = \tag{5.6}$$

$$= 2 \times 10^8\, L \left(\frac{\sigma\, \varrho_c'\, z}{S\, s\, V^{2/3}}\right)^{1/2}. \tag{5.7}$$

On the other hand, if B_m and z are fixed, for minimum R, say R_0',

$$\frac{\mathrm{d}R}{\mathrm{d}V} = 0, \quad \text{and}$$

$$V = \left(\frac{5\, qz}{3\, p\, B_m^4}\right)^{3/8}, \tag{5.8}$$

making the iron and copper losses in a ratio of 5 : 3 and the total

$$R_0' = 1{\cdot}936\,(p^5\,q^3\,z^3\,B_m^4)^{1/8} = \tag{5.9}$$

$$= 3{\cdot}26 \times 10^6 \left(\frac{\sigma^5\,\varrho_c^3\,L^6\,z^3\,B_m^4}{I^4\,s^3\,S}\right)^{1/8}. \tag{5.10}$$

6. *Minimum volume and other considerations*

6.1. *Case where maximum resistance is determining factor*

In most practical cases we are not concerned with obtaining minimum losses, but rather with obtaining the minimum cost or total volume while keeping the losses and the linearity factor β below some given limits.

Now the minimum cost will be obtained, at least approximately, when $V_c + SxV$ is a minimum for the given design requirements. Here a weight factor x has been introduced representing the ratio of the cost of the core material per unit volume to the cost of the windings per unit volume of conductor.

Considering first only the maximum tolerable resistance R_{max}, and examining equation (5.4), it may be seen that z must be expressed as a function of y, that is a function of V_c/V, to obtain an expression that may be used to find the conditions for minimum cost.

It is difficult to find a simple analytical expression relating y to the corresponding best value of z, at least for the various shapes discussed in sections 3 and 4 above. However, if the values are plotted it may be seen that, in all cases in which we are interested, namely when y is not greater than the value giving the lowest of all z-values for the particular general core type

$$\frac{dy}{dz} < 0. \tag{6.1.1}$$

Now from (2.2.6)

$$V_c + S\,x\,V = 2\,s\,V\,y + S\,x\,V, \tag{6.1.2}$$

and it is evident from (5.4) that we should take $R = R_{max}$ in order to be able to decrease the volume V as far as possible, or alternatively to decrease V_c and therefore decrease y and increase z as far as possible.

In this case solving (5.4) for z gives

$$z = \frac{(R_{max} - p\,V\,B_m^2)\,B_m^2\,V^{5/3}}{q}. \tag{6.1.3}$$

Then for minimum cost

$$\frac{d\,(V_c + S\,x\,V)}{d\,B_m} = 0, \tag{6.1.4}$$

$$\frac{d\,(V_c + S\,x\,V)}{d\,V} = 0. \tag{6.1.5}$$

The first condition (6.1.4) leads, with (6.1.2), to

$$\frac{dy}{dB_m} = \frac{dy}{dz} \cdot \frac{dz}{dB_m} = 0 . \tag{6.1.6}$$

But owing to the nature of the y-z-curves dy/dz is zero only at very large values of z; so considering only

$$\frac{dz}{dB_m} = 0 , \tag{6.1.7}$$

we then find

$$\frac{d^2y}{dB_m^2} = \frac{d^2y}{dz^2}\left(\frac{dz}{dB_m}\right)^2 + \frac{dy}{dz} \cdot \frac{d^2z}{dB_m^2} , \tag{6.1.8}$$

and this is positive, showing that a minimum and not a maximum has been obtained for y, since the first term is zero, from (6.1.7), dy/dz is negative, from (6.1.1), and d^2z/dB_m^2 is proportional to $R_{\max} - 6p\,VB_m^2$ which will also be found to be negative.

So we now have from (6.1.7) and (6.1.3)

$$B_m^2 = \frac{R_{\max}}{2\,p\,V} . \tag{6.1.9}$$

Here again we have the iron and copper losses equal.

The second condition (6.1.5) and equation (6.1.2) give

$$2\,sy + Sx + 2\,sV \frac{dy}{dz} \cdot \frac{dz}{dV} = 0 . \tag{6.1.10}$$

From (6.1.3) and (6.1.9)

$$\frac{dz}{dV} = \frac{B_m^2}{3\,q}\left(5\,R_{\max}\,V^{2/3} - 8\,p\,B_m^2\,V^{5/3}\right) = \frac{R_{\max}^2}{6\,p\,q\,V^{1/3}} . \tag{6.1.11}$$

Putting this value in (6.1.10) we have

$$y + \frac{R_{\max}^2\,V^{2/3}}{6\,qp} \cdot \frac{dy}{dz} = -\frac{Sx}{2\,s} . \tag{6.1.12}$$

Here again using (6.1.9) and (6.1.3) to eliminate V gives finally

$$y + \frac{2z}{3} \cdot \frac{dy}{dz} = -\frac{Sx}{2\,s} . \tag{6.1.13}$$

Since from (6.1.1) dy/dz is negative, this relation has a useful solution, which will be further discussed in section 6.3 below.

6.2. *Case where linearity is determining factor*

The value of V determined from equations (6.1.13), (6.1.3), and (6.1.9)

must not be greater than V_{max} found from equation (2.3.7), or β will be larger than the tolerable value.

If the V found is too great we must proceed in a different manner, as follows.

The value of V used must now be brought down to V_{max} and V is no longer independent of B_m. Putting in V_{max} from (2.3.7) equation (6.1.3) now becomes

$$z = \frac{\left(R_{max} - \frac{p\,w}{u} B_m^2\right) B_m^2 \left(\frac{w}{u}\right)^{2/3}}{q}, \tag{6.2.1}$$

and (6.1.2) becomes

$$V_c + S\,x\,V = (2\,s\,y + S\,x)\frac{w}{u}. \tag{6.2.2}$$

We may now select only B_m to make $V_c + S\,x\,V$ a minimum.

Therefore, using condition (6.1.4) we obtain

$$y + \frac{S\,x}{2\,s} = u\frac{\mathrm{d}\,y}{\mathrm{d}\,z}\cdot\frac{\mathrm{d}\,z}{\mathrm{d}\,u} = z\frac{\mathrm{d}\,y}{\mathrm{d}\,z}\left(\frac{u}{z}\cdot\frac{\mathrm{d}\,z}{\mathrm{d}\,u}\right). \tag{6.2.3}$$

The factor $(u/z)\cdot(\mathrm{d}z/\mathrm{d}u)$ is found from (6.2.1), and by substituting

$$\frac{B_m}{u}\cdot\frac{\mathrm{d}\,u}{\mathrm{d}\,B_m} = 1 + \frac{B_m^3}{u}\cdot\frac{\mathrm{d}^2 H_m}{\mathrm{d}\,B_m^2}, \tag{6.2.4}$$

and writing

$$\frac{p\,w\,B_m^2}{R_{max}\,u} = \frac{8\,\pi\times 10^7\,\sigma\,L\,(\beta_{max}-1)\,B_m^2}{R_{max}\,u} = \alpha, \tag{6.2.5}$$

where α is the ratio of the iron losses to the total losses, we finally obtain

$$\frac{u}{z}\cdot\frac{\mathrm{d}\,z}{\mathrm{d}\,u} = \frac{1-4\,\alpha-(5-8\alpha)\frac{B_m^3}{u}\cdot\frac{\mathrm{d}^2 H_m}{\mathrm{d}\,B_m^2}}{3\,(1-\alpha)\left(1+\frac{B_m^3}{u}\frac{\mathrm{d}^2 H_m}{\mathrm{d}\,B_m^2}\right)} = \frac{\delta}{3}\,(\text{say}), \tag{6.2.6}$$

and from (6.2.3)

$$y + \frac{S\,x}{2\,s} = \frac{\delta}{3}\,z\frac{\mathrm{d}\,y}{\mathrm{d}\,z}. \tag{6.2.7}$$

For $\alpha = 0{\cdot}5$ this reduces to a value $\delta = -2$, and (6.2.7) is then identical with (6.1.13).

6.3. *Case where current density is determining factor*

In some cases it may be found that V, calculated as in section 6.1 above,

is less than V_{min}, the lower limit set by equation (2.4.3). Then to keep the current density not higher than D_{max} we must bring V up to the value V_{min}, and here again V and B_m are no longer independent of each other.

By a procedure analogous to that for the case where linearity is the controlling factor it is now found that the condition for minimum cost is

$$y + \left\{\frac{2y}{3} + \frac{Sx}{2s}\left(\frac{1-2\alpha'}{\alpha'}\right)\right\}\frac{z}{y}\cdot\frac{dy}{dz} = -\frac{Sx}{2s}, \tag{6.3.1}$$

where α' is the ratio of iron losses to total losses, and is given by

$$\alpha' = \frac{p\, B_m^{5/4}\, G^{3/4}\, z^{3/8}}{R_{max}\, y^{3/8}}. \tag{6.3.2}$$

Examination of (6.3.1) shows that for $\alpha' = 0{\cdot}5$ the relation is again identical with (6.1.13), but that for $0{\cdot}5 < \alpha' < 1$ there may be occasions where a positive value of dy/dz is needed to satisfy the equation, that is the region of high y-values would be used.

6.4. *Solution of minimum-cost equations*

Equations (6.1.13) and (6.2.7) give the conditions for minimum cost, but to use them, an analytical expression is needed for the B-H curve of the magnetic core material. A convenient form of expression is that suggested by Rader and Litscher [3])

$$H = k_1 B + k_2 B^\gamma. \tag{6.4.1}$$

The other functions involved in the preceding equations then become

$$u = k_2 (\gamma - 1) B_m^{\gamma+1}, \tag{6.4.2}$$

$$\frac{B_m^2}{u} = \frac{1}{k_2 (\gamma - 1) B_m^{\gamma-1}}, \tag{6.4.3}$$

$$\frac{B_m^3}{u}\cdot\frac{d^2 H_m}{dB_m^2} = \gamma, \tag{6.4.4}$$

$$\frac{1}{\mu} = k_1 + k_2 B_m^{\gamma-1}, \tag{6.4.5}$$

$$\beta - 1 = \frac{(\gamma - 1)\left(\frac{EI'}{E'I} - 1\right)}{\left(\frac{E'}{E}\right)^{\gamma-1} - 1}, \tag{6.4.6}$$

$$\delta = \frac{1 - 5\gamma - \alpha(4 - 8\gamma)}{(1-\alpha)(1+\gamma)}. \tag{6.4.7}$$

Now δ is -2 for equation (6.1.13) and only extends as far as -5 when α is very small and γ is large; so by selecting δ values in this range and using the y-z-curves calculated as in preceding sections the solution of (6.2.7) is obtained for each particular core type. A similar procedure could be adopted if the solution of (6.3.1) is required in practice.

The solution of (6.1.13) and (6.2.7) is given in tabular form in tables III to VI inclusive, and it enables the correct value of z to be chosen for various given values of x and α when s and S are known.

Since it is z that must later be used to find the various shape ratios, tables III to VI have been prepared using incremental steps of z, and the corresponding ratios are included.

In converting the figures for the shell type to the core type the transformation for x, corresponding to equations (4.5), was found by examining (6.2.7) to be

$$x = 2x' . \tag{6.4.8}$$

6.5. *Design procedure*

A brief outline will now be given of the method of using the preceding data to design a choke with given properties, while typical numerical examples are given in the Appendix.

1) Design requirements assumed to be given are values of

$$E, I, L, R_{max}, \beta_{max}, f.$$

2) Other data to be ascertained are values of

$$x, s, S, k_1, k_2, \gamma, \sigma, \varrho_c .$$

These depend upon the particular materials to be used.

3) A general choice of shape must be made between the shell- and core-types, with or without some restriction on the stacking height.

4) From the appropriate table (tables III to VI) find the value of z, say z_1, corresponding to the known $Sx/2s$ for $\delta = -2$.

5) Calculate p and q from equations (5.2) and (5.3) and then V_1 from

$$V_1 = \frac{8}{R_{max}^3} (pqz_1)^{3/4} \tag{6.5.1}$$

(this is derived from (6.1.3) and (6.1.9)).

6) Find w from (2.3.8), B_m from (6.1.9), u from (6.4.2), and then V_{max} from (2.3.7).

7) If V_{max} found is not less than V_1, proceed at once to step 12) below, otherwise the intermediate steps are necessary, using a method of successive approximation.

8) If V_1 was greater than V_{max}, take a value of α lower than 0·5, say

$\alpha_2 = 0{\cdot}4$, find the corresponding δ_2 from (6.4.7) and the corresponding z_2 from the table as before.

9) Now calculate α_3 from

$$\alpha_3 = \frac{p\,w}{R_{\max}\,k_2\,(\gamma-1)}\left[\frac{k_2\,(\gamma-1)}{R_{\max}{}^3\,w}\left\{\frac{p\,q\,z_2}{\alpha_2\,(1-\alpha_2)}\right\}^{3/2}\right]^{\frac{\gamma-1}{\gamma+1}} \tag{6.5.2}$$

(derived from (6.2.1), (6.4.2), (6.2.5) and (6.4.3)).

10) Use $\alpha_4 = \frac{1}{2}\,(\alpha_2 + \alpha_3)$ to find the corresponding δ_4 and z_4 as in step 8), and repeat steps 9) and 8) until the successive values of the α's found, approach each other reasonably closely; usually it is unnecessary to go beyond α_7, that is, steps 8) and 9) are performed three times.

11) Using the suffix n to denote the final values found, find V from

$$V = V_{\max} = \left[\frac{p\,q\,z_n}{R_{\max}{}^2\,\alpha_n\,(1-\alpha_n)}\right]^{3/2} \tag{6.5.3}$$

(derived from (2.3.7) and (6.4.2)),

and then B_m from

$$B_m = \left\{\frac{w}{V\,k_2\,(\gamma-1)}\right\}^{\frac{1}{\gamma+1}} \tag{6.5.4}$$

(derived from (6.2.1), (6.4.2), and (6.5.3)).

12) The remaining unknowns are finally obtained first by using the appropriate table (tables III to VI) to find the size ratios corresponding to z_n, and then finding the choke dimensions from (2.2.6) and N, K, and a from (2.1.5), (2.1.7), and (2.1.2) respectively.

6.6. *Choke families*

Having designed a particular choke, it is possible to convert it to other values of the design parameters by applying a dimensional procedure [1]). For the equations developed in this paper the following conversion rules have been found.

For all chokes of the one family, that is with the same shape and values of x, y, z, and so on, and with the same constants for the core material and conductor, the parameters may be varied so that

$$\frac{(EI)^{3\gamma-1}\,(\beta-1)^2}{(I^2\,R_{\max})^{4\gamma-2}} = \text{constant}, \tag{6.6.1}$$

and the conversions for V and B_m are

$$V \propto \frac{(EI)^{\frac{3\gamma-3}{4\gamma-2}}}{(\beta-1)^{\frac{3}{2\gamma-1}}}, \tag{6.6.2}$$

$$B_m \propto (\beta - 1)^{\frac{2}{2\gamma-1}} (EI)^{\frac{1}{4\gamma-2}}. \qquad (6.6.3)$$

The above conditions hold for all values of α, but with $\alpha < 0{\cdot}5$ the β used must of course be β_{max}.

Further, we see that in the normal case where γ is large, if β is considered fixed, a family of similar chokes results when $I^2 R_{max}$, the watts loss, is varied as $(EI)^{1/4}$. Also in this case V must be varied as $(EI)^{1/4}$ and B_m is practically constant.

7. *Application to chokes carrying direct current*

Throughout this paper the mathematical treatment has been limited to chokes not carrying any direct-current component. However, the results obtained may be applied also to filter chokes used in connection with rectifiers to smooth out alternating-current ripples, provided that the meanings of some of the symbols employed are suitably modified. These modifications may be derived as follows.

When direct current is present the inductance is determined by the incremental [5]) permeability μ_i where

$$\mu_i = \frac{\Delta B}{\Delta H}, \qquad (7.1)$$

and equations (2.1.1) and (2.1.2) must be replaced by

$$L = \frac{4\pi N^2 \mu_i A S}{10^9 \, l \, K_i} \qquad (7.2)$$

and

$$K_i = 1 + \frac{\mu_i a S}{l}. \qquad (7.3)$$

If I_0 amp is the direct current and μ is still the ordinary permeability, equation (2.1.3) should be replaced by

$$0{\cdot}4\pi N I_0 = K H_0 l, \qquad (7.4)$$

where K is given by (2.1.2) and H_0 is the direct part of the magnetic field strength.

Then from (7.2) and (7.4)

$$N = \frac{10^8 L I_0 l K_i}{V S \mu_i K H_0}. \qquad (7.5)$$

Comparing (7.5) with (2.1.6) it is evident that B_m is now replaced by a factor C given by

$$C = \frac{\sqrt{2} K \mu_i H_0}{K_i}. \qquad (7.6)$$

Rewriting (7.5) in the form

$$C\left(\frac{N I_0}{l}\right) = \frac{\sqrt{2} \times 10^8}{S}\left(\frac{L I_0^2}{V}\right), \qquad (7.7)$$

it is seen that the terms in brackets are those used by Hanna [8]) in developing his method of choosing the best air-gap for any particular choke. He and several other writers [6]) [7]) have shown how this may be done when the shape and one other variable have been chosen, provided that the incremental permeability and other properties of the core material are known.

If, then, we choose the shape and V, l is fixed and Hanna's method enables the maximum, and therefore the best, C to be chosen, so N is also fixed for given values of L and I_0.

In the common case, where the alternating current in the choke is small compared with the direct current, we may neglect the magnetic loss resistance R_m in comparison with the copper resistance R_c, which is now

$$R_c = \frac{4 \times 10^{16}\, \varrho_c L^2 I_0^2 z}{C^2 S^2 s V^{5/3}}, \qquad (7.8)$$

and this replaces (2.2.8).

For $R_c = R_{\max}$ the condition for minimum cost will then be

$$y + \frac{S x}{2 s} = -\frac{5 z}{3} \cdot \frac{dy}{dz}, \qquad (7.9)$$

which is (6.2.7) with $\delta = -5$ corresponding to $\alpha = 0$, since iron losses are neglected.

The procedure for the solution of (7.9) will be similar to that already given in section 6.5 as far as step 5) but using $\delta = -5$. Then since V is found, together with z and the corresponding best shape, C is obtained by Hanna's method.

8. *Discussion and summary.*

In the foregoing sections it has been shown that the particular value required for the shape factor z automatically determines all the relative dimensions of the choke, and these proportions are valid for all chokes of a given general type when z is chosen.

For purposes of illustration the variations of some of the parameters are shown in *fig. 3* for the choke calculated in the examples *A* and *B* of the Appendix, but with various values of $R_{\max}$ and the corresponding V. It may be seen that as $R_{\max}$ is increased V decreases progressively until the region of equality of iron and copper losses ($\alpha = 0{\cdot}5$) ends where β has risen very sharply to the permitted value $\beta_{\max}$. As V decreases further, the α chosen must then be decreased and the copper losses take an increasingly

large share of the total losses permitted, while V_c decreases more rapidly than V.

The task with which the designer of a choke for a fluorescent lamp is confronted can briefly be formulated as follows.

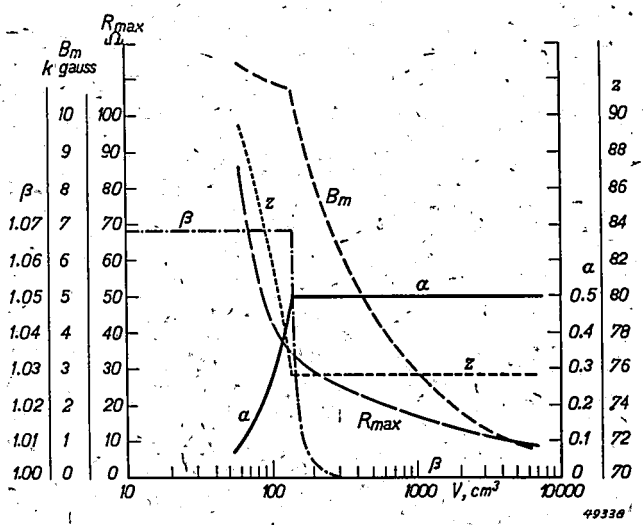


Fig. 3. Variations of parameters for a typical choke.

He must find the cheapest choke given the operating current I, the terminal choke voltage E, and the linearity β desired, that is the percentage increase of the current for 1% increase of the terminal voltage; he is given, furthermore, the price ratio x of iron and copper, the space factors s and S for copper and iron, and either the maximum permissable choke losses (maximum equivalent total resistance R_m) or he must suggest a reasonable value of R after taking into consideration cost and size.

The choke design is uniquely determined by its

a) shape,

b) total iron volume V, and

c) flux density B_m.

It turns out that the optimum shape is largely independent of the chosen R. The shape of the choke is best defined by three ratios:

1) window height/window width $= n$,

2) stacking height/centre limb width $= t/m$,

3) total volume occupied by copper/total volume occupied by iron $= 2y$.

The margins for the first two optimum ratios are relatively very narrow indeed. They are:

$$2{\cdot}2 < n < 3,$$
$$2 < t/m < 2{\cdot}67.$$

It is mainly the third ratio y that is responsible for the minimum cost. The largest possible value for the optimum y is $\sqrt{3}$ (if current density is not limited), corresponding with very large values of x, and y decreases towards zero with decreasing x.

Advisedly the designer will plot a curve R *versus* V (similar to fig. 3) for his type of choke in order to make a reasonable suggestion as to the choice of losses and size, cost, or weight. Starting with a very large V (small R), it is to be seen that the ratio α of the iron losses to the total losses is 0·5 and as long as $\alpha = 0{\cdot}5$, the ratio y is independent of R and is determined solely by x. Roughly (at least in the greater part of the region)

$$y \approx x^{1/3},$$

and so an error of 1 % in estimating x causes only $\frac{1}{3}$ % error in determining y. In the region $\alpha = 0{\cdot}5$ the linearity β is better than desired, a decrease of V causes an increase of both R and β. One per cent increase of R saves 3% of materials. Once the point has been reached when α becomes smaller than 0·5, 1% increase of R saves much less than 3% of materials, but at the same time β remains constant. In the region $\alpha < 0{\cdot}5$ the ratio y depends slightly on R as well (decreases with increasing R).

The general choice between the core- and shell-types may be made from the following considerations. If we take $V_c + SxV$ to represent the total cost of the choke, then, at least for the region of equal losses ($\alpha = 0{\cdot}5$), it will be found that $Sx/2s = 0{\cdot}5$ is the point where the shell- and its equivalent core-type of choke have equal costs, while with $Sx/2s < 0{\cdot}5$ the shell-type is cheaper, and with $Sx/2s > 0{\cdot}5$ the core-type is better. The difference in cost amounts to about 23% at $Sx/2s = 6$ but in practice this gain may often be more than lost again due to the need for two separate windings on the core-type, which of course necessitates the taking of different x-values for the two cases. In general then, the shell-type will be preferable where the cooling surface is not a major factor.

Comparing now the various stacking heights, it may be seen from a few trial calculations that the relative additional costs incurred by the use of restricted heights compared with the unrestricted case, are approximately: 16% for $h = d$, 5% for $h = 1{\cdot}5d$, and 1% for $h = 2d$. Even if winding difficulties can be overcome, there is little to be gained then in going nearer to the ideal case than $h = 2d$.

It must be pointed out that certain difficulties may occur in applying the design method outlined. Thus care must be taken in those parts of calculations involving such terms as B_m^{γ} since γ is usually large, say 10 or 12, as otherwise serious errors may occur in the final result. The final design should, of course, be checked against the actual B-H curve for the core material, particularly as equation (6.4.1) will not provide a good

fit for this curve at fairly low values of B. Further, although S may usually be taken as constant, s and x will vary to some extent with changes of shape of the winding window. If required, this difficulty may be overcome, if details of such changes can be obtained from practical data on costs and winding methods, by making the necessary choke calculations first with an approximate value of $Sx/2s$ and then with a value more accurately estimated for the particular shape arrived at from the first calculation.

Not included in the treatment is the choice of the proper electric steel (silicon content). Since the silicon content of commercial steels varies in steps (and not continuously), the designer will in most cases have the choice between not more than two or at the utmost three steels and it will then be advisable to calculate the design according to the treatment outlined for all steels in question and then make a comparison.

Finally, the writers's treatment does not take account of any limitations that may be imposed on the design, either (a) by considerations of heat dissipation and possible increase of the resistance of the windings with rise of temperature, or (b) by the possible necessity to fit the choke into a box of awkward shape.

Sydney, Philips Electrical Industries of Australia Pty Limited,
December 1946

REFERENCES

1) Massachusetts Institute of Technology, Department of Electrical Engineering, "Magnetic Circuits and Transformers", (1943), N.Y., John Wiley and Sons, Inc.
2) N. Partridge, Wireless World **48**, 286, 1942.
3) L. T. Rader and E. C. Litscher, Trans. Amer. Instn Elect. Engrs **63**, 133, 1944.
4) E. G. Reed, "The Essentials of Transformer Practice", (1927), N.Y., D. Van Nostrand Company, Inc.
5) T. Spooner, Trans. Amer. Instn Elect. Engrs **42**, 340, 1923.
6) J. Minton and I. G. Maloff, Proc. Inst. Radio Engrs **17**, 1021, 1929.
7) H. M. Turner, Proc. Inst. Radio Engrs **17**, 1822, 1929.
8) C. R. Hanna, Trans. Amer. Instn Elect. Engrs **46**, 155, 1927.

APPENDIX

Example A

A shell-type choke is required with $h = d$ and the following properties:

$E = 185$ volt $\quad \beta_{max} = 1{\cdot}068$

$I = 0{\cdot}41$ amp $\quad R_{max} = 36{\cdot}9$ ohm

$L = 1{\cdot}434$ henry $\quad f = 50$ c/sec

Known winding and core-material data are

$x = 0{\cdot}433$ $\quad \gamma = 12{\cdot}6$

$s = 0{\cdot}302$ $\quad \sigma = 1{\cdot}91 \times 10^{-10}$ watt/cm^{3}gauss^2

$S = 0{\cdot}92$ $\quad \varrho_c = 1{\cdot}78 \times 10^{-6}$ ohm.cm

$k_2 = 4{\cdot}46 \times 10^{-52}$ $\quad k_1 = 1{\cdot}84 \times 10^{-4}$

Step 4) $Sx/2s = 0{\cdot}66$, and from table IV

$z_1 = 73{\cdot}6$.

Step 5) $p = 1{\cdot}046 \times 10^{-9}$

$q = 9{\cdot}64 \times 10^{10}$

so $V_1 = 101{\cdot}7$ cm^3.

Steps 6) and 7) $w = 4{\cdot}48 \times 10^6$

$B_m = 13\,160$ gauss

$u = 5{\cdot}44 \times 10^5$

so $V_{max} = 8{\cdot}2$ cm^3

and the intermediate steps are required in this case.

Step 8) $\alpha_2 = 0{\cdot}4$, $\delta_2 = -2{\cdot}85$, $z_2 = 78{\cdot}4$.

Step 9) $\alpha_3 = 2{\cdot}38 \times 10^{-4} \left\{\dfrac{z^2}{\alpha_2 (1-\alpha_2)}\right\}^{1{\cdot}2794} = 0{\cdot}391.$

Step 10) $\alpha_4 = 0{\cdot}396$, $\delta_4 = -2{\cdot}88$, $z_4 = 78{\cdot}7$,

so $\alpha_5 = 0{\cdot}395$ which is close enough to α_4.

Step 11) Since $\alpha_n = 0{\cdot}395$, $\delta_n = -2{\cdot}89$, $z_n = 78{\cdot}7$,

$B_m = 10\,800$ gauss,

$V = V_{max} = 120{\cdot}5$ cm^3,

and of course, in this case

$\beta = \beta_{max} = 1{\cdot}068$.

Step 12)

$y = 0{\cdot}382$		$V^{1/3} = 4{\cdot}94$ cm		$l = 16{\cdot}40$ cm	
$m = 1{\cdot}98$		$d = 2{\cdot}72$ cm		$A = 7{\cdot}40$ cm^2	
$n = 2{\cdot}99$		$b = 4{\cdot}11$ cm		$N = 1134$	
$t = 1{\cdot}98$		$h = 2{\cdot}72$ cm		$\mu = 4720$ gauss/oersted	
$r = 0{\cdot}278$		$c = 1{\cdot}37$ cm		$K = 22{\cdot}1$	
				$a = 0{\cdot}080$ cm.	

Check: from (2.2.8)

$R_m = 14{\cdot}7$ ohm

$R_c = 22{\cdot}1$ ohm

$R = 36{\cdot}8$ ohm

in agreement with the value of 36·9 taken for R_{max}.

Example B

A choke with all requirements as in example A but with $R_{max} = 30$ ohm. Winding and core-material data as in Example A.

Step 4) $z_1 = 73{\cdot}6$.

Step 5) $p = 1{\cdot}046 \times 10^{-9}$

$q = 9{\cdot}64 \times 10^{10}$

so $V_1 = 189{\cdot}3$ cm^3.

Steps 6) and 7) $w = 4{\cdot}48 \times 10^6$

$B_m = 8700$ gauss

$u = 1{\cdot}956 \times 10^3$

so $V_{max} = 2290$ cm^3

and we may proceed to step 12):

Step 12) $y = 0{\cdot}473$ $V^{1/3} = 5{\cdot}74$ cm $A = 9{\cdot}43$ cm^2

$m = 1{\cdot}71$ $d = 3{\cdot}07$ cm $l = 20{\cdot}14$ cm

$n = 2{\cdot}91$ $b = 5{\cdot}21$ cm $N = 1104$

$t = 1{\cdot}71$ $h = 3{\cdot}07$ cm $\mu = 5370$ gauss/oersted

$r = 0{\cdot}312$ $c = 1{\cdot}79$ cm $K = 24{\cdot}7$

$a = 0{\cdot}097$ cm.

Check: from (2.2.8)

$R_m = 15{\cdot}0$ ohm

$R_c = 15{\cdot}0$ ohm

$R = 30{\cdot}0$ ohm.

TABLE I

Typical calculated optimum ratios for shell-type choke without restrictions

m	n	t	r	y	z
0·2	2·09	0·569	1·101	7·74	63·4
0·4	2·17	1·084	0·686	2·44	56·6
0·5	2·20	1·333	0·587	1·710	56·2
0·6	2·23	1·573	0·517	1·288	56·9
0·7	2·26	1·812	0·463	1·014	58·2
0·8	2·29	2·05	0·421	0·828	59·6
1·0	2·33	2·50	0·359	0·592	63·1
1·25	2·38	3·06	0·305	0·426	69·0
1·5	2·43	3·60	0·266	0·324	75·4
1·75	2·46	4·13	0·237	0·258	82·0
2·0	2·50	4·67	0·214	0·211	88·3
2·25	2·53	5·18	0·195	0·177	96·9
2·5	2·56	5·71	0·180	0·151	106·3
2·75	2·58	6·27	0·166	0·130	114·0
3·0	2·60	6·73	0·155	0·114	122·6

TABLE II

Typical ratios for "wasteless"-lamination choke

$m = 2$ $n = 3$

t	r	y	z
0·051	0·938	10	182
0·102	0·741	5	145
0·174	0·621	5	125
0·211	0·583	2·5	118
0·267	0·539	2·0	110
0·364	0·486	1·5	102
0·571	0·418	1·0	91·7
1·333	0·315	0·5	80·6
2·000	0·275	0·375	79·2
2·86	0·244	0·3	80·3
6·67	0·184	0·2	94·1

TABLE III

Data for shell-type choke without restrictions

z	y	m	n	t	r	$Sx/2s$			
						$\delta = -2$	$\delta = -3$	$\delta = -4$	$\delta = -5$
56·2	1·71	0·50	2·20	1·33	0·587				
57	1·28	0·61	2·23	1·60	0·512	13·1			
58	1·06	0·69	2·26	1·79	0·470	5·8	9·3	12·7	
60	0·787	0·82	2·29	2·11	0·412	3·3	5·4	7·4	9·4
62	0·647	0·94	2·32	2·37	0·375	1·70	2·9	4·1	5·2
64	0·558	1·03	2·34	2·59	0·350	1·02	1·81	2·6	3·4
66	0·499	1·13	2·36	2·78	0·331	0·71	1·32	1·92	2·5
68	0·448	1·21	2·38	2·97	0·313	0·55	1·05	1·56	2·1
70	0·411	1·30	2·39	3·14	0·298	0·47	0·90	1·34	1·78
72	0·373	1·39	2·40	3·31	0·285	0·41	0·81	1·20	1·60
74	0·341	1·47	2·42	3·47	0·273	0·36	0·72	1·07	1·42
76	0·317	1·56	2·43	3·65	0·263	0·32	0·63	0·95	1·26
78	0·297	1·64	2·44	3·81	0·254	0·28	0·56	0·84	1·13
80	0·279	1·72	2·45	3·98	0·245	0·23	0·49	0·74	1·00
82	0·258	1·79	2·46	4·14	0·237	0·20	0·43	0·66	0·89
84	0·240	1·85	2·48	4·29	0·228	0·17	0·38	0·59	0·79
86	0·224	1·92	2·49	4·47	0·220	0·15	0·34	0·52	0·71
88	0·213	1·98	2·50	4·64	0·212	0·12	0·28	0·44	0·61
90	0·203	2·04	2·51	4·77	0·204	0·09	0·24	0·38	0·53
92	0·194	2·10	2·52	4·89	0·195	0·07	0·20	0·33	0·47
94	0·186	2·16	2·52	5·02	0·187	0·06	0·19	0·30	0·42

TABLE IV

Data for shell-type choke with restricted stacking height, $h = d$

z	y	m	n	t	r	$Sx/2s$			
						$\delta = -2$	$\delta = -3$	$\delta = -4$	$\delta = -5$
59·5	2·15	0·695	2·51	0·695	0·629				
60	1·77	0·752	2·56	0·752	0·587				
61	1·34	0·858	2·61	0·858	0·527	10			
62	1·16	0·966	2·65	0·966	0·490	5·3	7·9	11	
64	0·944	1·13	2·72	1·13	0·430	3·1	4·9	6·7	9·2
66	0·803	1·26	2·77	1·26	0·398	2·15	3·6	5·0	6·6
68	0·682	1·39	2·81	1·39	0·369	1·57	2·67	3·8	4·9
70	0·561	1·51	2·85	1·51	0·343	1·14	2·02	2·87	3·7
72	0·509	1·62	2·89	1·62	0·324	0·84	1·52	2·17	2·87
74	0·464	1·73	2·92	1·73	0·308	0·63	1·15	1·70	2·28
76	0·426	1·84	2·95	1·84	0·294	0·47	0·91	1·36	1·83
78	0·392	1·94	2·98	1·94	0·282	0·37	0·75	1·11	1·52
80	0·363	2·04	3·01	2·04	0·272	1·33	0·68	1·03	1·37
82	0·340	2·13	3·03	2·13	0·261	0·27	0·58	0·88	1·19
84	0·318	2·22	3·05	2·22	0·253	0·25	0·53	0·81	1·09
86	0·299	2·31	3·07	2·31	0·245	0·23	0·49	0·75	1·02
88	0·281	2·40	3·09	2·40	0·238	0·22	0·47	0·73	0·98
90	0·268	2·48	3·11	2·48	0·231	0·21	0·45	0·69	0·93
92	0·253	2·56	3·12	2·56	0·225	0·19	0·40	0·62	0·84
94	0·240	2·65	3·14	2·65	0·218	0·15	0·35	0·55	0·75
96	0·228	2·73	3·15	2·73	0·213	0·13	0·31	0·49	0·67
98	0·217	2·81	3·17	2·81	0·208	0·11	0·28	0·45	0·61
100	0·208	2·90	3·18	2·90	0·203	0·10	0·25	0·40	0·56
105	0·187	3·10	3·22	3·10	0·192	0·07	0·20	0·33	0·46
110	0·170	3·30	3·24	3·30	0·183	0·06	0·17	0·28	0·40

TABLE IV

(Continued)

$h = 1{\cdot}5\ d$

z	y	m	n	t	r	$Sx/2s$			
						$\delta = -2$	$\delta = -3$	$\delta = -4$	$\delta = -5$
57·1	2·1	0·570	2·37	0·865	0·638				
58	1·46	0·717	2·43	1·08	0·539				
59	1·13	0·820	2·47	1·23	0·486	8·6	13·5		
60	0·966	0·902	2·50	1·35	0·452	4·2	6·7	9·2	11·8
62	0·782	1·05	2·56	1·58	0·404	2·38	4·0	5·5	7·1
64	0·660	1·17	2·60	1·76	0·371	1·61	2·75	3·9	5·0
66	0·569	1·28	2·63	1·92	0·346	1·14	2·00	2·86	3·7
68	0·504	1·38	2·65	2·07	0·327	0·83	1·49	2·15	2·82
70	0·462	1·48	2·67	2·22	0·310	0·61	1·14	1·67	2·21
72	0·413	1·57	2·70	2·36	0·295	0·48	0·93	1·38	1·82
74	0·381	1·66	2·72	2·49	0·283	0·39	0·77	1·15	1·54
76	0·354	1·74	2·73	2·61	0·272	0·31	0·64	0·98	1·31
78	0·330	1·82	2·75	2·73	0·263	0·27	0·57	0·87	1·18
80	0·308	1·90	2·77	2·85	0·254	0·25	0·53	0·81	1·09
82	0·287	1·98	2·78	2·97	0·245	0·24	0·51	0·77	1·04
84	0·271	2·07	2·80	3·11	0·237	0·22	0·47	0·72	0·96
86	0·252	2·15	2·81	3·23	0·230	0·21	0·44	0·67	0·90
88	0·236	2·24	2·82	3·36	0·223	0·18	0·39	0·59	0·80
90	0·223	2·32	2·83	3·48	0·216	0·15	0·34	0·52	0·71
92	0·211	2·40	2·84	3·60	0·210	0·11	0·27	0·43	0·59
94	0·201	2·47	2·85	3·70	0·205	0·09	0·23	0·38	0·52
96	0·193	2·55	2·86	3·82	0·200	0·08	0·21	0·35	0·48
98	0·185	2·62	2·87	3·93	0·196	0·07	0·19	0·32	0·44
100	0·178	2·69	2·88	4·04	0·192	0·06	0·18	0·30	0·42

TABLE IV (Continued)

$h = 2d$

z	y	m	n	t	r	$Sx/2s$			
						$\delta = -2$	$\delta = -3$	$\delta = -4$	$\delta = -5$
56·25	2·11	0·50	2·27	1·00	0·642				
57	1·33	0·66	2·34	1·32	0·522				
58	1·04	0·77	2·37	1·54	0·468	6·7	10·6		
60	0·810	0·895	2·40	1·79	0·416	2·83	4·6	7·5	8·3
62	0·685	1·00	2·43	2·00	0·384	1·62	2·78	3·9	5·1
64	0·586	1·10	2·45	2·20	0·357	1·24	2·16	3·1	4·0
66	0·513	1·20	2·47	2·40	0·333	0·86	1·55	2·24	2·92
68	0·461	1·295	2·49	2·59	0·315	0·65	1·21	1·77	2·32
70	0·417	1·38	2·51	2·76	0·300	0·54	1·02	1·49	1·97
72	0·377	1·46	2·53	2·92	0·286	0·43	0·84	1·24	1·64
74	0·348	1·54	2·54	3·08	0·275	0·35	0·70	1·05	1·40
76	0·323	1·62	2·55	3·24	0·264	0·30	0·60	0·92	1·23
78	0·301	1·70	2·56	3·40	0·255	0·26	0·54	0·82	1·10
80	0·279	1·78	2·57	3·56	0·245	0·22	0·47	0·72	0·97
82	0·263	1·855	2·58	3·71	0·238	0·20	0·43	0·66	0·89
84	0·248	1·93	2·59	3·86	0·230	0·17	0·37	0·58	0·78
86	0·230	2·005	2·60	4·01	0·223	0·14	0·33	0·52	0·70
88	0·220	2·07	2·60	4·14	0·217	0·12	0·28	0·45	0·62
90	0·209	2·135	2·61	4·27	0·212	0·08	0·23	0·38	0·52
92	0·200	2·20	2·62	4·40	0·207	0·05	0·18	0·30	0·43
94	0·192	2·265	2·63	4·53	0·202	0·03	0·14	0·25	0·36

TABLE V

Data for core-type choke without restrictions

z	y	m	n	t	r	$Sx/2s$			
						$\delta = -2$	$\delta = -3$	$\delta = -4$	$\delta = -5$
44·6	3·42	0·25	2·20	0·065	0·932				
45·2	2·56	0·305	2·23	0·80	0·813				
46·0	2·12	0·345	2·26	0·90	0·746	11·6			
47·6	1·57	0·41	2·29	1·06	0·655	6·6	10·8	4·8	
49·1	1·29	0·47	2·32	1·18	0·596	3·4	5·8	8·2	10·4
50·7	1·12	0·515	2·34	1·30	0·556	2·0	3·6	5·2	6·8
52·3	1·00	0·565	2·36	1·39	0·526	1·42	2·6	3·8	5·0
53·9	0·896	0·605	2·38	1·48	0·497	1·10	2·1	3·1	4·2
55·5	0·822	0·65	2·39	1·57	0·473	0·94	1·80	2·7	3·6
57·1	0·746	0·70	2·40	1·66	0·453	0·82	1·62	2·4	3·2
58·6	0·682	0·74	2·42	1·74	0·433	0·72	1·44	2·1	2·8
60·2	0·634	0·78	2·43	1·82	0·418	0·64	1·26	1·90	2·5
61·8	0·594	0·82	2·44	1·90	0·403	0·56	1·12	1·68	2·3
63·4	0·558	0·86	2·45	1·99	0·389	0·46	0·98	1·48	2·0
65·0	0·516	0·90	2·46	2·07	0·377	0·40	0·86	1·32	1·78
66·6	0·480	0·92	2·48	2·14	0·362	0·35	0·76	1·18	1·58
68·2	0·448	0·96	2·49	2·24	0·350	0·29	0·68	1·04	1·42
69·8	0·426	0·99	2·50	2·32	0·337	0·23	0·56	0·88	1·22
71·3	0·406	1·02	2·51	2·38	0·324	0·18	0·48	0·76	1·06
72·9	0·388	1·05	2·52	2·44	0·310	0·14	0·40	0·66	0·94
74·5	0·372	1·08	2·52	2·51	0·297	0·12	0·36	0·60	0·84

TABLE VI

Data for core-type choke with restricted stacking height, $h = d$

z	y	m	n	t	r	$Sx/2s$			
						$\delta = -2$	$\delta = -3$	$\delta = -4$	$\delta = -5$
47·2	4·3	0·348	2·51	0·348	0·998				
47·6	3·54	0·376	2·56	0·376	0·932				
48·4	2·68	0·429	2·61	0·429	0·837	20			
49·1	2·32	0·483	2·65	0·483	0·778	10·6	15·8	22	
50·7	1·89	0·565	2·72	0·565	0·683	6·2	9·8	13·4	18·4
52·3	1·61	0·63	2·77	0·63	0·632	4·3	7·2	10·0	13·2
53·9	1·36	0·695	2·81	0·695	0·586	3·1	5·3	7·6	9·8
55·5	1·12	0·755	2·85	0·755	0·545	2·28	4·0	5·7	7·4
57·1	1·02	0·81	2·89	0·81	0·514	1·68	3·0	4·3	5·7
58·6	0·928	0·865	2·92	0·865	0·489	1·26	2·30	3·4	4·6
60·2	0·852	0·92	2·95	0·92	0·467	0·94	1·82	2·72	3·7
61·8	0·784	0·97	2·98	0·97	0·448	0·74	1·50	2·22	3·0
63·4	0·726	1·02	3·01	1·02	0·432	0·66	1·36	2·06	2·74
65·0	0·680	1·06	3·03	1·06	0·415	0·54	1·16	1·76	2·38
66·6	0·636	1·11	3·05	1·11	0·402	0·50	1·06	1·62	2·18
68·2	0·598	1·16	3·07	1·16	0·389	0·46	0·98	1·50	2·04
69·8	0·562	1·20	3·09	1·20	0·378	0·44	0·94	1·46	1·96
71·3	0·536	1·24	3·11	1·24	0·367	0·42	0·90	1·38	1·86
72·9	0·506	1·28	3·12	1·28	0·358	0·38	0·80	1·24	1·68
74·5	0·480	1·32	3·14	1·32	0·346	0·30	0·70	1·10	1·50
76·1	0·456	1·36	3·15	1·36	0·339	0·26	0·62	0·98	1·34
77·7	0·434	1·40	3·17	1·40	0·330	0·22	0·56	0·90	1·22
79·3	0·416	1·45	3·18	1·45	0·323	0·20	0·50	0·80	1·12
83·3	0·374	1·55	3·22	1·55	0·305	0·14	0·40	0·66	0·92
87·3	0·340	1·65	3·24	1·65	0·291	0·12	0·34	0·56	0·80
91·2	0·310	1·74	3·27	1·74	0·277	0·09	0·28	0·48	0·68
95·2	0·288	1·84	3·30	1·84	0·265	0·08	0·24	0·42	0·60
99·2	0·267	1·93	3·32	1·93	0·254	0·07	0·22	0·38	0·54
103·1	0·248	2·02	3·34	2·02	0·246	0·06	0·20	0·34	0·50

TABLE VI

(Continued)

$h = 1{\cdot}5\ d$

z	y	m	n	t	r	$Sx/2s$			
						$\delta = -2$	$\delta = -3$	$\delta = -4$	$\delta = -5$
45·3	4·2	0·285	2·37	0·432	1·013				
46·0	2·92	0·385	2·43	0·54	0·856				
46·8	2·26	0·410	2·47	0·615	0·771	17·2			
47·6	1·93	0·451	2·50	0·675	0·718	8·3	13·4	18·4	
49·1	1·56	0·525	2·56	0·79	0·640	4·8	7·9	11·1	14·2
50·7	1·32	0·585	2·60	0·88	0·590	3·2	5·5	7·8	10·1
52·3	1·14	0·64	2·63	0·96	0·550	2·28	4·0	5·7	7·4
53·9	1·01	0·69	2·65	1·04	0·519	1·66	3·0	4·3	5·6
55·5	0·924	0·74	2·67	1·11	0·493	1·22	2·28	3·3	4·4
57·1	0·826	0·785	2·70	1·18	0·469	0·96	1·86	2·76	3·6
58·6	0·762	0·83	2·72	1·24	0·449	0·78	1·54	2·30	3·1
60·2	0·708	0·87	2·73	1·30	0·432	0·62	1·28	1·96	2·62
61·8	0·660	0·91	2·75	1·36	0·417	0·54	1·14	1·74	2·36
63·4	0·616	0·95	2·77	1·42	0·403	0·50	1·06	1·62	2·18
65·0	0·574	0·99	2·78	1·48	0·390	0·48	1·02	1·54	2·08
66·6	0·542	1·04	2·80	1·56	0·376	0·44	0·94	1·44	1·92
68·2	0·504	1·08	2·81	1·62	0·365	0·42	0·88	1·34	1·80
69·8	0·472	1·12	2·82	1·68	0·354	0·36	0·78	1·18	1·60
71·3	0·446	1·16	2·83	1·74	0·343	0·30	0·68	1·04	1·42
72·9	0·422	1·20	2·84	1·80	0·333	0·22	0·54	0·86	1·18
74·5	0·404	1·24	2·85	1·85	0·326	0·18	0·46	0·76	1·04
76·1	0·386	1·28	2·86	1·91	0·318	0·16	0·42	0·70	0·96
77·7	0·370	1·31	2·87	1·96	0·311	0·14	0·38	0·64	0·88
79·3	0·356	1·34	2·88	2·02	0·305	0·12	0·36	0·60	0·84
83·3	0·322	1·43	2·90	2·14	0·289	0·10	0·32	0·53	0·72
87·3	0·292	1·52	2·92	2·27	0·276	0·09	0·28	0·48	0·68
91·2	0·268	1·60	2·93	2·40	0·265	0·08	0·24	0·42	0·58
95·2	0·248	1·68	2·95	2·51	0·254	0·07	0·21	0·36	0·52
99·2	0·230	1·76	2·96	2·63	0·243	0·06	0·19	0·32	0·48

TABLE VI

(Continued)

$h = 2d$

z	y	m	n	t	r	$Sx/2s$			
						$\delta = -2$	$\delta = -3$	$\delta = -4$	$\delta = -5$
44·6	4·2	0·25	2·27	0·50	1·019				
45·2	2·66	0·33	2·34	0·66	0·828				
46·0	2·10	0·38	2·37	0·77	0·742	13·4			
47·6	1·62	0·45	2·40	0·895	0·660	5·7	9·2	15·0	16·6
49·1	1·37	0·50	2·43	1·00	0·609	3·2	5·6	7·8	10·2
50·7	1·17	0·55	2·45	1·10	0·567	2·48	4·3	6·2	8·0
52·3	1·03	0·60	2·47	1·20	0·529	1·72	3·1	4·5	5·8
53·9	0·922	0·65	2·49	1·295	0·500	1·30	2·42	3·5	4·6
55·5	0·834	0·69	2·51	1·38	0·476	1·08	2·04	2·98	3·9
57·1	0·754	0·73	2·53	1·46	0·454	0·86	1·68	2·48	3·3
58·6	0·696	0·77	2·54	1·54	0·436	0·70	1·40	2·10	2·80
60·2	0·646	0·81	2·55	1·62	0·419	0·60	1·20	1·84	2·46
61·8	0·602	0·85	2·56	1·70	0·405	0·52	1·08	1·64	2·20
63·4	0·558	0·89	2·57	1·78	0·389	0·44	0·94	1·44	1·94
65·0	0·526	0·93	2·58	1·855	0·378	0·40	0·86	1·32	1·78
66·6	0·496	0·96	2·59	1·93	0·365	0·33	0·74	1·16	1·58
68·2	0·460	1·00	2·60	2·005	0·354	0·29	0·66	1·04	1·40
69·8	0·440	1·04	2·60	2·07	0·345	0·23	0·56	0·90	1·24
71·3	0·418	1·07	2·61	2·135	0·337	0·17	0·46	0·76	1·04
72·9	0·400	1·10	2·62	2·20	0·329	0·10	0·35	0·60	0·86
74·5	0·384	1·13	2·63	2·265	0·321	0·05	0·27	0·50	0·72

Philips Res. Rep. 2, 313-319, 1947

ON THE AMORPHOUS AND CRYSTALLINE OXIDE LAYER OF ALUMINIUM

by A. J. DEKKER and W. CH. VAN GEEL

541.138.2:546.621

Summary

Aluminium can be covered electrolytically with a porous layer of aluminium oxide, and oxidation afterwards in boric acid gives rise to the formation of a crystalline layer. The experiments described below show undoubtedly that this layer only fills up the holes of the amorphous Al_2O_3. Moreover, there is a correlation between the current density in oxalic acid and the porosity of the amorphous layer thus formed.

1. *Introduction*

By anodic oxidation a piece of aluminium can be covered either with an amorphous or with a crystalline oxide layer.

The amorphous layer obtained in a solution of oxalic, sulphuric or phosphoric acid has a porous structure, the pores having a diameter of about 10^{-5} cm and being mainly oriented at right angles to the surface on which the layer is deposited (see electron-microscope picture, *fig. 1*). During the process of formation the porous layer is saturated with electrolyte solution and offers only a very slight resistance to the electric current passing through it. Hence, if the oxidation is carried out at a constant voltage, the current, too, will be practically constant, and the layer can be made as thick as we desire; there is only a slight current decrease during the first few seconds of the oxidation.

A crystalline layer of Al_2O_3 can be produced by using an aqueous solution of a borate or boric acid, a succinate or a citrate. This layer has pronounced insulating properties so that during formation at constant voltage the current steadily decreases. Hence, there is a limit to the thickness of a crystalline layer which is determined by the so-called "spark potential" which depends on the specific resistance of the electrolyte solution [1]); for instance, the maximal voltage obtainable in a solution of a borate with specific resistance of 40 Ω cm is about 400 volts giving a maximal thickness of 0·44 μ.

Now, it has been observed by Van Geel and Emmens [2]) that an insulating crystalline layer can be formed on an aluminium surface that has previously been coated with an amorphous layer. They held the view that the crystalline layer is in this case formed underneath the amorphous layer.

Recent experiments, however, have led us to the conclusion that this is not correct; the formation of a crystalline layer is taking place in the

pores of the amorphous layer present. These experiments will be described in the present paper.

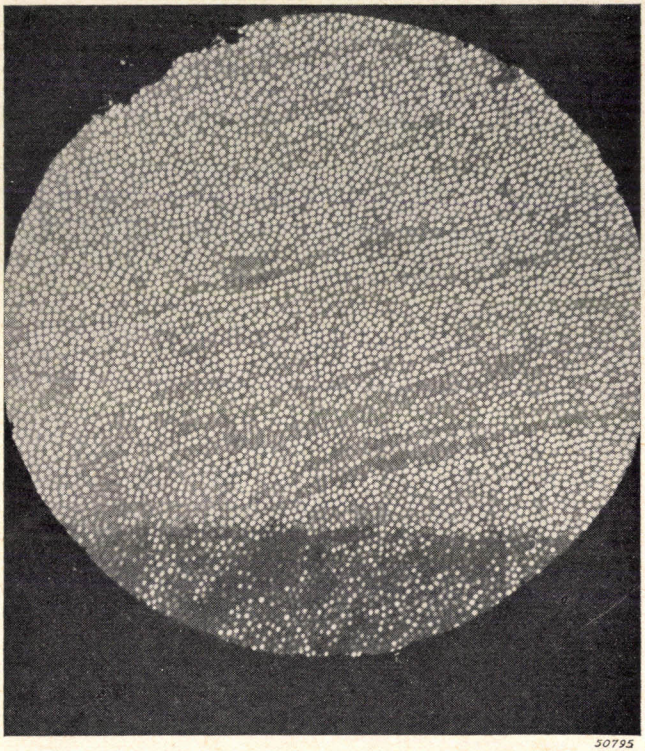


Fig. 1. Electron-microscope micrograph of a porous layer of 1·8 μ thickness with the electron beam (250 kV) perpendicular to the surface (4200 ×).

2. *Experimental results*

A) An aluminium plate (purity 99·99%) with an area of 440 cm² was anodically coated with a crystalline layer in boric acid at a constant current of 200 mA. Both the voltage V and the capacity C were measured during the process of formation; we give some of the results here:

Time (min)	Voltage (volts)	$1/C$ (cm^{-1})
0	10	1·00
5	70	2·62
15	210	8·5
30	375	16·2

$\times 10^{-8}$

(see *fig. 2* and *fig. 3*, curves I)

$dV/dt = 12{\cdot}2$ volts min^{-1},

$d(C^{-1})/dt = 0{\cdot}54.10^{-8}$ $cm^{-1}min^{-1}$.

The increase of weight after 30 minutes of oxidation appeared to be 28·75 mg; 1 amp. hour giving 298·4 mg of oxygen; the calculated current efficiency becomes 96%. This was in agreement with the fact that no gas development at the anode could be observed. During the formation of the oxide layer both V and $1/C$ vary linearly with time, as expected. The capacity at $t = 0$ is due to the oxidation of aluminium by air, and the thickness of this air layer is of the order of 1 mμ.

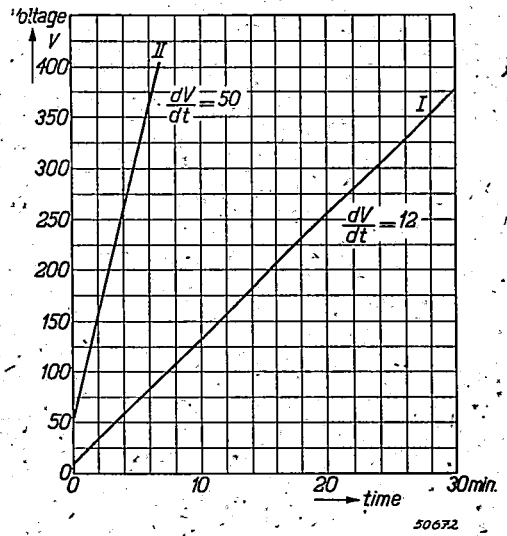


Fig. 2. Voltage as a function of time for an Al plate of 440 cm² during anodical oxidation in boric acid with 200 mA; curve I without an amorphous layer (A), curve II with an amorphous layer of 3·5 μ (B).

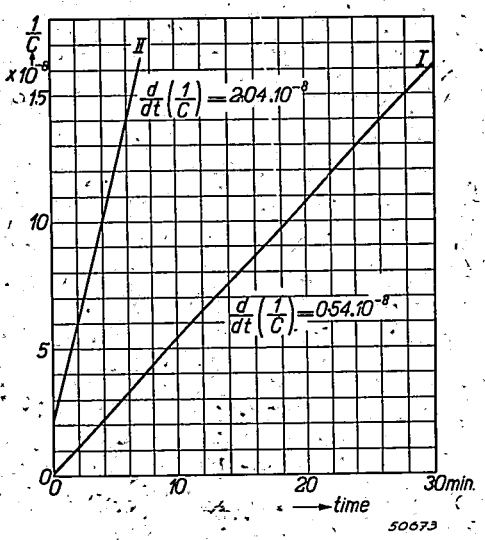


Fig. 3. $1/C$ in cm^{-1} as a function of time for an Al plate of 440 cm² during anodical oxidation in boric acid with 200 mA; curve I without an amorphous layer (A), curve II with an amorphous layer of 3·5 μ (B).

By a well-known formula, the capacity of a plate condenser is given by:

$$C = \frac{\varepsilon S}{4\pi d},$$

where ε stands for the dielectric constant of the oxide layer, S for the surface area, and d for the thickness. Writing ϱ for the specific weight we can express d in the increase of weight, and from this it follows that $\varepsilon \varrho = 24{\cdot}4$. This is in very good agreement with the numerical values given by Burgers, Claassen and Zernike [3]), *i.e.* $\varepsilon = 8$ and $\varrho = 3{\cdot}1$.

B) The same experiment was carried out with an aluminium plate (area 440 cm²) *covered with an amorphous layer* by oxydation in oxalic acid at 4A during 10 minutes, the thickness of the amorphous layer then being about 3·5 μ. Before the second oxidation in boric acid, the plate was

heated at 450 °C during 15 minutes, to remove residuals of oxalic acid and water. V and C were again determined as a function of time, the current being kept at a constant value of 200 mA. Results follow below (see fig. 2 and fig. 3, curves II):

Time (min)	Voltage (volts)	$1/C$ (cm^{-1})
0	54	$2{\cdot}24.10^{-8}$
6	360	$14{\cdot}5\ .10^{-8}$

$dV/dt = 50$ volts min^{-1}
$d(C^{-1})/dt = 2{\cdot}04.10^{-8}$ $cm^{-1}min^{-1}$.

It may be noticed that, as referred to in the introduction, the oxidation in oxalic acid produces a thin insulating layer which is responsible for the zero-capacity of 49·7 μF. The increase of weight accompanying the oxidation in boric acid amounted to 6·0 mg, and from this it follows that the current efficiency is just 100%. Comparing the values of dV/dt and $d/dt(1/C)$ of the curves I and II we see that, *due to the presence of the amorphous layer, these quantities are about four times as great as in the first experiment.*

Calculating the value of $\varepsilon\varrho$ for this case, on the assumption that the crystalline layer is formed *underneath* the amorphous layer, we find $\varepsilon\varrho = 6{\cdot}7$, and it is evident that this assumption is wrong. Bearing in mind, however, that the amorphous layer has an open structure containing pores, it is natural to suppose that the much greater velocity of formation, in case an amorphous layer is present, is due to the filling-up of the pores; then only a certain part of the volume has to be filled up with crystalline Al_2O_3, in order to produce a layer with insulating properties.

This hypothesis can be verified in the following way:

In the experiment just described we used an amorphous layer of about 3·5 μ, and the oxidation in boric acid to 350 volts gives rise to a crystalline layer of about 0·35 μ; hence, according to the hypothesis, only a small fraction of the amorphous layer is filled up with crystalline Al_2O_3. But using, instead, an amorphous layer of, say, 0·2 μ we may expect a sudden decrease in the slope of the $V(t)$-curve during oxidation in boric acid, because at a certain time t the amorphous layer will be wholly filled up, and from this time on the entire surface area has to be covered with crystalline Al_2O_3.

C) The following experiments were carried out to check this point of view. Five aluminium plates with areas of 165 cm^2 were covered with an amorphous layer in oxalic acid at a current of 200 mA during 2, 4, 6, 8 and 10 minutes, respectively, and in this way each plate was covered with an amorphous layer of a thickness proportional to the time of oxidation. The plates were again heated at 450 °C during 15 minutes and then anodically oxidized in boric acid at a current of 100 mA to build up

a crystalline layer; V and C were determined as a function of time; in *fig. 4* the results for V have been reproduced, those for $1/C$ being of the same shape.

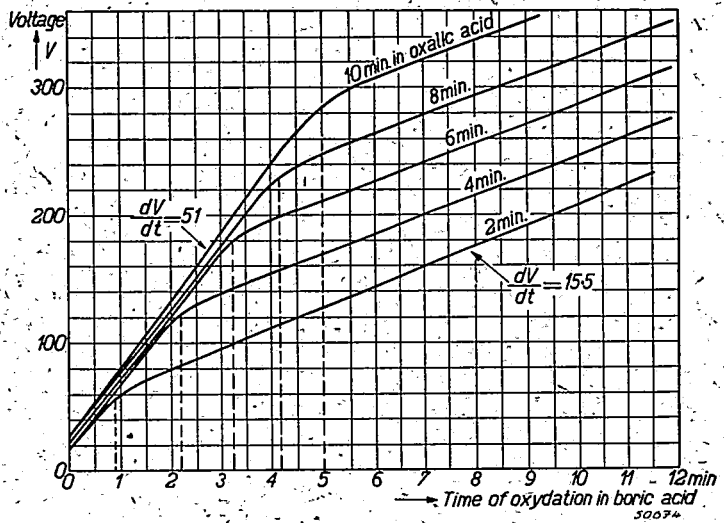


Fig. 4. Voltage as a function of time for Al plates of 165 cm² covered with layers of amorphous Al_2O_3 with different thicknesses during anodical oxidation in boric acid at 100 mA (C).

As expected, the slope of the curves are very steep in the start (the period of filling up the holes of the amorphous layer), but after a certain time a marked decrease of the slope takes place.

The two straight portions of the curves in fig. 4 have been extrapolated until they intersect, and the time corresponding to this point of intersection has been taken as the time at which the pores of the amorphous layer are completely filled up with crystalline material. As shown in *fig. 5*, these time intervals are proportional to the time of oxidation in oxalic acid, as we should expect in virtue of our hypothesis.

From the two extreme values of the slope in fig. 4 one may conclude that the volume of the pores in the amorphous layer amounts to 30% of the total volume. This is in rather good agreement with the specific weights of amorphous and crystalline Al_2O_3, these values being, respectively, 2·5 and 3·1 from which follows a relative volume of the pores of about 25%.

The same result is obtained by noticing that in fig. 5 the slope is just 0·5, and, bearing in mind that the current in oxalic acid was 200 mA and in boric acid 100 mA, we see that a porous layer formed by x amp. minutes is just filled up in boric acid by $^1/_4\ x$ amp. minutes; this leads to 25% for the relative pore volume.

3. *The rate of change of voltage as a measure of the porosity of the amorphous layer*

Let us now compare the rate of change of the voltage during oxidation in boric acid in the various experiments described above. The rate of change in the voltage will, of course, be proportional to the current density, so that to render our data mutually comparable we should divide dV/dt by the current density i. It will be easily verified that we then have

experiment A: $$\frac{dV/dt}{i} = 26{\cdot}4 \frac{V\,\mathrm{min}^{-1}}{\mathrm{mA\,cm}^{-2}},$$

second stage of experiment C: $$\frac{dV/dt}{i} = 25{\cdot}6 \frac{V\,\mathrm{min}^{-1}}{\mathrm{mA\,cm}^{-2}},$$

in satisfactory agreement, since in both cases the entire surface area was being coated with a crystalline layer.

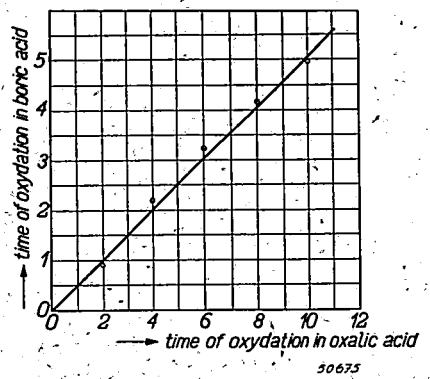


Fig. 5. The time necessary for filling up the amorphous layer at 100 mA as a function of the time of oxidation in oxalic acid at 200 mA, *i.e.* as a function of the thickness of the amorphous layer.

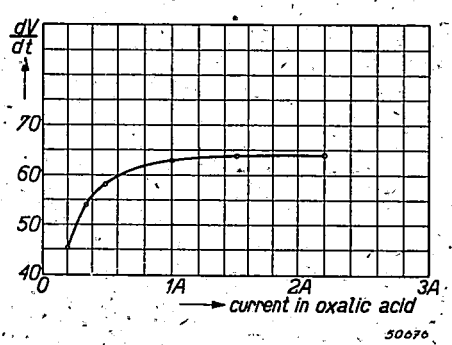


Fig. 6. The velocity of oxidation for 100 mA in boric acid as a function of the oxidation current in oxalic acid for Al plates of 165 cm².

On the other hand we have

in experiment B: $$\frac{dV/dt}{i} = 110 \frac{V\,\mathrm{min}^{-1}}{\mathrm{mA\,cm}^{-2}},$$

and in the initial stage of experiment C: $$\frac{dV/dt}{i} = 83 \frac{V\,\mathrm{min}^{-1}}{\mathrm{mA\,cm}^{-2}}.$$

In these two instances the crystalline matter was being formed in the pores of an amorphous layer. The increased rate of change of voltage must then be due to a current density in the pores higher than the current density

calculated from the apparent surface area. If so, our data lead to the conclusion that the porosity of the amorphous layer was of the order of 25 to 30%. This is in keeping with density observations, the specific weight of an amorphous and a crystalline layer being 2·5 and 3·1 respectively.

The above comparison between experiment B and the initial stage of experiment C also suggest a difference in the degree of porosity in these two cases. This may be a consequence of different current densities used in the formation of the amorphous layer, 9·1 mA/cm² and 1·2 mA/cm² respectively. This has led us to investigate the rate of change $\frac{dV/dt}{i}$ as a function of the current density used in forming the amorphous layer, with the results tabulated below and plotted in *fig. 6*.

Current density in oxalic acid (mA/cm²)	$\frac{dV/dt}{i}$ in boric acid ($V\,min^{-1}$/mA cm^{-2})
1·21	76
2·22	89
3·04	95·5
6·05	104
9·10	106
13·35	106

There is indeed a marked dependence, a higher current density in oxalic acid producing a denser structure of the amorphous coating.

It may finally be noted that the capacity of an aluminium surface that has wholly been oxidized in boric acid is the same as the capacity of a surface that has first been oxidized in oxalic acid and subsequently in boric acid. This shows that the dielectric constants of amorphous and crystalline Al_2O_3 must be equal.

Eindhoven, June 1947

REFERENCES

1) W. Ch. van Geel, Physica **1**, 989-995, 1934.
2) W. Ch. van Geel and H. Emmens, Physica **1**, 415-416, 1935.
3) W. G. Burgers, A. Claassen and J. Zernike, Z. Phys. **74**, 593-603, 1932.

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

Reprints of the majority of these papers can be obtained on application to the Administration of the Research Laboratory, Kastanjelaan, Eindhoven, Netherlands. Those papers of which no reprints are available in sufficient number are marked with an asterisk (*).

1704*: C. J. Bouma: Kleuren en kleurindrukken (Philips technische bibliotheek; uitg. Meulenhoff & Co N.V. Amsterdam 1946, 320 pages, 113 fig.) (Colours and colour impressions).

In this book, an English translation of which is forthcoming, a survey in simple language is given of the fundamentals of colour science and the methods of colorimetry. In the exposition of the basic facts ample use is made of colour space. Additional chapters deal with colour vision, the historical development of colour science, thresholds, the estimation of colour differences and related problems, technical and scientific applications. The appendix comprises 15 tables and an extensive list of references.

1705*: H. A. Klasens and M. E. Wise: Decay of zincsulfide-type phosphors (Nature London **158**, 483, 1946).

The bimolecular decay of zincsulphide phosphors, assuming retrapping in the metastable phosphorescence centra, is decribed by a second-order non-linear differential equation. In the special case where the two unknown constants (α = retrapping coefficient, β = bimolecular recombination constant) are equal, the equation can be solved and curves may be drawn for the intensity versus time at different temperatures and for different intensities of excitation. These curves generally display a steep initial decay due to the bimolecular recombination, and a more or less prolonged tail due to the phosphorescence mechanism.

VOL. 2 NO. 5 OCTOBER 1947

Philips Research Reports

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

R 53 *Philips Res. Rep. 2, 321-330, 1947*

METHOD OF MEASUREMENT OF NOISE RATIOS AND NOISE FACTORS

by A. van der ZIEL 621.396.822:621.3.08

Summary

In this paper methods are described as applied in this laboratory for the measurement of noise ratios of impedances and of the noise factor of receivers. Instead of standard signal generators, saturated diodes are used as "standard noise generators". The noise voltages are amplified in a linear amplifier having a relatively small bandwidth (50-100 kc/sec), and detected by a thermocouple. The results obtained by these methods will be dealt with in subsequent papers.

1. *Introduction*

Several methods have recently been described for the measurement of the noise ratio of crystals [2] [4] and of the noise factor of receivers [3]. These methods, in which saturated diodes are used as standard noise generators, have much in common with the method to be described here though several details are different. As some of our measurements will subsequently be published in these Reports, it is considered necessary first to describe our measuring equipment.

The difference with other methods is twofold. In measuring noise ratios of impedances (*e.g.* crystals) the saturated diode is also used for the measurement of the resistance itself. In measuring noise factors of single receiver stages two saturated diodes are used, one at the input and one at the output of the stage; the latter is sometimes necessary in order to make a correction for the noise of the subsequent stages of the receiver.

2. *Measuring equipment*

The noise ratio of impedances and the noise factor of receiver stages are measured with a high-gain amplifier feeding a thermocouple. Instead of a standard signal generator a saturated diode is used as standard noise generator. It is well known that for a frequency interval $\Delta\nu$ a saturated diode having an anode current I_e behaves as a generator of noise current i_d such that:

$$\overline{i_d^2} = 2eI_e \Delta\nu \qquad (1)$$

(e denoting the electronic charge).

The high-gain amplifier has to be linear. This may be investigated by applying a saturated diode across the input terminals of the amplifier, and plotting the mV-output of the thermocouple against the diode current. The amplifier is linear if the points measured lie on a straight line (*fig. 1*) *).

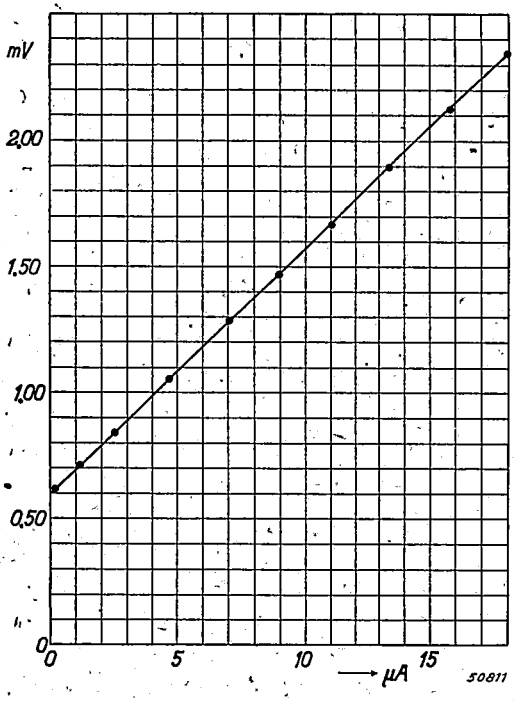


Fig. 1. mV-output of thermocouple *versus* μA-anode current of the saturated diode. The measured points lie on a straight line.

With respect to linearity it must be borne in mind that for equal r.m.s. values a noise voltage contains much higher peak values than an alternating voltage. Hence, linearity for A.C. signals is not always a good criterion for a high-gain amplifier used in noise measurements.

For many measurements it is important to know the tuned input impedance R of the input circuit of the linear amplifier. If the input capacity C of this amplifier is calibrated, R may be evaluated from the noise

*) The input noise power is proportional to $\overline{i_d^2}$, and if the amplifier is linear, the output voltage of the thermocouple is proportional to the input noise power.

resonance curve (*fig. 2*) on which the output of the thermocouple is plotted against the capacity C.

According to Nyquist's theorem the thermal noise of a tuned circuit of impedance R at α-times normal room temperature T (T in degrees Kelvin) and for a small frequency interval $\Delta\nu$ may be represented by a noise e.m.f. v_R in series with R (*fig. 3a*):

$$\overline{v_R^2} = \alpha\, 4\, kTR\, \Delta\nu\,, \tag{2a}$$

or by a noise current generator i_R in parallel to R (*fig. 3b*) as:

$$\overline{i_R^2} = \alpha\, \frac{4\, kT\, \Delta\nu}{R}\,. \tag{2b}$$

With the help of equation (2b) *noise ratios* may be defined for all kinds of impedances. Suppose that we have an admittance Y such that for a small frequency interval around a central frequency:

$$Y = \frac{1}{R} + jB \tag{3}$$

(*e.g.* Y may consist of a complicated network containing thermionic valves

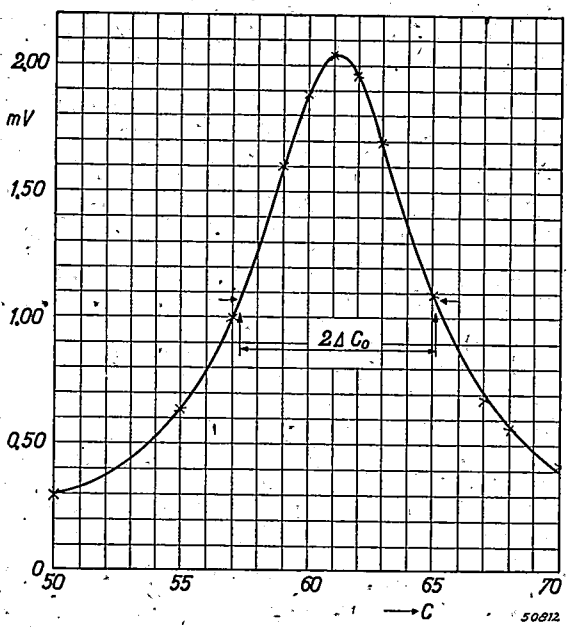


Fig. 2. Noise resonance curve of an amplifier at 50-m wavelength; mV-output of thermocouple plotted against the capacity of the input circuit. One scale division of the abscissa corresponds to 0·425 $\mu\mu$F. Measured points are denoted by crosses; the full-drawn curve is of the general form (4) having $V_0 = 0{\cdot}09$ mV and $V_1 = 2{\cdot}04$ mV.

and embodying feedback), and let the spontaneous fluctuations of electricity in this admittance, for a small frequency interval $\Delta\nu$, be described by a noise current generator i in parallel to Y (*fig. 3c*), then the noise ratio α_0 of the admittance Y is defined as:

$$\alpha_0 = \frac{\overline{i^2}R}{4kT\Delta\nu} \tag{3a}$$

(compare equation (2b)).

We now turn again to our noise resonance curve. As a consequence of Fourier analysis we may consider noise currents as ordinary alternating currents, inasmuch as a noise resonance curve has exactly the same shape as a resonance curve obtained by means of a signal generator. The only condition which has to be satisfied is that the bandwidth of the input circuit is large in comparison with the bandwidth of the linear amplifier *). The bandwidth of our amplifiers was 50-100 kc/sec.

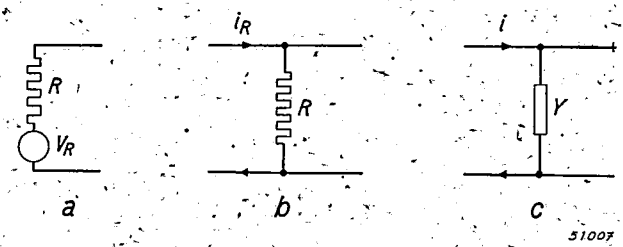


Fig. 3. *a*) Thermal noise of a resistance R, described by an e.m.f. v_R in series with R such that

$$\overline{v_R^2} = 4kTR\Delta\nu .$$

b) Thermal noise of the same resistance R, described by a current generator i_R in parallel to R such that

$$\overline{i_R^2} = \frac{\overline{v_R^2}}{R^2} = \frac{4kT\Delta\nu}{R} .$$

c) Noise of an admittance Y, described by a noise current generator i in parallel to Y.

Let ω denote the angular frequency at which the linear amplifier is tuned, and C_0 the capacity of the tuned input circuit. Let V_0 denote the thermocouple output voltage when the input circuit is short-circuited, and V_1 its value when the input circuit is tuned. If $\Delta C = (C - C_0)$, we have for the output voltage of the thermocouple (as a function of ΔC), in analogy to the case of an alternating input signal, because a thermocouple is a quadratic detector:

$$V = V_0 + \frac{V_1 - V_0}{1 + (\omega\Delta CR)^2} . \tag{4}$$

*) This condition is not a very critical one; with the help of equations (2a) and (2b) it may be shown that (4) is valid within 1% if the bandwidth of the input circuit is seven times the bandwidth of the amplifier.

V_0 is chiefly due to the noise of the first valve whilst the second term in (4) is caused by the noise of the input circuit.

R may be calculated from (4) in the following way.

Let $(V - V_0) = \frac{1}{2}(V_1 - V_0)$ for $\Delta C = \pm \Delta C_0$, then according to (4):

$$\omega \Delta C_0 R = 1. \qquad (4a)$$

As $(2\,\Delta C_0)$ may be accurately derived from the resonance curve (compare fig. 2), R may be calculated from (4a).

After having calculated R from (4), we proceed to the measurement of the noise ratio α_0 of the input circuit. Let again the saturated diode be connected to the input circuit of the amplifier, and let the saturated current I_0 cause an increase of the output voltage of the thermocouple from V_1 to V_2. As the input-circuit noise corresponds to $(V_1 - V_0)$, and the saturated diode current I_0 to $(V_2 - V_1)$, we may define an equivalent saturated diode current I_e such that the input-circuit noise and the current I_e give rise to an equal contribution to the output voltage of the thermocouple. Obviously:

$$I_e = \frac{(V_1 - V_0)}{(V_2 - V_1)} I_0, \qquad (5)$$

or, by equating (1) and (2b),

$$2eI_e \Delta\nu = \alpha_0 \frac{4\,kT\Delta\nu}{R},$$

so that:

$$\alpha_0 = \frac{e}{2kT} I_e R = 20\, I_e R, \qquad (6)$$

in practical units (I_e in amps, R in Ω).

For a normal input circuit of a linear amplifier at 50-m wavelength one finds $\alpha_0 = 1$ within the limit of experimental error (except for very large values of the tuned input impedance R); this may be considered as a good proof of the reliability of the equipment. At 7 m or lower, however, one finds $\alpha_0 > 1$ (*e.g.* $\alpha_0 = 2$-5); this is due to induced grid noise [6]). Moreover, the noise resonance curve of the input circuit becomes sometimes asymmetric; this is due to the *correlation* between the shot-effect noise of the first valve and the induced grid noise. In a subsequent paper it will be shown how this asymmetry may be quantitatively explained in the above way, and that it may be largely reduced by a proper design of the input stage of the linear amplifier. The best results were obtained by using an EFF 50 in which both systems were connected in parallel.

3. *Measurement of noise ratios of impedances*

The noise ratio of a resistance R_1 (*e.g.* the internal resistance of a crystal

diode, a non-saturated diode, or a triode) or the output noise ratio of a mixer circuit may now be measured in the following way.

After having carried out the measurements of the preceding section the resistance R_1 is connected across the input circuit and the latter is tuned again. The input impedance of the tuned circuit then is:

$$R' = \frac{R R_1}{R + R_1}. \tag{7}$$

Let V_3 be the output voltage of the thermocouple for zero diode current, and V_4 the output voltage for a saturated diode current I_1.

As the noise current due to the saturated diode current I_0 was flowing into the resistance R and the noise current due to I_1 into the resistance R', we have:

$$2eI_0\Delta\nu R^2 : 2eI_1\Delta\nu R'^2 = (V_2 - V_1) : (V_4 - V_3),$$

so that

$$\left(\frac{R'}{R}\right) = \sqrt{\left(\frac{V_4 - V_3}{V_2 - V_1}\right)\frac{I_0}{I_1}} \tag{8}$$

and

$$\left(\frac{R_1}{R}\right) = \left[\sqrt{\left(\frac{V_2 - V_1}{V_4 - V_3}\right)\frac{I_1}{I_0}} - 1\right]^{-1}. \tag{8a}$$

This method is inaccurate for large values of R_1.

R' and hence R_1 may also be obtained from the new resonance curve, except for small values of R_1 because the resonance curve then becomes very flat; this method too is inaccurate for large values of R_1. The results of the two methods agree within the experimental error.

As a saturated diode current I_1 corresponds to $(V_4 - V_3)$, and the noise of R' to $(V_3 - V_0)$, we have for the corresponding equivalent saturated diode current I'_e (compare (5)):

$$I'_e = \frac{V_3 - V_0}{V_4 - V_3} I_1. \tag{9}$$

As I'_e is the equivalent saturated diode current of the parallel connection of R_1 and R, while I_e is the equivalent saturated diode current of R, we have for the equivalent saturated diode current I''_e of R_1:

$$I''_e = I'_e - I_e = \frac{V_3 - V_0}{V_4 - V_0} I_1 - \frac{V_1 - V_0}{V_2 - V_1} I_0. \tag{10}$$

Hence we obtain for the noise ratio α_1 of R_1:

$$\alpha_1 = 20\, I''_e R_1 \tag{11}$$

(compare equation (6)).

This method was applied for the measurement of the noise ratio of the internal resistance of a diode at 7-m wavelength in the cut-off region [5]).

The method described in sections 2 and 3 is only valid if the output of the thermocouple is due to noise components that have to be added quadratically. Two noise components have to be added quadratically under the following conditions:

a. *The components are independent*

The output noise voltage is chiefly due to the noise voltage across the input circuit and to the noise of the first valve. For a linear amplifier at 50-m wavelength the noise voltage across the input circuit is due to thermal noise which is independent of the noise of the first valve. At 7-m wavelength, however, the situation is the opposite, the noise voltage across the input circuit being to a large extent due to induced grid noise which is correlated to that part of the noise of the first valve which is due to shot effect.

b. *The components are correlated, but their phase difference is about 90 degrees*

This is the case for a linear amplifier tuned at 7-m wavelength where the induced grid noise and the shot effect of the first valve have a phase difference of about 90 degrees.

4. *Measurement of noise factors of receiver stages*

In a similar way noise factors of receivers may be measured. Let the antenna be replaced by a dummy having the same impedance. For a tuned half-wave dipole antenna this may consist of a resistance R_a of about 70 Ω (*fig. 4a*). Let a saturated diode be connected in parallel to R_a,

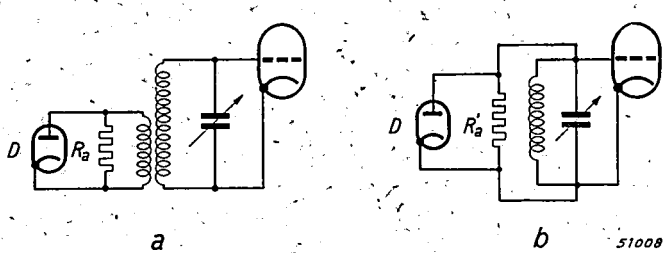


Fig. 4. *a*) Saturated diode and dummy-antenna resistance R_a connected across the input terminals of a receiver in order to measure its noise factor F.
b) Saturated diode D and transformed antenna resistance R_a' connected across the input circuit of a receiver in order to measure the noise factor F as a function of R_a'.

and let I_e be the anode current of the saturated diode for which the noise output power of the amplifier is doubled.

We then put:

$$(2eI_e \Delta\nu)\, R_a^2 = F.4kTR_a \Delta\nu,$$

and obtain in accordance with equation (6):

$$F = \frac{e}{2kT} I_e R_a = 20\, I_e R_a. \tag{12}$$

The factor F is usually called the *noise factor* of the receiver [1]). Its physical meaning is that the noise of receiver + antenna is such as if the receiver were noiseless and the antenna resistance R_a had the "equivalent noise temperature" FT. As T is the noise temperature of the dummy antenna, the part

$$a = (F-1) \tag{13}$$

is due to the receiver.

In this determination of the noise factor no account had to be taken of the correlation between the various noise components; neither had anything to be assumed about the bandwidth of the amplifier.

In many cases it is important to measure the noise factor as a function of the antenna coupling, for a receiver is often designed in such a way that the antenna is matched to the input of the receiver, but this does not necessarily mean that the noise factor then has a minimum value.

Hence it is necessary to measure the noise factor as a function of the transformed antenna resistance. For a given value of R_a the corresponding value of the transformed antenna resistance might be varied by variation of the transformation ratio, but it is much easier to replace the transformed antenna resistance by a carbon resistor *) R_a' connected across the input circuit with the saturated diode in parallel to it (*fig. 4b*). In this way the noise factor F may be determined as a function of the transformed antenna resistance R_a' in the same way as before, and from that its minimum value may be found. On the other hand the signal output of the receiver is a maximum when R_a' is chosen equal to the input impedance R_e of the receiver.

The input impedance R_e of the receiver may be measured in the following way. For $R_a' = \infty$ a saturated current I_e' gives rise to a change ΔV in the output of the thermocouple whilst for a given finite value of R_a' a saturated current I_e'' may give rise to the same change ΔV in the output.

*) For carbon resistors between 50 Ω and 10 000 Ω the resistance at 7-m wavelength is practically equal to its D.C. value.

We then have:

$$I'_e R_e^2 = I''_e \left(\frac{R_e R'_a}{R_e + R'_a} \right)^2, \qquad (14)$$

from which R_e may be calculated.

So far we have only mentioned cases in which noise factors of complete receivers were measured, but it is often necessary to measure the noise factor of a single stage. In that case the stage is connected between the dummy antenna and the noise amplifier described previously, and it is measured what part of the noise output is due to the stage itself (input circuit + valve), and what part is due to the noise amplifier. This is carried out (*fig. 5*) by using a diode noise generator D_1 across the input of the stage and a second diode noise generator D_2 across the input of the amplifier. We assume a transformed antenna resistance R'_a to be connected across the input of the stage.

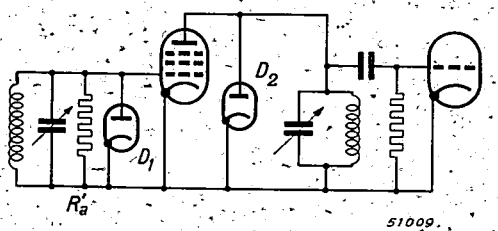


Fig. 5. Measurement of the noise factor of a single amplifier stage. A saturated diode D_1 and the transformed antenna resistance R_a' are connected across the input circuit of the stage to be measured, and a saturated diode D_2 is connected across the output circuit of the stage, in order to make a correction for the noise of the subsequent amplifier.

We first measure an output V_0 when the input of the noise amplifier is short-circuited (this part is chiefly due to the first valve of the amplifier). Next we remove the short-circuiting lead and connect the stage to the noise amplifier without applying any voltages to the various electrodes of the valve; we tune and measure an output V_1, and furthermore we observe that an anode current I_1 of D_2 gives rise to a change in output of ΔV. We then apply the necessary voltages to the various electrodes of the valve, tune the input circuits of both the stage and the amplifier, and measure an output V_2 whilst now an anode current I_2 of D_2 gives rise to a change in output of ΔV. Obviously the part

$$V = V_2 - V_0 - \frac{I_1}{I_2}(V_1 - V_0) \qquad (15)$$

of the output of the amplifier is due to the stage (input circuit + valve).

We finally measure that an anode current I_3 of D_1 gives rise to a change in output of ΔV. Hence we obtain for the equivalent diode current I_e of the amplifier stage (antenna + input circuit + valve):

$$I_e = \frac{I_3}{\Delta V}\left\{V_2 - V_0 - \frac{I_1}{I_2}(V_1 - V_0)\right\}, \tag{16}$$

and for the noise factor F of the stage:

$$F = (\alpha + 1) = 20\, I_e\, R'_a\,, \tag{17}$$

whilst

$$\alpha = (20\, I_e\, R'_a - 1) \tag{17a}$$

is the measure of the contribution of the receiver stage itself to the noise factor.

In many cases V_0 and V_1 are so small as to be negligible; the measurement then becomes equivalent to the one described in the first part of this section. Sometimes, however, the above more elaborate method has to be used because V_1 or even V_0 are too large to be ignored.

Our method is independent of the matching or the mismatching between the output of the stage and the input of the noise amplifier. Moreover, the gain of the receiver stage does not enter into our calculations explicitly (though the gain might be calculated with the help of our measurements).

Eindhoven, July 1947

REFERENCES

1) H. T. Friis, Proc. Inst. Radio Engrs **32**, 419-422, 1944.
2) P. H. Miller, Proc. Inst. Radio Engrs **35**, 252-256, 1947.
3) J. M. Pettit, Proc. Inst. Radio Engrs **35**, 257-265, 1947.
4) S. Roberts, Proc. Inst. Radio Engrs **35**, 302-306, 1947.
5) A. van der Ziel and A. Versnel, Nature, May 10, 1947.
6) C. J. Bakker, Physica **8**, 23-43, 1941.

Philips Res. Rep. **2**, 331-339, 1947

REFLECTIONS IN ELECTRON TUBES

by J. L. H. JONKER

537.543:621.385

Summary

The characteristics of some electron tubes show irregularities as a result of reflected electrons. By studying the peculiarities of these reflections at low potentials, it is possible to deduce their influence on the electric field between the electrodes, and in this way to explain the irregularities.

1. *Introduction*

In the characteristics of some electron tubes irregularities occur as a result of fast reflected electrons. Various data can be found in the literature dealing with this subject. Van der Pol and Weyers [1]) examined the characteristics of diodes by measuring the higher derivatives, and they found a considerable deviation for diodes with anode voltages in the neighbourhood of 10 volts (*fig. 1*). This becomes apparent from the passing

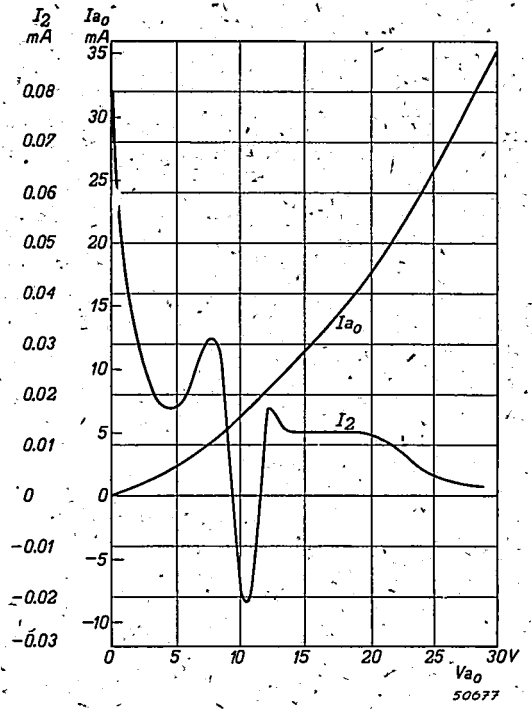


Fig. 1. I_a-V_a-characteristic of a diode (Ia_0) and curve of second derivative (I_2) according to Van der Pol and Weyers.

through zero of the second-order derivative which indicates two points of inflection in the characteristic. The authors mentioned above ascribe this phenomenon to reflected electrons.

Kleen [2]) also found irregularities in the characteristics of a diode at approximately 10 volts. He ascribes these to an interchange of electrons between the anode and the screen plates which he used in his diode in order to maintain a linear electric field. It is not quite clear how this effect could possibly explain the irregularities. In addition, this explanation is not conclusive because, according to our own observations, the said irregularities also occur in diodes without screen plates, as can be seen in the characteristic reproduced in *fig. 2*. This was measured with a tube

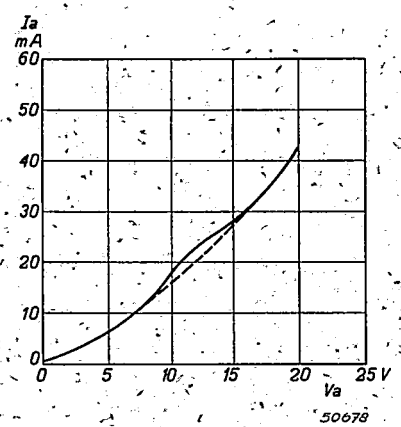


Fig. 2. I_a-V_a-characteristic of a diode. Irregularities are caused by reflected and secondary electrons from the anode.

consisting of an indirectly heated oxide cathode surrounded by a cylindrical nickel anode. The remarkable thing about this irregularity is that the characteristic shows a local *increase of current*. If reflected electrons would appear at the corresponding potential (approximately 10 volts), then in the space between cathode and anode an increased space charge would occur which would result in a *decrease of current*.

Thompson and others [3]) found that the noise in a diode exceeds the amount that can be expected from fluctuation in space charge, which they contribute to reflected electrons.

Irregularities also occur with multi-grid tubes, as Van der Pol and Weyers found by measuring the higher derivatives. We also found deviations in the anode-current anode-voltage characteristics of penthodes by direct measurements of the characteristics at low potentials, as can be seen in *fig. 3*. In this case the filament voltage V_f was kept low so as to reduce the influence of space charge; see also [4]).

Generally the suppressor grid in a penthode is so constructed that the potential minimum which this grid must cause between the anode and the screen grid, practically disappears at low anode potentials. In that case the anode potential becomes equal to or lower than the average potential in the suppressor grid [5]). In this case all secondary electrons from the anode,

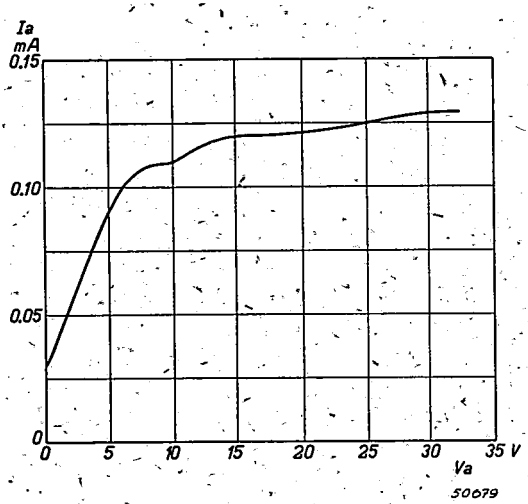


Fig. 3. I_a-V_a-characteristic of a penthode. Irregularities are caused by reflected as well as by real secondary electrons.

the slow (the so-called real secondary electrons) as well as the fast (reflected electrons), can pass through the suppressor grid and arrive at the screen grid, thus causing undesired irregularities in the characteristics.

The above phenomena can be explained by means of the characteristics of real secondary and reflected electrons at low potentials, of which a short survey may follow.

2. *Reflection of electrons by solids at low potentials*

The method mostly used for measuring these reflections is that of the central field in which a small plate of the material to be examined is placed at the centre of a spherical electrode. If the sphere is slightly positive with regard to the plate which is bombarded by electrons, all secondary electrons will arrive at the sphere. If a negative potential is applied the velocity distribution of the secondary electrons can be examined [6]).

Now we shall first describe for different materials how the ratio

$$\delta = \frac{\text{number of secondary electrons}}{\text{number of primary electrons}}$$

varies at low potentials V_p of the plate. The potential of the sphere is kept at such a value that all secondary electrons reach the sphere.

In *fig. 4* this is shown for the metals Ni, Cu, and Pt, and it appears that, at the potentials used, the curves do not run smoothly but show irregularities. These irregularities can be compared with those of light reflections in the case of $h\nu$ being of the order of some electron volts; they are caused by the fact that electrons can only dissipate definite quantities of energy on account of quantal processes [7]).

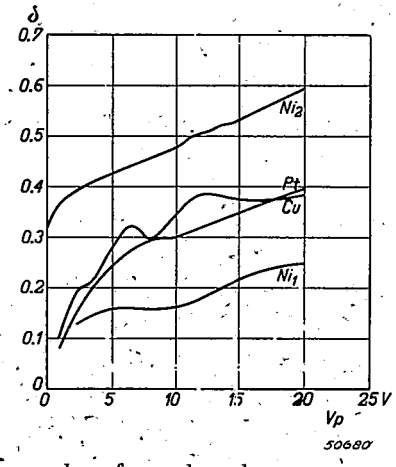


Fig. 4. $\delta = \dfrac{\text{number of secondary electrons}}{\text{number of primary electrons}} = f(V_p)$.

for pure Cu, Pt, Ni (Ni_1), and for Ni coated with gas (Ni_2) (H. F. Farnsworth, Phys. Rev. **20**, 358-374, 1922 and **25**, 41-57, 1925).

In technical electron tubes the metal surfaces are seldom clean, so that we also have to study the behaviour of metals that are covered with a thin film of gas (see fig. 4) or of salts or oxides. When an anode is placed opposite an oxide cathode, O_2, Ba, and BaO will evaporate, and in

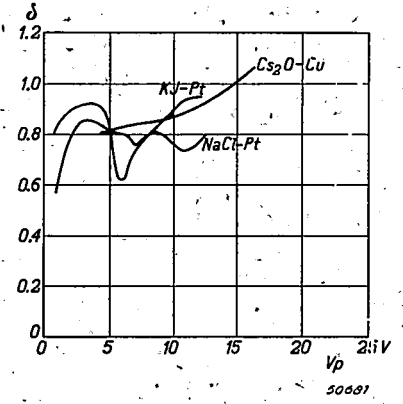
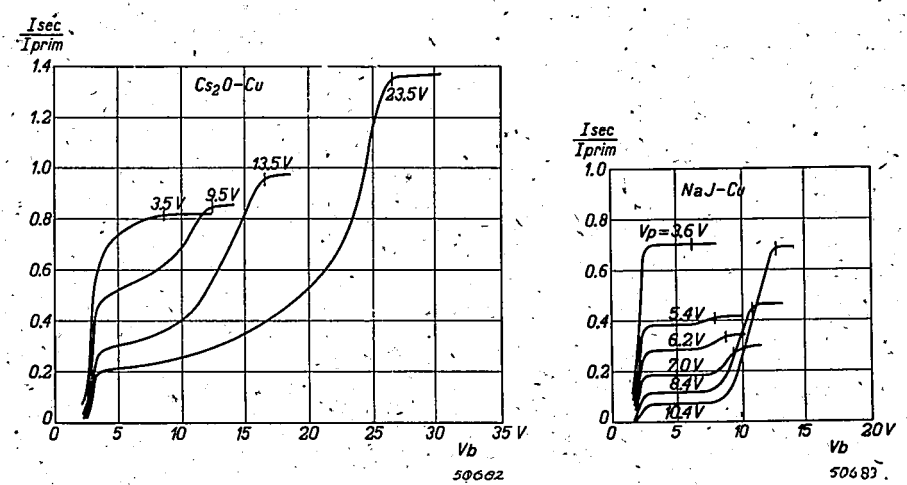


Fig. 5. $\delta = f(V_p)$ for KJ, NaCl, and Cs_2O, deposited on Pt, Pt, and Cu, respectively.

addition the electrode may be soiled by a Ba- or Mg-getter. As an example *fig. 5* shows $\delta = f(V_p)$ for different metals on which a thin film is deposited [8]). It appears that under these circumstances the irregularities may exceed those of pure metals.

By applying a central counterfield it is possible to obtain a picture of the velocity distribution of secondary electrons at low potentials. For this purpose the potential of the spherical electrode V_b is varied while the



Figs 6 and 7. Bundle of countervoltage curves I sec/I prim $= f(V_b)$ with the potential V_p of the bombarded plate as a parameter.

potential of the plate V_p is kept constant. In this way we arrived at the countervoltage curves of *figs 6* and *7* [8]) from which the energy distribution can be derived. At the vertical dashes the potentials of the sphere and the plate are equal (with corrections for the contact potential). The slope of the curve at a given voltage is a measure for the number of secondary electrons that have a velocity corresponding with this voltage (the contact potential is taken into account). It appears that in the case of salts, at very low potentials, practically only reflected electrons occur; at slightly higher values of V_p the absorption of primary electrons increases and slow secondary electrons appear. In most cases the picture is less clear than in fig. 7. With pure metals the transition is more gradual.

The above measurements enable us to determine the number of fast reflected electrons as a function of the primary voltage V_p (full-drawn curves in *fig. 8*). It appears that the number of reflected electrons decreases rapidly with increasing potential. When the energy is sufficiently large the primary electrons can lose energy by lifting an electron from a full band into an empty band [9]). From there this electron can escape from the crystal. This explains why in certain cases (salts) the following can be

clearly observed: at the lowest voltages practically all primary electrons are reflected without loss of velocity; at a certain value of the primary voltage the absorption of the primary electrons increases considerably so that the number of reflected electrons decreases, and the real secondary emission commences (dashed curves in fig. 8). For instance, in the case of certain salts, absorption of primary electrons starts sharply in the region of 5 to 10 volts where also absorption of ultra-violet light starts corresponding with a wavelength of 250-100 mμ, as can be shown by comparing data from the literature [7]).

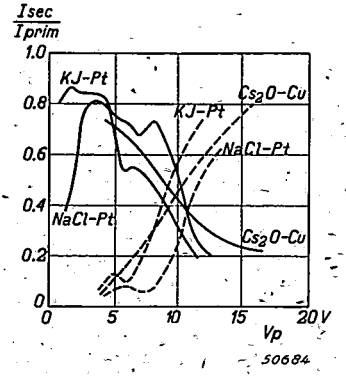


Fig. 8. Slow secondary electrons (dashed curves) and fast reflected electrons (full-drawn curves) as a function of the primary voltage V_p.

As a matter of fact the behaviour of metals is less clear, but even there we find a corresponding increase in light absorption and a rapid decrease of the number of reflected electrons.

3. *Diode characteristics*

If the secondary electrons from the anode of a diode leave the anode with a certain velocity, the electrons proceed in the discharge space and cause an increase in space charge. The electric field between cathode and anode will vary and, on account of that, the potential minimum in the neighbourhood of the cathode will increase, with the result that a larger proportion of the electrons emitted by the cathode will return to the cathode so that the cathode current decreases. This effect will be more noticeable when the initial velocity of the secondary electron is higher so that they penetrate closer to the cathode.

In order to obtain a clearer picture of this phenomenon, we represent in *fig. 9* the imaginary case of a flat diode in which the initial velocity of all secondary electrons is equal and at right angles to the surface of the anode. Little is known about the actual distribution over various directions of secondary electrons. It is generally assumed that they leave the material approximately according to the cosine law.

As an example we take a secondary-emission factor of 2, and the initial energy of the secondary electrons in the direction of the cathode successively 0, $\frac{1}{4}$, $\frac{1}{2}$, $\frac{3}{4}$, and 1-times the energy of the primary electrons at the anode. After having proceeded towards the cathode over a certain distance, depending on their initial energy, the secondary electrons will turn back and return to the anode. According to Poisson's equation, $d^2V/dx^2 = 4\pi\varrho$, we can calculate the space-charge density at every point of the space between anode and cathode. By numerical integration we can compute the potential curves of fig. 9. The anode current decreases and, for the above-mentioned energies of the secondary electrons, will amount to I, 0·872 I, 0·689 I, 0·478 I, and 0·2 I, respectively. From this we can conclude that electrons returning from the anode with high energy contribute considerably more to the reduction of the cathode current than those with less energy.

When the anode consists of Ni partially coated with BaO, with an anode voltage of a few volts the returning electrons will practically all be fast reflected electrons which can cause a considerable decrease of anode current. With higher anode voltages, when absorption sets in, the number of reflected electrons will decrease rapidly while real secondary electrons are not yet emitted. The reduction of anode current decreases which explains the lower portion of the deviation of the diode characteristic. When

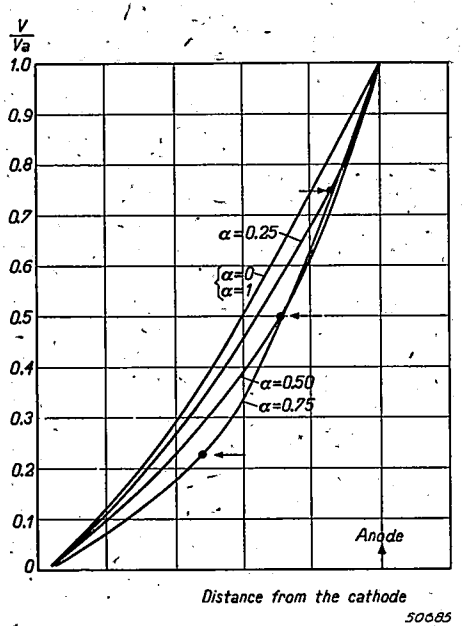


Fig. 9. Potential curves in a flat diode. The coefficient of secondary emission δ of the anode is 2. The velocities of the secondary electrons are assumed at right angles to the surface, and their initial energy is $\alpha = 0, {}^1/_4, {}^1/_2, {}^3/_4$, and 1-times the energy of the primary electrons at the anode. On each curve the returning point of the secondary electrons is marked.

the anode voltage is sufficiently raised, the number of real secondary electrons increases rapidly which again causes a decline of the characteristic. There exists therefore an upward deviation of the diode characteristic which is actually caused by a reduction of the influence of the returning electrons, in which case the fast electrons are of considerably more importance.

With regard to the experimental curve of fig. 2, the calculated curve (disregarding secondary emission) lies higher, such in contradiction with Kleen's supposition. This explanation is also confirmed by the fact that the additional noise caused by the secondary emission shows a minimum in the region of the deviation [10]).

From this it is obvious that tubes showing a tendency towards this effect can be improved by cleaning the anode with extreme heat so that the secondary emission of pure nickel comes instead of that of the soiled. As metals show a more gradual transition to the effect of secondary emission, it is to be expected that the deviation of the characteristic is only slight or altogether absent. In effect, the hump in the curve at approximately 10 volts decreases but does not disappear altogether. It appears that even very small traces of impurity are sufficient to cause the effect. With the use of oxide cathodes it cannot be avoided completely. Only by using a W-filament and a Ni-anode without applying any getter at all, did the irregularities in the characteristic disappear.

4. *The penthode characteristic*

In section 1 it was mentioned that the I_a-V_a-characteristic of a penthode also showed irregularities. In the region of the very low anode potentials where the potentials are lower than the average potential of the suppressor grid, the electric field near the anode is so directed that all secondary electrons of the anode reach the screen grid. It is clear that the irregularities in the curve $\delta = f(V_p)$ (figs 4 and 5) in this voltage region manifest themselves in the anode current (fig. 3).

When the anode voltage rises to a certain value, it will become higher than the average potential of the suppressor grid. In that case the field in front of the anode will be so directed that only those secondary electrons which have a sufficient velocity in the direction of the screen grid will reach this grid. From the countervoltage curves of fig. 6 we can derive what proportion of the total number of secondary electrons will have a sufficient velocity. Of course, the distribution over various directions (cosine law) must be taken into account.

For a given configuration of the electrodes in a penthode *fig. 10* shows to what extent the anode current, which is otherwise supposed to be constant, varies on account of electrons returning to the screen grid. The curve is calculated with the aid of countervoltage curves, as in fig. 6,

taking the anode as if made of pure Ni. It appears that the reflected electrons which can pass through the suppressor grid, are still apparent in the I_a-V_a-characteristic even at higher potentials. Their irregular behaviour

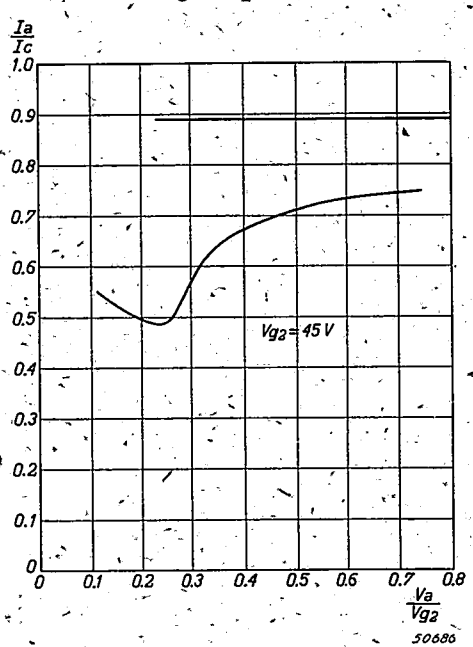


Fig. 10. Calculated I_a-V_a-characteristic of a penthode; for determining the current distribution between screen grid and anode only the secondary electrons are taken into account (I_c = cathode current, V_{g2} = screen-grid voltage).

can also be noticed as they cause an increased screen-grid current and a decreased internal resistance of the tube. Furthermore, it appears that the transition of the real secondary electrons and the reflected electrons to the screen grid at lower voltages causes a bend in the characteristic which contributes to the well-known knee or upper bend in this important characteristic of the penthode.

Eindhoven, February 1947

REFERENCES

1) Balth. van der Pol and Th. J. Weyers, Physica, 's-Grav. **1**, 481-496, 1934.
2) W. Kleen, Telef. Röhre, Heft 9, 1937.
3) B. J. Thompson, D. O. North and W. A. Harris, R.C.A. Rev. 4, 269-285, 1940.
4) E. Schulże, Hochfrequenztech. Elektroak. **44**, 118-125, 1934.
5) J. L. H. Jonker, Wireless Engr **16**, 274-286 and 344-349, 1939.
6) H. Bruining and J. H. de Boer, Physica, 's-Grav. **5**, 17-30, 1938.
7) R. Hilsch, Z. Phys. **77**, 427-436, 1932.
8) O. Krenzien, Wiss. Veröff. Siemens Werke, **20 (2)**, 91-108, 1942.
9) H. Bruining and J. H. de Boer, Physica, 's-Grav. **6**, 834-839, 1939.
10) According to measurements of Dr C. J. Bakker of this laboratory.

R 55 Philips Res. Rep. 2, 340-348, 1947

THE TEMPERATURE DEPENDENCE OF THE FLUORESCENCE OF TUNGSTATES AND MOLYBDATES IN RELATION TO THE PERFECTION OF THE LATTICE

by F. A. KRÖGER

535.37

Summary

The quenching temperature of the photoluminescence of tungstates and molybdates is dependent on the perfection of the crystals, imperfect crystals having a low quenching point. This property may be explained in terms of the quenching theories of Mott, and Seitz, as well as in terms of the theories of Peierls, Frenkel, and Möglich and Rompe.

1. *Introduction*

Among the factors that determine whether a solid shows photoluminescence or not, the state of crystallization is of primary importance. Tiede and Schleede[1]) have shown that calcium tungstate, when made by precipitation from a solution, is obtained in the form of extremely small, rather imperfect crystals which are practically not luminescent but which become so upon recrystallization. It has since been recognized, however, that the crystalline state is not a general condition for the appearance of photoluminescence. Glasses, for instance, have been developed that show luminescence with a fairly high efficiency. Therefore there must be a special reason why systems like the calcium tungstate mentioned are non-luminescent in the semi-amorphous state. Since the absorption spectrum of the system is not markedly changed upon recrystallization, the phenomenon cannot be due to a decreased absorption of the exciting radiation, but it must be connected with the dissipation of the absorbed energy. In investigations on the temperature dependence of luminescence of a large number of systems it has been found that the dissipation generally increases at higher temperatures. Therefore it was expected that the badly crystallized tungstates and molybdates might also show luminescence, though at a lower temperature than the well-crystallized products. In order to check this view $CaWO_4$, $CdWO_4$, $MgWO_4$, and $CaMoO_4$ have been prepared in various degrees of crystallization, and the temperature dependence of luminescence has been studied.

2. *Method of preparation* *)

$CaWO_4$, $CaMoO_4$, and $CdWO_4$ were made by precipitation from aqueous

*) The preparations have been carried out by Mr A. Bosman.

solutions of calcium and cadmium nitrate by means of ammonium tungstate or molybdate. $CaWO_4$ was precipitated in the presence of an excess of ammonium ions to prevent the formation of Ca_3WO_6 [3]). The precipitates were filtrated, thoroughly washed with hot water, and dried at 80 °C *in vacuo*. Fractions of these products were heated in air for one hour at different temperatures below 1200 °C. The heated cadmium tungstates were slightly coloured by cadmium oxide, but this colour could be removed by washing with dilute acetic acid. Since $MgWO_4$ is highly soluble in water, this substance was made by a semi-dry method. MgO was mixed with the right amount of H_2WO_4 together with a small quantity of water. The mixture was dried on a water bath and subsequently heated at temperatures above 700 °C.

3. *Crystal structures*

The crystal structures were checked with the aid of X-ray diffraction photographs.

$CaWO_4$ was precipitated in the scheelite structure and did not change its structure upon heating. The X-ray patterns, however, grew gradually sharper in going from the precipitated products to the products heated at increasing temperatures. Besides, lines of higher orders, which were hardly visible on the diffraction photographs of the precipitate and of the products heated below 400 °C, appeared in the patterns after heating at higher temperatures. These effects indicate an increase both of the size and of the perfection of the crystals.

$CaMoO_4$ behaves similarly, the scheelite structure being found for all products.

For $CdWO_4$ the structure of the precipitate differs from that of the products heated at 300 °C or higher. Probably the tungstate is precipitated either in an unstable form, or as a hydrate. The effect of heating at 300 or 400 °C is the transformation of the precipitate into the normal structure; at higher temperatures, up to 1100 °C, the crystal size increases and the crystals grow more perfect. Heating at still higher temperatures causes a slight decomposition as indicated by a brownish colour of the crystals. Heating with addition of fluxes, *e.g.* $CdCl_2$, produces crystallization effects at temperatures that are much lower than those required to obtain the same effect without fluxes.

$MgWO_4$ prepared at 700 or 800 °C mainly showed the X-ray pattern of the high-temperature α-modification, with a faint indication of the normal fluorescent β-form. Probably the unstable α-modification is formed as a consequence of Ostwald's law of successive states. The product made at 900 °C was a mixture of nearly equal amounts of α and β-$MgWO_4$, while a product made at 1050 °C was pure β. A product heated at 1320 °C was

pure α, but could be transformed into the β-form by subsequent heating below the transition point (1250 °C), *viz.* at 1020 °C.

4. *Fluorescence*

Although at room temperature only the products heated at a comparatively high temperature show photoluminescence, all products do so at low temperatures. The spectral distributions of the emissions at — 180 °C are shown in *figs 1, 2, 3* and *4*.

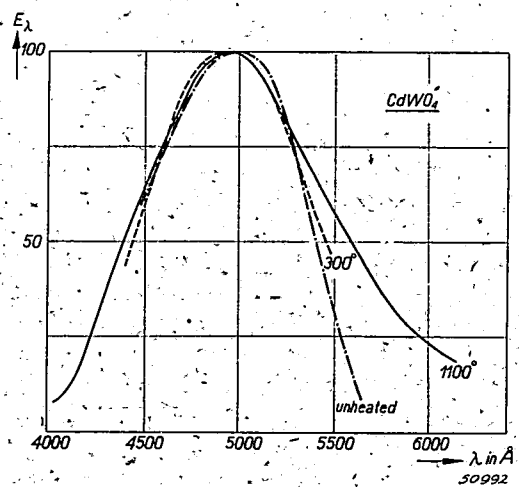


Fig. 1. Spectral distribution of the fluorescence of cadmium tungstate, made at different temperatures, for excitation by $\lambda = 2537$ Å at — 180 °C.

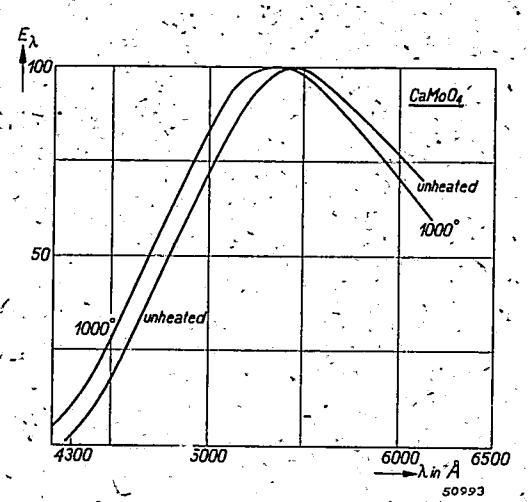


Fig. 2. Spectral distribution of the fluorescence of calcium molybdate, made at different temperatures, for excitation by $\lambda = 2537$ Å at — 180 °C.

For $CdWO_4$ (fig. 1) the unheated precipitate and the heated products which — as we have seen — have a different crystal structure, show fluorescence of nearly the same spectral distribution.

For $CaMoO_4$, where no change of structure occurs, the spectral distributions of all products are also nearly identical.

Although $MgWO_4$ occurs in different structures, α and β, all products have the same spectral distribution of the emission which is identical with that of normal fluorescent β-$MgWO_4$. Since the α-form is known to show

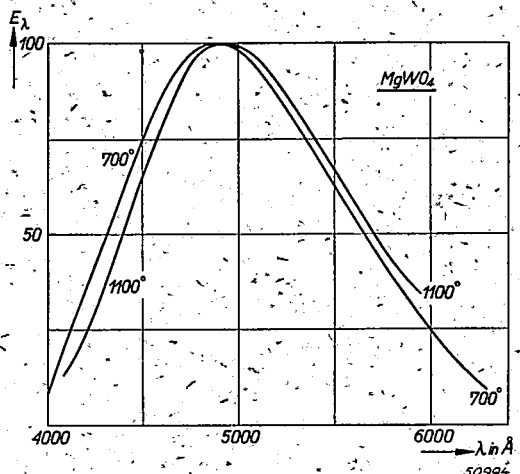


Fig. 3. Spectral distribution of the fluorescence of magnesium tungstate, made at different temperatures, for excitation by $\lambda = 2537$ Å at -180 °C.

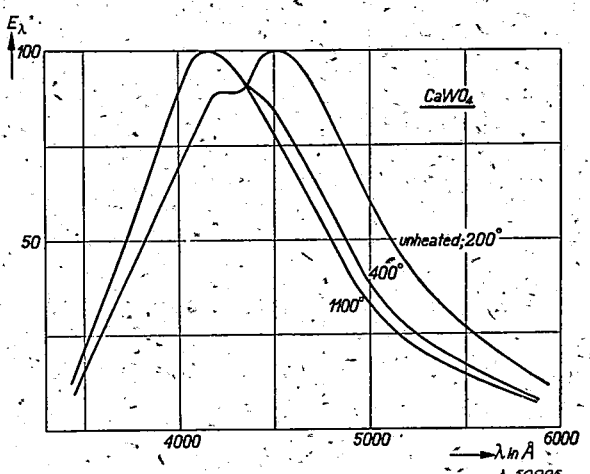


Fig. 4. Spectral distribution of the fluorescence of calcium tungstate, made at different temperatures, for excitation by $\lambda = 2537$ Å at -180 °C.

an extremely faint blue luminescence at — 180 °C [2]), the blue-green fluorescence of all products must be attributed to β-$MgWO_4$.

$CaWO_4$ provides the only example of a system in which the spectral distribution of products heated at different temperatures is markedly different, though the crystal structure is the same. The unheated precipitate has a fluorescence with a maximum at 4500 Å, the products heated at 1100 °C have a maximum at 4200 Å, while the products heated at intermediate temperatures show the two bands together in varying proportions (fig. 4).

The temperature dependence of fluorescence for the various products is shown in *figs 5, 6, 7* and *8*, where curves have been reduced to such a scale that their maxima are equal to 100. The curves show the same character for all products: the fluorescence has a constant maximum intensity over a range of low temperatures, but above a certain temperature it decreases rapidly with increasing temperature. The temperature range in which the quenching occurs ("quenching range") is in all cases shifted towards higher temperatures with increased crystallization. In most cases the highest quenching temperatures are found with the products made at the highest temperatures. In one case, however, namely that of $CdWO_4$, heating at a temperature above 1100 °C caused the quenching range to shift back to lower temperatures.

With $CaWO_4$ different results have been obtained, dependent on the method of precipitation. When the precipitation is carried out with extremely dilute solutions, scheelite crystals of intermediate size and perfection are formed, which show the fluorescence described above. Fig. 5 (and also table I) refer to a sample made in this way. Precipitation from concentrated solutions, however, produces much smaller crystals, which are not luminescent, even at — 180 °C. Probably the quenching point of the

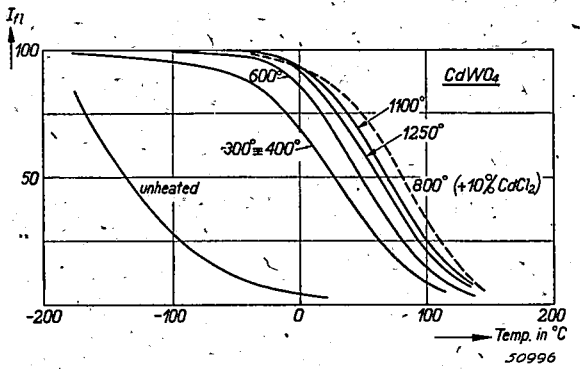


Fig. 5. Temperature dependence of the fluorescence for cadmium tungstate, made at different temperatures.

fluorescence of the latter precipitate lies below —180 °C. Fluorescence appears with increasing crystallization in the same way as with the other precipitate.

Table I gives the maximum intensities for $CaWO_4$ and $MgWO_4$ in their true relative value. For $CaWO_4$ the intensity of the badly crystallized products is of the same order as that of well-crystallized products. The low value for the sample heated at 800 °C is due to a colouring occurring particularly at this temperature. For $MgWO_4$ the intensity of the products prepared at low temperatures is considerably lower than of those made at higher temperatures. This is probably due to absorption of part of the exciting ultra violet by α-$MgWO_4$ present in the former samples.

5. *Discussion*

In the four systems that have been investigated the temperature range in which the fluorescence is quenched is located at high temperatures for well-crystallized products, but at lower temperatures for badly crystallized products. Crystallization consists in two effects: (1) the grain size is increased, and (2) the structure grows more perfect. *A priori* both effects might be held responsible for the variation of the quenching range. Two experimental results provide evidence in favour of the view that it is the perfection rather than the size of the crystals that is the most important factor. In the first place we observed a decrease of the quenching range for $CdWO_4$ by the increase of the heating temperature from 1100 to 1250 °C. In this case the grain size certainly increased but the perfection decreased, as is indicated by the colouring of the crystals. In the second place we found that the quenching curves for $MgWO_4$ and $CaWO_4$ were not affected by ball-milling of the products. This treatment reduces the grain size and may, if the properties of the crystals allow this, also cause deformation. Since we did not find any variation of the quenching point, the reduction in size which certainly occurs has apparently no effect on the temperature quenching *).

Generally heating at a high temperature increases the perfection of the crystals. As is shown by figs 5-8, the quenching curves for products made by heating at various temperatures converge towards a limit. Since normally methods of preparation are used that lead to fairly-well crystallized products, this convergence explains why different authors have observed the same quenching range. The quenching range is more or less

*) This observation proves further that application of unidirectional pressure on these tungstates breaks the crystals into pieces rather than deforms them. Crystals of other, softer substances like ZnO, ZnS and the alkaline and earth-alkaline sulphides and oxides behave differently. These substances are coloured by milling or grinding and their luminescence is killed (pressure quenching, german: "*Druckzerstörung*").

characteristic of a certain emission and may be useful as a means of identification of emission bands in cases in which the spectral distributions are identical *).

TABLE I

Maximum intensities of fluorescence

Substance	Precipitated unheated	Heating temperatures							
		200	300	600	700	800	900	1100	1320; 1250
$CaWO_4$	59	74	97·5	87·5	—	53	—	100	—
$MgWO_4$	—	—	—	—	26	31	61	91	100

Although cooling always causes the appearance of luminescence, the maximum efficiency generally remains below that of well-crystallized products. Apart from the trivial cases in which different crystal forms are present (*e.g.* α and β-$MgWO_4$), the cause of this decrease must probably be sought in an ineffective absorption connected with the same disturbances that are responsible for the decrease of the quenching point.

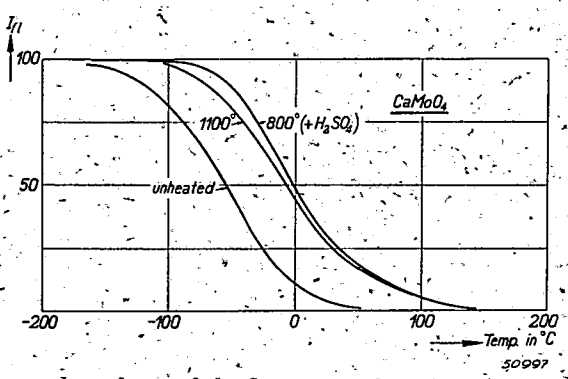


Fig. 6. Temperature dependence of the fluorescence for calcium molybdate, made at different temperatures.

To explain the remarkable influence of lattice disturbances on the quenching range we have to extend the present theories for the temperature quenching. The systems under discussion are systems in which the processes of absorption of ultra-violet radiation and the emission of fluorescence are closely related, the electronic transitions connected with these two phenomena being in principle the same, though in opposite directions. In con-

*) This method has been used in the systems $(Ca, Pb)WO_4$, $(Sr, Pb)WO_4$, $(Ba, Pb)WO_4$ [2]).

nection with this, the efficiency η can be expressed by the simple relation:

$$\eta = \frac{K_{\text{lum}}}{K_{\text{lum}} + K_{\text{diss}}} = \frac{1}{1 + \frac{K_{\text{diss}}}{K_{\text{lum}}}}$$

in which K_{lum} and K_{diss} are the probabilities of the processes in which the excited electrons return to the ground state with the emission of light, or with dissipation of the excitation energy. This formula can account for a temperature dependence of luminescence as found, if K_{lum} varies only slightly with the temperature, while K_{diss} increases rapidly with the temperature.

The dissipation process has been treated theoretically in two different ways. One group of theories, that of Peierls, Frenkel, Möglich and Rompe, assumes that the excitation energy is converted in one process into a large number of heat vibrations of the whole crystal [4] [5] [6]. Another theory, that of Mott and Seitz, has sought the cause of the

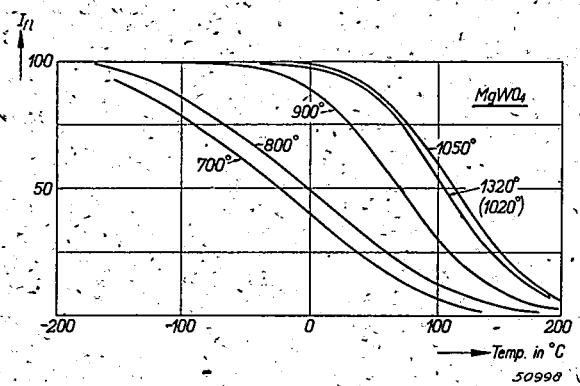


Fig. 7. Temperature dependence of the fluorescence for magnesium tungstate, made at different temperatures.

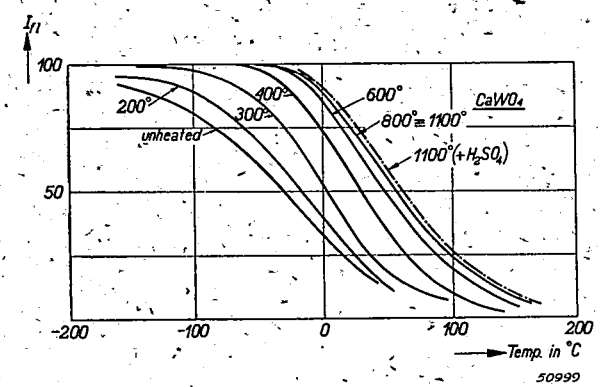


Fig. 8. Temperature dependence of the fluorescence for calcium tungstate, made at different temperatures.

temperature dependence of the dissipation process in an activation energy hindering the transition of a centre and its nearest surroundings from the electronic excited state to a high vibrational state of the electronic ground state. A third theory for the dissipation, proposed by Klasens [8]), needs not be considered here, since it only applies to cases in which excitation separates electrons from their "holes"; which is probably not the case in tungstates and molybdates.

In the frame-work of the first theory the disturbances must be supposed to influence the properties of the lattice as a whole, in the same way as has been assumed for disturbances caused by high activator concentrations [9]). In terms of the Mott-Seitz model, the disturbances must be assumed to affect the properties of separate centres. In the simplest case the temperature-dependence curve observed is a superposition of the curves of disturbed and undisturbed centres. It may also occur, however, that the excitons, travelling through the lattice, are preferentially annihilated in the disturbed centres. In that case a comparatively small amount of disturbed sites may cause a considerable shift of the quenching towards lower temperatures. This model may also account for the variation in spectral distribution of the fluorescence as observed with $CaWO_4$: changed surroundings cause not only a different vibrational interaction between the centre and these surroundings, but also a variation of the energy levels or the transition probabilities in the centre itself. This effect cannot be easily accounted for on the basis of the first of the two theories mentioned above. Since the variation of the emission spectrum, however, occurs only in one out of the four cases investigated, we are not certain whether this is an essential point. Therefore a definite choice between these two theories cannot be made on the basis of the observations described in this paper.

Eindhoven, January 1947

REFERENCES

1) E. Tiede and A. Schleede, Z. Elektrochem. **29**, 304-312, 1923.
2) F. A. Kröger, Some Aspects of the Luminescence in Solids, Amsterdam 1947, in the press, Chapter IV.
3) Riza Berkum, C. R. Paris, **210**, 731-733, 1940.
4) R. Peierls, Ann. Phys., Lpz. **13**, 905-952, 1932.
5) J. Frenkel, Phys. Z. Sowjet Union **9**, 158-186, 1936.
6) F. Möglich and R. Rompe, Z. Phys. **115**, 707-728, 1940; **117**, 119-124, 1940/41. Naturw. **29**, 105-113, 129-134, 1941. Phys. Z. **41**, 236-242, 1940.
7) N. F. Mott and R. W. Gurney, Electronic Processes in Ionic Crystals, Oxford 1940, p. 218.
F. Seitz, Modern Theory of Solids, New-York—London 1940, p. 450.
The same is also to be found in Trans. Faraday Soc. **35**, 69-73, 74-85, 1939; especially pp. 71 and 79.
8) H. A. Klasens, Nature **158**, 306-307, 1946.
9) K. Birus, F. Möglich and R. Rompe, Phys. Z. **44**, 122-129, 1943.
M. Schön, Forsch. Fortschritte **19**, 4-9, 1943.

Philips Res. Rep. 2, 349-351, 1947

SOME CHARACTERISTICS OF TRIGONAL SELENIUM CRYSTALS OBTAINED FROM THE VAPOUR PHASE

by F. de BOER

546.23:548.5

Summary

Trigonal selenium crystals of a size sufficient for conduction measurements have been prepared from the vapour phase. The majority of these crystals had the form of hollow needles showing slight deviations from the ideal monocrystal structure. A small fraction, however, consisted of thick and massive twin crystals, amongst which we found one rather well-developed monocrystal. All well-developed planes were prism planes.

From his measurements of electric conductivity Müller [1]) concluded that selenium crystals prepared from the vapour phase are not true monocrystals but are built up of "*schuppenformige Mikrokristalle*". Some experiments made to check this conclusion by direct X-ray examination will be described in the present note; corresponding conductivity observations will be discussed in the next Report, *R 57*.

The selenium crystals were prepared by a method developed by Saunders [2]): 20 grams of selenium were placed on the bottom of a pyrex tube 30 cm long and 2 to 5 cm wide. This tube was evacuated, filled with A or Ne to a pressure of 1 to 5 cm, and then sealed off. Subsequently the lower end of the tube was heated for about a fortnight to 520 °C, the upper portion being kept at room temperature.

As a rule the selenium deposit in the upper part of the tube consisted of a more or less massive base from which needle-shaped crystals were sticking out into the interior of the tube (*fig. 1a*). These needles had a tapering base on top of which a hexagonal cylinder had developed (fig. 1*b*). The cylindrical part was hollow, and the tapering base was massive as indicated in the cross-section in fig. 1*c*. The form of these crystals is nearly identical with that described by Straumanis [3]), but, in contrast with his observations, all the well-developed planes that we could distinguish were prism planes.

Nearly all the prism planes of the larger crystals were composed of different parts slightly differing in the direction of their c-axis (by not more than 2°). As a result of the high reflectivity the boundaries between these differents regions could be easily seen without a microscope. Even in cases, however, where such boundaries were not directly visible, X-ray photographs showed the existence of slight variations in the directions of the c-axis up to 30′. Goniometer measurements also indicated that the six prism

planes were not exactly parallel to the hexagonal axis of the crystal; the deviations observed were of the same order of magnitude as those recorded above.

Though the tapering bases of the needles had no regular side planes they did not show large deviations from a monocrystal structure.

It may also be noted that some of our crystals showed such incomplete structures as illustrated in figs 1*d* and 1*e*: fig. 1*d* presents a case with six prism planes that do not form a closed surface; in the example of fig. 1*e* only one half of the hexagon had been developed.

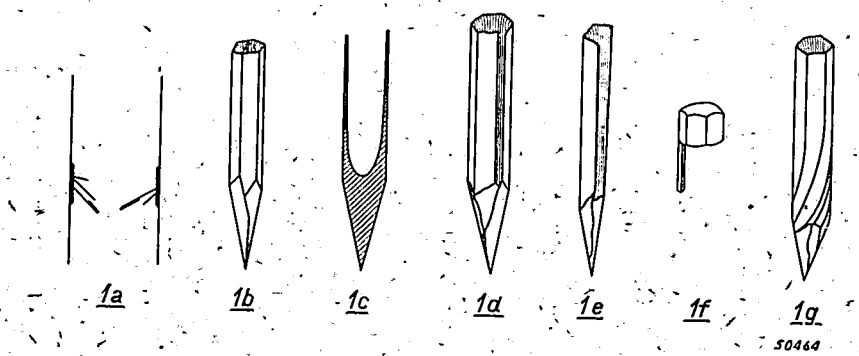


Fig. 1. *a*) A section of the sublimation tube, showing the position of the needles on the polycrystalline base; *b*) selenium crystal needle; *c*) section of a selenium needle; *d*) and *e*) selenium crystals that show deviations from the hexagonal prism; *f*) thick selenium crystal, grown on needle; *g*) needle crystal, showing deformation.
The figures *b*) to *g*) except *f*) show needles with a thickness of 0·1 tot 0·7 mm; the form shown in *f*) was about 2 mm thick.

The main product of our experiments consisted of needles as described above, but in one tube crystals in the shape of massive short hexagonal prisms of a much greater diameter had developed on top of the needle, as illustrated in fig. 1*f*. These were so crowded that they had grown together to twin crystals with their neighbours. Only one single crystal was found amongst them, and this one proved to be a monocrystal though it had no regular planes except prism planes. This crystal could easily be split along a plane parallel to the c-axis.

The crystal structure of the trigonal form indicates very long selenium spiral molecules with essentially covalent bonds, the spirals being connected by much weaker forces. This difference in bond strength explains the preference for needle-shaped crystals as resulting from the tendency to saturate the free valencies at the ends of the spirals. The weakness of the forces connecting the different spirals also explains the observed distortions of the crystal (see fig. 1*g*) and their fragility. This last property will be further enhanced by the isotropic character of the Van der Waals

forces. The position of the cleaving plane is obviously determined by the inequality of the binding forces in the principal directions.

Müller described his crystals as "*schlanke Pyramide*". Presumably our needles which possess well-developed prism planes are nearer to the monocrystal structure. As, however, the less perfect tapering bases of our crystals were also found to be monocrystals we are inclined to conclude that Müller's pyramids, too, must have been of a monocrystal structure. The consequences in connection with conduction measurements will be discussed in Report *R 57*.

Eindhoven, March 1947

REFERENCES

1) Th. Müller, Sitz. Ber. Phys. Med. Soz. Erlangen **70**, 7, 1938.
2) A. P. Saunders, J. Phys. Chem. **4**, 423, 1900.
3) M. Straumanis, Z. Krist. **102**, 482, 1940.

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

Reprints of the majority of these papers can be obtained on application to the Administration of the Research Laboratory, Kastanjelaan, Eindhoven, Netherlands. Those papers of which no reprints are available in sufficient number are marked with an asterisk (*).

1706: C. J. Bouwkamp: A note on singularities occurring at sharp edges in electromagnetic diffraction theory.
(Physica, 's-Grav. **12**, 467-474, 1946).

Wave functions u describing the diffraction by plane screens can be divided into two classes according as u or $\partial u/\partial n$ vanishes at the surface of the screen. Differentiation with respect to the coordinate in the normal-direction alters the character of the wave function, whilst tangential differentiation does not change the class.

Another effect (Rayleigh) of differentiation is the appearance of singularities at the edge of the screen. This is demonstrated explicitly in case of Sommerfeld's solution of the electromagnetic diffraction by a half-plane. Typical difficulties in electromagnetic diffraction theory are mentioned. Möglich's solution of the electromagnetic diffraction by a circular screen is critisized.

(Continued on page 356)

Philips Res. Rep. 2, 352-356, 1947

ON THE ELECTRICAL CONDUCTIVITY OF SELENIUM CRYSTALS

by F. de BOER

537.311.33:546.23

Summary

For the specific resistance of selenium monocrystals parallel to the c-axis values ranging from 2.10^4 to 5.10^4 Ω cm were observed; it is made plausible that the specific resistance of a pure monocrystal is lying near the lower end of this range. The specific resistance at right angles to the c-axis was measured to be 2.10^6 Ω cm. In addition some experiments on the influence of temperature and pressure on the conductivity were carried out.

For explaining the electrical properties of selenium rectifiers a knowledge of the conductivity of pure selenium is of fundamental importance.

The observations that have been made on polycrystalline annealed selenium are too divergent to be trustworthy; and the measurements on monocrystals too are unrealiable owing to the difficulty of providing good electrical contacts [1]). Hence we have carried out some further measurements, which will be communicated below.

The crystallographic properties of the crystals used have been described in a previous note [2]).

These crystals are fitted with electrodes of gold or of aquadag. The gold electrodes were prepared by evaporation *in vacuo*; the aquadag electrodes were painted on the crystals and then made to dry for a fortnight at least before the observations were carried out. By these techniques heating of the crystals was avoided; measurements with gold and aquadag electrodes yielded concordant results which indicates that the electrical contact was sufficiently good in both cases.

When the potential across a "virgin" crystal is gradually raised the current increases in due proportion, but when this voltage is subsequently diminished the relation between voltage and current is found to be curved, owing to the formation of a blocking layer *).

To eliminate the influence of the blocking layer from the later observations, the curves giving the voltage as a function of the current were represented by

$$V = RI + a\sqrt{I},$$

*) This blocking layer differs from that in a selenium rectifier, since the latter is developed by the very first quantities of electricity flowing, which cause an immediate decrease in the current.

R being the true resistance and $a\sqrt{I}$ representing the effect of the blocking layer. It could not be ascertained, however, that the whole of the last term is due to a blocking layer near the electrodes; part of it may be due to a transition resistance at internal crystal boundaries which may have existed, since our monocrystals were not ideally perfect.

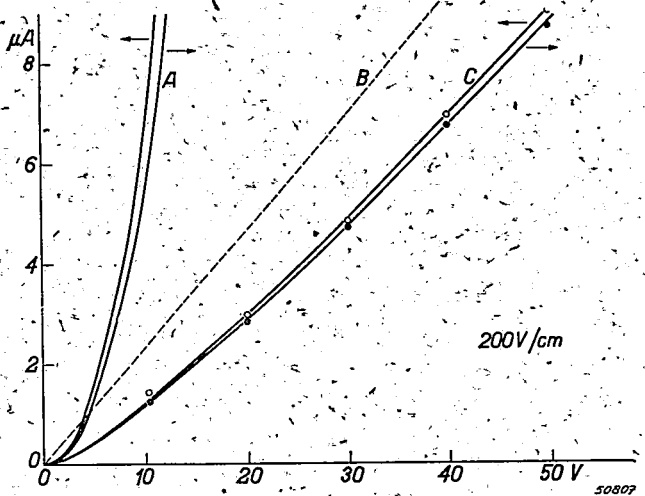


Fig. 1. Current-voltage curves for conductivity parallel to the c-axis, using a needle crystal. Direct current was used, and a second series of measurements was made with the direction of the current reversed as indicated by the arrows. C indicates the total voltage measured, A the voltage over the "blocking layer", and B the voltage over the selenium layer.

Our measurements yielded the following results.

a) The specific resistance of hollow well-developed needle-shaped crystals parallel to the c-axis varied from 2·5 to $4 \cdot 0.10^4$ Ω cm for the first series of measurements, and from 3·2 to $5 \cdot 2.10^4$ Ω cm when computed from later series by means of eq. (1).

b) Observations on the one massive rectangular block we had at our disposal, prepared by cleaving the crystal mentioned under c), yielded values between 1·5 and $3 \cdot 0.10^4$ Ω cm, the uncertainty being due to a relatively strong influence of the blocking layer in this case.

c) For the specific resistance at right angles to the c-axis a value of 2.10^6 Ω cm was measured; this observation was carried out on the short and massive crystal shown in fig. 1e of the preceding paper [2]).

In a parallel direction Müller [1]) measured specific resistances of 2.10^4 to 5.10^5 Ω cm; his crystals, however, exhibited such a complex electrical behaviour that he concluded that they were not pure monocrystals,

but aggregates. No X-ray data are mentioned in Müller's paper but, as the majority of his crystals had the shape of pyramids rather than of prisms, it seems likely that the deviations from a pure monocrystal structure were more marked in his case than in ours. The specific resistance of a pure monocrystal should then be sooner near the lowest than near the highest value observed by Müller.

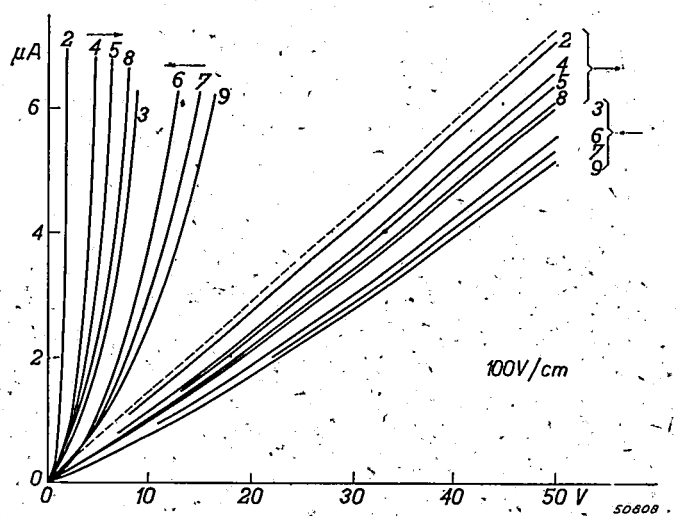


Fig. 2. Current-voltage curves for parallel conductivity in both directions (indicated by arrows as in fig. 1), using a needle crystal. The numbers given in the figure are the serial numbers of the observations — each series involved the flowing of about 100 μC through the crystal, and in general the time between two series was one or two days. The voltages over the blocking layer have been constructed as mentioned in fig. 1. The dashed line gives the voltage across the selenium layer thus derived; the resistance of this layer appears to be unchanged by the measurements.

It cannot be denied that the measurements recorded in this paper are subject to similar inaccuracies; the area of the cross-section of hollow crystals, for instance, could only roughly be estimated, and the massive crystals exhibited pronounced blocking-layer effects.

Broadly the specific resistances that we observed show a tendency to decrease when the crystals approach the pure monocrystal structure; in view of this tendency we are inclined to estimate the correct value for a pure monocrystal to be $2{\cdot}5.10^4\ \Omega$ cm, a value being close to the $2{\cdot}0.10^4\ \Omega$ cm that Dr Addink *) of this laboratory determined for a highly purified polycrystalline sample of selenium.

*) N. W. H. Addink. To be published shortly.

Our value of 2.10^6 Ω cm for the transverse conductivity is in accord with the value observed by Müller on a rather irregular crystal; this indicates that the transverse conduction is much less influenced by irregularities than the conduction in a parallel direction. This suggests an important difference in the conductivity mechanism in the parallel and the transverse direction.

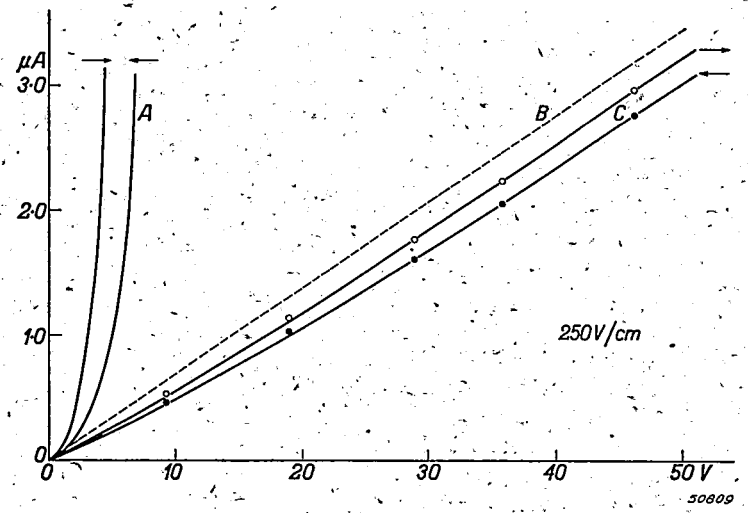


Fig. 3. Current-voltage curves for conductivity perpendicular to the c-axis, using a massive crystal. C = total voltage measured, A = voltage across the blocking layer and B = voltage across the selenium layer.

Some preliminary investigations on the influence of the temperature and of the pressure on the resistance confirmed Müller's statement that the influence of the temperature is considerably larger on the transverse than on the parallel conduction. For the direction parallel to the c-axis a

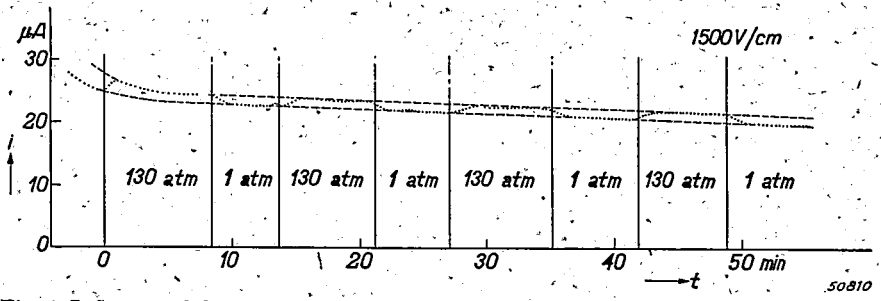


Fig. 4. Influence of the pressure on the conductivity perpendicular to the c-axis, measured with a D.C. field strength of 1500 V/cm.

decrease of the resistance amounting to 4% per degree was found, a value that, on substitution in the formula $R = R_0\, e^{E/kT}$, gives $E = 0{\cdot}3$ eV for the activation energy. Measurements perpendicular to the c-axis gave a change in the slope of the log R *vers.* $1/T$ curve at about 70 °C; below this temperature this slope corresponds to an energy of activation of $E = 0{\cdot}15$ eV, and above it to $E = 0{\cdot}8$ eV.

The influence of pressure on the parallel resistance could not be determined, owing to deformation of the hollow crystals; for the transverse resistance the measurements shown in *fig. 4* give $1/R\ (\mathrm{d}R/\mathrm{d}p) = 4.10^{-4}$ cm²/kg, where R indicates the resistance, p the pressure.

Measurements of the capacity of the needles confirmed Müller's statement that this is a small quantity; it does not exceed 5 pF for a needle with a section of 0·2 mm² and a coated surface of 2 mm².

Eindhoven, March 1947

REFERENCES

1) Th. Müller, Sitz. Ber. Phys. Med. Soz. Erlangen, **70**, 7, 1938.
2) F. de Boer, Philips Res. Rep. **2**, 352, 1947.

ABSTRACTS

(Continued from page 351)

1707: W. Elenbaas: The hypothesis of minimum voltage in the theory of the arc. (Physica, 's-Grav. **12**, 491-498, 1946).

The author critisizes the explanation of the contraction of the high-pressure discharge in the vicinity of the cathode, making use of the theory of minimum voltage. Although this theory and the exact theory may be put in such a form that they yield very similar formulae, there is a discrepancy, as regards the cross-section of the arc, by a factor that at ordinary current densities may be as high as 50 and that, moreover, depends on the temperature. At very high current densities (almost complete ionisation) the principle is considered to give good results.

1708: P. C. van der Willigen: Contact arc welding. (The Welding Journal Research Supplement **25**, 313*S*-320*S*, 1946).

This paper describes the properties and the advantages of a new type of welding electrode (developed by Philips), which combines the properties of touch welding, self starting and reignition, while also some other advantages are obtained (for full particulars see Philips Technical Review **8**, 161-167, 304-309, 1946).

(Continued on page 381)

Philips Res. Rep. **2**, 357-381, 1947

ELASTIC RELAXATION AND SOME OTHER PROPERTIES OF THE SOLID SOLUTIONS OF CARBON AND NITROGEN IN IRON *)

by L. J. DIJKSTRA

539.389.3:541.123

Summary

The theory of elastic relaxation in α-iron caused by carbon and nitrogen [19]) in solid solution, as given for the first time by Snoek [6]) and later elaborated by Polder [9]), predicts a strong anisotropy for the various crystal directions. This theoretical conclusion has been confirmed in a series of experiments carried out on prepared single crystals of iron in the [100]- and [111]-directions. For carbon we were able to check the theoretical absolute magnitude of the effect. The question of the most probable place of the dissolved particles in the iron lattice is also discussed.

Finally at 20 °C the rate of segregation taking place in the form of a carbide or nitride has been determined by measuring the decrease in magnitude of the elastic relaxation in the course of time.

1. *The* Fe-C *and* Fe-N *diagrams*

In metallography it is a well-known fact that the influences of small quantities of C and N on the properties of iron are similar in many respects [1]) [2]). This similarity is already expressed in their binary equilibrium diagram with Fe. *Fig. 1* shows a part of the Fe-C diagram in the region of low C content. There are only two stable types of solid solution, the cubic face-centred γ-phase or austenite and the cubic body-centred α-phase or ferrite.

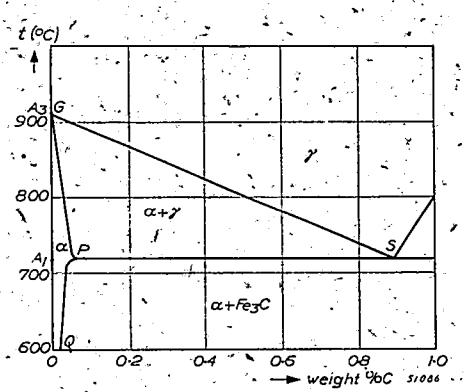


Fig. 1. The Fe-C diagram.

*) A short account of our experiments was given already in J. L. Snoek, New developments in ferromagnetic materials, Elsevier's Publishing Company, Amsterdam-New York, 1947; especially pp. 46-54.

As is evident from the diagram the solubility of C in the γ-phase is considerably larger than in the α-phase. The solubility in the α-phase is very small and does not exceed 0·05% by weight. The curve *G-P* comprising a temperature interval between 910 °C (A_3-point) and the eutectoid temperature of 720 °C (A_1-point) is the solubility curve of C in α-iron in equilibrium with austenite; *P-Q* represents the solubility of C in α-iron in equilibrium with cementite (Fe_3C). The solubility reaches a maximum value of about 0·05% at 720 °C. The only reliable data on the curve *P-Q* determined by Köster [1]) are given in the next table. All contents in this paper are expressed in per cent by weight.

t (°C)	300°	500°	600°	650°	650°	680°	695°	710°
% C	0·005	0·010	0·014	0·020	0·022	0·025	0·031	0·043

According to Yensen [3]) the solubility of C at 20 °C is about 0·006%.

For low N content the Fe-N diagram has quite the same aspect. Now the eutectoid temperature has the somewhat lower value of 580 °C. The following values are available for the solubility of N in α-iron in equilibrium with iron nitride (Fe_4N): [2])

t (°C)	100°	200°	300°	400°
% N	0·001	0·005	0·01	0·02

Thus the solubility at 20 °C does not exceed 0·001%. In the γ-phase the solubilities of C and N in equilibrium with, respectively, Fe_3C and Fe_4N nearly always exceed 1%. A supersaturated ferritic solid solution at 20 °C can easily be obtained by quenching from a higher temperature (*cf.* section 5).

As metastable structures not represented in the equilibrium diagram the tetragonal and cubic martensite must be mentioned. These two modifications, however, are not generally considered as solid solutions of C in Fe (*cf.* section 2).

2. *The character of the solid solutions of* C *and* N *in the* Fe *lattice*

Density measurements have shown that the solid solutions of C and N in Fe are of the interstitial type. As to the atomic state of the dissolved particles only a few data are available from some experiments on the behaviour of the solid solution in the γ-phase in an electrostatic field. From these experiments it appears that in the γ-phase C is present, at least partly, as a multivalent positive ion [14]), N probably as a neutral atom or as a negative ion [15]).

The next question here interesting us is: what is the exact location of the C and N particles in the Fe lattice? It is not easy to answer this question. X-ray analysis hardly gives any information about this point, the C or N content being too small and the atomic number too low. Therefore we have to fall back on geometrical considerations of the lattice combined with certain experimental facts.

As a first approximation the lattice may be considered as an assembly of touching rigid spheres, the radii of which correspond to the dimensions ("range of action") of the atoms or ions constituting the crystal lattice. If we then calculate the dimensions of the available interstices in the pure Fe lattice, the latter appear to be always much smaller than the dimensions of the C or N atoms. At 20 °C the lattice parameters a_γ and a_α for pure Fe are:

$$a_\gamma = 3{\cdot}54 \text{ Å},$$
$$a_\alpha = 2{\cdot}86 \text{ Å}.$$

The corresponding radii of the Fe atom are:

$$R_\gamma^{Fe} = 1{\cdot}26 \text{ Å} \text{ and } R_\alpha^{Fe} = 1{\cdot}24 \text{ Å}.$$

The γ-phase

The face-centred elementary cell is shown in *fig. 2*. In it we have to consider two kinds of interstices as a probable position for the dissolved particles. Firstly the central interstice (with crystallographic indices $\frac{1}{2}\ \frac{1}{2}\ \frac{1}{2}$) with the coordination number 6; secondly the centres (*e.g.* with indices $\frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}$) of the tetrahedrons formed by the Fe atoms in a corner of the elementary cell and in the centres of the three faces joining in this corner. These centres ($\frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}$) correspond with the coordination number 4.

The radii of the inscribed spheres of these interstices are:

$$R_\gamma^6 = 0{\cdot}52 \text{ Å} \text{ and } R_\gamma^4 = 0{\cdot}28 \text{ Å}.$$

Per Fe atom we find one R_γ^6- and two R_γ^4-holes. (The upper index denotes the coordination number).

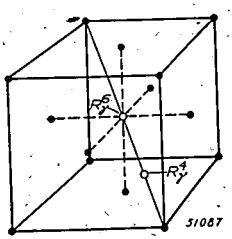


Fig. 2. The elementary cell of the cubic face-centred lattice.

The α-phase

The elementary cell of the body-centred lattice is given in *fig. 3*. Here too the matter is rather complicated, because of the existence of more than one kind of interstices that must be taken into consideration.

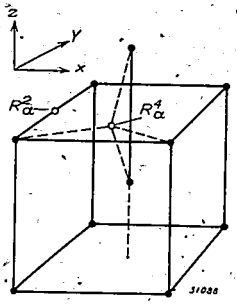


Fig. 3. The elementary cell of the cubic body-centred lattice.

Firstly we have to regard the middles of the edges of the elementary cell (*e.g.* 0 0 $\frac{1}{2}$) with coordination number 2. Secondly we must look at the centre (*e.g.* $\frac{1}{2}$ $\frac{1}{4}$ 0) of the tetrahedron (not regular) constituted by two corners and two cell-centres (fig. 3). For this interstice the coordination number is 4.

We easily calculate:

$$R_\alpha^2 = 0{\cdot}19 \text{ Å} \text{ and } R_\alpha^4 = 0{\cdot}36 \text{ Å}.$$

In this case we have per Fe atom three R_α^2- and six R_α^4-interstices. Each R_α^2-interstice is directly connected with four adjacent R_α^4-interstices.

According to the data of Goldschmidt[16]) and of Pauling[17]) the atomic and ionic radii of C and N are:

$$R^C = 0{\cdot}80 \text{ Å}; \; R^{C^{4+}} \approx 0{\cdot}18 \text{ Å}.$$
$$R^N = 0{\cdot}70 \text{ Å}; \; R^{N^{5+}} \approx 0{\cdot}13 \text{ Å}.$$

C^{++++} and N^{+++++} are ions with a complete He core, with two electrons. We do not know the exact atomic state of the dissolved particles, but for mono- to trivalent C ions and mono- to quadrivalent N ions the ionic radii will only slightly differ from the atomic radius. They will be somewhat smaller because of the higher effective nuclear charge in the ionic state.

On comparing these radii with the dimensions of the interstices in the pure γ- and α-iron lattice it seems reasonable to expect a considerable deformation of the elementary cell of the Fe lattice as a result of the intro-

duction of a C or N particle. For the R_γ^6- and R_γ^4-interstices this deformation is symmetrical with respect to the three cubic axes as is denoted by the three crystallographical indices. As to the R_α^2- or R_α^4-interstices, however, we can distinguish between three classes most easily qualified as R_α^x-, R_α^y- and R_α^z-interstices, *e.g.* those indicated in fig. 3 are R_α^y-interstices. If now C or N is introduced in the α-lattice, let us say on the R_α^z-places, this will result in a tetragonal lattice-deformation in the direction of the z-axis. It is clear that for the $R_\alpha^{2,z}$-configuration the tetragonal axis will be the long one; for the $R_\alpha^{4,z}$-configuration the situation is less simple.

We wish to emphasize here that this tetragonality measured by the difference in lattice constants along the z- and x-axes will certainly be much greater for the $R_\alpha^{2,z}$- than for the $R_\alpha^{4,z}$-structure; firstly because for the latter the magnitude of the deformation is much smaller, as is evident from the values of the radii mentioned before; secondly because for a structure with coordination number 2, the asymmetry will be more pronounced than for the structure with coordination number 4.

Now we shall discuss the various types of solid solutions.

Austenite. In this stable modification the R_γ^6-interstices are by far the largest ones and so are generally assumed to be the most probable position for the dissolved particles. This assumption has got experimental evidence from an X-ray investigation of Petch [18]) of the austenite structure by means of the powder method. The C atoms were found to be situated at centres of the unit cells.

Tetragonal martensite. The martensitic carbon steel is formed from austenite if the latter contains a certain amount of C and is quenched at a sufficiently low temperature. Martensite is metastable and not proof against heating. It is generally regarded as a solid solution of C in Fe.

The cubic γ-lattice can also be considered as a tetragonal body-centred lattice as is shown in *fig. 4*. According to Bain [5]) the $\gamma \rightarrow \alpha$ transition is most simply realized by assuming that the elementary cell of the body-centred γ-lattice is transformed into the elementary cell of the body-

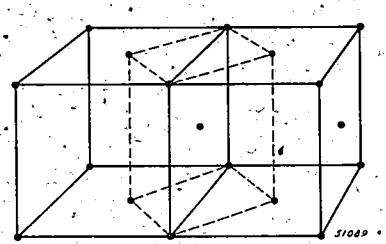


Fig. 4. This figure shows how a cubic f.c. lattice can be considered as a tetragonal b.c. lattice. Part of the atoms in the centres of the faces has been omitted for the sake of clearness.

centred cubic α-lattice. It is evident from the cell dimensions that this directly results from a contraction in the direction of the tetragonal axis and a small expansion along the other axes. During this transition, with which only a small volume effect of about 1% is associated, the R_γ^6- and R_γ^4-interstices transform into, respectively, the R_α^2- and R_α^4-interstices of one definite class, for convenience called the z-class.

The condition for the formation of the tetragonal martensite must now be understood in the following way. The martensite transition cannot be prevented by very rapid quenching. The presence of C in the lattice, however, causes the transition from the cubic γ-phase to be incomplete, if care is taken that the rate of quenching is sufficiently high and the final temperature sufficiently low to prevent the C from migrating away from their original places.

From this picture we conclude a most characteristic property of the tetragonal martensite. Considering the latter as a (tetragonal) body-centred lattice, the C must be located at the R_α-interstices of one definite class, further denoted as R_α^z-interstices. In this way the observed tetragonality (in the z-direction) of the martensite lattice has found a very natural explanation.

Cubic martensite. This phase is formed for instance by annealing tetragonal martensite at 150 °C for some time. By this treatment the tetragonal lattice-deformation disappears and only a small cubic deviation from the pure α-lattice, linear with the C content, remains [8]).

Cubic martensite too is sometimes regarded as a solid solution of C in α-iron. There is no general accordance as to the structure of cubic martensite. It is regarded both as a solid solution of C in α-iron, and as a two-phase structure in which most of the C has precipitated as highly dispersed cementite [8]). Just like tetragonal martensite it is not proof against heating.

Fig. 5 shows the cell dimensions in dependence on the C content of tetragonal martensite as measured by Honda [7]) and by Hägg [8]). Curve *a* represents the lattice constant along the tetragonal axis; *b* along the two other axes. Curve *c* is the lattice constant of cubic martensite according to Hägg [8]).

Remark. If, as generally assumed for austenite, the C is located at the R_γ^6-interstices, and if the picture of the $\gamma \rightarrow \alpha$ transition according to Bain is right, then in tetragonal martensite, considered as a tetragonal body-centred structure, the $R_\alpha^{2,z}$- and not the $R_\alpha^{4,z}$-interstices will be occupied. If cubic martensite is considered as a solid solution we can imagine that the temperature at which it is formed is just high enough to provide the C particles with sufficient mobility for sliding into the large R_α^4-places, without being able to diffuse over a large range and to leave the lattice by the formation of Fe_3C. In this way the cubic structure

of cubic martensite and the fact that the lattice constant deviates only slightly from that of the pure Fe lattice can be understood. Numerically these concepts are not quite unreasonable.

Again by approximating the lattice by a conglomeration of touching rigid spheres we can evaluate the "range of action" of the C atom (ion) in martensite from the cell dimensions given in fig. 5. The acceptable value $R_C = 0{\cdot}80$ Å is found to give a good agreement with the experimental curves *a*, *b* and *c* in fig. 5, on the condition that for the tetragonal martensite the $R_a^{2,z}$-, for cubic martensite the R_a^4-structure is assumed.

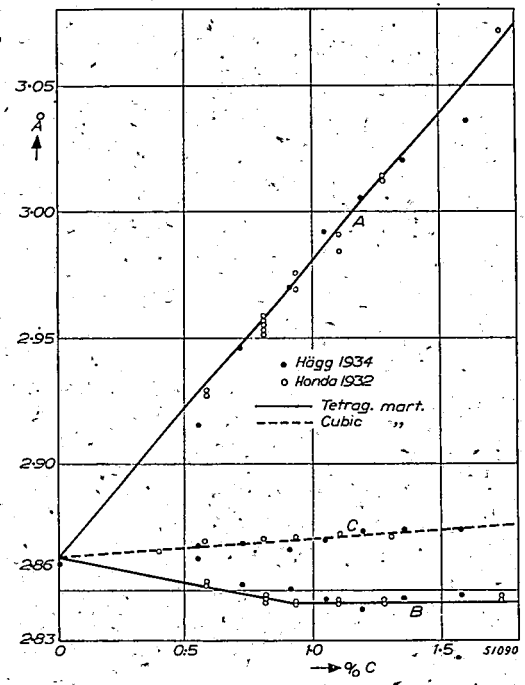


Fig. 5. Cell dimensions of tetragonal and cubic martensite.

Ferrite. By slowly cooling down from the γ-phase or by tempering martensite above 200 °C the dissolved C is given the opportunity to migrate away from the interstices and to precipitate for the greater part from the solid solution as cementite (Fe_3C). The result is a stable, cubic α-modification which is called ferrite and contains only a very small quantity of C (or N) in solid solution [1]) [2]). The way in which a supersaturated ferritic solid solution of C or N is obtained, will be discussed in section 5.

Up till now nothing has been said about the exact location of the dissolved particles in the ferrite lattice. The reason is that the C or N contents

involved are too small to produce any observable effect in X-ray diagrams or on the density (*cf.*, however, section 9). In literature the dissolved particles are assumed to be located at the R_a^4-places [4]) as well as at the R_a^2-places [10]) without further explanation. We should expect ferrite to have the R_a^4-structure since the lattice deformation will in this case be much smaller than for the R_a^2-structure.

3. *Theory of elastic relaxation caused by* C *and* N *in α-iron*

In general, elastic relaxation may be briefly characterized as follows. If a material showing elastic relaxation is suddenly subjected to an external force at the time $t = 0$, the resulting deformation consists of two different parts. These two parts are the normal, purely elastic component λ_1 being established within an immeasurably short time interval, and a relaxation component λ_2, which is only established in the course of some time.

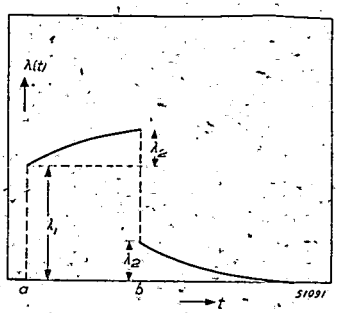


Fig. 6. Deformation of a substance showing elastic relaxation under influence of an external constant force. At the moment *a* the force is suddenly applied, at *b* removed.

In the most simple case of one single relaxation time τ the component λ_2 as a function of the time t can be represented by the expression $\lambda_2(1-e^{-t/\tau})$; that is, on exposing the test object to an external constant force at $t = 0$, the total deformation $\lambda(t)$ is given by:

$$\lambda(t) = \lambda_1 + \lambda_2(1-e^{-t/\tau}).$$

On removing the external force (at the time $t = 0$) we find:

$$\lambda(t) = \lambda_2 e^{-t/\tau} \quad (fig.\ 6).$$

For small applied forces p the deformations λ_1 and λ_2 will be proportional to p. Therefore we introduce the constants ε_1 and ε_2 according to $\lambda_1 = \varepsilon_1 p$ and $\lambda_2 = \varepsilon_2 p$. Then, $1/\varepsilon_1$ is the elastic modulus E; the ratio $\lambda_2/\lambda_1 = \varepsilon_2/\varepsilon_1$ is called the relative magnitude of the effect.

The elastic relaxation effects that will be discussed in this paper are of the particular type caused by the presence of very small quantities of C or N in solid solution in the α-iron lattice. This ferritic solid solution is characterized by:

1) a component of relatively large mobility (C or N), and

2) a dissolved C or N particle located at a R_a^2- or R_a^4-interstice which causes a tetragonal deformation of the elementary cell.

These conditions are sufficient to produce elastic relaxation; for, in addition to the instantaneous, purely elastic deformation caused by an external load (*e.g.* by unidirectional stretching or pressing of the test object) we have to expect on thermodynamical grounds a relaxation effect resulting from a gradual transfer of some of the dissolved particles between adjacent interstices of different classes. The direction of this short-range material transfer will be such that it will cause the substance further to give way to the applied external load. This we proceed to explain in more detail.

Snoek's theory [6]) of these elastic relaxation effects is based upon the following crystallographic picture. We assume the test object to consist of a single crystal of α-iron, and for simplicity we suppose C or N to be located at the R_a^2-interstices. The reasoning, however, will be quite the same for the R_a^4-structure. A continuous transition of the C or N atoms (ions) takes place between the x-, y- or z-positions as a consequence of their thermal agitation. In the absence of any stress the x-, y- and z-positions are on the average equally occupied and the lattice as a whole remains cubic. If now the material is undirectionally strained, *e.g.* pressed in the direction of the z-axis, the thermodynamical equilibrium will be disturbed. The C or N will preferably migrate from the z- to the x- and y-interstices since by such a movement they give way to the external load.

The counteracting factor in this migration process is the decrease in entropy resulting from this rearrangement. On account of these two counteracting tendencies the thermodynamical equilibrium is shifted towards a surplus of occupied x- and y-interstices. The equilibrium concentrations of the x-, y- and z-particles in the stationary state can be calculated by the rules of elementary thermodynamics, if the tetragonal deformation of the elementary cell by the dissolved particles is known [6]) [9]).

It will be clear that in a homogeneous crystal this microdiffusion process will be characterized by one single relaxation time τ, the order of magnitude of which corresponds to the average lifetime of an occupied x-, y- or z-position. The macroscopic effect will be that, in addition to the momentaneous elastic deformation λ_1, a sudden change in the external load produces a relaxation component λ_2 which will be fully established only after a time lag of the order of magnitude τ (fig. 6).

The following important conclusion may be drawn at once. The relaxation effect will be strongly anisotropic in this sense that the ratio λ_2/λ_1 will depend on the direction of the external stress with respect to the crystal axes. If the material is stressed in the direction of the cubic axis (*e.g.* [100]-direction) the effect will have its maximum value; in the direction of the trigonal axis ([111]-direction) the effect will rigorously disappear, since for this direction there is no difference between x-, y- and z-positions.

The quantitative evaluation of the ratio λ_2/λ_1 has been given by Snoek [6]) only for the simple case of a single crystal stressed in the [100]-direction. A general theory based on the picture outlined in this section was given by Polder [9]). Here we shall only mention his final results. If the material is strained in a quite arbitrary direction by a periodic force

$$p = p_0 e^{i\omega t}, \tag{1}$$

the corresponding deformation λ can be written:

$$\lambda = \lambda_0 e^{i\omega t}, \tag{2}$$

with

$$\lambda_0 = \lambda_1 + \lambda_2 \frac{1 - i\tau\omega}{1 + \tau^2\omega^2}. \tag{3}$$

The relaxation time

$$\tau \approx \frac{a^2}{D_\alpha}, \tag{4}$$

a being the lattice constant and D_α the coefficient of diffusion of the dissolved particles in α-iron. The relative magnitude is given by

$$\frac{\lambda_2}{\lambda_1} = \tfrac{2}{5} \frac{EV}{RT} C_0 \left(\frac{e_{zz} - e_{xx}}{c_z}\right)^2 \left\{1 - 3\,(\alpha_1^2\alpha_2^2 + \alpha_2^2\alpha_3^2 + \alpha_3^2\alpha_1^2)\right\}, \tag{5}$$

where

$\alpha_1, \alpha_2, \alpha_3$ = the direction cosines with respect to the cubic axes;
E = elastic modulus in the direction α_1, α_2, α_3;
V = volume of one mole of Fe;
R = gas constant per mole;
T = absolute temperature;
$3\,C_0$ = number of gram atoms of C or N per gram atom of Fe;
c_z = number of gram atoms of C or N on the z-places per gram atom of Fe;
$e_{zz}(e_{xx})$ = relative deformation in the z-(x-)direction caused by c_z on the condition that the crystal is not externally stressed and $c_x = c_y = 0$.

Formula (5) still remains valid for the R_a^4-structure. Only, in this case, according to section 2, the tetragonality factor

$$F = \frac{e_{zz} - e_{xx}}{c_z}$$

is sure to be much smaller in absolute value than for the R_a^2-type.

Remark.

We regret not being able to give any exact experimental support of this assertion. We can only assume that tetragonal martensite is of the $R_a^{2,z}$-, cubic martensite of the R_a^4-type. If we are only interested in the order of magnitude we probably make no great error by replacing F for cubic martensite by the slope of curve c in fig. 5. On these assumptions we find from the data in fig. 5 that, as to the order of magnitude, λ_2/λ_1 for the R_a^4-structure will be only about 0·1% of λ_2/λ_1 for the R_a^2-structure, C_0, T and α_1, α_2, α_3 being the same.

Our main purpose in the experimental work dealt with in this paper was to ascertain the expected anisotropy of the elastic relaxation effect and to verify the numerical value of the relative magnitude as predicted by formula (5).

4. *Method of measuring the elastic relaxation*

Simple and exact determinations may be carried out by observing the damping of the free oscillations of a thin strip of the material, the presence of C or N in solid solution functioning as a cause of internal friction [10]). The three dimensions of the rectangular strips were about 50 mm, 5 mm and 0·15 mm. These strips were fixed at one end, the free end being loaded with an extra mass by which the period of oscillation could be adjusted so as to be of the order of one second. The oscillating strips were suspended in a space that could be kept at a constant temperature ranging from about —100 to 300 °C. In order to eliminate magneto-mechanical hysteresis effects all measurements were performed in a longitudinal magnetic field. During the oscillation both halves of the strips on either side of the neutral zone are exposed to pressure and stretch in turn. On account of the relaxation a phase difference δ occurs between the elastic force p and the elongation λ. Taken over one period this results in a value of $\oint p \, d\lambda$ differing from zero; elastic energy is lost and reappears in the form of heat. The equation of motion for any damped free oscillation is given by

$$m \frac{d^2\lambda}{dt^2} + p = 0, \text{ or, putting } \lambda = p\varepsilon,$$

$$m\frac{d^2\lambda}{dt^2}+\frac{\lambda}{\varepsilon}=0\,. \tag{6}$$

Here m represents the effective moving mass. Moreover, ε is complex because of the phase difference δ between p and λ. Therefore we write $\varepsilon = \varepsilon' - i\varepsilon''$.

We shall suppose $\varepsilon'' \ll \varepsilon'$, so that $\tan\delta \equiv \varepsilon''/\varepsilon' \ll 1$. Under these circumstances the solution of (6) is

$$\lambda(t) \approx \lambda_0\, e^{-(\varepsilon''/2\varepsilon')\omega t}\, e^{i\omega t}\,. \tag{7}$$

The natural frequency $\omega = 2\pi\nu = \sqrt{1/\varepsilon' m}$.

From (7) the logarithmic decrement Δ is readily found to be

$$\Delta = \pi\frac{\varepsilon''}{\varepsilon'} = \pi\tan\delta\,. \tag{8}$$

In this paper we are only interested in the special case in which the damping is caused by a relaxation effect with the characteristic time τ. If also $\varepsilon_2 \ll \varepsilon_1$, then it can be shown for $\varepsilon_2 \ll \varepsilon_1$ that $\varepsilon' \approx \varepsilon_1$,

$$\varepsilon'' = \varepsilon_2\frac{\tau\omega}{1+\tau^2\omega^2},$$

so that $\tan\delta$ is reduced to

$$\tan\delta \approx \frac{\varepsilon_2}{\varepsilon_1}\frac{\tau\omega}{1+\tau^2\omega^2}. \tag{9}$$

Tan δ reaches a maximum value for $\omega = 1/\tau$, and

$$\tan\delta_{\max} \approx \tfrac{1}{2}\frac{\varepsilon_2}{\varepsilon_1}. \tag{10}$$

This enables us to determine both the relaxation time τ and the relative magnitude $\lambda_2/\lambda_1 = \varepsilon_2/\varepsilon_1$ by measuring the damping peak in the Δ—ω curve. From a technical point of view it is rather difficult to vary ω in a sufficiently large interval. Therefore, following Snoek [19]), we kept ω constant and varied τ by varying the temperature. The relaxation time τ is strongly dependent on T and can be represented by the expression

$$\tau = \tau_0\, e^{U/kT}. \tag{11}$$

The activation energy U corresponds to a temperature of 10000°. In view of the symmetrical character of formula (9) in τ and ω it follows that $\tau = 1/\omega$ at that temperature T at which Δ reaches its maximum value, assuming that $\varepsilon_2/\varepsilon_1$ varies only slightly with T (*cf.* formula (5)).

5. *The preparation of the samples*

For a study of the anisotropy of the effect it was necessary to prepare single crystals of iron. The method of annealing-strained iron, originally proposed by Edwards and Pfeil[11]), proved succesful. The starting material consisted of strips of rolled Armco iron 0·15 mm thick which were critically strained about 3% beyond the elastic limit. By subsequent annealing for about eight hours at a temperature just below the A_3-point (fig. 1) we succeeded in preparing a large number of single crystals having various orientations. Among them there were some of 50×30 mm^2. Out of these crystals rectangular strips of $50 \times 5 \times 0{\cdot}15$ mm^3 were cut.

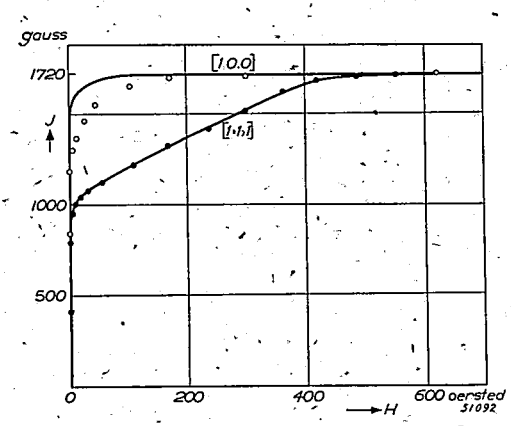


Fig. 7. Magnetization curves for single crystals of Fe in the [100]- and [111]-directions.

No X-ray diagrams have been made, but for a determination of the orientation of the various crystals we have made use of the well-known magnetic anistropy of α-iron. On measuring the magnetization curves of a large number of single crystals the curves of some specimens (measuring points in *fig. 7*) turned out to be nearly identical with the curves characteristic for the [100]- and [111]-orientations as recorded in literature[12]) and drawn in fig. 7. The coercive force was 0·15 oersted. The crystals were not perfectly homogeneous but they always contained a certain amount, never exceeding 5%, of inclusions of a different orientation.

For measurements on polycrystalline material the basic material used was carbonyl iron.

In order to purify the test objects of all traces of C or N they were previously heated for some hours at a temperature of 850 °C in an atmosphere of moist hydrogen. After this heat treatment the damping curve of the purified specimen was determined in a range from —50 to +50 °C. This damping turned out to be very small ($\Delta \approx 0{\cdot}005$) and remained practically

constant at temperatures up to 50 °C. From the fact that after this purification process no trace of a damping peak is observed we may conclude (*cf.* section 6) that by this heat treatment C and N have been rigorously eliminated.

To obtain a solid solution of C in α-iron we proceeded in the following way. The purified strips were annealed in an atmosphere of hydrogen for at least one hour at 720 °C, the temperature of the eutectic at which the solubility of C in equilibrium with Fe_3C reaches a maximum value (fig. 1). The hydrogen was purified in a glass tube containing silicagel and cooled by liquid nitrogen. Before reaching the specimens the hydrogen passed a tube containing solid naphthalene and kept at a fixed temperature in the neighbourhood of 20 °C; by varying this temperature the partial naphtalene pressure and so the amount of C added could be adjusted. By subsequently quenching in water of 20 °C the C in solid solution is prevented from segregating, and in this way a supersaturated solid solution of C in α-iron was obtained. Even at 20 °C the rate of segregation was still considerable (fig. 12).

N was added in solid solution by annealing the specimens for one or two hours at about 580 °C in a mixture of purified hydrogen and 2-6% of NH_3 by volume. After this heat treatment quenching in paraffin oil proved to be adequate to prevent the dissolved nitrogen from segregating. By quenching, N can be kept much easier in supersaturated solid solution than C. This may be partly explained by remembering that (see section 2) the solubility of N (in contrast with C) is still considerable at a rather low temperature at which the rate of segregation has already become relatively small. Moreover, at 20 °C (*cf.* section 10) we observed that at the same temperature C segregates more rapidly than N.

6. *Determination of the* C *and* N *content in solid solution*

The next problem was to determine the exact quantity of C or N present in solid solution. On account of our small concentrations (about 0·01%) this was not without some difficulty, the increase in weight of the relative light strips (about 0·3 gram) being too low to be measured with accuracy. Now, in our region of small contents a linear relationship proved to exist between C or N concentration and the electrical resistance of the material, and changes in the latter, which were of the order of 1%, could easily be measured. The relation between C and N content and change in electric alresistance was determined on a polycrystalline carbonyl-iron bar, about 17 grams of weight, which had been carburized or nitrided in the same way as the strips used in the relaxation experiments. As quenching in water was not possible on account of the disturbing formation of a thin layer of iron oxide on the surface of the bars, we chose paraffin oil instead.

The observed proportionality between increases in weight and in resistance proves that with oil-quenching the rate of cooling is sufficiently large for the small quantities that were added here. The electrical resistance was simply determined by current-voltage measurements carried out in a bath of carbon tetrachloride kept at a constant temperature (25 °C).

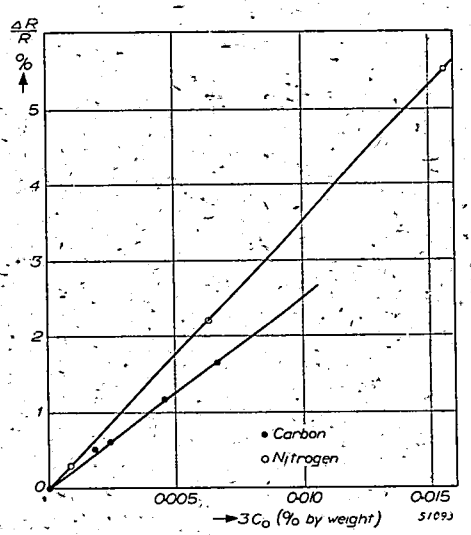


Fig. 8. Relative increase in electrical resistance, $\Delta R/R$, as a function of the C or N content of their solid solution in α-iron.

The relation between concentration and relative increase in resistance $\Delta R/R$ has been recorded in table I and plotted in *fig. 8*, both for C and N. It was very satisfactory that, after each purification process in moist hydrogen, weight and resistance exactly reached their original value. There is no reason for supposing $\Delta R/R$ to be anisotropic for a crystal structure of cubic symmetry. (This cubic symmetry disappears *e.g.* if the material is magnetized in some direction).

According to Köster [1]) 0·01% of C in solid solution corresponds to $\Delta R/R = 2{\cdot}5\%$. The content of C present as Fe_3C was determined by the colorometric method. By subtracting this quantity from the total C amount Köster obtained the content of C in solid solution. As shown by the data mentioned in table I our own measurements on the effect of C check well with Köster's.

For N his results are widely different from ours [2]). He observed that, on segregating from solid solution at 100 °C, 0.01 % of N only causes $\Delta R/R \approx 1\%$. At 100 °C N segregates as a highly dispersed nitride, and the much lower value of $\Delta R/R$ obtained by Köster compared with ours

is most probably explained by the fact that this nitride too is responsible for an increase in electrical resistance.

Quenching alone, without C or N being added, had no effect on the resistance.

TABLE I

Carbon		Nitrogen	
$3\,C_0$ (%)	$\frac{\Delta R}{R}$ (%)	$3\,C_0$ (%)	$\frac{\Delta R}{R}$ (%)
0	0	0	0
0·0018	0·53	0·0009	0·29
0·0024	0·61	0·0063	2·21
0·0046	1·17	0·0156	5·49
0·0066	1·64		

0·01 % of carbon → $\Delta R/R = 2{\cdot}5$ %.
0·01 % of nitrogen → $\Delta R/R = 3{\cdot}5$ %.

7. *The anisotropy measurements*

The decrement Δ as a function of the temperature has been measured for the [100]- and [111]-orientations for different concentrations. The time of oscillation was 1·6 sec. The maximum in the damping curve was reached for N at about 20 °C; for C the maximum lies somewhat higher, at 36 °C. The results are given in table II. In *fig. 9*, $\lambda_2/\lambda_1 = 2/\pi$. Δ_{max} has been plotted against $\Delta R/R$.

TABLE II

Carbon			
[100]		[111]	
$\frac{\Delta R}{R}$ (%)	$\frac{\lambda_2}{\lambda_1}$ (%)	$\frac{\Delta R}{R}$ (%)	$\frac{\lambda_2}{\lambda_1}$ (%)
0	0		
0·5	1·0	4·9	0·47
0·6	1·6		
0·8	1·2		
2·1	3·8		
3·3	5·5		
3·9	6·6[5]		

$\Delta R/R = 1$ % → $\lambda_2/\lambda_1 = 1{\cdot}72$ %. $\Delta R/R = 1$ % → $\lambda_2/\lambda_1 = 0{\cdot}1$ %.

Nitrogen			
[100]		[111]	
$\frac{\Delta R}{R}$ (%)	$\frac{\lambda_2}{\lambda_1}$ (%)	$\frac{\Delta R}{R}$ (%)	$\frac{\lambda_2}{\lambda_1}$ (%)
0	0		
5·3	4·8	6·0	0·41
7·4	6·8		
10·1	9·4		

$\Delta R/R = 1\ \% \rightarrow \lambda_2/\lambda_1 = 0{\cdot}92\ \%$. $\Delta R/R = 1\ \% \rightarrow \lambda_2/\lambda_1 = 0{\cdot}07\ \%$.

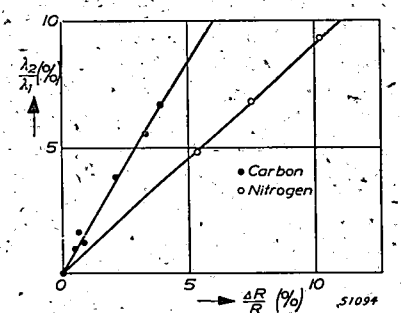


Fig. 9. The relative magnitude of the elastic relaxation as a function of the relative increase in electrical resistance caused by C or N in solid solution in α-iron.

The values of Δ_{max} have been corrected by subtracting the small damping observed at the purified material.

Thus from tables I and II the final results are:

1) λ_2/λ_1 is proportional to the concentration $3C_0$ for low values:

2) [100]-orientation: 0·01 % carbon $\rightarrow \lambda_2/\lambda_1 = 4{\cdot}3$ %
0·01 % nitrogen $\rightarrow \lambda_2/\lambda_1 = 3{\cdot}2$ %
[111]-orientation: 0·01 % carbon $\rightarrow \lambda_2/\lambda_1 = 0{\cdot}25$ %
0·01 % nitrogen $\rightarrow \lambda_2/\lambda_1 = 0{\cdot}24$ %.

From deflection measurements we got the following values for the elastic modulus.

$$E[100] \approx 1{\cdot}5.10^{12}\ \text{dyne/cm}^2;\ E[111] \approx 2{\cdot}4.10^{12}\ \text{dyne/cm}^2.$$

We do not pretend these values to be very accurate. The best data from literature are:

$$E[100] = 1{\cdot}3.10^{12}\ \text{dyne/cm}^2;\ E[111] = 2{\cdot}8.10^{12}\ \text{dyne/cm}^2.$$

8. *Results on polycrystalline material, and the effect of internal stresses on the elastic relaxation*

This research would not be complete without observations on polycrystalline material. In this case the relative magnitude λ_2/λ_1 has been measured for various C contents. The results are very irregular, as is evident from table III.

TABLE III

$\frac{\Delta R}{R}$ (%)	C content (%)	$\frac{\lambda_2}{\lambda_1}$ (%)
1·7	0·0068	2·3
1·8	0·0072	1·4
2·6	0·0104	3·4
3·2	0·0128	1·8 *)
4·0	0·0160	5·6
5·5	0·0220	3·0 *)
6·0	0·0240	4·6
6·3	0·0252	5·1
6·7	0·0268	4·8

Here we must remark that all values belong to different strips except the two marked with an asterisk which have been measured on the same specimen. All specimens, however, have been cut parallel to each other from the rolled carbonyl iron which served as basic material. In this way we are reasonably sure that the influence of texture is the same for all strips; after the heat treatment these were still of a fine-grained polycrystalline structure. We are not able to offer any explanation of the large scatter in the results, which was not observed with different [100]-strips cut from a common single crystal. At any rate it is striking that the two marked points which belong to the same specimen indeed show a linear relation. For the polycrystalline material quenching alone, without C being dissolved, gave rise to $\Delta R/R \approx 1$ %. In table III this increase has been corrected for.

To investigate the influence of internal stresses we chose as a test object a polycrystalline strip of carbonyl iron containing about 2 % of N. We preferred N to C as N segregates less rapidly (*cf.* section 10). In *fig. 10* the logarithmic decrement Δ is plotted against the temperature t. The curve with the higher maximum is characteristic for the freshly annealed material but, after straining it 10 % beyond the elastic limit, the lower curve was recorded. The data have not been corrected in the usual way for the damping observed on purified material, as this correction is of no interest here.

From the two curves in fig. 10 it is evident that internal stresses of the order of magnitude of 10 kg/mm², which we can expect after such a deformation, have only a very small effect on both the relaxation time and relative magnitude. The curve for the plastically deformed strips seems to be slightly smoother. The lower maximum too may be due to the fact that the N has segregated to some degree during the mechanical treatment.

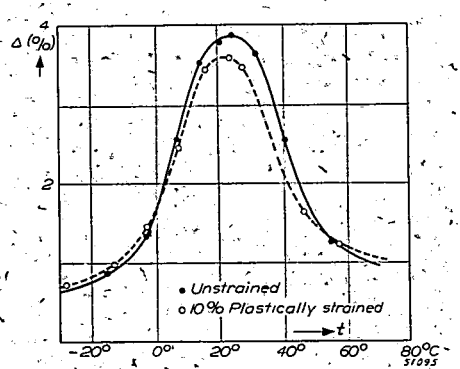


Fig. 10. Damping curve of an unstrained and a plastically strained iron strip loaded with nitrogen.

9. *Discussion*

We will discuss now our results in connection with the general theory of the elastic relaxation as given by Polder [9]). According to the general formula (5) the effect should rigorously disappear in the [111]-direction and reach a maximum value for the [100]-orientation. Experimentally the [111]-crystal only showed a very weak damping, the effect in the [100]-direction being 17 times as large for C and 14 times as large for N. A satisfactory result, in view of the fact that both crystals contained about 5 % of inclusions of an orientation differing from that of the basic crystal. Moreover the [100]- and [111]-orientation were only approximately correct. A more complete investigation would require the verification of the anisotropy factor $[1-3\ (\alpha_1^2\alpha_2^2 + \alpha_2^2\alpha_3^2 + \alpha_3^2\alpha_1^2)]$ for other directions too, but in view of the great difficulties in preparing the samples we have made no attempts in this direction.

In order to check the absolute value according to (5) the factor F^2, which is characteristic for the tetragonal deformation of the elementary cell caused by an interstitial foreign particle, must be known. Snoek [6]) has already pointed out that for C this factor can be derived from the cell dimensions of the tetragonal martensite as a function of the C content as

plotted in fig. 5. For in section 2 we saw that in this modification all C particles were located at interstices of one definite class. From fig. 5 we get

$$\left(\frac{e_{zz}}{c_z}\right)_{\substack{\text{tetr.}\\\text{mart.}}} = 0{\cdot}90 \text{ and } \left(\frac{e_{xx}}{c_z}\right)_{\substack{\text{tetr.}\\\text{mart.}}} = 0{\cdot}12\,.$$

Substituting these values in formula (5) and putting $V = 7{\cdot}15$ cm^3, $E[100] = 1{\cdot}3.10^{12}$ dyne/cm^2, $3C_0 = 0{\cdot}01$ % by weight, we get $\lambda_2/\lambda_1[100] = 4{\cdot}0$ %, in complete agreement with the experimental value of 4·3 %. We should, however, not attach too much weight to this numerical accordance, as the factor F^2 of course is not very accurate and the [100]-orientation is only approximated.

Nevertheless the result may point to the fact that in tetragonal martensite and in ferrite the dissolved C is located at crystallographically identical interstices. We base this conclusion upon the large difference in numerical value of the factor F^2 for the R_a^2- and R_a^4-structure. Since for tetragonal martensite we regard the $R_a^{2,z}$-structure as the most likely one, the ferritic solid solution of C, too, should be of the R_a^2-type. As mentioned before we did not expect this result for a stable phase like ferrite *).

Data from density measurements in the region of very small C contents can also furnish some information as to the exact location of the dissolved C particles in the ferrite lattice. So far as we know, such measurements, which must be made with great accuracy, have only been carried out by Köster on a carbon steel of a total C amount of about 0·2 % [1]). The C was previously present as cementite (Fe_3C), and by quenching from various temperatures the C content in solid solution was varied. The change in density resulting from the reaction $Fe + 3C \lesssim FeC_3$ is practically due to the volume change of the elementary cell of α-iron as connected with the transition of C into or from the solid solution.

In *fig. 11* the relative volume effect $\Delta V/V$ has been plotted against the C content in solid solution according to the data of Köster. On the other hand $\Delta V/V$ can easily be calculated for a definite structure from the dimensions of the elementary cell, if the lattice deformation by the dissolved particles is known from X-ray diagrams. For tetragonal and cubic martensite these dimensions were given in fig.5. The value of $\Delta V/V$ based on the tetragonal martensite structure is represented by the solid line in fig. 11. For comparison the dashed line gives the value of $\Delta V/V$ as derived from the data on cubic martensite. These curves too suggest that in ferrite

*) We discussed the difficulty with Mr J. D. Fast who suggested that both in tetragonal martensite and ferrite the large R_a^4-interstices should be occupied by C. In cubic martensite the C should be precipitated as highly dispersed Fe_3C. In that case, however, the large magnitude of the lattice deformation of tetragonal martensite can not easily be understood from the dimensions of the iron lattice and the C atom.

and tetragonal martensite the C is situated at crystallographically identical interstices.

No elastic relaxation effect can be expected either in tetragonal martensite or in austenite. In the former no microdiffusion is possible between interstices of different classes, whereas in the latter the lattice deformation by the dissolved particles will be symmetrical with regard to the cubic axes. In cubic martensite taken as a solid solution the occurrence of a small effect is not excluded, as microdiffusion may be possible to a certain degree.

For N the data necessary to evaluate the factor F are not available.

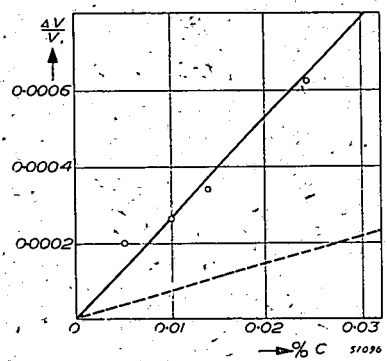


Fig. 11. Relative volume effect, $\Delta R/R$, of α-iron as a function of the C content of the solid solution according to Köster.

According to the theory of Polder the proportionality of the relative magnitude λ_2/λ_1 and the concentration C_0 is based on the assumption that a possible mutual interaction between the dissolved particles may be neglected. From figs 8 and 9 it is clear that this assumption is justified, at least in the region of low contents.

Further investigations are needed to clarify the irregular behaviour of the polycrystalline material. On the average, 0·01 % of C corresponds to a magnitude λ_2/λ_1 of about 2 %.

We should like to point out the possibility of using the elastic relaxation for the determination of extremely small quantities of C or N in α-iron. A magnitude $\lambda_2/\lambda_1 = 0{\cdot}1$ %, that is, a content of 0·001 % could easily be measured while chemical methods completely fail in this region. Apart from the irregularities in our data on the polycrystalline material the possibility of textures, often present in technical materials, prevents an exact quantitative determination in view of the strongly anisotropic character of the effect; but as a qualitative indication of extremely small quantities of C or N it should be a very sensitive method.

Formula (5) does not account for the effects occurring when an external constant stress is superposed on the periodic force. Theoretically [9]) a stress in the direction of one of the cubic axes, let us say the x-axis, of $p_z \approx 1$ kg/mm^2 causes a relative change in concentration $\Delta c_z/c_z \approx 1$ %.

We should therefore expect that internal stresses of the order of magnitude 1-10 kg/mm^2 will only have a small effect on the magnitude of λ_2/λ_1 (fig. 10).

Moreover from the damping curves in fig. 10 it is evident that the activation energy U in formula (11) is practically unaffected by internal stresses of this order of magnitude.

10. *The rate of segregation of* C *and* N *from a supersaturated solid solution in α-iron at 20 °C*

It is a matter of interest to know how rapidly and in what degree the segragation of C and N from their supersaturated solid solution in α-iron takes place. As we know, this segregation occurs as a carbide (Fe_3C) or nitride (Fe_4N).

This rate of segragation is most easily measured by suspending the strip loaded with C or N in the apparatus and by following the behaviour of the damping in the course of time. According to our conceptions this damping is a direct and quantitative indication of C or N in solid solution, and the relaxation effect provides a very convenient method for separating the amount of C or N dissolved from the total amount present; as such it should be preferred to other methods based upon the change in electrical resistance, or in mechanical or magnetic properties, since these last are also affected by segregation, *e.g.* depend on the degree of dispersion of the segregated substance.

As a basic material we used a polycrystalline carbonyl-iron strip for the study of C and a [100]-crystal of Armco iron in the case of N. The final constant values resulting after a time of observation of more than a month are much smaller than the correction which must be applied, and therefore these rest values are not very accurate. The results have been plotted in *fig. 12*, where for convenience a logarithmic time scale has been adopted.

The general aspect on comparing the curves for C and N is that a supersaturated solid solution of C precipitates much more easily. After quenching C starts segregating at once, N in observable quantities only when some 24 hours have passed. After a lapse of some days C has precipitated for the greater part, but for N this requires some weeks. For C the residual solubility at 20 °C is practically reached within two weeks; for N a period of four weeks is needed. The N content was about 2 %, the C content about 0·2 %.

For N, $\Delta_{20°} \approx \Delta_{max}$; from the shape of the damping curve it follows that $\Delta_{20°} \approx 1/3\ \Delta_{max}$ for C. Assuming a possible absolute error of 0·1 % in the

value of the residual damping, we conclude from the preceding data in section 7 that the solubility at 20 °C both of C and N in equilibrium with their carbide or nitride does not exceed 0·001 %. According to quite different methods the upper limit of this residual solubility was found by other authors to be 0·006 % for C [3]) and 0·001 % for N [2]).

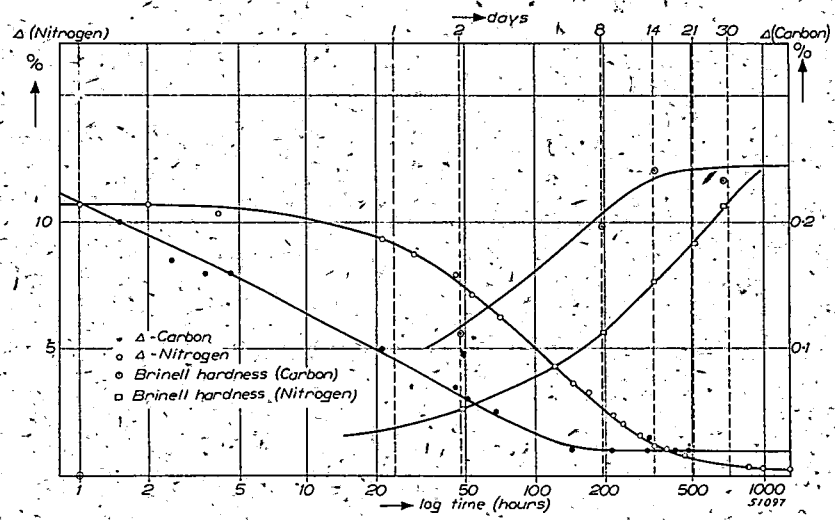


Fig. 12. The logarithmic decrement for an oscillating iron strip loaded with C (Δ-scale on the right) or N in solid solution plotted against log time.

Analytically the precipitation process for N can very well be characterized by the expression:

$$Q - Q_0 = \text{const. } e^{-A_N t} \qquad (12)$$

as long as the equilibrium state is not approached too closely. This means that

$$\frac{dQ}{dt} = A_N (Q - Q_0),$$

where Q is the quantity of dissolved N, Q_0 the residual solubility, and t the time. This is quite a reasonable result.

In *fig. 13*, $\ln(Q - Q_0)$ has been plotted against t, for N up to 336 hours, for C up to 69 hours, the scale being different in these two cases. For C the measurements are not very accurate in view of the small quantities involved, but if we still represent them by (12) we find $A_C \approx 5\, A_N$. Thus C segregates more rapidly than N, though from the values of τ deduced before we conclude that the rate of migration is much larger for N than for C.

It is well known that the segregation of C and N is accompanied by a change in mechanical properties such as an increase in Brinell hardness and tensile strength, a phenomenon usually called mechanical aging. The mechanical C an N aging process at 20 °C as observed by Köster [1] [2]) proceeds parallel in time with the segregation process as registered by our method based on elastic relaxation (fig. 12), which corroborates our observations in a satisfactory manner.

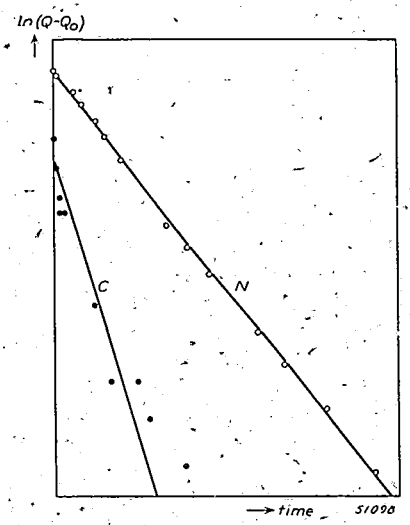


Fig. 13. ln $(Q-Q_0)$ plotted against time.

We admit hat our study has been incomplete as for instance we did not investigate the influence on the rate of segregation of C and N of other foreign elements in the Fe lattice such as Mn and Si, always present in technical iron. According to Köster [13]) the precipitation of N is highly affected by the presence of C. In our experiments C and N were never present together.

Finally during the N segregation-process the behaviour of the electrical resistance R was followed. For the purified [100]-crystal $R = 0{\cdot}00258\ \Omega$; for the nitrided crystal immediately after quenching $R = 0{\cdot}00277\ \Omega$,

20 days after quenching $R = 0{\cdot}00269\ \Omega$,

53 days after quenching $R = 0{\cdot}00268\ \Omega$.

So the limit value of R of the nitrided strip proves to be much higher than the original resistance of the purified strip containing no N at all. We have already remarked that the difference between these two values is due to the highly dispersed nitride.

Eindhoven, December 1946

REFERENCES

1) W. Köster, Arch. Eisenhüttenw. **2**, 503, 1928/29.
2) W. Köster, Arch. Eisenhüttenw. **3**, 637, 1929/30.
3) C. F. D. Yensen, J. Am. Inst. elect. Engrs **43**, 455, 1924.
4) C. H. Johansson, Arch. Eisenhüttenw. **11**, 241, 1937.
5) E. C. Bain, Trans. Am. Inst. Mining Met. Engr. **70**, 25, 1924.
6) J. L. Snoek, Physica, 's-Grav. **9**, 862, 1942.
7) K. Honda and N. Nishiyama, Sci. Rep. Tohoku Univ. **21**, 306, 1932.
8) G. Hägg, Stahl und Eisen **54**, 1328, 1934.
9) D. Polder, Philips Res. Rep. **1**, 5, 1945.
10) J. L. Snoek, Physica, 's-Grav. **8**, 711, 1941.
11) C. A. Edwards and L. B. Pfeil, J. Iron and Steel Inst. **109**, 129, 1924.
12) K. Honda and S. Kaya, Sci. Rep. Tohoku Univ. **15**, 721, 1926.
13) W. Köster, Arch. Eisenhüttenw., **4**, 145, 1930/31.
14) W. Seith and O. Kubaschewski, Z. Electrochem. **41**, 551, 1935.
15) W. Seith and Th. Dauer, Z. Electrochem. **44**, 256, 1938.
16) V. M. Goldschmidt, Chem. Ber. **60**, 1263, 1927.
17) L. Pauling, J. Am. Chem. Soc. **49**, 765, 1927.
18) N. J. Petch, J. Iron and Steel Inst. **145**, 111, 1942.
19) J. L. Snoek, Physica, 's-Grav. **5**, 663, 1938; **6**, 161, 1939; **6**, 591, 1939; **8**, 734, 1941; Chem. Weekblad **39**, 1, 1942.

ABSTRACTS

(Continued from page 356)

1709: N. G. de Bruijn: On the zeros of a polynomial and its derivative (Proc. Kon. Ned. Akad. Wet. Amsterdam **49**, 1037-1044, 1946).

It is proved that the sum of the absolute values of the imaginary parts of the roots of $f'(z)$ is equal to or less than $(1-1/n)$ times the corresponding sum for $f(z)$, in which $f(z)$ is a real polynomial of degree n in z, and $f'(z)$ its derivative. This theorem is extended and a few specializations are considered. It is unknown whether the inequality holds for polynomials with complex coefficients. It does so if all the roots are assumed to lie on the imaginary axis.

1710: H. B. G. Casimir and D. Polder: Influence of retardation on the London-van der Waals forces (Nature, London **158**, 787, 1946).

In the course of work on the stability of colloidal solutions, Overbeek arrived at the conclusion that, in order to obtain agreement between theory and experiment, it is necessary to assume that the London-van der Waals energy decreases more rapidly than R^{-6}. He pointed out that the retardation of the electrostatic forces might be responsible for such an effect, and that deviations from the R^{-6} law should become effective at a distance comparable with the wavelength λ corresponding to the excitation energies of the interacting atoms. Calculations with the aid of quantum electrodynamics prove that this suggestion holds true, the energy being proportional to R^{-7} rather than to R^{-6} for $R \ll \lambda$. Details of the quantum-mechanical calculation will be published elsewhere.

(Continued on page 399)

Philips Res. Rep. 2, 382-398, 1947

THE DISSOCIATION OF NITROGEN IN THE WELDING ARC

by J. D. FAST

541.123:621.791.75

Summary

For temperatures in the range from 5000-10000 °K the dissociation of nitrogen is computed on the basis of spectroscopic data, and for three different values of the dissociation energy, *viz.* 7·383, 8·573 and 9·764 electron volts; in view of a recent investigation by Gaydon and Penney the highest value is probably the correct one.
At 10000 °K and a pressure of one atmosphere the degree of dissociation is 99 % at least; at 5000 °K 25·1 % of the nitrogen will be dissociated if the dissociation energy amounts to 7·383 eV, but only 1·63% will be dissociated if the dissociation energy is 9·764 eV (see table VIII). Two cruder methods of computation give results that differ only slightly from these "exact" values (see table IX).

1. *Introduction*

Though it has repeatedly been pointed out [1]) that the dissociation of nitrogen is an important factor in the welding arc, reliable data concerning the degree of dissociation are still lacking. It may be true that Zeise [2]) calculated the dissociation of nitrogen at high temperatures on the basis of the spectroscopic values of the molar free enthalpy given by Giauque and Clayton [3]), but these calculations were not extended beyond 5000 °K; and besides, Zeise adopted 7·34 eV for the dissociation energy of nitrogen, a value that is today no longer universally accepted.

To supply more detailed information on this problem we shall in this paper compute the degree of dissociation of nitrogen between 5000 and 10000 °K for three different values of the dissociation energy, of which the highest, 9·764 eV, is probably the correct one according to Gaydon and Penney [4]). Calculations have also been carried out at 1800 and at 2750 °K, that is, in the neighbourhood of the melting and boiling points of iron, in view of their importance in electric steel-welding problems.

2. *The temperature of the welding arc*

Let us first recapitulate our present state of knowledge of the temperature in an electric arc.

At very low pressures the mean free path of the electrons in a gas discharge is large, and collisions with atoms are relatively seldom; between two collisions the electrons will be able to pick up a large amount of energy from the electric field, and, because of the scarcity of collisions and the pronounced differences in mass between electrons and atoms, only a small amount

of this energy will be passed on to the atoms *). The final result will be a stationary state such that the electron temperature is very high and the gas temperature comparatively low.

But the number of collisions, and consequently the transfer of energy from electrons to atoms, rapidly increases with the pressure, thus producing a drop in the electron temperature and a corresponding rise in the gas temperature. At a pressure of 1 atmosphere the exchange of energy is already so complete that in a monatomic gas the electron temperature is only a few per cent higher than the gas temperature. For the positive column of a mercury discharge the changes of both temperatures with the pressure are schematically represented in *fig. 1* [5]).

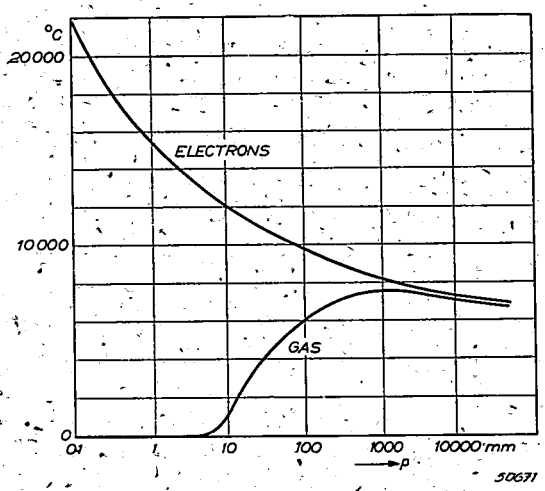


Fig. 1. Schematic representation of the electron and gas temperatures in the positive column of a mercury discharge as functions of the pressure.

For a diatomic gas, equality of electron and gas temperatures will be reached at even lower pressures, since in this case the existence of rotational and vibrational energy levels with their low energies of excitation will strongly promote the exchange of energy between molecules and electrons. Hence it may reasonably be assumed that no appreciable difference between electron and gas temperature will occur in a welding arc in atmospheric air.

*) If m_e represents the mass of the electron and M that of the atom, the fraction of the energy transferred from electron to atom in a single elastic collision is on the average, approximately,

$$f = \frac{2\,m_e}{M}.$$

From the distribution of intensities amongst the individual lines of CN bands, Ornstein and his associates [6]) deduced temperatures that for a carbon arc in air at atmospheric pressure varied between 6000 and 7000 °K. These determinations have subsequently been repeated by Lochte-Holtgreven and Maecker [7]) who took great care in eliminating the disturbing influence of self-absorption. In a quiescent arc — current in amps about equal to the diameter of the carbon electrodes in millimetres — of a length of 3 mm they observed a maximum temperature of 7600 °K *).

An increase in current was found to produce a decrease in the temperature to a minimum of circa 6800 °K, which is followed by a sharp rise; at this stage the arc is no longer quiescent but starts hissing. When the current had reached a value in amps of about twice the diameter of the electrodes in mm the temperature had risen to 9000 °K (current 20-30 amps, electrode diameter 14 mm).

The carbon arc has been investigated at still higher currents by Finkelnburg [8]) who observed an increasing contraction of the column of the arc above 80 amps, owing to the magnetic action of the current; this phenomenon which is independent of the diameter of the electrodes is fully developed at a current of about 130 amps, the current density in the contracted column then being in the neighbourhood of 3000 amps/cm^2 and the temperature circa 15000 °K [9]).

Since the currents used in welding practice are usually greater than 100 amps, it may safely be assumed that the temperature in the welding arc will be higher, and locally even much higher, than 5000 °K. This is corroborated by the experiments of Suits [10]) who used a rotating mirror camera to determine the velocity of sound waves passing an arc. For a 125-amps arc 1 cm long, and between a 3/16″ (0·5 cm)-coated welding rod as cathode and a large steel plate as anode, the observed average sound velocity was $2{\cdot}08 \times 10^5$ cm sec^{-1}, which corresponds to an average temperature of 6020 °K; locally even higher temperatures may be expected.

From these data it may be concluded that in studying the welding arc it will be of value to know the dissociation of nitrogen from 5000 °K upwards to temperatures at which the dissociation is nearly complete.

3. *Statistical-mechanical basis of the calculations*

The degree of dissociation of nitrogen can be computed as soon as we know either the free energy (Helmholtz free energy F) or the free enthalpy (Gibbs free energy G) of both molecular and atomic nitrogen. These quantities will be calculated on the basis of spectroscopic data in the following sections.

*) At 6 mm length the temperature was about 1000 °K lower.

The Helmholtz free energy of 1 mole of an ideal gas with reference to its value at absolute zero is represented by

$$f - u_0 = -kT \ln \frac{Z^N}{N!}, \tag{1}$$

where $Z^N/N!$ is the partition function (state sum) for the entire system, and

$$Z = \Sigma g e^{-E/kT} \tag{2}$$

the partition function for a single molecule.

In these expressions, $u_0 (= f_0)$ is the energy (free energy) at absolute zero, k is the Boltzmann constant, T the absolute temperature, N the Avogadro number, and g the degree of degeneration of the energy level E.

Now the general partition function Z can be resolved into the product of a partition function Z_{tr} for translation and Z_i for the so-called internal degrees of freedom of the molecule:

$$Z = Z_{tr} Z_i, \tag{3}$$

in which

$$Z_{tr} = \frac{v}{h^3} (2\pi mkT)^{3/2}, \tag{4}$$

where v is the volume containing the gas, h the Planck constant, and m the mass of a single molecule.

By substituting (3) and (4) in (1), and using Stirling's approximation

$$N! \approx \left(\frac{N}{e}\right)^N, \tag{5}$$

we obtain

$$f - u_0 = -kT \ln \left[\frac{ev}{Nh^3} (2\pi mkT)^{3/2} Z_i\right]^N, \tag{6}$$

e being the base of the neperian logarithms.

Using the relations

$$v = \frac{kNT}{p} \tag{7}$$

and

$$m = \frac{M}{N}, \tag{8}$$

where M is the molecular weight, we can rewrite the molar free enthalpy,

$$g - u_0 = f - u_0 + pv = f - u_0 + RT, \tag{9}$$

in the form

$$-\frac{g-u_0}{T} = \tfrac{5}{2} R \ln T + \tfrac{3}{2} R \ln M - R \ln p + R \ln \frac{2^{5/2} \pi^{3/2} e k^{5/2}}{h^3 N^{3/2}} - R + R \ln Z_i, \tag{10}$$

$R = kN$ being the gas constant. Inserting the standard values of the universal constants [11]), and going over to base-10 logarithms we obtain for a gas at 1 atmosphere (= $1{\cdot}01325 \times 10^6$ dynes/cm^2)

$$-\frac{g^0-u_0^0}{T} = 11{\cdot}437 \log T + 6{\cdot}862 \log M - 7{\cdot}281 + 4{\cdot}575 \log Z_i$$

cal degree^{-1} mole^{-1}, (11)

the superscript 0 indicating the standard value of the pressure (1 atm) and the subscript 0 the zero of the absolute temperature. This equation holds true for all gases at temperatures and pressures at which ideal-gas laws apply.

4. *The partition function of atomic nitrogen*

For the particular case of atomic nitrogen ($M = 14{\cdot}008$) equation (11) changes into

$$-\frac{g^0-u_0^0}{T} = 11{\cdot}437 \log T + 0{\cdot}585 + 4{\cdot}575 \log Z_i \text{ cal degree}^{-1} \text{ mole}^{-1}. \tag{12}$$

Up to a temperature of 10 000 °K appreciable contributions to the partition function are only coming from those states of the atom in which the 7 electrons have a $1s^2\ 2s^2\ 2p^3$ configuration, corresponding to a ^4S, a ^2D and a ^2P state. Of these the ^4S term is the ground level; the ^2D term, which consists of a fourfold $^2D_{1^1/_2}$ and a sixfold $^2D_{2^1/_2}$ level, is lying 19202 cm^{-1} higher [12]).

The ^2P term is located still higher, at 28808 cm^{-1} [12]), and is composed of a twofold $^2P_{1/_2}$ and a fourfold $^2P_{1^1/_2}$ state. Both at ^2D and ^2P the doublet separation is so small that it can safely be neglected.

From the foregoing we infer that the internal partition function of atomic nitrogen may simply be represented by

$$Z_i(\mathrm{N}) = 4 + 10\, e^{-19202\, hc/kT} + 6\, e^{-28808\, hc/kT}. \tag{13}$$

For the range of temperatures in which we are interested the values computed from this equation have been compiled in table I together with the negative, reduced *), molar free enthalpies at 1 atmosphere as given by (12).

*) In thermodynamics and statistical mechanics the term "reduced" is frequently used to indicate division by the absolute temperature.

TABLE I

Internal partition function Z_i and standard values of the negative, reduced, molar free enthalpy of atomic nitrogen

Temp. (°K)	Z_i(N)	$-\left(\frac{g^0-u_0^0}{T}\right)_N$ (cal mole^{-1} degree^{-1})
1800	4·000	40·570
2750	4·000	42·675
5000	4·041	45·665
6000	4·106	46·603
7000	4·209	47·417
8000	4·350	48·146
9000	4·525	48·809
10000	4·727	49·419

5. *The partition function for molecular nitrogen*

Applying equation (11) to molecular nitrogen ($M = 28{\cdot}016$) we obtain

$$-\frac{g^0-u_0^0}{T} = 11{\cdot}437 \log T + 2{\cdot}651 + 4{\cdot}575 \log Z_i \text{ cal degree}^{-1}\text{ mole}^{-1}. \quad (14)$$

The evaluation of Z_i for molecules is more cumbersome than it is for atoms. In the following pages four different methods of computation will be indicated, three of which will actually be used.

a. *The completely exact method*

For the nitrogen molecule the internal partition function can, in principle, be evaluated, as it was done for the nitrogen atom, by computing and adding the terms $ge^{-E/kT}$ for each of the energy levels deduced from spectral data.

However, the very large number of rotational and vibrational energy levels that have to be taken into account renders this method too laborious for practical purposes.

b. *The "exact" method*

Mathematical simplifications have, however, led to a method of computation that is nearly as accurate as the completely exact method, but involves much less labour. This method will be discussed in section 6.

c. *The crudest approximation*

For a rough approximation it may be assumed that the internal partition function is the product of three different partition functions, one for the electronic, one for the vibrational, and one for the rotational states:

$$Z_i \approx Z_{el} Z_v Z_r . \quad (15)$$

In so doing only the electronic ground state of the molecule is taken into account, and the interaction between vibration and rotation is disregarded. Within the frame of this approximation it is also customary to consider the molecule as a harmonic vibrator and as a rigid rotator [13]). The vibrational frequency, ω, and the moment of inertia, I, are taken to be those corresponding to the equilibrium position of the atoms in the classical (naive) model of the molecule, *i.e.* the position corresponding to the minimum potential. These values will be indicated by ω_e and by I_e (suffix e for equilibrium). *)

d. *The approximate method*

Since, for the high temperatures which we are considering, the preceding method is very inaccurate we will follow a slightly different course in this section, applying a comparatively crude correction to account for the deviations from harmonicity, and for the influence of the vibration and of the centrifugal forces on the moment of inertia.

We shall start from the fact that at 7000 °K still about 90% of the molecules occupy one of the first five vibrational energy levels of the electronic ground state. To calculate these levels we made use of the approximate formula

$$E_v = \omega_e(v + \tfrac{1}{2}) - x_e\,\omega_e(v + \tfrac{1}{2})^2 \ \text{cm}^{-1}, \qquad (16)$$

where, as already stated in different words above, ω_e is the classical frequency of a vibration with an infinitesimal amplitude, v is the vibrational quantum number, and $x_e\,\omega_e$ is the so-called anharmonicity constant which is added to account for the fact that the vibrations are not truly harmonic. By (16) the mutual distances between the first five vibrational levels are

$$\begin{aligned} 0 \to 1 \quad & \omega_e - 2\,x_e\omega_e\,, \\ 1 \to 2 \quad & \omega_e - 4\,x_e\omega_e\,, \\ 2 \to 3 \quad & \omega_e - 6\,x_e\omega_e\,, \\ 3 \to 4 \quad & \omega_e - 8\,x_e\omega_e\,. \end{aligned}$$

To simplify the calculations we will assume that in the region from 5000-10 000 °K the deviations from harmonicity can roughly be taken into account by considering the molecule as a harmonic oscillator but with the smaller frequency $\omega_e - 5\,x_e\omega_e$ cm^{-1}, a value corresponding to the average distance between the first five vibrational levels.

Locating the zero of the energy scale at the lowest vibrational level (zero-point vibration) the partition function for the vibration will be

$$Z_v = e^{-0/kT} + e^{-(\omega_e - 5\,x_e\omega_e)hc/kT} + e^{-2\,(\omega_e - 5\,x_e\omega_e)hc/kT} + \ldots$$

*) The symbol ω designates *not* the frequency ν in sec^{-1} but the quantity $\omega = \nu/c$ in cm^{-1}; in the following this quantity will be called the frequency.

or

$$Z_v = \frac{1}{1 - e^{-(\omega_e - 5\, x_e\omega_e)hc/kT}} \tag{17}$$

To evaluate the rotational levels for each of the vibrational states we shall make use of the approximate expression

$$E_r = B_v J(J+1) - DJ^2(J+1)^2\,, \tag{18}$$

where

$$B_v = \frac{h}{8\pi^2\, cI_v} \text{ cm}^{-1} \tag{19}$$

is a measure of the inverse value of the moment of inertia I; J is the rotational quantum number, and D may be considered as a quantity representing the influence of the centrifugal forces ($D \ll B$). Thus D accounts in an approximate way for the fact that the molecule is not a rigid rotator.

In (18) we have still to account for the influence of the vibration on the moment of inertia or, in other words, for the interaction between vibration and rotation, and this we shall do by writing

$$B_v = B_e - \alpha(v + \tfrac{1}{2})\,, \tag{20}$$

in which

$$B_e = \frac{h}{8\pi^2 cI_e},$$

I_e, as defined above, being the moment of inertia that the molecule would possess if its nuclei were fixed in their equilibrium positions; α is a positive constant small in comparison with B_e. As this equation shows, B_v decreases, and consequently the moment of inertia I increases, with increasing values of the vibrational quantum number v.

We shall now again use the same principle of approximation that was adopted in computing the partition function for the vibrations; we shall suppose that for temperatures ranging from 5000 to 10 000 °K the changes in the moment of inertia can be accounted for by taking B_v constant but equal to its average value for the five lowest vibrational levels; in other words, we use the value of B corresponding to an average vibrational quantum number $v = 2$. By (20) this value is

$$B_2 = B_e - 2{\cdot}5\,\alpha. \tag{21}$$

Having regard to a symmetry factor 2, the partition function for the rotation will then at high temperatures approximately be given by

$$Z_r = \frac{kT}{2hcB_v} + \left(\frac{D}{B_v}\right)\left(\frac{kT}{hcB_v}\right)^2 + 6\left(\frac{D}{B_v}\right)^2\left(\frac{kT}{hcB_v}\right)^3, \tag{22}$$

where $v = 2$.

Since the electronic ground level of an N_2 molecule is a singlet Σ state, X ($^1\Sigma_g^+$), the electronic partition function will be

$$Z_{el} = 1 , \tag{23}$$

since in this method of computation the electronic ground level is the only electronic level that has to be taken into account.

In virtue of (15) the total internal partition function is equal to the product of (17), (22) and (23), and by inserting the numerical data of the second column of table IV we obtain the values that have been entered in the second column of table II. For comparison the values computed according to the crudest method of approximation described on page 387 ($x_e\omega_e$, α, and D neglected) have been added in column 3, and the values computed according to the "exact" method (see next section) have been added in column 4.

TABLE II

Values of the internal partition function of N_2 for temperatures ranging from 5000 to 10 000 °K

Temperature °K	$Z_i(N_2)$ approximate method	$Z_i(N_2)$ crudest approx.	$Z_i(N_2)$ "exact" method
5000	1857	1757	1833
6000	2552	2406	2530
7000	3359	3156	3346
8000	4279	4007	4287
9000	5312	4961	5362
10000	6460	6019	6581

6. *The "exact" evaluation of the partition function for molecular nitrogen*

In an "exact" calculation of the internal partition function of N_2 we have to account more accurately for the interaction between vibration and rotation and for the deviations from harmonicity. This can be done by computing with the aid of (20) and (22) the partial rotational partition function, $Z_r(v)$, separately for each of the vibrational levels, v, and combining these into a rotational-vibrational partition function by means of the equation

$$Z_{rv} = Z_r(0) + e^{-E_1hc/kT} \cdot Z_r(1) + e^{-E_2hc/kT} \cdot Z_r(2) + \ldots \tag{24}$$

where E_1, E_2, . . ., in cm^{-1}, represent the energies (above the level of the zero-point vibration) of the successive vibrational levels for a non-rotating molecule. These energies can be computed with the aid of (16). By proceeding in this manner it is possible to arrive in a much shorter time at practically the same results as would be obtained by the completely exact term-by-term summation mentioned at the beginning of the foregoing section.

However, at the high temperatures with which we are concerned so many vibrational levels contribute to the partition function that even the simplified method of equation (24) becomes too laborious for practical purposes. It is fortunate therefore that the mathematical simplification can be pushed one step further by combining the vibrational constants in one series with the rotational constants [14])[15]). This may most conveniently be done by making use of the theoretical relations

$$\frac{a}{B_e} = \left(\frac{6B_e}{\omega_e}\right)\left[\left(\frac{x_e\omega_e}{B_e}\right)^{1/2} - 1\right], \tag{25}$$

$$\frac{D}{B_e} = 4\left(\frac{B_e}{\omega_e}\right)^2, \tag{26}$$

by which a and D are fixed in terms of ω_e, B_e and $x_e\omega_e$. For the electronic ground level these formulae yield very satisfactory results, the values computed from (25) and (26) being $a = 0{\cdot}017$ cm^{-1} and $D = 5{\cdot}8.10^{-6}$ cm^{-1} as compared with the empirical values of $0{\cdot}018$ and $5{\cdot}77.10^{-6}$ cm^{-1} given in table IV.

We will write the total contribution to the internal partition function by the electronic ground level in the form

$$Z_i(gr) = Z_{el}\, Z_v\, Z_r\, Z_c\,. \tag{27}$$

in which Z_v refers to a harmonic oscillation with a frequency corresponding to the distance $\omega_e - 2\,x_e\omega_e$ between the two lowest vibrational levels; Z_r refers to a rigid rotator with a moment of inertia corresponding to that of the zero-point vibration; and Z_c is a correction factor embodying the three different corrections that have already been specified before, *viz.* anharmonicity, centrifugal forces and interaction between vibration and rotation.

By the formulae of the previous section we obtain for Z_v and Z_r

$$Z_v = (1 - e^{-u})^{-1}, \tag{28}$$

where

$$u = \frac{hc(\omega_e - 2\,x_e\omega_e)}{kT}, \tag{29}$$

and

$$\ln Z_r = -\ln 2\sigma + \sigma/3\,, \tag{30}$$

where

$$\sigma = \frac{hc(B_e - a/2)}{kT}. \tag{31}$$

The first term on the right-hand side of (30) corresponds to the same term in (22), and the second term $\sigma/3$ provides a small correction resulting from substituting a sum by an integral. The correction terms involving D in (22) are missing in (30), but these terms have been transferred to Z_c together with all the other corrections. Z_c is given by [16])

$$\ln Z_c = \frac{1}{u}(2\gamma + 6\gamma^{1/3}x^{1/3} + 2x) + (3\gamma - 3\gamma^{1/3}x^{1/3} - 2x) + \\ + \frac{u}{6}(-3\gamma + 3\gamma^{1/3}x^{1/3} + 5x) - \frac{u^2}{6}x + \frac{u^3}{120}(\gamma - \gamma^{1/3}x^{1/3} + x), \quad (32)$$

in which

$$\gamma = \frac{B_e}{\omega_e}, \quad (33)$$

$$x = \frac{x_e\omega_e}{\omega_e}. \quad (34)$$

With the aid of equations (27) to (34) and the numerical data of the second column of table IV the contribution of the electronic ground level to the partition function given in table III has been computed.

TABLE III

"Exact" contributions of the electronic ground level to the partition function of N_2

T (°K)	$Z_i[N_2; X(^1\Sigma_g^+)]$	T (°K)	$Z_i[N_2; X(^1\Sigma_g^+)]$
1800	373·5	7000	3345
2750	688	8000	4283
5000	1833	9000	5347
6000	2530	10000	6538

So far we have only been considering the electronic ground level, and the contributions of the higher electronic levels have still to be taken into account. These contributions are so small, however, that the departure from harmonicity, the interaction between vibration and rotation, and the influence of the centrifugal forces on the moment of inertia can all be neglected in these higher terms. Z_c in (27) will then be equal to unity, and Z_v and Z_r can be computed from equations (28) to (31).

At 10 000 °K only three states have to be considered *viz.*: a $^3\Sigma$ state, A($^3\Sigma_u^+$), a $^3\Pi$ state, B($^3\Pi_g$), and a $^1\Pi$ state, a($^1\Pi_u$). The chief molecular con-

stants for these states have been collected in table IV, A_0 representing the energy level of the zero-point vibration above the corresponding level of the electronic ground state. *)

TABLE IV

Molecular constants of N_2 in cm^{-1} [17])

	$X(^1\Sigma_g^+)$	$A(^3\Sigma_u^+)$	$B(^3\Pi_g)$	$a(^1\Pi_u)$
A_0	0	49774	59328	68957
ω_e	2359·61	1460·4	1732·84	1692·3
$x_e\omega_e$	14·445	13·93	14·44	13·32
B_e	2·007	1·440	1·643	1·642
α	0·018	0·013	0·018	0·021
D	$5·77.10^{-6}$			

Since the triplet state $A(^3\Sigma_u^+)$ is located at 49774 cm^{-1} above the ground level its contribution to Z_e will be

$$Z_e;\ A(^3\Sigma_u^+) = 3\ e^{-49774hc/kT}, \qquad (35)$$

the factor 3 accounting for the triplet character of this level. Consequently the total contribution of this electronic state to the partition function will be

$$Z_i;\ A(^3\Sigma_u^+) = 3\ e^{-49774hc/kT}\ Z_v Z_r\ , \qquad (36)$$

which by equations (28) to (31), and by use of the numerical data collected in table IV, reduces to

$$Z_i;\ A(^3\Sigma_u^+) = 3\ e^{-71600/T}\left(\frac{1}{1-e^{-2061/T}}\right) 0{\cdot}242\ T. \qquad (37)$$

For low values of J the $B(^3\Pi_g)$ state is split up into three levels, $^3\Pi_0$, $^3\Pi_1$ and $^3\Pi_2$, each of which is a doublet state in itself owing to "Λ-doubling". The mutual distances between these $^3\Pi$ levels are so small that they may be disregarded, so that we have

$$Z_i;\ B(^3\Pi_g) = 6\ e^{-59328hc/kT}\ Z_v Z_r\ , \qquad (38)$$

which by (28) to (31) and table IV becomes

$$Z_i;\ B(^3\Pi_g) = 6\ e^{-85340/T}\left(\frac{1}{1-e^{-2451/T}}\right) 0{\cdot}213\ T. \qquad (39)$$

*) In table IV we omitted a $a'(^1\Sigma_u^-)$ state since the corresponding value of A_0 is unknown. Herzberg [20]) places this level above 100 000 cm^{-1}, while Gaydon [21]) assumes it to be as low as circa 60 000 cm^{-1}. Even if this last assumption should be correct the contribution to the partition function would not exceed 2 at 10 000 °K.

For the $a(^1\Pi_u)$ state each level is a doublet (Λ-doubling) so that we must have

$$Z_i\,;\,a(^1\Pi_u) = 2\,e^{-68957hc/kT}\,Z_v Z_r \qquad (40)$$

or proceeding as above

$$Z_i\,;\,a(^1\Pi_u) = 2\,e^{-99200/T}\left(\frac{1}{1-e^{-2396/T}}\right)0{\cdot}213\,T\,. \qquad (41)$$

By adding the contributions computed from (37), (39), and (41) to the values given in table III we may now calculate the total internal partition function of N_2; the result has been recorded in table V.

TABLE V

Total internal partition function of N_2 at different temperatures according to the "exact" method

T (°K)	1800	2750	5000	6000	7000	8000	9000	10000
$X(^1\Sigma_g^+)$	373·5	688	1833	2530	3345	4283	5347	6538
$A(^3\Sigma_u^+)$	0	0	0	0	1	3	11	30
$B(^3\Pi_g)$	0	0	0	0	0	1	4	12
$a(^1\Pi_u)$	0	0	0	0	0	0	0	1
	373·5	688	1833	2530	3346	4287	5362	6581

It will be seen from this table that the higher electronic states contribute only a small amount to the total internal partition function, even at temperatures as high as 10 000 °K.

By inserting the values of table V in (14) we finally computed the negative, reduced, molar free enthalpies of molecular nitrogen at a pressure of one atmosphere as given in table VI.

TABLE VI

Internal partition function Z_i and standard values of the negative, reduced, molar free enthalpies of molecular nitrogen for different temperatures

Temp. (°K)	Z_i (N_2) "exact"	$-\left(\frac{g^0-u_0^0}{T}\right)_{N_2}$ (cal mole^{-1} degree^{-1})
1800	373·5	51·650
2750	688	54·968
5000	1833	59·886
6000	2530	61·431
7000	3346	62·752
8000	4287	63·908
9000	5362	64·937
10000	6581	65·868

7. *The dissociation of* N_2 *in the temperature range from* 5000 *to* 10 000 °K

For the reaction

$$N_2 \rightleftharpoons 2\,N \tag{42}$$

the reaction constant

$$K_p = \frac{p_N^2}{p_{N_2}} \tag{43}$$

will be given by the well-known thermodynamical relation

$$-RT \ln K_p = \Delta G^0 , \tag{44}$$

where ΔG^0 is the change in free enthalpy produced by converting 1 mole ideal gaseous N_2 at a pressure of 1 atmosphere into 2 moles of ideal gaseous atomic N at the same pressure. Consequently we shall have

$$\Delta G^0 = 2g^0(N) - g^0(N_2) , \tag{45}$$

in which the symbol g^0 designates the molar free enthalpies of the ideal gases at the standard pressure.

Instead of (44) and (45) we may also use the equivalent equations

$$-R \ln K_p = 2\frac{g^0(N) - u_0^0(N)}{T} - \frac{g^0(N_2) - u_0^0(N_2)}{T} + \frac{2\,u_0^0(N) - u_0^0(N_2)}{T} \tag{46}$$

or

$$\log K_p = \frac{g^0(N_2) - u_0^0(N_2)}{4{\cdot}575\ T} - 2\,\frac{g^0(N) - u_0^0(N)}{4{\cdot}575\ T} - \frac{\Delta U_0^0}{4{\cdot}575\,T} \tag{47}$$

The first two terms on the right-hand side of this last equation have been computed in the foregoing sections, and in the third term ΔU_0^0 designates the dissociation energy at absolute zero which can in principle be deduced from spectroscopic data. Up till a short time ago 7·383 eV was generally accepted, but recently this value has been criticized by various authors.

On the basis of comparative spectroscopical studies Schmid and Gerö[18]) concluded that the dissociation energy should be 5 volts only, but this must be considered as far too low both on spectroscopic and chemical grounds. If, for instance, we estimate the dissociation energy, according to the method of Birge and Sponer[17]), by extrapolating the vibrational levels of the electronic ground state, that is, by means of the equation

$$\Delta U_0^0 = \frac{\omega_e^2}{4\,x_e\omega_e} - \frac{\omega_e}{2} , \tag{48}$$

the numerical data of table IV lead to a value of 11·8 eV. Though the values obtained by this method are known to be always too high, they hardly ever exceed the true value by more than 2 eV; there is, therefore, even reason to assume that the former value of 7·383 eV is still too low.

A more important contribution to this problem is due to Gaydon and Penney [4]). According to the interpretation that they brought forward, certain pre-dissociation phenomena [19]) are compatible with only three possible values of the dissociation energy, namely 7·383, 8·573 and 9·764 eV; but the value of 7·383 eV must be discarded since it violates the non-crossing rule of Hund, by which potential curves belonging to the same electronic species should not intersect. Gaydon and Penney hold the view that of the two remaining values the higher, 9·764 eV, is most probably correct.

In view of the existing uncertainty we have preferred to compute the reaction constant (from (47) and tables I and VI) separately for each of the three values of the dissociation energy just mentioned. The results are entered in table VII.

TABLE VII

Reaction constant K_p for the reaction $N_2 \rightleftharpoons 2N$ for three different values of the dissociation energy ΔU_0^0

Temp. (°K)	$\Delta U_0^0 = 7{\cdot}383$	$\Delta U_0^0 = 8{\cdot}573$	$\Delta U_0^0 = 9{\cdot}764$
1800	$5{\cdot}83.10^{-15}$	$2{\cdot}75.10^{-18}$	$1{\cdot}26.10^{-21}$
2750	$1{\cdot}27.10^{-7}$	$8{\cdot}45.10^{-10}$	$5{\cdot}51.10^{-12}$
5000	0·268	0·0170	0·00107
6000	5·51	0·554	0·0551
7000	49·5	6·90	0·956
8000	266	47·5	8·42
9000	1020	220	47·2
10000	3050	768	192

From K_p it is now comparatively easy to compute the degree of dissociation of nitrogen. If we have a total of n moles of N_2 of which a fraction α is dissociated, we will have a mixture containing $n(1-\alpha)$ moles of N_2 and $2n\alpha$ moles atomic N. Since in an ideal gas the partial pressures are proportional to the number of moles, we shall have

$$p_N = C\,2n\alpha \tag{49}$$

$$p_{N_2} = Cn(1-\alpha) \tag{50}$$

$$p_{tot} = Cn(1+\alpha) \tag{51}$$

where C is some constant; therefore

$$K_p = \frac{p_N^2}{p_{N_2}} = C\,\frac{4n^2\alpha^2}{n(1-\alpha)}. \tag{52}$$

Eliminating C from (51) and (52) we obtain

$$K_p = p\,\frac{4\alpha^2}{1-\alpha^2} \tag{53}$$

or

$$\alpha = \sqrt{\frac{K_p}{K_p + 4p}}, \tag{54}$$

p being the total pressure.

For a pressure of 1 atmosphere the values of α computed in this manner have been recorded in table VIII.

TABLE VIII

Degree of dissociation, α, for N_2 at a total pressure of 1 atmosphere for three different values of the dissociation energy ΔU_0^0

Temp. °K	$\Delta U_0^0 = 7{\cdot}383$	$\Delta U_0^0 = 8{\cdot}573$	$\Delta U_0^0 = 9{\cdot}764$
1800	$3{\cdot}82.10^{-8}$	$8{\cdot}29.10^{-10}$	$1{\cdot}77.10^{-11}$
2750	$1{\cdot}78.10^{-4}$	$1{\cdot}45.10^{-5}$	$1{\cdot}17.10^{-6}$
5000	0·251	0·0651	0·0163
6000	0·761	0·349	0·117
7000	0·962	0·796	0·439
8000	0·993	0·960	0·823
9000	0·998	0·991	0·960
10000	0·999	0·997	0·990

8. *Discussion of the results*

As we see from table VIII, 99% at least of the nitrogen is dissociated at a temperature of 10 000 °K and a pressure of 1 atmosphere; at 5000 °K still 25·1% will be dissociated if the dissociation energy is 7·383 eV, but only 1·63% if the dissociation energy is 9·764 eV.

For the purpose of comparison the degree of dissociation, α, calculated according to the three different methods of computation ("crudest", "approximate" and "exact") is represented in table IX; the computations have been made for three different temperatures and for the three different values of the dissociation energy. It goes without saying that all methods of computation will lead to the same values of α at very low and at very high temperatures, *viz.* 0% and 100% dissociation respectively. The table shows that in the intermediate region the relative error will never exceed 3%, even if the crudest method is adopted.

TABLE IX

Degree of dissociation, α, for N_2 at a total pressure of 1 atmosphere according to three different methods of calculation and for three different temperatures and three different values of the dissociation energy

Temp. °K	Method	$\Delta U_0^0 = 7{\cdot}383$	$\Delta U_0^0 = 8{\cdot}573$	$\Delta U_0^0 = 9{\cdot}764$
5000	"crudest"	0·256	0·0664	0·0167
	"approximate"	0·249	0·0646	0·0162
	"exact"	0·251	0·0651	0·0163
7000	"crudest"	0·964	0·804	0·450
	"approximate"	0·962	0·795	0·438
	"exact"	0·962	0·796	0·439
10000	"crudest"	0·999	0·998	0·991
	"approximate"	0·999	0·997	0·990
	"exact"	0·999	0·997	0·990

Eindhoven, February 1947

REFERENCES

1) Compare D. Séférian, Thesis, Paris 1935; R. F. Wyer, Gen. Elect. Rev. **42**, 170-172, 1939.
2) H. Zeise, Z. Elektrochem. **40**, 885-890, 1934.
3) W. F. Giauque and J. O. Clayton, J. Am. Chem. Soc. **55**, 4875-4889, 1933.
4) A. G. Gaydon and W. G. Penney, Proc. Roy. Soc. A **183**, 374-388, 1945.
5) Handbuch der Lichttechnik, part I, Berlin 1938.
6) L. S. Ornstein and W. R. van Wijk, Proc. Kon. Ned. Akad. Wetenschappen Amsterdam **33**, 44-46, 1930.
L. S. Ornstein and H. Brinkman, Physica, 's-Grav. **1**, 797-824, 1934.
7) W. Lochte-Holtgreven and H. Maecker, Z. Physik **105**, 1-15, 1937.
8) W. Finkelnburg, Z. Physik **112**, 305-325, 1939; **113**, 562-581, 1939; **114**, 734-746, 1939; **116**, 214-234, 1940; **117**, 344-357, 1941; **119**, 206-222, 1942.
9) W. Finkelnburg, Naturwissensch. **32**, 105-111, 1944.
10) C. G. Suits, Physics **6**, 315-322, 1935.
11) W. de Groot, Ned. Tijdschr. Natuurk. **9**, 497-505, 1942; R. T. Birge, Rev. Modern Physics **13**, 233-239, 1941.
12) R. F. Bacher and S. Goudsmit, Atomic Energy States, New York and London 1932.
13) Compare *e.g.* R. H. Fowler and E. A. Guggenheim, Statistical Thermodynamics, Cambridge 1939.
14) L. S. Kassel, J. Chem. Phys. **1**, 576-585, 1933.
15) A. R. Gordon and C. Barnes, J. Chem. Phys. **1**, 297-307, 1933.
16) J. E. Mayer and M. G. Mayer, Statistical Mechanics, New York 1940.
17) H. Sponer, Molekülspektren I, Berlin 1935; G. Herzberg, Molekülspektren und Molekülstruktur, I. Zweiatomige Moleküle, Dresden und Leipzig 1939.
18) Hungarian publications, cited by J. G. Valatin in J. Chem. Phys. **14**, 568, 1946.
19) Compare A. van der Ziel, Physica, 's-Grav. **1**, 353-362, 1934, and Thesis, Groningen 1934.
20) G. Herzberg, Phys. Rev. **69**, 362-365, 1946.
21) A. G. Gaydon and R. E. Worley, Nature **153**, 747, 1944.
A. G. Gaydon, Proc. Roy. Soc. A **182**, 286-301, 1944.

ABSTRACTS

(Continued from page 381)

1711: P. J. Bouma: Die Grassmannschen Gesetze der Farbmischung (Physica, 's-Grav. **12**, 545-552, 1946) (Grassmann's laws of additive colour mixing).

The author proposes a new formulation of Grassmann's laws of additive colour mixing, which is axiomatically as pure as possible and which does not contain more than is necessary for the construction of elementary colorimetry. The correctness of the formulation is proved by working out this construction in rough lines. Finally the questions of continuity, arising in connexion herewith, are treated and are reduced to the existence of thresholds.

1712: E. J. W. Verwey (+ J. E. Asscher): Lattice structure of the free surface of alkalihalide crystals (Rec. Trav. Chim. Pays-Bas **65**, 521-528, 1946).

Calculations have been made of the lattice distortion at the free crystal surface of the cube face of alkali halide crystals on the basis of Born's lattice theory, with the result that the displacement of the positive surface ions in a direction perpendicular to the surface differs considerably from that of the negative. A kind of ionic double layer is formed with the negative charge pointing away from the crystal. The corresponding dipole is more or less compensated by the dipole moment induced in the negative ions. This distortion lowers the surface energy as is illustrated by the calculation for the case of NaBr.

1713: J. Th. G. Overbeek and P. W. O. Wijga: On electro-osmosis and streaming potentials in diaphrams (Rec. Trav. Chim. Pays-Bas **65**, 556-563, 1946).

It is stated that the streaming potential E/P and the volume of liquid transported by electro-osmosis v/i are equal, independent of the structure of the diaphragm and independent of its surface conductance. The value of these ratios which, in the absence of surface conductance, equal $\varepsilon\zeta/4\pi\eta\lambda$ (ε = dielectric constant, η = viscosity, λ = electric conductivity) diminishes when surface conductance is present. This decrease can be accurately taken into account when the diaphram consists of a single capillary of constant diameter. In the case of real diaphragms, consisting of a network of capillaries of different shape and dimensions, the correction factor for surface conductance cannot be computed! It has been the com-

mon practice to estimate the correction factor from the ratio of the observed to the calculated electric resistance of the diaphragm. This procedure is shown to be erroneous, leading to values of the ζ-potential that are essentially too low.

1714: J. M. Stevels: The physical porperties of glass in relation to its structure (J. Soc. Glass Technology **30**, 31-53, 1946).

A fairly extensive account is given of the knowledge we have at present of the structure of glass in general. On this basis some physical properties of glass, *viz.* the density and the electric conductivity, are discussed in detail. The density of "normal" glasses can be calculated satisfactorily by means of a formula containing only two constants. One of these constants gives valuable information about the structure of the glass, especially about the way in which the "excess" of oxygen is taken up by the network. The relationship between electric conductivity and temperature is briefly discussed.

1715: J. A. Haringx: The Notch-Impact Test according to Schnadt (De Ingenieur **58**, Mk 15-17, 1946).

A new type of notch-impact bar has been introduced by Schnadt. With this bar the bending pressure does not act upon the material itself but on a hardened steel pin. Since the centre line of this pin functions more or less as the axis of rotation and the material itself is thus entirely subjected to a tensile load, this is really a sort of impact-tensile test. The greatest advantage lies in the fact that every test bar breaks, even if no other notch is made in it. This makes it possible, with the same cross-section of fracture (3×10 mm^2), to make several notches with greatly different radius of curvature. Consequently a series of widely divergent stress conditions can be created.

1716*: H. A. Klasens: The light output of zinc sulphide on irradiation with alpha rays (Trans. Faraday Soc. **52**, 666-668, 1946).

A critical review of the literature dealing with the light output of ZnS excited with alpha rays shows that in most measurements values are found of 10-15% and of the same order of magnitude as those found with cathode rays, in accordance with the theories of Thomson, Bethe, and others.

The author refutes the much higher value (80%) given by Riehl without giving accurate data about his measurements.

VOL. 2 NO. 6 DECEMBER 1947

Philips Research Reports

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

R 60 *Philips Res. Rep. 2, 401-419, 1947*

THE DIODE AS CONVERTER AND AS DETECTOR *)

by J. HAANTJES and B. D. H. TELLEGEN

537.543:621.396.622.71

Summary

The current through a diode, to which is applied, in addition to an A.C. voltage and a D.C. voltage, a small extra voltage v, can be developed into a power series of v whose coefficients are Fourier series. The magnitude of these coefficients is calculated for a diode that has a linear characteristic in the pass direction. By confining ourselves to the term linear in v, four-pole equations and equivalent circuits can be set up for the diode as converter and as detector. From these equivalent circuits various properties and quantities can be deduced. The fluctuations of the diode as converter can also be represented with the help of the equivalent circuit. Upon conversion as well as upon detection it is found favourable to give the diode a small internal resistance.

1. *The general expression for the diode current*

In this article the properties of the diode as converter and as detector will be investigated. Firstly, in both cases an A.C. voltage $V_h \cos \omega_h t$ and a D.C. voltage V_0 act on the diode (*fig. 1*). When the diode is used as a converter this A.C. voltage corresponds to the oscillator voltage; when it is used as a detector this voltage corresponds to the voltage of the carrier. The voltage on the diode, v_d, is then equal to $V_h \cos \omega_h t - V_0$. The current through the diode, i_d, will change periodically with the time, namely with the angular frequency ω_h. If we confine ourselves to frequencies at

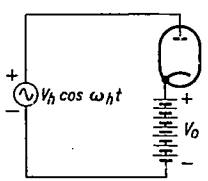


Fig. 1. Diode circuit with A.C. voltage and D.C. voltage; the signs + and — indicate the sense in which the voltages are considered positive.

*) Published in Dutch in Tijdschr. Ned. Radiogenootschap **10**, 237, 1943. Investigations in the same line were recently carried out by Peterson and Llewellyn [1]).

which the influence of the finite transit time of the electrons in the diode may be neglected, the current, like the voltage, will be an even function of the time and may thus be represented by the Fourier series

$$i_d = i_0 + i_1 \cos \omega_h t + i_2 \cos 2\omega_h t + \ldots + i_n \cos n\omega_h t + \ldots, \quad (1)$$

where

$$i_0 = \frac{1}{2\pi} \int_{-\pi}^{+\pi} i_d \, \mathrm{d}\omega_h t$$

and

$$i_n = \frac{1}{\pi} \int_{-\pi}^{+\pi} i_d \cos n\omega_h t \, \mathrm{d}\omega_h t \quad (n \geqq 1).$$

The quantities $i_0, i_1, \ldots, i_n, \ldots$ will depend upon V_h, upon V_0, and upon the characteristic of the diode.

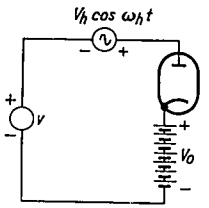


Fig. 2. Diode circuit in which a small extra voltage is introduced.

When a small variation is introduced in the D.C. voltage V_0, the magnitude of the diode current can be developed into a Taylor series of this variation ΔV_0. In *fig. 2* this variation of the bias is indicated separately by the extra voltage v. The polarity is so chosen that

$$\Delta V_0 = -v. \quad (2)$$

The series then becomes:

$$\left.\begin{aligned} i_d = {} & i_0 - \frac{\mathrm{d}i_0}{\mathrm{d}V_0} v + \frac{1}{2!} \frac{\mathrm{d}^2 i_0}{\mathrm{d}V_0^2} v^2 - \frac{1}{3!} \frac{\mathrm{d}^3 i_0}{\mathrm{d}V_0^3} v^3 + \ldots \\ & + \left(i_1 - \frac{\mathrm{d}i_1}{\mathrm{d}V_0} v + \frac{1}{2!} \frac{\mathrm{d}^2 i_1}{\mathrm{d}V_0^2} v^2 - \frac{1}{3!} \frac{\mathrm{d}^3 i_1}{\mathrm{d}V_0^3} v^3 + \ldots \right) \cos \omega_h t + \ldots \\ & + \left(i_n - \frac{\mathrm{d}i_n}{\mathrm{d}V_0} v + \frac{1}{2!} \frac{\mathrm{d}^2 i_n}{\mathrm{d}V_0^2} v^2 - \frac{1}{3!} \frac{\mathrm{d}^3 i_n}{\mathrm{d}V_0^3} v^3 + \ldots \right) \cos n\omega_h t + \ldots \end{aligned}\right\} \quad (3)$$

The coefficients of (3) can thus be calculated from those of (1). It is possible to arrange (3) according to ascending powers of v whereupon it takes on the form

$$\left.\begin{aligned} i_d = {} & A_0 + A_1 \cos \omega_h t + A_2 \cos 2\omega_h t + \ldots + A_n \cos n\omega_h t + \ldots \\ & + (\alpha_0 + \alpha_1 \cos \omega_h t + \alpha_2 \cos 2\omega_h t + \ldots + \alpha_n \cos n\omega_h t + \ldots)\, v + \\ & + (\beta_0 + \beta_1 \cos \omega_h t + \beta_2 \cos 2\omega_h t + \ldots + \beta_n \cos n\omega_n t + \ldots)\, v^2 + \\ & + (\gamma_0 + \ldots + \gamma_n \cos n\omega_h t + \ldots)\, v^3 + \ldots \end{aligned}\right\} \quad (4)$$

This expression now remains correct when v is an A.C. voltage. The momentary value of the current through the diode, as long as the influence of the transit time of the electrons may be neglected, is determined exclusively by the momentary value of the voltage on the diode. Thus when in (4) the momentary value of the A.C. voltage is substituted for v, the correct value of the anode current is obtained. Expression (4) for the diode current will accordingly be used in the following.

2. *The diode with a linear characteristic*

To obtain an estimate of the magnitude of the coefficients occurring in (4) these coefficients will be calculated for a certain case, namely for a diode that has a constant internal resistance R in the pass direction and an infinitely high internal resistance in the opposite direction (*fig. 3*). When such a diode is placed into the circuit of fig. 1, then

$$\left.\begin{aligned} & i_d = \frac{V_h \cos \omega_h t - V_0}{R} && \text{for} \quad V_h \cos \omega_h t > V_0 \\ \text{and} \quad & i_d = 0 && \text{for} \quad V_h \cos \omega_h t < V_0 . \end{aligned}\right\} \quad (5)$$

In *fig. 4* the voltage on the diode is represented as a function of the time. The current is proportional to the height of the shaded area which represents the quantity $V_h \cos \omega_h t - V_0$ for the times that it is positive. In order to simplify the calculation an auxiliary angle φ is introduced which is determined by those values of $\omega_h t$ for which the current becomes equal to zero, thus by

$$V_0 = V_h \cos \varphi . \quad (6)$$

The desired Fourier coefficients then become

$$i_0 = \frac{1}{2\pi} \int_{-\pi}^{+\pi} i_d \, d\omega_h t = \frac{1}{\pi} \int_0^{\varphi} \frac{V_h(\cos \omega_h t - \cos \varphi)}{R} \, d\omega_h t = \frac{V_h}{\pi R} (\sin \varphi - \varphi \cos \varphi).$$

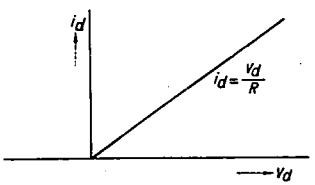


Fig. 3. Linear diode characteristic.

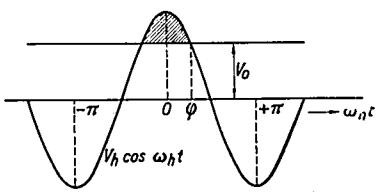


Fig. 4. The voltage on the diode.

By similar calculations the following is found:

$$i_1 = \frac{V_h}{\pi R}\left(\varphi - \frac{\sin 2\varphi}{2}\right),$$

$$i_n = \frac{V_h}{\pi R}\left\{\frac{\sin(n-1)\varphi}{n(n-1)} - \frac{\sin(n+1)\varphi}{n(n+1)}\right\} \quad (n \geqq 2).$$

The coefficients of series (1) are hereby determined for this case. Now the coefficients of series (4) can be determined. The A's are identical with the already calculated values of the i's. The coefficients α are calculated as follows:

$$\alpha_0 = -\frac{\mathrm{d}i_0}{\mathrm{d}V_0} = \frac{1}{V_h \sin\varphi}\frac{\mathrm{d}i_0}{\mathrm{d}\varphi} = \frac{\varphi}{\pi R}.$$

In a similar way it is found that

$$\alpha_n = -\frac{\mathrm{d}i_n}{\mathrm{d}V_0} = \frac{2\sin n\varphi}{n\pi R} \quad (n \geqq 1).$$

In an analogous manner the β- and γ-coefficients can be determined. The results finally have the following form:

$$\left.\begin{aligned}
A_0 &= \frac{V_h}{\pi R}(\sin\varphi - \varphi\cos\varphi),\\
A_1 &= \frac{V_h}{\pi R}\left(\varphi - \frac{\sin 2\varphi}{2}\right),\\
A_n &= \frac{V_h}{\pi R}\left\{\frac{\sin(n-1)\varphi}{n(n-1)} - \frac{\sin(n+1)\varphi}{n(n+1)}\right\} \quad (n \geqq 2),\\
\alpha_0 &= \frac{\varphi}{\pi R},\\
\alpha_n &= \frac{2\sin n\varphi}{n\pi R} \quad (n \geqq 1),\\
\beta_0 &= \frac{1}{2\pi R V_h}\frac{1}{\sin\varphi},\\
\beta_n &= \frac{1}{\pi R V_h}\frac{\cos n\varphi}{\sin\varphi} \quad (n \geqq 1),\\
\gamma_0 &= -\frac{1}{6\pi R V_h^2}\frac{\cos\varphi}{\sin^3\varphi},\\
\gamma_n &= -\frac{1}{6\pi R V_h^2}\frac{(n+1)\cos(n-1)\varphi - (n-1)\cos(n+1)\varphi}{\sin^3\varphi} \quad (n \geqq 1).
\end{aligned}\right\} \quad (7)$$

It may be seen from these coefficients that the A's are proportional to V_h, that the α's do not contain V_h, while the β's are inversely proportional to V_h, and the γ's inversely proportional to V_h^2. When it is desired to use the diode as converter or as detector the α's are of primary importance. The terms with β and γ cause in general undesired non-linear effects. With a given φ and v an increase in V_h will thus decrease the magnitude of these effects.

3. *The substitution for the D.C. voltage source of a capacitively shunted resistance*

The D.C. voltage source V_0 of figs 1 and 2 can be replaced by a resistance R_0 shunted by a large capacitance, so that only the D.C. resistance remains. Since the D.C. through the diode amounts to A_0, the D.C. voltage then becomes

$$V_0 = A_0 R_0 = \frac{V_h R_0}{\pi R} (\sin \varphi - \varphi \cos \varphi) = V_h \cos \varphi,$$

thus

$$\tan \varphi - \varphi = \frac{\pi R}{R_0}, \tag{8}$$

which establishes the relation between R, R_0, and φ.

The current of the frequency ω_h amounts to $A_1 \cos \omega_h t$. The conductance G_h that the diode forms for this frequency, thus amounts to

$$G_h = \frac{A_1}{V_h} = \frac{1}{\pi R} \left(\varphi - \frac{\sin 2\varphi}{2} \right). \tag{9}$$

The currents of the frequencies $n\omega_h$ ($n \geqq 2$) are of little interest to us. We shall assume that there are no impedances for them present in the diode circuit, so that they do not cause voltages of those frequencies.

With (8), (6), and (9), φ, V_0/V_h, and $G_h R_0/2$ can be calculated as functions of R/R_0. These are drawn in *fig. 5*. It is striking that at very small values of R/R_0 the φ takes on relatively large values. For $R/R_0 = 0{\cdot}01$, for example, φ is already 25°; V_0/V_h and $G_h R_0/2$ are thereby about 10 per cent smaller than at $R = 0$. For small values of R/R_0 we have approximately:

$$\varphi \approx \left(\frac{3\pi R}{R_0} \right)^{1/3} \left\{ 1 - \tfrac{2}{15} \left(\frac{3\pi R}{R_0} \right)^{2/3} \right\},$$

$$G_h \approx \frac{2}{R_0} \left\{ 1 - \tfrac{3}{5} \left(\frac{3\pi R}{R_0} \right)^{2/3} \right\},$$

$$V_0 \approx V_h \left\{ 1 - \tfrac{1}{2} \left(\frac{3\pi R}{R_0} \right)^{2/3} \right\}.$$

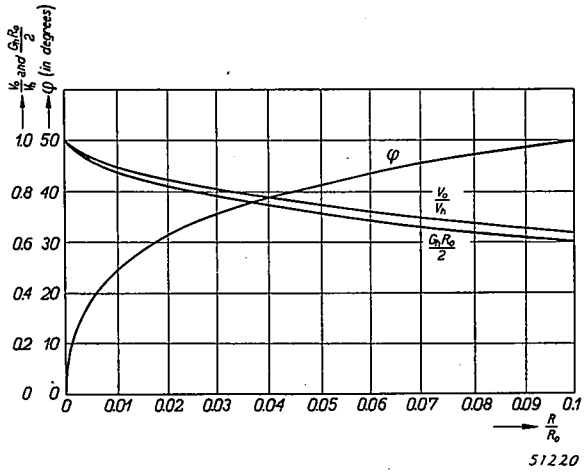


Fig. 5. The angle φ, the D.C. voltage V_0, and the conductance G_h as functions of R/R_0.

THE DIODE AS CONVERTER

4. *The equivalent circuit of the diode as converter*

If the diode is to be used as converter, a high-frequency circuit is introduced in the diode circuit, as indicated in *fig. 6*, which is tuned to the signal and to which signal voltage is supplied, and an intermediate-frequency circuit on which the intermediate-frequency current due to conversion causes an intermediate-frequency voltage. The diode with oscillator voltage and D.C. voltage can now be conceived of as a four-pole by means of which the H.F. circuit and the I.F. circuit are coupled. We shall set up the equations of this four-pole.

When a small signal voltage of frequency ω_i is applied to the diode, the terms with the α's in the diode current i_d are of primary importance. These contain, in addition to a component with the frequency ω_i, also components with the frequencies $\omega_h \pm \omega_i$, $2\omega_h \pm \omega_i$, *etc.* Since we are only interested in the behaviour of the circuit for the signal frequency

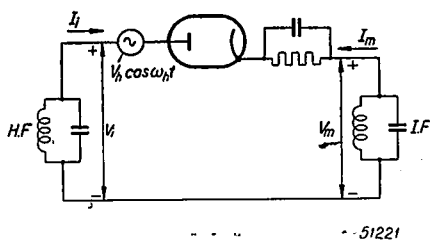


Fig. 6. Diode-conversion-circuit diagram.

and for the intermediate frequency $\omega_h + \omega_i$, $\omega_h - \omega_i$, or $\omega_i - \omega_h$, we may confine ourselves to the terms with α_0 and α_1 of (4) and thus write

$$i_d = (\alpha_0 + \alpha_1 \cos \omega_h t)v\,. \tag{10}$$

The voltage v on the diode consists of H.F. voltage and I.F. voltage. If the H.F. voltage is $V_i \cos(\omega_i t + \varphi_i)$ and the I.F. voltage is $V_m \cos(\omega_m t + \varphi_m)$ then (see fig. 6 for the sign)

$$v = V_i \cos(\omega_i t + \varphi_i) - V_m \cos(\omega_m t + \varphi_m)\,.$$

In the input current we are only interested in the frequencies in the neighbourhood of the signal frequency ω_i (amplitude I_i, phase ψ_i). We shall assume that for the other frequencies in the diode current no impedances are present in input or output circuit, so that they do not cause any voltages to appear. In particular we disregard the impedance which the H.F. circuit could exhibit for the so-called image frequency $2\omega_h - \omega_i$.

By considering the terms with the frequencies ω_i and ω_m separately two new equations are obtained from (10). We assume first that $\omega_m = \omega_i + \omega_h$ or that $\omega_m = \omega_i - \omega_h$. For the input current the following then holds:

$$I_i \cos(\omega_i t + \psi_i) = \alpha_0 V_i \cos(\omega_i t + \varphi_i) - \underset{\text{freq. } \omega_i}{\{\alpha_1 \cos \omega_h t\,.\, V_m \cos(\omega_m t + \varphi_m)\}} =$$

$$= \alpha_0\, V_i \cos(\omega_i t + \varphi_i) - \frac{\alpha_1}{2} V_m \cos(\omega_i t + \varphi_m).$$

For the output current

$$-I_m \cos(\omega_m t + \psi_m) = -\alpha_0 V_m \cos(\omega_m t + \varphi_m) + \underset{\text{freq. } \omega_m}{\{\alpha_1 \cos \omega_h t\,.\, V_i \cos(\omega_i t + \varphi_i)\}} =$$

$$= -\alpha_0\, V_m \cos(\omega_m t + \varphi_m) + \frac{\alpha_1}{2} V_i \cos(\omega_m t + \varphi_i)\,.$$

The two equations can be written in the form

$$\left.\begin{aligned} \mathbf{I}_i &= \alpha_0 \mathbf{V}_i - \frac{\alpha_1}{2}\mathbf{V}_m\,, \\ \mathbf{I}_m &= -\frac{\alpha_1}{2}\mathbf{V}_i + \alpha_0 \mathbf{V}_m\,, \end{aligned}\right\} \tag{11}$$

in which we have passed over to the method of complex notation for the A.C. quantities. Equations (11) give a relation between the amplitudes and phases, and do not contain the frequency. They are entirely analogous to the equations of a four-pole, although they here give the relation between amplitudes and phases of different frequencies. We may conceive of these equations as the four-pole equations of the converter circuit. This

four-pole can be represented by three conductances connected in a triangle (*fig. 7*).

For the case where $\omega_m = \omega_h - \omega_i$ the calculation is subject to a slight alteration. In that case the equation for the signal current and the intermediate-frequency current are

$$I_i \cos(\omega_i t + \psi_i) = \alpha_0 V_i \cos(\omega_i t + \varphi_i) - \frac{\alpha_1}{2} V_m \cos(\omega_i t - \varphi_m)$$

and

$$-I_m \cos(\omega_m t + \psi_m) = -\alpha_0 V_m \cos(\omega_m t + \varphi_m) + \frac{\alpha_1}{2} V_i \cos(\omega_m t - \varphi_i).$$

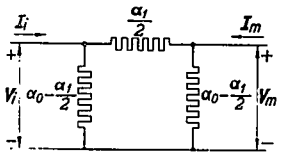


Fig. 7. Equivalent circuit of the diode converter.

The phase angles of the currents due to conversion are reversed in sign. When we pass over to the complex notation, therefore, the conjugate-complex quantities must be used for these components which will be marked by an asterisk. Instead of (11) thus we now obtain

$$\left.\begin{aligned} \mathbf{I}_i &= \alpha_0 \mathbf{V}_i - \frac{\alpha_1}{2} \mathbf{V}_m^*, \\ \mathbf{I}_m &= -\frac{\alpha_1}{2} \mathbf{V}_i^* + \alpha_0 \mathbf{V}_m. \end{aligned}\right\}$$

In the last equation all the quantities can also be replaced by their conjugate-complex quantities, whereupon we obtain

$$\left.\begin{aligned} \mathbf{I}_i &= \alpha_0 \mathbf{V}_i - \frac{\alpha_1}{2} \mathbf{V}_m^*, \\ \mathbf{I}_m^* &= -\frac{\alpha_1}{2} \mathbf{V}_i + \alpha_0 \mathbf{V}_m^*. \end{aligned}\right\} \qquad (12)$$

From (12) it is evident that the equivalent circuit of fig. 7 may now also be used, provided we replace the current and the voltage at the output terminals by the conjugate-complex quantities.

5. *Use of the equivalent circuit for the calculation of several quantities*

From the equivalent circuit of fig. 7 all kinds of quantities can now immediately be derived. If, for example, the output terminals are connected to an I.F. circuit with a conductance G_m, the *input conductance* for the signal frequency amounts to

$$a_0 - \frac{a_1}{2} + \frac{a_1/2(a_0 - a_1/2 + G_m)}{a_1/2 + a_0 - a_1/2 + G_m} = \frac{a_0^2 - a_1^2/4 + a_0 G_m}{a_0 + G_m}. \tag{13}$$

In the same way, the *output conductance* for the intermediate frequency when the input terminals are connected to a H.F. circuit with the conductance G_i, amounts to

$$\frac{a_0^2 - a_1^2/4 + a_0 G_i}{a_0 + G_i}. \tag{14}$$

From the equivalent circuit it follows also that when a H.F. voltage $\mathbf{V}_i$ is applied to the input terminals, the following voltage occurs on the I.F. circuit:

$$\mathbf{V}_m = \frac{a_1/2}{a_1/2 + a_0 - a_1/2 + G_m} \mathbf{V}_i,$$

so that one may speak of a *voltage-conversion factor*

$$\frac{a_1/2}{a_0 + G_m}. \tag{15}$$

When a H.F. circuit is connected to the input terminals (*fig. 8*), and to it a H.F. current $\mathbf{I}$ is applied while the output terminals are short-circuited, there occurs in it an I.F. current

$$\mathbf{I}_m = -\frac{a_1/2}{a_1/2 + a_0 - a_1/2 + G_i} \mathbf{I}, \tag{16}$$

so that one may speak of a *current-conversion factor*

$$F_c = \frac{a_1/2}{a_0 + G_i}. \tag{17}$$

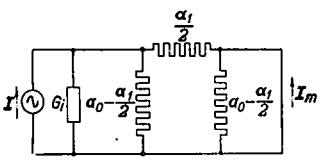


Fig. 8. Equivalent circuit of the conversion circuit for the calculation of the current conversion.

We may further inquire as to the voltage on the I.F. circuit with a conductance G_m when a H.F. current **I** is applied in parallel to the H.F. circuit with the conductance G_i. With the help of the diagram of *fig. 9* it can be found that

$$\mathbf{V}_m = \frac{\alpha_1/2}{\alpha_0^2 - \alpha_1^2/4 + \alpha_0(G_i + G_m) + G_i G_m}\mathbf{I}. \qquad (18)$$

When attention is paid not only to the frequencies to which the circuits are tuned, but also to the neighbouring frequencies, the expressions (13)-(18) can still be used if G_i in them is replaced by the admittance $\mathbf{Y}_i$, and if, when $\omega_m = \omega_i \pm \omega_h$, G_m is replaced by the admittance $\mathbf{Y}_m$ and, when $\omega_m = \omega_h - \omega_i$, G_m is replaced by the admittance $\mathbf{Y}_m^*$.

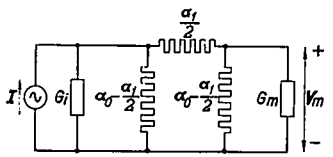


Fig. 9. Equivalent circuit of the conversion circuit for the calculation of the relation between output voltage and input current.

At first glance this seems a somewhat strange result for the last case, but it becomes understandable when the following is kept in mind. When a frequency is considered that is higher than the frequency to which the H.F. circuit is tuned, the intermediate frequency corresponding to this is lower than the frequency to which the I.F. circuit is tuned. The I.F. admittance is thus inductive for that frequency, so that its conjugate complex is capacitive.

The reaction of the I.F. circuit on the H.F. circuit thus always behaves like a normal circuit that is connected to the H.F. circuit via the equivalent circuit of the diode. The circuit then behaves as if it were tuned to the same frequency as the H.F. circuit, and as if the admittance had the same variation in the vicinity of the signal frequency as that of the I.F. circuit in the vicinity of the intermediate frequency. If the I.F. circuit is conceived of as a parallel connection of a conductance, a capacitance, and an inductance, and this, moreover, has a small relative band-width (*i.e.* band-width small compared with the resonance frequency), this can be achieved by replacing the I.F. circuit by a H.F. circuit with the same conductance and the same capacitance, but whose inductance is altered so much that the circuit is H.F.-tuned.

Instead of the H.F. equivalent circuit thus obtained, an I.F. equivalent circuit can also be set up by replacing the H.F. circuit in an analogous manner by an I.F. circuit.

6. *The diode with a linear characteristic as converter*

It may now be asked how the expressions derived look when the values from (7) are substituted for the α's, and use is made of (8). It is then found that in the equivalent circuit of the diode the parallel resistances become approximately equal to $2R_0$ for not too large values of φ, while the series resistance becomes approximately equal to $3^{-1/3}\,\pi^{2/3}\,R^{2/3}\,R_0^{1/3}$. In general it will be desired to keep the series resistance small and the parallel resistance large. This corresponds to a small value of R and a large value of R_0. From this it is therefore evident that it is always favourable to work with a diode that has a small internal resistance. If the impedances of the H.F. circuit and the I.F. circuit are given, one must thus on the one hand choose R_0 so large that the parallel resistance no longer has much influence, and on the other hand R_0 must not be chosen too large because the series resistance would then have too much influence.

When by a correct choice of the diode and the resistance R_0 the ideal state has been succesfully approached, the circuit behaves almost like two circuits in parallel; and the input conductance approaches G_m, the output conductance G_i, and the two conversion factors approach unity.

The effect of a large conductance in series, $G_s = \alpha_1/2$, and a small conductance in parallel, $G_p = \alpha_0 - \alpha_1/2$, can easily be understood by reference to the equivalent circuit. The following expressions are thus found approximately:

input conductance $$G_m\left(1-\frac{G_m}{G_s}\right)+2G_p, \tag{13a}$$

output conductance $$G_i\left(1-\frac{G_i}{G_s}\right)+2G_p, \tag{14a}$$

voltage-conversion factor $$1-\frac{G_m}{G_s}, \tag{15a}$$

current-conversion factor $$1-\frac{G_i}{G_s}. \tag{17a}$$

Expression (18) for the output voltage as a function of the input current then gives

$$\mathrm{V}_m \approx \frac{\mathrm{I}}{G_i+G_m}\left(1-\frac{2G_p}{G_i+G_m}-\frac{1/G_s}{1/G_i+1/G_m}\right). \tag{18a}$$

Thus two loss terms occur, one due to the branches in parallel and one to the branch in series.

7. *Fluctuations of the diode converter*

The fluctuations of an admittance $Y = G + jB$ can be characterized by giving the fluctuation current when this admittance is short-circuited. The expression for this is

$$\overline{di^2} = 4kTG\,dv; \tag{19}$$

k is Boltzmann's constant, T the absolute temperature, and dv the frequency region considered.

For the fluctuation current of the unsaturated diode there exists an analogous expression [2])

$$\overline{di^2} = 4bkT_cS\,dv\,. \tag{20}$$

T_c is the cathode temperature, S the slope at the point where the diode is set, while b is a constant amounting to about 0·6.

When the diode is used as converter, however, the slope is not constant, but according to (4) equals

$$a_0 + a_1 \cos \omega_h t + a_2 \cos 2\omega_h t + \ldots$$

In order to ascertain the fluctuation current in a certain frequency region, the average over time of this expression must be taken. With no H.F.- or I.F.-tuned circuit in the diode circuit the fluctuation current is thus given by

$$\overline{di^2} = 4bkT_c\,a_0\,dv\,. \tag{21}$$

This fluctuation current contains H.F. as well as I.F. components. When we include a H.F.-tuned circuit in the diode circuit, the H.F. fluctuation component of the diode current will cause a H.F. fluctuation voltage thereon, which by conversion will again give rise to an I.F. fluctuation current. This I.F. fluctuation current will exhibit a certain correlation with the original I.F. fluctuation current of the diode.

In order to investigate this correlation, one can first examine the result of a fluctuation peak in the diode current at the moment $t = \tau$ (*fig. 10*). All frequency components are present to an equal extent and have a maximum at $t = \tau$. When the H.F. component is $a \cos \omega_i(t-\tau)$, the I.F. component will thus be $-a \cos \omega_m(t-\tau)$ (the minus sign occurs because we consider the I.F. current positive in the opposite direction as the H.F. current; see fig. 6). The H.F. diode current $a \cos \omega_i(t-\tau)$ corresponds to a current I of $-a \cos \omega_i(t-\tau)$ applied to the H.F. circuit (compare fig. 8). This current, according to (16) and (17), results in an I.F. current whose amplitude amounts to $F_c a$ and whose phase is determined by

$$[\cos \omega_h t \cos \omega_i(t-\tau)] = \underset{\text{freq. } \omega_m}{[\cos \{\omega_h(t-\tau) + \omega_h\tau\} \cos \omega_i(t-\tau)]} = \underset{\text{freq. } \omega_m}{} = \tfrac{1}{2} \cos \{\omega_m (t-\tau) \pm \omega_h\tau\}\,.$$

The plus sign refers to $\omega_m = \omega_h \pm \omega_i$, the minus sign to $\omega_m = \omega_i - \omega_h$. The two I.F. components $-a \cos \omega_m(t-\tau)$ and $F_c\, a \cos \{\omega_m(t-\tau) \pm \omega_h\tau\}$ can now be combined into a single I.F. component, the square of whose amplitude amounts, according to the cosine law, to

$$a^2(1-2F_c\cos\omega_h\tau + F_c^2).$$

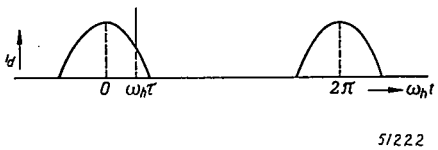


Fig. 10. Current through the diode with fluctuation peak at $t = \tau$.

The contributions from peaks that appear at different moments in the diode current, must now be summed. Because these peaks are not correlated, the summation of the squares of the amplitudes is permissible. According to (20), the average value of a^2 at a certain τ is proportional to the momentary value of the diode slope. In our case the momentary value of the diode slope is given by

$$a_0 + a_1 \cos\omega_h\tau + a_2 \cos 2\omega_h\tau + \ldots$$

The I.F. fluctuation current due to the diode will thus be proportional to

$$\frac{1}{2\pi}\int_{-\pi}^{+\pi} (a_0 + a_1\cos\omega_h\tau + a_2\cos 2\omega_h\tau + \ldots)(1-2F_c\cos\omega_h\tau + F_c^2)\,d\omega_h\tau = a_0 - a_1F_c + a_0F_c^2.$$

When we short-circuit the H.F. circuit, F_c becomes equal to zero and the above expression becomes equal to a_0. From this it follows that

$$\frac{\text{mean square of the I.F. diode fluctuation current with H.F. circuit}}{\text{mean square of the I.F. diode fluctiation current with no H.F. circuit}} = 1 - \frac{a_1}{a_0}F_c + F_c^2. \quad (22)$$

The I.F. fluctuation current due to the diode with H.F. circuit is thus, in connection with (21), determined by

$$d\overline{i^2} = 4bkT_c\,(a_0 - a_1F_c + a_0F_c^2)\,d\nu\,. \quad (23)$$

For ascertaining how large this fluctuation current is in a given case, the equivalent circuit can again be used. It is possible to verify that expression (23) for the fluctuation current also occurs when it is assumed

that each conductance of the equivalent circuit has fluctuations corresponding to the temperature bT_c. The fluctuations of each conductance G can then be represented by a fluctuation current in parallel with this conductance whose mean square is given by

$$\overline{di^2} = 4bkT_c\, G\,dv .$$

If one calculates thus, with the help of the equivalent circuit, the fluctuation current in the short-circuited output circuit when the conductance G_i is included in the input circuit, expression (23) is found when the fluctuation currents of the various conductances are considered to be independent of each other. From this equivalent circuit it is immediately clear that the fluctuations due to the diode become smaller, the closer the ideal state is approached in which a large series conductance and a small parallel conductance are present.

THE DIODE AS DETECTOR

8. *Detection of a single-side-band signal*

A case of detection that corresponds completely to conversion is obtained when to the diode a voltage is applied consisting of a carrier with one side-band. The carrier then corresponds to the oscillator voltage, the side-band to the signal, and the L.F. signal, due to detection, to the I.F. signal. The equivalent circuit that was derived for conversion can thus immediately be employed in this case.

When a carrier of the frequency ω with a side-band frequency $\omega + p$ is detected, a voltage of the frequency p will occur, if for it there is a certain impedance, which may be complex, in the diode circuit. When p approaches zero, this impedance will approach the D.C. resistance R_0 of the diode circuit. The H.F. reaction of this L.F. impedance thus corresponds to that of a H.F. impedance that, in the vicinity of the carrier frequency, has the same variation as to modulus and argument as the L.F. impedance in the vicinity of zero frequency. The circuit of *fig. 11*, where the L.F. impedance consists of R_0 and C_a in parallel, leads in this manner to the H.F. equivalent circuit of *fig. 12*, where C_a is replaced by $\frac{1}{2}C_a$, which is tuned to the carrier with an inductance in parallel.

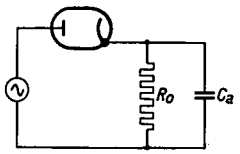


Fig. 11. Detector-circuit diagram.

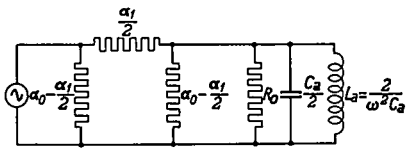


Fig. 12. High-frequency equivalent circuit of the circuit of fig. 11 upon single-side-band detection.

This manner of making a H.F. impedance from a L.F. impedance can be extended to any desired L.F. impedance. The rule is that resistances are not changed, that capacitances are halved and tuned with inductances in parallel, and that inductances are halved and tuned with capacitances in series [3]). This method leads, however, to the here desired relation between H.F. and L.F. impedances only when the H.F. signal has a small relative band width, *i.e.* when $p \ll \omega$, which condition is usually satisfied.

If we are dealing with a diode with a linear characteristic, the equivalent circuit is also correct for the carrier. If we replace G_m by $1/R_0$ in expression (15) for the voltage-conversion factor, and then express all the quantities in terms of φ and R, we come to $\cos \varphi$. If we do the same in expression (13) for the input conductance, we find the input conductance (9) for the carrier.

In all these considerations on single-side-band detection we have tacitly assumed that only H.F. impedances are present in the diode circuit for the frequencies ω and $\omega + p$, so that in the H.F. voltage on the diode only those frequencies occur. It is assumed in particular that no impedance for the frequency $\omega - p$ is present in the diode circuit. In the diode current a component of the frequency $\omega - p$ is formed which would otherwise cause the appearance of a voltage of that frequency. We shall not go deeper into this.

9. *Detection of a double-side-band signal*

For the investigation of the detection of a double-side-band signal it is assumed that in addition to the carrier $V \cos \omega t$, the signal $V_i \cos (pt + \varphi_i) \cos \omega t$ is present, while the L.F. voltage that occurs is $V_a \cos (pt + \varphi_a)$. In this case

$$v = V_i \cos (pt + \varphi_i) \cos \omega t - V_a \cos (pt + \varphi_a) .$$

From (4) two equations now follow, namely for the currents of the frequencies $\omega \pm p$ and for those of the frequency p. These are

$$I_i \cos (pt + \psi_i) \cos \omega t = a_0 V_i \cos (pt + \varphi_i) \cos \omega t - a_1 \cos \omega t \,.\, V_a \cos (pt + \varphi_a) +$$
$$+ a_2 \{\cos 2\omega t \,.\, V_i \cos (pt + \varphi_i) \cos \omega t\}_{\text{freq. } \omega \pm p} =$$
$$= \left(a_0 + \frac{a_2}{2}\right) V_i \cos (pt + \varphi_i) \cos \omega t - a_1 V_a \cos (pt + \varphi_a) \cos \omega t,$$
$$- I_a \cos (pt + \psi_a) = -a_0 V_a \cos (pt + \varphi_a) + a_1 \{\cos \omega t \,.\, V_i \cos (pt + \varphi_i) \cos \omega t\}_{\text{freq. } p} =$$
$$= -a_0 V_a \cos (pt + \varphi_a) + \frac{a_1}{2} V_i \cos (pt + \varphi_i).$$

In the complex notation the four-pole equations are thus as follows:

$$\left.\begin{aligned} \mathbf{I}_i &= \left(a_0 + \frac{a_2}{2}\right) \mathbf{V}_i - a_1 \mathbf{V}_a , \\ 2\,\mathbf{I}_a &= -a_1 \mathbf{V}_i + 2a_0 \mathbf{V}_a . \end{aligned}\right\} \qquad (24)$$

This four-pole can be represented by the equivalent circuit of *fig. 13*. The occurrence of the current $2\mathbf{I}_a$ in the equivalent circuit means that the output current that follows from the equivalent circuit, is twice the current that actually occurs. If the output terminals are connected via an admittance $\mathbf{Y}_a$, the output terminals in the equivalent circuit must be connected via the admittance $2\mathbf{Y}_a$ in order to obtain the correct relation between current and voltage (*fig. 14*).

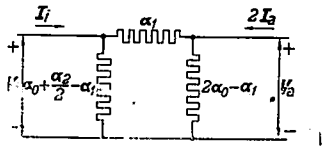


Fig. 13. Equivalent circuit of the diode upon double-side-band detection.

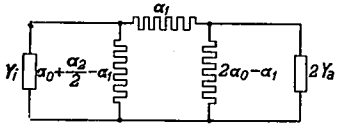


Fig. 14. Equivalent circuit of the circuit for double-side-band detection.

As an example the H.F. equivalent circuit of the detector circuit of *fig. 15* is given in *fig. 16*. The admittance over which the output terminals must be connected is thus now twice that in the case of single-side-band detection. So the rule is now that resistances of the L.F. admittance are halved, that capacitances are not changed and are tuned with inductances in parallel, and that inductances are decreased by a factor four and tuned with capacitances in series — it again being assumed that $p \ll \omega$.

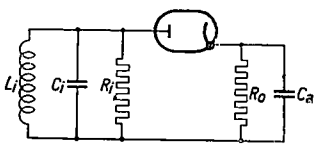


Fig. 15. Detector-circuit diagram.

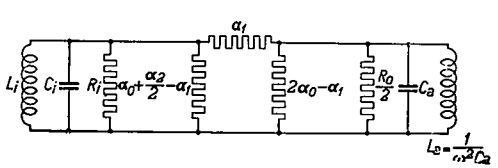


Fig. 16. High-frequency equivalent circuit of the circuit of fig. 15 upon double-side-band detection.

The two equations (24) can also be divided by 2, giving

$$\left.\begin{aligned} \tfrac{1}{2}\mathbf{I}_i &= \left(\frac{\alpha_0}{2}+\frac{\alpha_2}{4}\right)\mathbf{V}_i - \frac{\alpha_1}{2}\mathbf{V}_a, \\ \mathbf{I}_a &= -\frac{\alpha_1}{2}\mathbf{V}_i + \alpha_0\mathbf{V}_a. \end{aligned}\right\} \qquad (25)$$

This leads to the equivalent diagram of *fig. 17*. The occurrence of the current $\frac{1}{2}\mathbf{I}_i$ means that when the input terminals are connected to an

admittance $\mathbf{Y}_i$, the input terminals in the equivalent circuit must be connected to an admittance $\frac{1}{2}\mathbf{Y}_i$ (*fig. 18*).

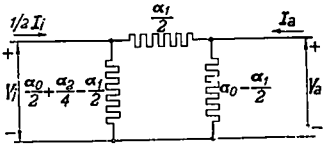


Fig. 17. Variant of the equivalent circuit of the diode upon double-side-band detection.

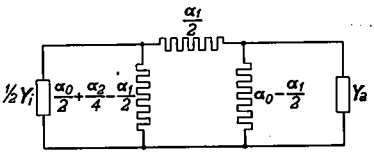


Fig. 18. Variant of the equivalent circuit of the circuit for double-side-band detection.

Instead of a H.F. equivalent circuit, a L.F. equivalent circuit can also be drawn, for which it is best to start from fig. 18. As an example the L.F. equivalent circuit of the detector circuit of fig. 15 is given in *fig. 19*. The H.F. circuit of fig. 15, consisting of a resistance and an inductance in parallel, is here replaced by the capacitance and twice the resistance in parallel.

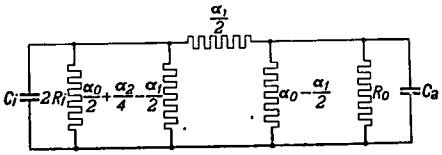


Fig. 19. Low-frequency equivalent circuit for the circuit of fig. 15 upon double-side-band detection.

10. *Use of the equivalent circuits for the calculation of several quantities*

Just as in the case of conversion here also various quantities can be derived from the equivalent circuits. The *input admittance* for the side-bands follows most simply from fig. 14, from which may be read off

$$a_0 + \frac{a_2}{2} - a_1 + \frac{a_1(2a_0 - a_1 + 2\mathbf{Y}_a)}{a_1 + 2a_0 - a_1 + 2\mathbf{Y}_a} = \frac{a_0^2 + a_0 a_2/2 - a_1^2/2 + (a_0 + a_2/2)\,\mathbf{Y}_a}{a_0 + \mathbf{Y}_a}. \tag{26}$$

In the same way it follows from fig. 18 that the *output admittance* has the value

$$\frac{a_0^2 + a_0 a_2/2 - a_1^2/2 + a_0 \mathbf{Y}_i}{a_0 + a_2/2 + \mathbf{Y}_i}. \tag{27}$$

From fig. 14 or fig. 18 it may be seen that when a H.F. voltage $\mathbf{V}_i$ is applied to the input terminals the following L.F. voltage occurs at the output terminals:

$$\mathbf{V}_a = \frac{\alpha_1/2}{\alpha_1/2 + \alpha_0 - \alpha_1/2 + \mathbf{Y}_a}\,\mathbf{V}_i;$$

so one may speak of a *detection factor*

$$\frac{\alpha_1/2}{\alpha_0 + \mathbf{Y}_a}. \tag{28}$$

Finally, in fig. 14 we may inquire about the L.F. voltage on the output terminals when a H.F. current **I** is applied in parallel to the input terminals. From the figure the following can be calculated for that:

$$\mathbf{V}_a = \frac{\alpha_1/2}{\alpha_0^2 + \alpha_0\,\alpha_2/2 - \alpha_1^2/2 + (\alpha_0 + \alpha_2/2)\,\mathbf{Y}_a + \alpha_0\,\mathbf{Y}_i + \mathbf{Y}_a\mathbf{Y}_i}\,\mathbf{I}. \tag{29}$$

11. *The diode with a linear characteristic as detector*

If the diode has a linear characteristic and we again confine ourselves to small values of φ, we find for the quantities of the equivalent circuit of fig. 13 approximately

$$\left.\begin{aligned} \alpha_0 + \frac{\alpha_2}{2} - \alpha_1 &\approx -\frac{1}{R_0}\left\{1 - \tfrac{3}{4}\left(\frac{3\pi R}{R_0}\right)^{2/3}\right\}, \\ 2\alpha_0 - \alpha_1 &\approx \frac{1}{R_0}\left\{1 - \tfrac{9}{20}\left(\frac{3\pi R}{R_0}\right)^{2/3}\right\}, \\ \alpha_1 &\approx \frac{6}{R_0}\left(\frac{R_0}{3\pi R}\right)^{2/3}, \end{aligned}\right\} \tag{30}$$

while the quantities of the circuit of fig. 17 amount to half these values. One of the conductances of these equivalent circuits is thus negative.

In order to avoid the occurrence of negative conductances, which make it difficult to understand the behaviour of the circuits, we may replace the circuits of figs 13 and 17, repectively, by the equivalent circuits of *figs 20* and *21*, in which two positive conductances and an ideal transformer with the transformer ratio u occur. For a diode with a linear characteristic,

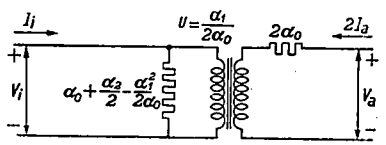


Fig. 20. Equivalent circuit of the diode upon double-side-band detection, equivalent to the diagram of fig. 13.

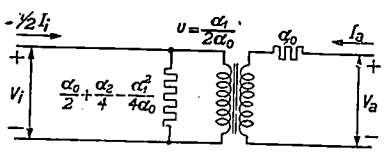


Fig. 21. Variant of the equivalent circuit of the diode upon double-side-band detection, equivalent to the diagram of fig. 17.

for small values of φ we find approximately the following for the quantities of the circuit of fig. 20:

$$\left.\begin{aligned} a_0 + \frac{a_2}{2} - \frac{a_1^2}{2a_0} &\approx \frac{2}{15R_0}\left(\frac{3\pi R}{R_0}\right)^{3/2}, \\ 2a_0 &\approx \frac{6}{R_0}\left(\frac{R_0}{3\pi R}\right)^{3/2}, \\ \frac{a_1}{2a_0} &\approx 1 - \frac{1}{6}\left(\frac{3\pi R}{R_0}\right)^{3/2}, \end{aligned}\right\} \qquad (31)$$

while the conductances of the circuit of fig. 21 amount to one half of these.

In general it will be desired to keep the series conductance large and the parallel conductance small, for which it is favourable to give the diode a small internal resistance. In the ideal state, therefore, the input admittance approaches $2\mathbf{Y}_a$, the output admittance $\frac{1}{2}\mathbf{Y}_i$, and the detection factor unity.

When we are concerned with a diode with a linear characteristic, then, as in single-side-band detection, the equivalent circuits also hold for the carrier. If in expression (28) for the detection factor, $\mathbf{Y}_a$ is replaced by $1/R_0$, and all the quantities are then expressed in φ and R, one obtains $\cos\varphi$. When the same is done in expression (26) for the input admittance one finds the input conductance (9) for the carrier.

Eindhoven, June 1943

REFERENCES

1) L. C. Peterson and F. B. Llewellyn, Proc. Inst. Radio Engrs **33**, 458-476, 1945.
2) D. O. North, R.C.A. Rev. **4**, 441-472, 1940; see also C. J. Bakker, Physica **8**, 23-43, 1941.
3) V. D. Landon, Proc. Inst. Radio Engrs **24**, 1582-1584, 1936.

R 61 Philips Res. Rep. 2, 420-425, 1947

RADIATION AND HEAT CONDUCTION IN LIGHT-SCATTERING MATERIAL

by H. C. HAMAKER 536.24:536.33

IV. VARIOUS EXTENSIONS AND A GENERALIZED THEORY

Summary

In previous papers[1]) only simplified problems have been considered. For a fuller discussion of many practical questions of a similar kind some extensions of the theory may be required which will be indicated below, though no attempts will be made to work them out in full detail.

16. *Optical temperature determinations*

Optical temperature determinations have already been considered in section 11 under the assumption, however, that the optical observations were made for total radiation. In practice this is never the case; on the contrary, the optical temperature observations are preferably carried out in a very limited region of the spectrum.

The simplest approximation to experimental circumstances seems to assume a temperature distribution in the layer according to equations (47), the constants a and s (or α and β) referring to total radiation. For this *given* temperature distribution the problem will then be to solve the equations

$$\begin{aligned} \frac{dI_\lambda}{dx} &= -(a_\lambda + s_\lambda) I_\lambda + s_\lambda J_\lambda + a_\lambda E_\lambda(T) \\ \frac{dJ_\lambda}{dx} &= (a_\lambda + s_\lambda) J_\lambda - s_\lambda I_\lambda - a_\lambda E_\lambda(T) \end{aligned} \qquad (115)$$

which are identical with equations (36) except for the index λ which indicates that all quantities now refer to the effective wavelength of the pyrometer used.

In a narrow temperature range the black-body radiation, $E_\lambda(T)$, may, as before, be considered as a linear function of the temperature:

$$E_\lambda(T) = E_{0\lambda}(T) + b_\lambda(T - T_0), \qquad (116)$$

and with this additional supposition the solution of equations (115) presents no essential mathematical difficulties, though the final equations become fairly involved.

Owing to the temperature gradient in the layer the optical temperature determinations will again be subject to errors (compare section 11). If

only the order of magnitude of these errors is of interest the problem can be further simplified by neglecting the exponential terms in the temperature distribution given by (47), which will make T a linear function of x.

Though I have derived the complete solution of the equations indicated above, I do not consider it of sufficient interest for detailed publication.

17. *Total radiation as a function of layer thickness*

According to Clausing and Ludwig's data [2]) the total radiation of an oxide-coated cathode goes through a maximum when the layer increases in thickness. Though their experimental methods may be open to criticism from our present point of view (see section 11), the changes in the total radiation are too pronounced to be accounted for by systematic errors alone.

Clausing and Ludwig explained their results by assuming that an increase in layer thickness produces an increase in radiation in one part of the spectrum, and a decrease in another part. To incorporate this hypothesis in our present theories we should have to adopt two different sets of constants, a_1, s_1 and a_2, s_2, pertaining to different spectral regions, and then to solve the equations

$$\begin{aligned}
\frac{dI_1}{dx} &= -(a_1 + s_1) I_1 + s_1 J_1 + a_1 E_1, \\
\frac{dJ_1}{dx} &= \quad (a_1 + s_1) J_1 - s_1 I_1 - a_1 E_1, \\
\frac{dI_2}{dx} &= -(a_2 + s_2) I_2 + s_2 J_2 + a_2 E_2, \\
\frac{dJ_2}{dx} &= \quad (a_2 + s_2) J_2 - s_2 I_2 - a_2 E_2, \\
k \frac{d^2T}{dx^2} &= +a_1(I_1 + J_1) + a_2(I_2 + J_2) = 2a_1 E_1 + 2a_2 E_2,
\end{aligned} \tag{117}$$

where the indices 1 and 2 refer to the two different regions of the spectrum. Assuming, as before, E_1 and E_2 to be linear functions of T, $E_1 = E_{10} + b_1(T - T_0)$, *etc.*, and introducing new dependent variables, $i_1 = I_1 - E_{10}$, *etc.*, these equations are reduced to a set of homogeneous linear equations analogous to (36, 37), which can be solved by making all dependent variables proportional to $e^{\sigma x}$.

As in section 7 we are led to a determinantal equation from which σ has to be evaluated, this time one of the sixth degree. The solution is found to split into two parts, *viz.*

$$\sigma^2 = 0\,, \tag{118}$$

and

$$k - \frac{2a_1b_1}{\sigma^2 - a_1(a_1+2s_1)} - \frac{2a_2b_2}{\sigma^2 - a_2(a_2+2s_2)} = 0\,. \qquad (119)$$

Equation (118) will give rise to linear terms, $i_1 = C_1x + D_1$, *etc.*, while (119) will lead to four different exponential terms proportional to $e^{\sigma_1 x}$, $e^{-\sigma_1 x}$, $e^{\sigma_2 x}$, and $e^{-\sigma_2 x}$, σ_1 and σ_2 being the roots of (119). Ultimately we get a solution similar to (47) but with four exponential terms instead of two, and with six constants to be fixed by boundary conditions. In view of the large number of arbitrary constants it is highly likely that from such expressions we might construct curves closely fitting Clausing and Ludwig's data. This would, however, involve much tedious calculation.

18. *Complete equations*

In the previous section we have been trying to partly remove the drawback of having disregarded the spectral distribution of the radiation and of the optical constants; from a theoretical point of view it would be more satisfactory to introduce the wavelength as a separate variable, and we shall now devote some attention to the equations then resulting. It will easily be seen that equations (36, 37) are in that case to be replaced by

$$\frac{dI(\lambda,x)}{dx} = -\{a(\lambda)+s(\lambda)\}I(\lambda,x) + s(\lambda)J(\lambda,x) + a(\lambda)E(\lambda,T)\,, \qquad (120a)$$

$$\frac{dJ(\lambda,x)}{dx} = \{a(\lambda)+s(\lambda)\}J(\lambda,x) - s(\lambda)I(\lambda,x) - a(\lambda)E(\lambda,T)\,, \qquad (120b)$$

$$k\frac{d^2T}{dx^2} + \int_0^\infty a(\lambda)\{I(\lambda,x)+J(\lambda,x)\}\,d\lambda = 2\int_0^\infty a(\lambda)E(\lambda,T)\,d\lambda\,, \qquad (120c)$$

where, as indicated, a, s, I, J, and E are now functions of the wavelength λ.

Introducing

$$E(\lambda,T) = E(\lambda,T_0) + b(\lambda,T_0)(T-T_0)\,, \qquad (121)$$

and

$$i(\lambda,x) = I(\lambda,x) - E(\lambda,T_0),\quad j(\lambda,x) = J(\lambda,x) - E(\lambda,T_0),\quad t = (T-T_0)\,, \qquad (122)$$

equations (120) transform into

$$\frac{di_\lambda(x)}{dx} = -(a_\lambda+s_\lambda)i_\lambda(x) + s_\lambda j_\lambda(x) + a_\lambda b_\lambda t\,, \qquad (123a)$$

$$\frac{dj_\lambda(x)}{dx} = (a_\lambda+s_\lambda)j_\lambda(x) - s_\lambda i_\lambda(x) - a_\lambda b_\lambda t\,, \qquad (123b)$$

$$k\frac{d^2t}{dx^2} + \int_0^\infty a_\lambda\{i_\lambda(x)+j_\lambda(x)\}\,d\lambda = 2t\int_0^\infty a_\lambda b_\lambda d\lambda\,, \qquad (123c)$$

where, to simplify the notation, the dependence on λ has been indicated by a suffix.

The third of these last equations expresses that heat is nowhere accumulated; consequently the transfer of energy through the layer must be independent of x, a physical condition that will provide a general integral of these equations. As a matter of fact, subtracting (123b) from (123a) we have

$$\frac{\mathrm{d}(i_\lambda - j_\lambda)}{\mathrm{d}x} = -a_\lambda (i_\lambda + j_\lambda) + 2a_\lambda b_\lambda t \,. \tag{124}$$

Integrating this equation with respect to λ and inserting the result in (123c) we obtain

$$-k\frac{\mathrm{d}^2 t}{\mathrm{d}x^2} + \int_0^\infty \frac{\mathrm{d}(i_\lambda - j_\lambda)}{\mathrm{d}x}\,\mathrm{d}\lambda = 0\,, \tag{125}$$

which by integration with respect to x yields

$$-k\frac{\mathrm{d}t}{\mathrm{d}x} + \int_0^\infty (i_\lambda - j_\lambda)\,\mathrm{d}\lambda = W. \tag{126}$$

The constant of integration, W, is evidently identical with the total transfer of energy in the direction of positive x.

The same method of integration can be carried one step further. Addition of (123a) and (123b) produces

$$\frac{\mathrm{d}(i_\lambda + j_\lambda)}{\mathrm{d}x} = -(a_\lambda + 2s_\lambda)(i_\lambda - j_\lambda)\,, \tag{127}$$

which upon introduction in (126) and integration with respect to x yields

$$-kt - \int_0^\infty \frac{(i_\lambda + j_\lambda)}{(a_\lambda + 2s_\lambda)}\,\mathrm{d}\lambda = Wx + H, \tag{128}$$

H being a second constant of integration.

In the simpler cases discussed in previous papers we have obtained particular solutions where all the dependent variables were linear functions of x, and equation (128) suggests that a similar solution may also exist in the present case. This turns out to be true; for it will be easily verified that the expressions

$$i_{1\lambda} = -\gamma b_\lambda \left(Wx - \frac{W}{a_\lambda + 2s_\lambda} + H\right), \tag{129a}$$

$$j_{1\lambda} = -\gamma b_\lambda \left(Wx + \frac{W}{a_\lambda + 2s_\lambda} + H\right), \tag{129b}$$

$$t_1 = -\gamma\,(Wx + H), \tag{129c}$$

satisfy all equations of this section, provided that

$$\frac{1}{\gamma} = k + \int_0^\infty \frac{2b_\lambda}{a_\lambda + 2s_\lambda}\,\mathrm{d}\lambda. \tag{130}$$

From what has been remarked in section 9 of the second paper it will be clear that this linear solution will again be valid in the interior of a thick layer at such distances from the boundaries that the specific surface conditions no longer play a role. Under these circumstances we have from (129c)

$$\frac{\mathrm{d}t_1}{\mathrm{d}x} = -\gamma\,W, \tag{131}$$

so that (130) can be rewritten as

$$W = -\frac{\mathrm{d}t_1}{\mathrm{d}x}\left\{k + \int_0^\infty \frac{2b_\lambda}{a_\lambda + 2s_\lambda}\,\mathrm{d}\lambda\right\}, \tag{132}$$

Here again (see section 9) the total transport of energy is represented as the sum of two parts, *viz*:

$$V = -k\frac{\mathrm{d}t_1}{\mathrm{d}x},$$

the heat transport by pure conduction, and

$$U = -\frac{\mathrm{d}t_1}{\mathrm{d}x}\int_0^\infty \frac{2b_\lambda}{a_\lambda + 2s_\lambda}\,\mathrm{d}\lambda, \tag{133}$$

which consequently represents the energy transported by radiation. On the assumption that a_λ and s_λ are independent of λ, equation (134) reduces to (55) since

$$b = \int_0^\infty b_\lambda \mathrm{d}\lambda; \tag{134}$$

this is as it should be.

In general, a set of linear differential equations can be simplified if a particular solution is known, and this also holds in the present instance. If instead of i_λ, j_λ and t we introduce a new set of dependent variables given by

$$\begin{aligned} i_\lambda' &= i_\lambda - i_{1\lambda}, \\ j_\lambda' &= j_\lambda - j_{1\lambda}, \\ t' &= t - t_1, \end{aligned} \tag{135}$$

where $i_{1\lambda}$, $j_{1\lambda}$ and t_1 are given by (129), i_λ', j_λ', and t' will satisfy equations (124) and (127) and the equation

$$-kt' - \int_0^\infty \frac{i_\lambda' + j_\lambda'}{a_\lambda + 2s_\lambda} \, d\lambda = 0, \tag{136}$$

which is derived from (128) by putting the right-hand side equal to zero. This set of equations, (124), (127) and (136), is equivalent to our original set (123).

From these three equations $(i_\lambda' - j_\lambda')$ and t' can be eliminated, and we obtain

$$\frac{d^2(i_\lambda' + j_\lambda')}{dx^2} = a_\lambda(a_\lambda + 2s_\lambda)\,(i_\lambda' + j_\lambda') + \frac{2a_\lambda b_\lambda(a_\lambda + 2s_\lambda)}{k} \int_0^\infty \frac{(i_\lambda' + j_\lambda')}{(a_\lambda + 2s_\lambda)} \, d\lambda. \tag{137}$$

If we now introduce a new independent variable defined by

$$\varphi = \int_0^\lambda \frac{2a_\lambda b_\lambda}{k} \, d\lambda, \tag{138}$$

and a dependent variable given by

$$u(\varphi, x) = \frac{k(i_\lambda' + j_\lambda')}{2a_\lambda b_\lambda(a_\lambda + 2s_\lambda)}, \tag{139}$$

equation (137) reduces to

$$\frac{d^2u(\varphi, x)}{dx^2} = \alpha(\varphi) \, . \, u(\varphi, x) + \int_0^\Phi u(\varphi, x) \, d\varphi \, , \tag{140}$$

where Φ is the greatest possible value of φ, and $\alpha(\varphi)$ is the product $a_\lambda(a_\lambda + 2s_\lambda)$ written as a function of φ. Solving the original set of equations has thus been reduced to finding a general solution of this integro-differential equation. I have attempted to construct a solution by splitting the spectrum into a number of small regions, and replacing (140) by a set of simultaneous equations, a summation taking the part of the integration. This conventional method leads, however, to a solution in the form of a sum of exponential terms that in the limit transforms into a divergent integral. Other methods should therefore be used in solving (140).

Eindhoven, April 1947

REFERENCES

1) H. C. Hamaker, Radiation and heat conduction in light-scattering material I, II and III, Philips Res. Rep. 2, 55-67, 103-111, 112-125, 1947.

2) P. Clausing and J. B. Ludwig, Physica, 's-Grav. 13, 193-205, 1933.

R 62 Philips Res. Rep. 2, 426-441, 1947

THE TIME-LAG IN STARTING A LOW-PRESSURE ARC ON A MERCURY OR GALLIUM CATHODE IN CONNECTION WITH FIELD EMISSION AND SURFACE DEFORMATION

by N. WARMOLTZ 537.525:621.314.653

Summary

At low vapour pressures an arc is struck by suddenly applying a tension of several kilovolts between a mercury or gallium surface as cathode and a tungsten anode a short distance above it.

The minimum field strength at the cathode required to start a discharge amounts to about 50 kV/cm, a value far too low to produce any appreciable field emission of electrons; hence the discharge cannot be initiated directly by field emission.

Oscillographic observations show the existence of a marked time-lag in starting the arc which decreases rapidly with increasing field at the cathode.

At field strengths above 250 kV/cm no macroscopic movement of the surface could be observed. In this region the time-lag corresponds to the time (as computed by Tonks) in which a microscopic disturbance on a liquid surface is drawn out into a sharp peak by the electric field, if we assume that the height of the initial hump is of the order of 10^{-8} cm. This is a roughness of the same order as that arising from thermal agitation according to a theory developed by Gans. We therefore assume that the discharge is initiated owing to a microscopic deformation of the surface which continues until the electric field at its peak has increased to such a high value as to produce an appreciable field emission of electrons. This emission current rapidly develops into an arc-discharge, a process that may be promoted by an enhanced evaporation of mercury from the peak of the deformation.

At field strengthts below 250 kV/cm a macroscopic movement of the mercury surface, which could be directly observed, precedes the mechanism described above and thereby lengthens the time-lag.

A clean anode surface proved to be an essential condition. If the surface is not cleaned by heating before each measurement, the discharge is probably started by mercury droplets which are dislodged from the anode by the electric forces.

The phenomena observed at extremely low vapour pressures obtained by using a gallium surface as cathode were essentially the same as those observed with mercury.

1. *Introduction*

In certain types of mercury-arc rectifiers so-called dielectric ignitors are used, consisting of a metallic core coated with a thin layer of insulating material and partly immersed in the mercury cathode, as is indicated in *fig. 1*.

By applying to the core a positive voltage of several kilovolts with respect to the mercury cathode, a spark is generated somewhere near the

boundary line between the mercury and the ignitor; this spark develops into an arc between the cathode and an auxiliary anode, and this arc subsequently develops into the main discharge.

It is generally assumed that the spark is initiated by field emission.

Now with ignitors of this type we have observed in the starting of the arc a marked time-lag which was particularly large — of the order of 10^{-2} sec — at low ignitor potentials. This seemed to indicate that mechanical inertia of the mercury surface played a part, since it can reasonably be assumed that under these starting conditions an arc discharge will very rapidly set in as soon as field emission at some spot on the cathode exceeds a critical value. If the potential applied is not sufficient to produce the required field off-hand, deformation of the mercury-cathode surface may be expected to occur, until the electric field has increased to such an extent as to start a discharge. The time-lag observed might well be explained by such a mechanism.

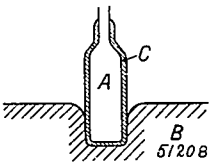


Fig. 1. Dielectric ignitor; A metallic core, C insulating layer, B mercury cathode.

These arguments have induced us to investigate the formation of an arc discharge between a mercury cathode and an anode a short distance above it. Compared with the dielectric ignitor this arrangement has the advantage that the electric field at the cathode surface can be computed with accuracy. Though the ultimate development of an arc discharge is certainly not identical in both cases, it was considered that the initial stages, *i.e.* those connected with field emission and causing the time-lag, would be the same.

2. *Survey of literature*

Before proceeding to describe our own observations we shall briefly recapitulate what is known from experiments by other investigators.

Field emission of electrons is described by the following theoretical formula [1])

$$i = \frac{1{\cdot}5.10^{-6}}{\varphi} F^2 \exp\left\{\frac{-6{\cdot}85\ 10^7\ \varphi^{3/2}}{F} f\left(\frac{3{\cdot}62\ 10^{-4}\sqrt{F}}{\varphi}\right)\right\},$$

where i is the current density in amp/cm^2, φ the work function in volts,

F the field strength in volts/cm, and f a function of F and φ which varies between 0 and 1 and which has a value of 0·8 to 0·9 in the examples adduced below.

Nearly all attempts to check this formula by experiment [2]) have produced appreciable emission currents at electric fields many times lower than would be expected from this formula; for instance, for $\varphi = 4{\cdot}4$ volts and $F = 10^7$ volts/cm equation (1) predicts a current density of 3×10^{-16} amp/cm^2, and for $F = 3 \times 10^7$ volts/cm a current density of 45 amp/cm^2, while measurable current densities have been observed at fields of 10^5 to 10^6 volts/cm, where, according to (1), they should be quite imperceptible.

Schottky [3]) ascribed this discrepancy to microscopic irregularities of the surface, the electric field at points and edges being may times higher than the field computed from macroscopic dimensions. This view was confirmed by the well-known experiments of Haefer [4]), who observed the field emission from tungsten points, the structure of which had been determined by an electron microscope with a resolving power of 10^{-6} cm; whenever the surface was microscopically smooth, equation (1) was found to be satisfied.

To be sure of a perfectly smooth cathode surface one may use a metal in the liquid state, and mercury readily presents itself for this purpose. This has led Beams [5]) to investigate field emission from a mercury surface by an indirect method.

Beams observed the electric field strength required to start an arc discharge between a mercury cathode and a spherical iron anode a short distance above it. Presumably a discharge is initiated as soon as the field-emission current has reached a critical value (of the order of 1 mA to 1 A), and, since field emission is very strongly dependent on the value of the electric field, F in equation (1), observations of this kind will enable us to check this equation approximately, even if the numerical value of this critical emission current is not exactly known.

A mercury surface, however, when exposed to an electric field of 10^6 volts/cm, will be drawn towards the anode by a pressure of the order of some kg/cm^2, so that continuous voltages cannot be applied. For that reason Beams used pulses obtained from condenser discharges for which the relaxation time had a value of $1{\cdot}2 \times 10^{-6}$ sec, a period so short that macroscopic movements of the mercury surface were excluded.

Under these circumstances an arc discharge was observed at a field of $3{\cdot}5 \times 10^5$ volts/cm for an impure mercury surface, and at a field of $1{\cdot}8 \times 10^6$ volts/cm after the surface had been purified by repeated distillation and overflowing, operations that could be carried out in the tube itself. Again these electric fields are considerably smaller than those to be

expected from the theory, which are of the order of 3×10^7 volts/cm. The discrepancy, for which Beams did not offer an explanation, can hardly be due to deformation of the cathode surface, since variation of the duration of the pulses did not have any influence on the value of the arcing potential.

These experiments were subsequently repeated by Quarles [6]) and by Moore [7]). Quarles determined the relation between the arcing potential and the work function of the mercury cathode, the latter being varied by varying the purity of the surface by repeated distillations. The arcing fields observed were of the order of 4 tot 5 $\times$ 10^5 volts/cm, independent of the duration of the pulses, which were varied from 10^{-6} to 10^{-7} sec; the relation between arcing field and work function did not agree with the theoretical expectation.

Finally Moore, by using an oscillograph in similar experiments, showed that the total time required for the development of an arc is not more than 2×10^{-7} sec.

These observations will be further discussed in section 5 together with our own experimental results.

3. *Description of our apparatus*

The reasons that led us to a renewed study of arc formation in a mercury discharge have already been stated in the introduction. The tubes used in these investigations were of a type similar to those with which Beams [5]) carried out his experiments. A narrow tube, widening into a funnel at its

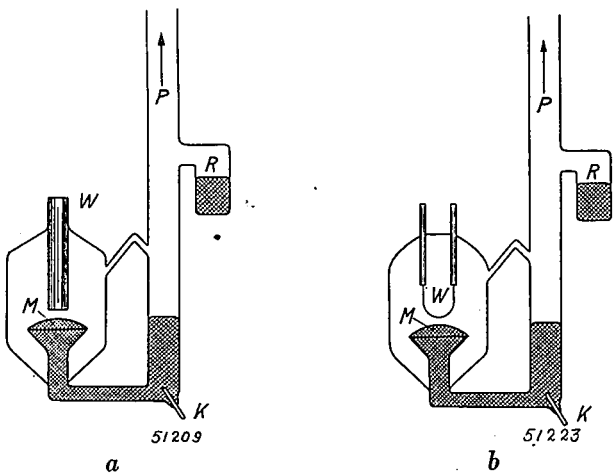


Fig. 2. Mercury-cathode tube for measuring the time-lag as a function of the field strength; M cathode, R mercury reservoir.
a) Tube with tungsten rod anode, W.
b) Tube with tungsten-ribbon anode, which can be heated by an electric current.

upper end, was sealed in a cylindrical bulb 5 cm high and 2 cm in diameter (see *fig. 2*); the funnel holds the cathode (M), the distance between the anode (W) and the cathode (M) being regulated by varying the amount of mercury; if necessary additional amounts of mercury could be supplied from a reservoir (R). The whole assembly was connected to the pump by means of a slightly flexible glass helix, so that the anode-cathode gap could be finally adjusted to any desired value by slightly tilting the tube; a fresh adjustment was made before each experiment.

The tube, which was made of pyrex glass, was baked out on the pump to about 550° C for some hours. Then mercury was first distilled from a vessel into an arc tube with mercury electrodes, and then distilled into the experimental tube by running the arc.

To avoid mechanical disturbances the tube was mounted on a heavy iron plate resting on soft rubber sheets. The anode consisted originally of a tungsten rod 1 mm in diameter, as shown in fig. 2*a*, but at a later stage it was replaced by a tungsten strip of 2 mm width, which could be heated by passing an electric current though it (fig. 2*b*); in this way the surface could be thoroughly cleaned, whenever desired.

Before each observation the gap between cathode and anode was adjusted to 0·25 mm for low fields and to 0·125 mm for high fields, this distance being checked by means of a telescope with micrometer reading. To get a proper view of the mercury profile an illuminated white screen was used as a background.

The cathode surface was cleaned by letting some of the mercury flow over the rim of the funnel, the overflowing mercury being collected on the bottom of the tube, where it did no harm.

A diagram of the electrical circuit is reproduced in *fig. 3*; the condenser C, of 10000 $\mu\mu$F, which may be charged to different potentials from a rectifier, is suddenly discharged via the resistance R (of 1000 ohms), and the experimental tube E by applying a positive pulse to the grid of the thyratron Th; the voltage across E is connected to the vertical deflection plates of an oscillograph via a small condenser S, which serves to cut down the tension of several kV to a tension of a few hundred volts required for the cathode-ray tube.

The same pulse which triggers the thyratron simultaneously operates a rectangular pulse-generator B; the pulse from B in turn releases the cathode ray and sets the single-sweep time-base T in motion. Since, in starting, the thyratron has a longer time-lag than the high vacuum valves of the time-base *etc.*, the latter will always be in operation before the voltage pulse reaches the experimental tube E.

The time-base used allowed of such variation that 1 cm along the time axis on the screen of the cathode-ray tube could be made to correspond to

values between 2×10^{-3} and 3×10^{-8} sec, while the time scale was calibrated by means of an oscillator of known frequency.

To avoid electrical disturbances the oscillograph and the pulse generator had both to be housed in copper-clad iron cages; for the same reason it was found advantageous to apply the pulse to the pulse-generator B via a diode and a small condenser of 10 $\mu\mu$F (not shown in the diagram).

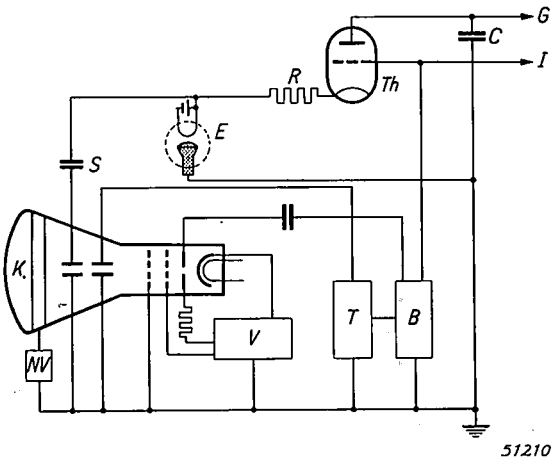


Fig. 3. Diagram of the oscillograph circuit used. K is a 5-kV cathode-ray tube (Philips DN 9-5); NV and V are high-tension rectifiers. The condenser C, charged to several kV, is discharged through the experimental tube E by triggering the thyratron Th. B is a rectangular-pulse generator operating the linear single-sweep time-base T and releasing the cathode ray.

4. *Observations*

That an arc discharge develops upon the application of a tension of several kilovolts to the tube described in the foregoing section, is evidenced by the light emitted from a sharply defined cathode spot on the mercury surface, and by a sharp drop of the potential across the tube to the value of the cathode fall of a mercury discharge. Since, for the tubes used in our experiments, the voltages were much too low to produce a glow discharge in mercury at room temperature *) it is likely that the discharge is initiated by field emission.

The main features of the oscillograms obtained by the methods described above are shown in *fig. 4*; the tension across the tube rises rapidly in about 10^{-7} sec, then remains constant for a much longer period, finally falling to the potential of the mercury arc discharge. (The spread of the time-lag with constant voltage amounts to only about 20%).

*) This holds in an even stronger degree for gallium, which at room temperature has a vapour pressure of only about 10^{-40} mm Hg.

From observations on the vacuum spark by Hull and Burger [8]), by Snoddy [9]), and by Chiles [10]), and from our own observations with tungsten spheres in vacuum [15]), we know that initial field emission very rapidly develops into an arc discharge, the time required being of the order of 10^{-7} sec. Consequently a time-lag as demonstrated in fig. 4 must be due to phenomena taking place before the field emission has reached a value high enough to start an arc discharge.

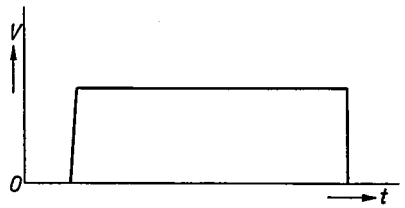


Fig. 4. Diagrammatic representation of the oscillogram giving the tension across the tube as a function of time.

Now, by closely watching the cathode surface during the discharge through the telescope used to adjust the anode-cathode gap, the following phenomena could be observed. When the electric field strength is 250 kV/cm or more the cathode spot can be seen exactly in the same place where the cathode surface was seen before the discharge started. At lower field strengths, however, the light from the cathode spot was observed somewhere between the anode and the original position of the cathode, the distance from this original position increasing with decreasing voltages; at the same time the cathode profile as seen through the telescope grew unsharp during the discharge.

These results can readily be explained by assuming a deformation of the cathode surface. For fields less than 250 kV/cm the cathode surface moves towards the anode until the field strength has increased to this critical value; and at that moment — by a different mechanism, probably a microscopic deformation of the cathode surface, as will be discussed in section 5 — the arc develops in a very short time and the cathode spot is observed. When, on the other hand, the electric field is at the outset higher than 250 kV/cm, the time required for starting the arc is so short that macroscopic movements of the cathode surface cannot take place.

The mean time-lag for the tube with tungsten-rod anode is represented in *fig.* 5 as a function of the electric field; at about 250 kV/cm, the point where macroscopic movement of the cathode surface ceases to be perceptible, the time-lag is seen to drop to a much lower value; whether this fall is discontinuous or continuous, though in that case very steep, could not be decided with certainty.

As far as the four observations at low fields are concerned, curve I is a singular case which has not repeated itself; the same tube yielded curve II after it had not been used for a period of two months. Curve III was obtained with a second tube of the same pattern, with which we also investigated the influence of the following treatments:

1. letting air into the tube to a pressure of 0·05 mm Hg, and evacuating again;
2. causing a discharges in air at atmospheric pressure, a treatment that produced a visible contamination of the cathode surface, and evacuating again.

As a result of these treatments the curves remained unchanged to the left of 250 kV/cm. To the right of that point, the curves varied after the treatment between curves I and II; after the last-mentioned treatment it lay slightly to the right of curve II. On the whole, however, the variations were not spectacular.

It may be added here that the time-lag for a given field strength was quite reproducible.

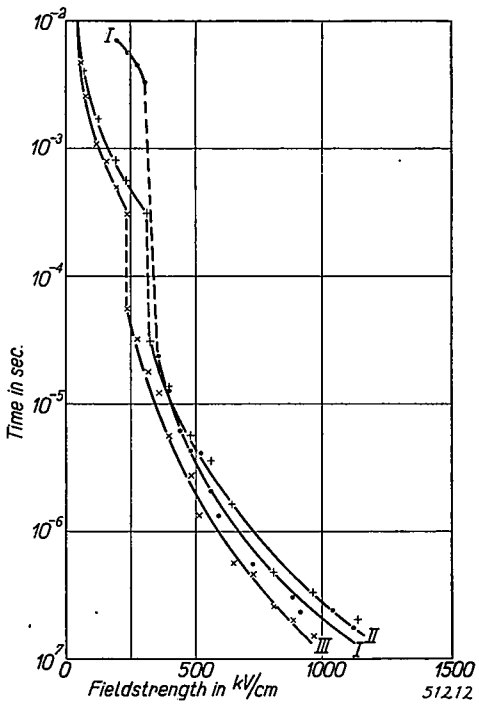


Fig. 5. The time-lag as a function of the field strength for tubes as sketched in fig. 2*a* (tungsten-rod anode).
Curves I and II: Two different series of observations on the same tube.
Curve III: Observations on a second tube.

In a similar series of experiments made with a chromium-iron anode the steep fall in the curves was observed at 300 kV/cm, the time-lag becoming practically imperceptible at 500 to 600 kV/cm.

The formation, during the experiments, of an almost hemispherical cap of mercury on the tungsten anode made it possible to study the time-lag between two mercury surfaces. The results of these observations are represented in *fig. 6*, curve I; though similar in shape, this curve is in certain

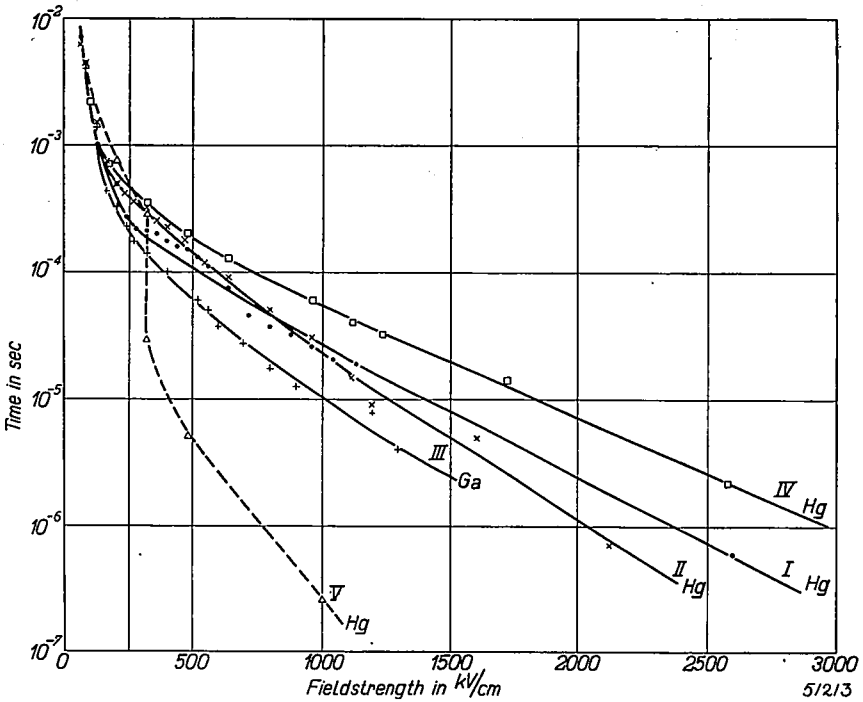


Fig. 6. The time-lag as a function of the field strength for clean anode surfaces.
Curves II and IV: Observations with a tungsten-ribbon anode heated before each experiment.
Curve I: Observations with a tungsten-rod anode coated with a cap of mercury.
Curve III: Observations with gallium under the same conditions as curve I.
Curve V: The average curve of fig. 5.

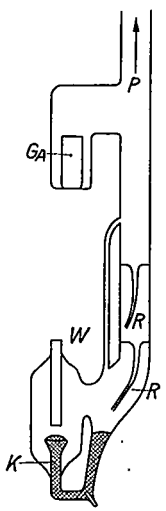
respects definitely different from the curves in fig. 5; the discontinuity at 250 kV/cm has disappeared and for field strengths above this value the time-lag is many times greater than in fig. 5.

This result induced us to construct the tube with a tungsten-ribbon anode (fig. 2*b*) which could be cleaned at will simply by passing a current through it. Curves II and IV in fig. 6 represent the data obtained when the tungsten anode was freshly cleaned before each measurement; these curves very closely resemble those obtained with a mercury coating on the anode.

However, as soon as systematic heating of the anode before each experiment was stopped, the measurements were found to follow such curves as are shown in fig. 5. From these experiments it became quite evident that the cleanliness of the anode surface had a profound influence on the value of the time-lag. We shall return to this point in section 5. The influence of a general contamination of the tube is sometimes very large, so that in performing these experiments extreme cleanliness is necessary.

Finally the experiments were repeated with gallium in order to eliminate a possible influence of vapour pressure, which, for mercury, still amounts to 0·002 mm at room temperature, as compared with only 10^{-40} mm for gallium. To this end a special tube was developed which has been sketched in *fig.* 7. Since gallium, owing to its low vapour pressure, cannot be purified by destillation at temperatures below the softening point of pyrex glass, cleansing was carried out by letting the gallium flow through narrow capillaries, on which an oxide film is left. After that treatment the cathode surface was finally cleansed by letting some of the gallium flow over the rim of the funnel holding the cathode. Though the melting point of gallium is at 30 °C, this metal exhibits pronounced under-cooling properties, so that measurements could actually be made at room temperature. In these experiments the anode consisted of a practically hemispherical cap of liquid gallium metal.

The results obtained with gallium are represented by curve III in fig. 6,



51214

Fig. 7. Tube especially designed for observations with gallium; K gallium cathode, W anode and G_A gallium reservoir.

which follows a course similar to that observed with a freshly cleaned tungsten-ribbon anode.

In the experiments with the heated anode or with a metal cap on the anode the elevation of the cathode at field strengths less than approximately 250 kV/cm, as described at the beginning of this section, was still observed. Therefore we shall make a systematic distinction between two different mechanisms: 1) a macroscopic movement below 250 kV/cm, and 2) a microscopic movement above this field strength. Under the conditions represented in fig. 6 there seems to be a more gradual change between these mechanisms than with those of fig. 5.

To conclude this section it may be mentioned that the electric field has invariably been computed by dividing the applied potential by the cathode-anode distance; this was permissible since the distance between the two electrodes (0·125 or 0·25 mm) was small in comparison with the diameters of anode (1 or 2 mm) and cathode.

5. *Discussion of the observations; comparison with a theory developed by Tonks.*

A) *The theory of Tonks*

The work of Beams [5]) induced Tonks [11]) to carry out a theoretical investigation concerning the microscopic deformation of a liquid surface under the influence of an electric field. First of all he succeeded in demonstrating that under the combined effects of gravity and surface tension a mercury surface will no longer be stable in an electric field, when the field strength has reached a value of 53 kV/cm at its surface; this result has also been obtained by Frenkel [12]), independently and by quite different arguments.

To predict exactly what will happen once the electric field strength has passed this critical value is, of course, a much more difficult matter and Tonks has only attempted to arrive at a general estimate of the time in which a small disturbance of the surface is drawn out by the field.

On the basis of the deformations of liquid droplets and bubbles which have been observed under the influence of an electric field [13]), Tonks prescribes the deformations of a mercury surface under similar circumstances. He assumes that a dome-shaped initial elevation, with a diameter many times greater than its height, is first drawn out into a hemi-spherical hump; this process is accompanied by an increase in volume and a decrease in diameter. In a second stage this hump is drawn out very rapidly at its top into a sharp peak without any further volume changes. The whole sequence of deformations is diagrammatically represented in *fig. 8.*

Once these geometrical deformations have been predetermined, the

time required for their development may be computed. This is done under the assumption that gravity may be ignored so that the electric force and surface tension are the only forces to be taken into consideration; this will substantially be correct when the field strength is greater than 158 kV/cm.

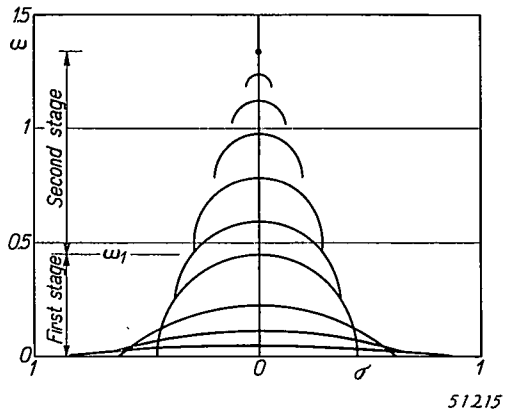


Fig. 8. Evolution of a hump on the cathode surface according to Tonks, $\sigma = r/A$ and $\omega = y/A$, $A = 8\pi T/E_0^2$, where E_0 is the field strength at the undisturbed surface.

The equations used are rather complicated and the numerical computations required have been made in a special instance only. We need not, however, go into details. Tonks's final results, as far as they are of interest to our problem, may be stated as follows:

Given an electric field of F kV/cm and an initial elevation of the surface of y_0 cm, the total time required for the series of deformations depicted in fig. 8 will be

$$t = 5{\cdot}4 \times 10^3 F^{-3} + 1{\cdot}2 \times 10^3 F^{-3} \log \frac{40F^{-2}}{y_0} \text{ sec.} \qquad (2)\ *)$$

This result is only valid within a certain range of values of y_0 and, moreover, the radius of the initial disturbance is made equal to

$$r_0 = 0{\cdot}93 \times 10^3 F^{-2} \text{ cm}. \qquad (3)\ **)$$

The data computed from equation (2) are represented for different values of y_0 in *fig. 9*, the values of the radius according to (3) having been added at the top of the figure.

*) This expression is a combination of Tonks's equations (33), (39) and (40).

**) Tonks carried out numerical computations only in a single case; his data do not allow a calculation of the time t for any arbitrary values of F, y and r, and hence this particular condition must be added.

B) *Discussion of our observations*

Let us now consider our observations. Two distinct types of curves have been observed, those of fig. 5 and those of fig. 6; for a better survey the average curve for mercury of fig. 6 has again been reproduced in fig. 9 (curve *E*). To explain these results three different mechanisms have to be taken into account which we shall now proceed to discuss.

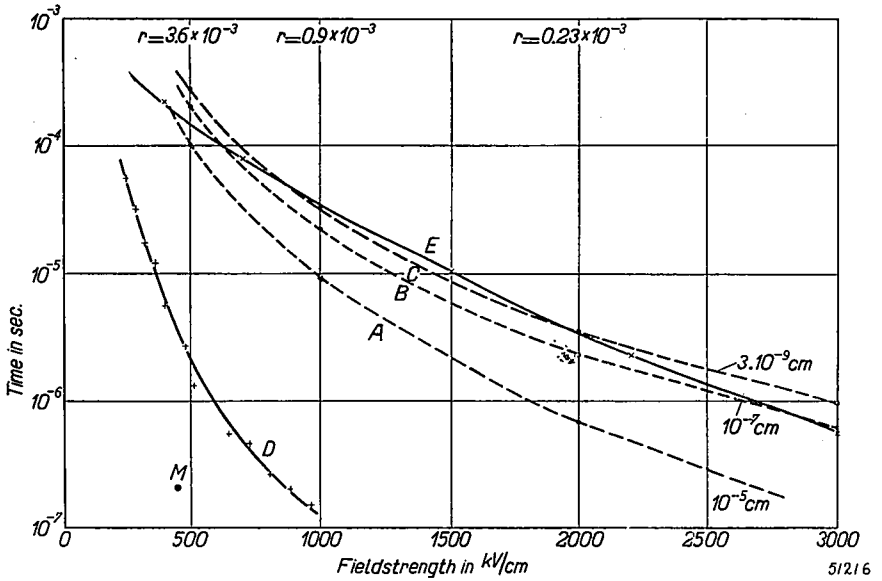


Fig. 9. The peaking time as computed from Tonks's theory compared with the time-lag of the discharge. Curves *A*, *B* and *C*: the peaking time as a function of the macroscopic field strength for three different heights of the initial disturbance. Curves *D* and *E*: The average of our own observations as given in figs 5 and 6 respectively. Point *M*: Moore's observation.

a) Below 250 kV/cm the time-lag is large and macroscopic deformation of the cathode surface has been observed in both cases. This part of our curves cannot be explained by such processes as considered by Tonks, for his theory refers only to microscopic phenomena. Hence we will assume that at low fields a major part of the time-lag is determined by a slow movement of the cathode surface towards the anode, which continues until the electric field has reached the value of about 250 kV/cm, a field still much too low to start an arc by field emission straight-away. Round about this point a different and more rapid mechanism comes into operation.

b) That this second mechanism, which above field strengths of 250 kV/cm is the only mechanism in operation, may well be of the type assumed in Tonks's theory is clearly demonstrated in fig. 9. The observations made

with a clean tungsten or a mercury anode fit well with the curves computed from Tonks's equations, especially with the curve valid for an initial elevation of 10^{-8} cm. Tonks has already mentioned the possibility that disturbances of this order of magnitude might be produced by thermal agitation, and this has been confirmed by a more rigorous analysis by Gans [14]).

By means of a Fourier analysis Gans computed the mean square of the amplitude of the thermal disturbances of a liquid surface. He concludes that the thermal elevation of the surface is of the order of magnitude of $C(k\Theta/T)^{1/2}$, where k is the Boltzmann constant, Θ the absolute temperature, and T the surface tension; C is a constant, which, according to Gans's equations in the case of mercury has a value of 2 to 3. At room temperature and for a mercury surface this formula actually yields values of the order of 10^{-8} cm, in satisfactory agreement with the conclusions drawn from a comparison of our observations with Tonks's theory.

For an initial elevation of 10^{-8} cm Tonks's theory assumes an initial diameter of the order of 10^{-3} cm; hence the initial hump on the surface covers a comparatively large area and is not of purely atomic dimensions, which would, according to Tonks's theory require much stronger fields to render them unstable (see also the legend of fig. 9).

A further check on Tonks's theory is possible by comparing our observations for mercury and for gallium, as plotted in fig. 6. Apart from depending on the magnitude of the initial disturbances and on the electric field strength, the time-lag, according to this theory, will be proportional to the surface tension and to the root of the density. Since the surface tension and the density of mercury are, respectively, about 1·3 and 2·3 times as high as those of gallium, we should expect a time-lag about twice as high in the case of mercury for the same value of the electric field; the actual ratio as read from fig. 6 is 2·5 which, in view of the approximative character of the theory, may be considered as a satisfactory agreement.

c) Finally we have still to explain the series of observations of fig. 5 above 250 kV/cm (curve V in fig. 6). These were made with tungsten anodes which could not be cleaned before each measurement, and also with the tungsten-ribbon anode when systematic cleaning was omitted. In our opinion these results must be attributed to contamination of the anode surface with minute mercury droplets. Upon each ignition of the arc, mercury will be evaporated and small droplets of mercury will be shot away from the cathode spot, while part of this mercury will no doubt condense on or be caught by the anode. In a subsequent experiment these droplets are shot away from the anode under the influence of the electric forces and, when striking the cathode, these charged droplets cause

considerable disturbances and thereby initiate an arc discharge. That our observations with the tungsten-ribbon anode fit in with the upper curves in fig. 6, when the anode was heated before each observation, and with the lower curve (V) when such heating was omitted, furnishes the most convincing evidence in favour of this point of view.

From this standpoint it can also easily be understood why the results obtained with a completely smooth coating of mercury or gallium on the anode coincide with those obtained with a clean tungsten anode, for small droplets on the anode surface will not have been present under these conditions.

Since Beams [5]), Quarles [6]), and Moore [7]) did not take special measures for cleaning their anodes it is highly probable that their observations must also be explained by mercury droplets on the anode surface. As a matter of fact Beams observed a time-lag of the order of 10^{-7} sec at 1800 kV/cm and Moore a similar value at 400 to 500 kV/cm; the corresponding point M in fig. 9 lies far below the curve representing our measurements taken with clean anodes (curve E).

Recapitulating we have arrived at the following picture:

By an electric field that is by itself far too weak to cause any appreciable field emission, a microscopic hump on the cathode surface is drawn out into a sharp point at which the field strength is high enough for the emission of electrons. The initial hump is probably of thermal origin.

For field strengths above 250 kV/cm the time-lag is completely accounted for by the time required for drawing out the hump as given by the theory of Tonks.

For field strengths below this value a macroscopic movement of the surface and the mechanism just mentioned, taken together, determine the time-lag.

Once field emission has started, the electron current rapidly develops into an arc discharge in a much shorter time.

When the anode is contaminated by minute droplets of mercury the discharge is probably initiated by these droplets being dislodged from the anode by the electric field and colliding with the cathode.

The same mechanism of starting is assumed to be operative in the dielectric ignitor mentioned at the beginning, since the time-lag for these ignitors is of the same order of magnitude. The starting mechanism of this ignitor has recently been reviewed in greater detail in Philips Technical Review [15]).

It may be mentioned that a fuller report of both investigations has been published in the author's thesis [16]).

Eindhoven, May 1947

REFERENCES

[1]) A. Sommerfeld and H. Bethe, Handbuch der Physik XXIV/2, p. 441.
[2]) i.a. R. A. Millikan and C. C. Lauritsen, Phys. Rev. **33**, 598—605, 1929. C. F. Eyring, S. S. Mackeown and R. A. Millikan, Phys. Rev. **31**, 900-910, 1928.
E. W. Müller, Z. Phys. **102**, 734-762, 1936.
[3]) W. Schottky, Z. Phys. **14**, 63-107, 1923.
[4]) R. Haefer, Z. Phys. **116**, 604-624, 1940.
[5]) J. W. Beams, Phys. Rev. **44**, 803-808, 1933.
[6]) L. R. Quarles, Phys. Rev. **48**, 260-265, 1935.
[7]) H. Moore, Phys. Rev. **50**, 344-348, 1936.
[8]) H. W. Hull and E. E. Burger, Phys. Rev. **31**, 1121, 1928.
[9]) L. B. Snoddy, Phys. Rev. **37**, 1687 (A), 1931.
[10]) J. A. Chiles, J. appl. Phys. **8**, 622-627, 1937.
[11]) L. Tonks, Phys. Rev. **48**, 562-569, 1935.
[12]) J. Frenkel, Physik. Z. Soviet Union **8**, 675-680, 1935.
[13]) L. Tonks, J. Franklin Inst. **221**, 613-621, 1936.
[14]) R. Gans, Ann. Phys., Lpz. **74**, 231-252, 1924.
[15]) N. Warmoltz, Philips Tech. Rev. **9**, 105-114, 1947.
[16]) N. Warmoltz, Over het mechanisme van den capacitieven ontsteker en van den weerstands-ontsteker bij kwikdamp-gelijkrichters, Delft 1946.

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

Reprints of the majority of these papers can be obtained on application to the Administration of the Research Laboratory, Kastanjelaan, Eindhoven, Netherlands. Those papers of which no reprints are available in sufficient number are marked with an asterisk (*).

1717: J. M. Stevels: The physical properties of glasses.
IV. The density of phosphate and germanate glasses.
(J. Soc. Glass Tech. **30**, 173-197, 1946).

The theory given in an earlier communication *) on the density of silicate glasses (as a function of their compostion) is extended to borate, boro-silicate and boro-aluminate glasses. It is shown that in these three cases there is an "accumulation region" (where the O ions form tetrahedral instead of triangular configurations round the B^3Z ions) and a "destruction region" (where the OZZ ions are taken up by the breaking of linkages between the network formers). For the boro-aluminates a "crystalline region" (where at least two phases are formed, of which at least one is crystalline) occurs as well. Methods are given for calculating the density. The values so derived are in satisfactory agreement with those obtained by experiment. The theory given is found to hold also for phosphate and germanate glasses.

*) J. M. Stevels, Rec. Trav. chim. Pays-Bas **60**, 85, 1941; **62**, 19, 1942.

(Continued on page 453)

Philips Res. Rep. **2**, 442-453, 1947

THE CONTINUOUS SPECTRUM OF THE HIGH-PRESSURE MERCURY DISCHARGE

by W. ELENBAAS

537.527:535.33

Summary

The intensity of the continuous spectrum (C.S.) of the high-pressure mercury discharge has been measured between $^1/_4$ and 25 atm for inputs from 20 to 60 watts/cm and for diameters between 6 and 60 mm. The ratio of the intensity of the continuum to the intensity of the yellow lines 5770/91 is almost independent of the input and increases linearly with the mean vapour density m/d^2. The ratio at $m/d^2 = 0$ is interpreted as being due to recombination of electrons with ions, whereas the contribution proportional to m/d^2 originates from molecular radiation.

A rough energy balance is given in which the line energy and the energies radiated in the two types of C.S. occur. Using this equation we find the absolute value of the intensity of the C.S. of the U.V.-standard to be more than twice as high as the value measured by Rössler.

The discrepancy between the measured and calculated gradient between 1 and 5 atm disappears on using this energy-balance equation instead of that formerly used in which the line- and the recombination spectrum were the only radiations taken into account. The remaining discrepancy at higher pressures is probably due to the diminution of the ionization potential at these high densities. The magnitudes of these diminutions necessary to account for the measured gradients are tabulated as a function of the mean vapour density.

1. *Introduction*

Though the origin of the continuous spectrum (C.S.) of the high-pressure mercury discharge has been discussed in many papers, a satisfactory solution has not been reached. The first calculations were made by Unsöld [1]), who supposed that the C.S. is caused by recombination of positive ions with electrons. Unsöld's calculations resulted in a C.S. with constant energy per unit of frequency, whereas experiments gave an intensity more or less independent of frequency, but about 10 times as high as the calculated value. Due to the uncertainty of the quantities used in the calculation, recombination could not, however, be definitely excluded as the main cause of the C.S. [2]).

Subsequently Schulz [3]), by adopting a higher arc-temperature, brought the intensity of the C.S., as calculated by Unsöld's method, in line with the measured intensity.

From their observations that the intensity of the C.S. is proportional to $I^{3/2}$ (I = current), Rompe, Schulz and Thouret [4]) likewise concluded that it originates from the recombination of positive ions and electrons. So far the recombination theory was thus most favoured.

Rössler [5]), on the other hand, determined the height of the initial level of the C.S. by comparing the decrease of the intensity of the C.S. at right angles to the axis of the discharge with that of mercury lines of different initial levels, and found a value of 8·6 volts for the apparent initial level of the continuum. If the C. S. were due to recombination, one would expect that the height of the initial level of the C.S. would be 10·4 volts (ionization potential of mercury). Though it is possible that at high pressures the ionization potential is smaller than that of an isolated atom, a diminution to 8·6 volts would imply that the yellow lines, of which the initial levels levels lie at 8·8 volts, would be practically absent, which is not the case. Moreover, the height of the initial level of the C.S., as measured by Rössler, does not diminish with the pressure. Rössler, therefore, favours the assumption that the C.S. is caused by the collision of an excited atom with a normal atom. Rössler [6]) sees a further argument for the molecular theory in the fact that the intensity of the C.S. increases linearly with the input per cm, as, by recombination of ions with electrons, the intensity of the C.S. would be proportional to the square of the current. Rössler mentions two difficulties for this "molecular" theory:

1°) The absolute value of the intensity of the C.S. is in agreement with the intensity calculated by Unsöld from the recombination theory, if the temperatures of Schulz [3]) are used. There is thus no place for a molecular C.S.

2°) The absolute value of the molecular C.S. as calculated by Rompe and Steenbeck [7]) is only one thousandth of the measured intensity.

According to Rössler this second objection is less serious, because it is possible that the number of excited molecules is much larger as a result of a flat minimum in the potential curve of the system: excited atom + normal atom.

In connection with these considerations the following points may be stressed:

A) In his own calculations of the intensity of the C.S. with the aid of the theory of Unsöld, the author did not use the temperature of the arc, since the temperature has a very pronounced influence and was not exactly known. Instead the intensity of the C.S. was calculated via the electrical conductivity of the arc. The main uncertainty in the result was caused by the value of the cross-section of the mercury atom towards electrons, for which $5 \cdot 10^{-15}$ cm^2 was used. Recently [8]) the author determined the temperature of the mercury arc from the intensity of the yellow lines and found for the cross-section of the mercury atom a value of 4×10^{-15} cm^2. With this slightly smaller value the calculated intensity of the C.S., which is already far below that observed, will become even smaller. Moreover, a value of 10·4 volts was used for the ionization potential of the mercury

atom, whereas at these high pressures a value ΔV_i lower should have been employed (see section 4 of this paper). This will lead to a further decrease in the calculated intensity of the C.S. The possibility that at these high pressures recombination causes an essential part of the C.S. seems, therefore, to be definitely ruled out; the temperatures used by Schulz [3]) must be far too high, and the first objection of Rössler against the molecular origin of the C.S. is invalid.

B) The behaviour of the intensity of the C.S. as a function of the current does not give any direct information as to the origin of the C.S. From their $I^{3/2}$-law Rompe, Schulz and Thouret [4]) concluded that the C.S. originates from recombination, whereas from the fact that the intensity of the C.S. is proportional to $(L-A)$ (where L represents the input per cm of arc length and A the loss per cm of arc length by conduction of heat) Rössler [6]) concludes that we are dealing with a molecular spectrum. However, neither of these conclusions is correct. The total radiation [9]) is proportional to $(L-A)$, and the distribution of this energy over the different wavelengths depends on the heights of the initial levels, on the transition probabilities and on the temperature, but the mechanism of the excitation is not involved. To the first approximation the intensity of the C.S. will thus be proportional to $(L-A)$, independently of the origin of the C.S. From a possibly larger or smaller increase, we may only conclude that the initial level lies higher or lower than the average one.

C) In figs 7 and 8 of Rössler's paper [6]), the intensity of the C.S. is plotted as a function of the electrical gradient G for a constant value of L. From the proportionality between the intensity and G, Rössler concludes that the intensity is proportional to the pressure,and therefore that the C.S. originates from the radiation of molecules. As, however, the gradient is proportional to p^q, where $q \approx \frac{1}{2}$, we should rather conclude that the intensity of the C.S. is approximately proportional to the square root of the pressure. As, to the first approximation, the intensity of the recombination spectrum is independent of the pressure, and the intensity of the molecular spectrum is proportional to the pressure, the figs 7 and 8 of Rössler's paper do not lead to any decision as to the origin of the C.S. other than the possibility of both processes playing a part.

2. *The measurements*

Using a double monochromater we have measured the intensity of the C.S. at 6400 and 4800 Å as a function of the input per cm length, of the tube diameter and of the amount of mercury evaporated per cm of length. In addition the joint intensity of the yellow lines 5770/91 Å was determined; these were corrected by subtracting the average intensity of the continuum on both sides of these lines. The tubes contained mercury

plus $2\frac{1}{2}$ atomic % of argon and were run on direct current in the vertical position. In table I the measurements on a tube of 13·8 mm inside diameter are given as an example.

TABLE I

Intensities, in arbitrary units, for varying input (only comparable at constant wavelengths)

L watts/cm	6400 Å	4800 Å	I_c	$I_{5770/91}$	$I_c/I_{5770/91}$
21·5	0·125	0·165	0·145	0·175	0·83
32·3	0·243	0·307	0·275	0·32	0·86
43·9	0·352	0·476	0·415	0·47	0·88[5]
54·8	0·470	0·595	0·530	0·66	0·80
67	0·620	0·790	0·705	0·81	0·87
					0·85

d = diameter = 13·8 mm; m = quantity of vapourized mercury per cm length = 6 mg/cm; $m/d^2 = 31{\cdot}4.10^{-3}$ mg cm^{-1} mm^{-2}; I_c = mean value of the intensity of the C.S. at 6400 Å and 4800 Å; $I_{5770/91}$ = joint intensity of the yellow lines.

The intensity ratio of the C.S. at 6400 Å and 4800 Å is fairly constant. As table I shows, the ratio $I_c/I_{5770/91}$ is practically independent of L, which indicates that the height of the initial level of the C.S. is about the same as the heigth of the initial level of the yellow lines, *i.e.* 8·8 volts. This agrees with the value of 8·6 volts found by Rössler [5] [6]).

In the same way the mean ratio $I_c/I_{5770/91}$ has been determined for a series of tubes differing in diameter and pressure; the results, together with the intensity of the C.S. for an input of 35 watts/cm have been collected in table II. For these different tubes the intensity ratio of the C.S. at 6400 Å and 4800 Å has not always the same value, so that their mean value provides but a poor estimate of *the* intensity of the C.S. $I_c/I_{5770/91}$, as given in table II, has been plotted in *fig. 1*, as a function of m/d^2, this last quantity being proportional to the mean vapour density. At each measurement the m-value is marked. For $m/d^2 < 0{\cdot}05$ (corresponding to pressures below about 6 atm) the points lie on a straight line, which intersects the $I_c/I_{5770/91}$-axis at a positive value. The line may be represented by the equation:

$$I_c/I_{5770/91} = 0{\cdot}24 + 24\, m/d^2. \qquad (1)$$

The absolute values of the numerics 0·24 and 24 have no significance owing to the relative nature of the intensity measurements; the factor 100 between the two coefficients is, however, of importance. Whether the deviation from the straight line for $m/d^2 > 0{\cdot}05$ is real or caused by errors in the measurements is not known.

TABLE II

Mean ratio of the intensities of the C.S. and of the yellow lines for a set of tubes of different diameters

d (mm)	m (mg/cm)	m/d^2	$I_c/I_{5770/91}$	I_c $\left(L = 35 \frac{\text{W}}{\text{cm}}\right)$
5·9	5·9	170×10^{-3}	2·93	0·42
6·05	4·6	126	2·93	0·51
5·95	4·0	113	2·20	0·44
6·05	4·1	112	2·85	0·55
6·0	3·65	101	2·02	0·45
9·8	6·0	62·5	1·59	0·36
15·25	9·25	40	1·28	0·38
14·7	7·95	36·8	1·13	0·35
15·3	7·55	32·3	1·00	0·35
13·8	6·0	31·4	0·85	0·30
21·6	10·0	21·4	0·80	0·30
29·8	18·2	20·5	$0·84^5$	0·32
$21·4^5$	8·9	19·3	0·85	0·33
29·4	10·4	12·0	0·54	0·25
47·5	$13·2^5$	$5·8^5$	0·44	0·22
32·8	6·0	5·6	0·41	0·20
58	5·8	1·72	$0·23^5$	0·11

In *fig. 2* the intensity I_c at $L = 35$ W/cm (table II) is plotted as a function of m/d^2. The m-values are again marked at each point.

In both figures the scattering of the points around the curves is rather large. Moreover, at the lowest values of m/d^2 the ratio of $I_c/I_{5770/91}$ seems

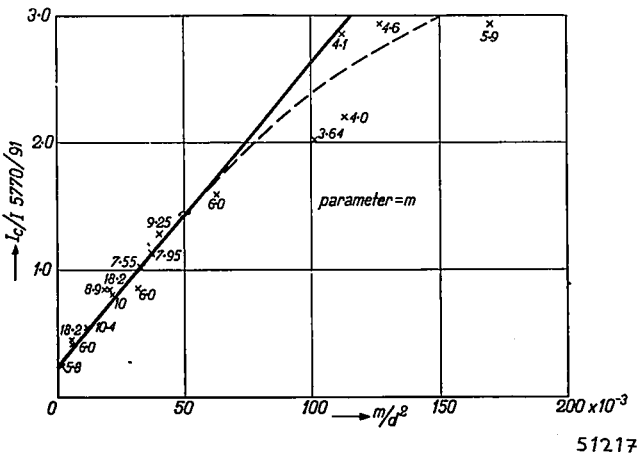


Fig. 1. The intensity of the continuous spectrum divided by the intensity of the yellow lines 5770/91 on an arbitrary scale as a function of the mean vapour density.

to increase with increasing L. The curves drawn in figs 1 and 2 are thus only a rough approximation to the real behaviour of the intensity of the C.S.

3. *Comparison of the intensity measurements with theory*

The intensity radiated per cm of arc length in the line spectrum may be written as:

$$I_l = C_l \, p T^{-1} \, \delta^2 \exp\left(-e\overline{V}/kT\right), \tag{2}$$

if $\overline{V}$ represents the average height [10]) of the initial levels, p the pressure, T the absolute temperature, and δ the diameter of the idealized uniform discharge.

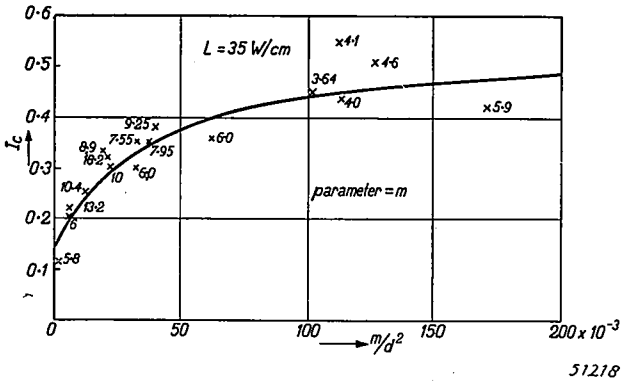


Fig. 2. The intensity of the continuous spectrum at 35 W/cm on an arbitrary scale as a function of the mean vapour density: × measured, —— calculated with (13) and (14).

The intensity of that part of the C.S. which is due to recombination of ions and electrons is, according to Unsöld [1]):

$$I_r = C_r \, p T \delta^2 \exp\left[-e(V_i - \Delta V_i)/kT\right], \tag{3}$$

where V_i represents the ionization potential and ΔV_i the correction to be applied to V_i at high pressures.

The intensity of the C.S. originating from the molecules is, according to Rössler [5]) [6]):

$$I_m = C_m \, p^2 T^{-3/2} \, \delta^2 \exp\left(-eV_m/kT\right), \tag{4}$$

where V_m is 8·6 volts.

In a first approximation the pressure p is proportional to m/d^2, but it depends on the temperature distribution too. This influence, however, as well as the influence of the T-powers in (2), (3) and (4), is small compared with the influence of T occurring in the exponential functions. We may therefore disregard these T-powers and write:

$$I_l = C_l' \delta^2 \frac{m}{d^2} \exp(-e\bar{V}/kT), \tag{5}$$

$$I_r = C_r' \delta^2 \frac{m}{d^2} \exp[-e(V_i - \Delta V_i)/kT], \tag{6}$$

$$I_m = C_m' \delta^2 \left(\frac{m}{d^2}\right)^2 \exp(-eV_m/kT). \tag{7}$$

Denoting the input per cm of arc length by L, and the loss by conduction per cm of arc length by A, the total radiation per cm of arc length is $(L-A)$. We may thus write:

$$L-A = C_l' \delta^2 \frac{m}{d^2} \exp(-e\bar{V}/kT) + C_r' \delta^2 \frac{m}{d^2} \exp[-e(V_i - \Delta V_i)/kT] + C_m' \delta^2 \left(\frac{m}{d^2}\right)^2 \exp(-eV_m/kT). \tag{8}$$

Assuming δ/d and A as constant, equation (8) determines the temperature T of the idealized, uniform discharge as a function of m and L.

As we will consider the influence of m/d^2 on the intensity distribution, T varying only slightly with the pressure, we shall in a first approximation assume:

$$\bar{V} \approx V_i - \Delta V_i \approx V_m \quad (\bar{V} \approx 7{\cdot}8 \text{ V};\ V_i - \Delta V_i \approx 10{\cdot}4 - 9 \text{ V};\ V_m \approx 8{\cdot}6 \text{ V}).$$

Denoting this common voltage by V, we may write (8) as:

$$L-A = \delta^2 \frac{m}{d^2} [C_l' + C_r' + C_m' m/d^2] \exp(-eV/kT). \tag{9}$$

For the ratio of the total intensity I_c of the C.S. to the total intensity of the lines I_l, we find from (9)

$$\frac{I_c}{I_l} = \frac{C_r' + C_m' m/d^2}{C_l'}. \tag{10}$$

This holds exactly only if the temperatures of the discharges under consideration are the same. We should thus choose the inputs in such a manner that the temperatures of the discharges are the same. As experimentally the ratio of the intensity of the C.S. to that of the yellow lines is, to the first approximation, independent of the input, we may write:

$$\frac{I_c}{I_{5770/91}} = \frac{C_r' + C_m' m/d^2}{C_{5770/91}} = \alpha + \beta m/d^2, \tag{11}$$

This is, however, exactly what fig. 1 shows. α is thus interpreted as due to

that part of the C.S. which is caused by recombination, whereas $\beta m/d^2$ represents the molecular contribution. The experiment gives $\beta/\alpha \approx 100$, so that too:

$$C'_m/C'_r \approx 100. \tag{12}$$

In order to determine C'_l in (9), we consider the intensity of the C.S. as a function of m/d^2 at a constant value of L (fig. 2). From (9), we find with (12) for the energy I_c in the C.S.:

$$I_c = (L-A)\frac{C'_r + 100\, C'_r\, m/d^2}{C'_l + C'_r + 100\, C'_r\, m/d^2}. \tag{13}$$

As L is constant in fig. 2, the course of I_c is determined by the fraction appearing in the right-hand side of (13). The curve in fig. 2 is drawn according to (13) with:

$$C'_l = 3\, C'_r, \tag{14}$$

the ordinate scale being chosen in such a way as to achieve the best agreement, which is allowed because the I_c-scale is arbitrary. Equation (9) may thus be written as:

$$L-A = \delta^2 \frac{m}{d^2} C'_r\, [3 + 1 + 100\, m/d^2] \exp(-eV/kT). \tag{15}$$

From the terms between brackets, the first gives the energy in the lines, the second the energy of the C.S. due to recombination, and the third term the energy of the C.S. due to the molecular radiation. Thus at low pressures the part of the molecular C.S. is small. In table II the smallest value of m/d^2 is $1{\cdot}72\ 10^{-3}$, so that at this pressure the ratio of the recombination contribution to the molecular contribution is as $1 : 0{\cdot}172$. At the corresponding pressure of about $\frac{1}{4}$ atm, the C.S. is thus chiefly due to recombination. In accordance herewith the ratio $I_c/I_{5770/91}$ increases with increasing L, in contradiction to what is found at higher pressures, because the initial level of the recombination spectrum lies at a height of 10·4 volts. In order to measure $I_c/I_{5770/91}$ at constant temperature (similar discharges), we should, according to (15), choose L so that $(L-A)/m[(1 + 25\ m/d^2)]$ is constant (fig. 1). Likewise we should in fig. 2 draw the curve through m-values for which $m(1 + 25\ m/d^2)$ is constant (for instance $m = 15$ at $m/d^2 = 0$ and $m = 2{\cdot}5$ at $m/d^2 = 200 \times 10^{-3}$). The scattering of the points is, however, such that this second approximation is without meaning.

B) From (13) and (14) we derive for the total energy L_c radiated in the C.S.:

$$I_c = (L-A)\frac{1 + 100\, m/d^2}{4 + 100\, m/d^2}, \tag{16}$$

and for the energy I_l radiated in the line spectrum:

$$I_l = (L-A)\frac{3}{4+100\,m/d^2}. \tag{17}$$

Rössler [11]) measured the total energy in the C.S. for the U.V.-standard [12]) and found a value of $\frac{1}{3}$ for the ratio of the energy in the C.S. to the energy in the lines. Since for the U.V.-Standard $m \approx 3{\cdot}7$ and $d = 18$, eqs (16) and (17) give for this ratio:

$$\frac{I_c}{I_l} = \frac{1+100\times 3{\cdot}7/324}{3} = 0{\cdot}71\,.$$

The calculated intensity of the C.S. is thus about twice as large as the measured intensity. We must, however, bear in mind that it is possible that the C.S. is partly radiated in a region of the spectrum (far ultraviolet or far infrared) where no measurements were made, so that the real ratio may be larger than the one observed.

4. *The electrical gradient*

For the electrical gradient G of the high-pressure mercury discharge we derived the formula [10]):

$$G \backsim \frac{L^{1/2}}{(L-A)^{1/3}}\,\frac{m^{7/12}}{d^{3/2}}. \tag{18}$$

This formula is quite correct for discharges at pressures of the order of 1 atm. At higher pressures, however, we found [13]) larger values for G than those calculated from (18), assuming a proportionality factor of 185, as found in the 1-atm pressure range. In *fig. 3* we reproduce the factor f, by which the measured gradient is larger than that computed from (18), plotted as a function of m/d^2 (fig. 2 of [13])). We have previously [13]) discussed several possible reasons for this discrepancy without reaching a satisfactory solution.

In deriving eq. (18), we used the energy-balance equation:

$$L-A = mq\,\exp\,(-eV/kT), \tag{19}$$

which only accounts for the radiation of the lines and the recombination part of the continuum. We shall now show that by using (15) instead of (19) in deriving the formula for the gradient, we find a close agreement with the experiment.

Using the formula of Saha and the mobility equation of Langevin, the current density i is given by:

$$i \backsim G\,p^{-1/2}\,T^{3/4}\exp\,[-e(V_i-\Delta V_i)/2kT]\,. \tag{20}$$

Neglecting the powers of T (the variation of T being small) and substituting m/d^2 for p, we find from (20):

$$G \backsim \frac{L^{1/2}\, m^{7/4} \exp\,[e(V_i - \Delta V_i)/4kT]}{\delta d^{1/2}}. \tag{21}$$

According to (15), we have:

$$\exp\,(eV/kT) \backsim \delta^2 \frac{m}{d^2} \frac{1 + 25\ m/d^2}{(L - A)}.$$

Using this eq. in (21), we obtain:

$$G \backsim \frac{L^{1/2}\, m^{1/4}}{\delta d^{1/2}} \left[\frac{\delta^2 m\ (1 + 25\ m/d^2)}{d^2(L - A)}\right]^{V_i/4\ V} \exp\,(-e\Delta V_i/4kT). \tag{22}$$

To achieve agreement with the experiments in the one-atmosphere range, where $\Delta V_i = 0$, we have to choose $V = \frac{3}{4}\ V_i$, so that (22) becomes:

$$G \backsim \frac{L^{1/2}}{(L - A)^{1/3}} \frac{m^{7/12}\,(1 + 25\ m/d^2)^{1/3}}{\delta^{1/3}\, d^{7/6}} \exp\,(-e\Delta V_i/4kT). \tag{23}$$

Comparing (23) with (18), we obtain for the factor f:

$$f \backsim \frac{d^{1/3}}{\delta^{1/3}} (1 + 25\ m/d^2)^{1/3} \exp\,(-e\Delta V_i/4kT).$$

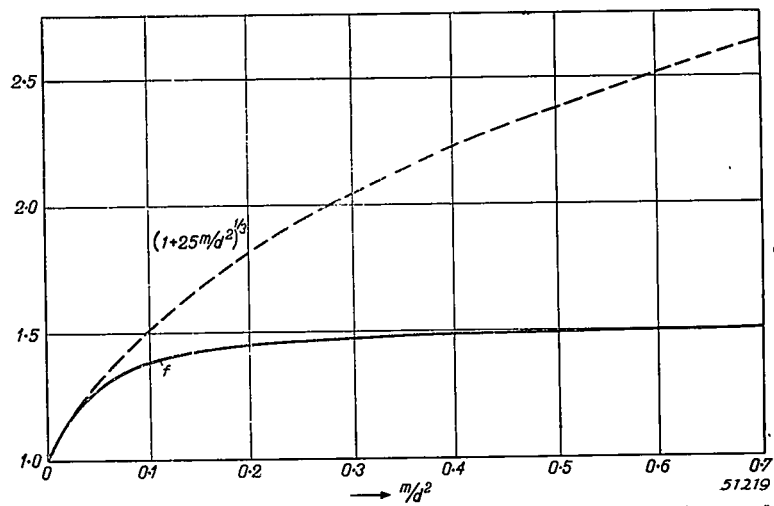


Fig. 3. The factor f by which the gradient calculated according to (18) must be multiplied, to arrive at the measured gradient. As a result of the molecular part of the continuous spectrum the gradient has to be multiplied by $(1 + 25\ m/d^2)^{1/3}$, so that for $m/d^2 < 0{\cdot}05$ the continuous spectrum explains the f-value. . . . .

The ratio of the diameter δ of the discharge to the inner tube-diameter d varies only slowly, so that $(d/\delta)^{1/3}$ is fairly constant. Neglecting the variation of $(d/\delta)^{1/3}$, we have:

$$f = (1 + 25\ m/d^2)^{1/3} \exp\ (-e\Delta V_i/4kT) \qquad (24)$$

(as, for small values of m/d^2, $\Delta V_i = 0$ and $f = 1$). The broken line in fig. 3 represents $(1 + 25\ m/d^2)^{1/3}$ as a function of m/d^2. For small values of m/d^2 this curve coincides with the f-curve. This forms a check on the factor 25 in the term $1 + 25\ m/d^2$. The combination of (12) and (14) thus just accounts for the higher value of G up to $m/d^2 = 0{\cdot}05$ (about 6 atm).*) At higher values of m/d^2, the gradient is smaller than $185\ (1 + 25\ m/d^2)^{1/3}\ L^{1/3}\ m^{7/12}/(L-A)^{1/3}\ d^{3/3}$, which may be accounted for by the factor $\exp\ (-e\Delta V_i/4kT)$ of eq. (24). We may calculate the value of ΔV_i necessary to bring eq. (24) in agreement with the experimental f-value. For this we need the value of T, but as T does not vary much, we will use a constant value of 5250 °K [8]). In table III these values of ΔV_i are tabulated for different values of m/d^2.

TABLE III

m/d^2	p (atm)	$(1+25\ m/d^2)^{1/3}$	f (measured value)	$\exp\ (11600\ \Delta V_i/4T)$	ΔV_i (volt)
0·025	3	$1{\cdot}17_5$	1·17	1·09	0
0·05	6	1·31	1·29	1·02	0·04
0·1	13	1·52	$1{\cdot}38_5$	1·10	0·18
0·2	25	1·82	1·45	$1{\cdot}25_5$	0·44
0·4	50	$2{\cdot}22_5$	1·49	1·49	0·76
0·7	90	$2{\cdot}64_5$	1·50	$1{\cdot}65_5$	1·09

These values of ΔV_i are reasonable, so that we may conclude that the f-value in our previous gradient formula may well be explained by the appearance of the molecular part of the C.S. and by the decrease of the ionization potential, the first being able to account for the discrepancy up to about 5 atm and the latter becoming perceptible above this pressure. At very high pressures the decrease of the height of the initial levels of the lines may also influence the gradient. As a result of a decrease of V, a smaller value of ΔV_i would suffice. It is therefore possible that for the highest pressures ΔV_i is somewhat smaller than indicated in table III.

The intensity measurements were carried out by Mr J. Riemens.

*) It is possible to write (14) as: $C_e' = 4\ C_r'$ without obtaining a calculated curve in fig. 2 that disagrees with the measurements. The number 25 then becomes 20. The curve representing $(1 + 20\ m/d^2)^{1/3}$ does not fit as well as the one drawn in fig. 3, but the deviation is smaller than the possible errors in the gradient measurements. The values of ΔV_i become smaller with this change and for the U.V.-standard I_e/I_l becomes 0·53 instead of 0·71 and thus in better agreement with the measured value of 0·33.

Eindhoven, June 1947

REFERENCES

1) A. Unsöld, Ann. Phys. Leipzig **33**, 607, 1938.
2) W. Elenbaas, Physica **6**, 299, 1939.
3) P. Schulz, Z. Phys. **119**, 167, 1942.
4) R. Rompe, P. Schulz & W. Thouret, Z. Phys. **112**, 369, 1939.
5) F. Rössler, Z. Phys. **112**, 667, 1939.
6) F. Rössler, Z. Phys. **122**, 285, 1944.
7) R. Rompe & M. Steenbeck, Ergeb. Exakt. Naturw. **18**, 257, 1939.
8) W. Elenbaas, Philips Res. Rep. **2**, 20, 1947.
9) W. Elenbaas, Physica **4**, 413, 1937.
10) W. Elenbaas, Physica **2**, 757, 1935.
11) F. Rössler, Ann. Phys. Leipzig **34**, 1, 1939.
12) H. Krefft, F. Rössler & A. Rüttenaner, Z. tech. Phys. **18**, 20, 1937.
13) W. Elenbaas, Physica **4**, 279, 1937.

ABSTRACTS

(Continued from page 441)

1719/1721: C. J. Bouwkamp: On the dissection of rectangles into squares I, II, III (Proc. Kon. Ned. Akad. Wetenschappen Amsterdam **49**, 1176-1188, 1946; **50**, 58-71, 72-78, 1947).

According to Brooks, Smith, Stone and Tutte, the semi-topological problem of the dissection of a rectangle into non-overlapping, unequal squares is reduced to a physical problem: that of the distribution of current in a planar electrical network. The networks required for squarings of order not exceeding 14 are shown. The corresponding squared rectangles are codified and classified. Construction of squared squares.

1722: H. J. Lindenhovius: Het meten van impedanties bij hoge frequenties en toepassingen van de staande-golfindicator (T. Ned. Radiogenootschap **12**, 60-82, 1947). (The measurement of impedance at h.f. and applications of the standing-wave indicator).

In this article a survey is given of the different methods used in measuring impedances. For frequencies below 300 Mc/s the conventional method is that which employs a tuned circuit while the impedance is determined from its damping and detuning effects on the circuit. For higher frequencies the lumped circuit may be replaced by a tuned transmission line, but to avoid certain difficulties it is preferable to use an untuned line and to determine the impedance from the voltage distribution along the line. This voltage distribution is characterized by the standing-wave-ratio and the position of the voltage minimum. A new diagram is designed which enables one to determine the impedance graphically in a most comprehensible way, even for very large values of the standing-wave-ratio. The standing-wave indicator used is described. Other applications of this instrument concern the measurement of the characteristic impedance and the attenuation constant of a transmission line, and the measurement of the net power flow along the line.

(Continued on page 474)

Philips Res. Rep. 2, 454-467, 1947

ABSOLUTE MEASUREMENT OF THE TIME CONSTANT OF RESISTORS

by J. W. L. KÖHLER and C. G. KOOPS

621.3.08:621.317.33

Summary

A new method is described for the accurate determination of the time constant of resistors in absolute measure, using as standards of reference a set of standard condensers with negligible losses as described in a previous paper.

1. *Introduction*

With the aid of a modern A.C. bridge, the determination of the difference between the time constants *) of a given resistor and a standard resistor has become a fairly simple measurement. This measurement is very similar to the comparison of the loss angles of two condensers, and the bridges used in both cases have to satisfy similar requirements when equal accuracies are demanded. For the final determination of the unknown time constant it is, of course, necessary that the time constant of the standard resistor is known. The present paper deals with the absolute determination of time constants, which constitutes a much more complicated problem than the simple comparison.

So far as the authors are aware, two methods have previously been used for the solution of this problem. The first consists in the construction of resistors of such geometrical properties that the self-inductance and the parallel capacitance are calculable, in which case the time constant $T(\approx L/R - RC)$ is known. For the other method two resistors are constructed of exactly equal geometrical properties while for both resistors materials with different conductivities are used. The resistances must be so low as to permit the omission of the influence of the parallel capacitance. In this case the resistance values are different but the inductances are equal. After measurement of the difference of the time constants both are known ($T_1 - T_2 = L/R_1 - L/R_2$), provided R_1 and R_2 have been measured separately. It is difficult to achieve a high degree of precision with either of these methods, and consequently the uncertainty in published figures is seldom smaller than 10^{-9} sec.

The problem under consideration is very similar to that of the absolute

*) The time constant T of a resistor is defined by the equation: $Z = R(1 + j\omega T)$, where Z denotes the impedance and R the effective resistance of the resistor at the angular frequency ω.

determination of the losses of a condenser. Since in previous work a set of standard condensers had been obtained, the loss angles of which were determined with great accuracy [1]), trying to use these condensers for the solution of the present problem seemed promising. As a matter of fact a method has been found allowing of the time constants of two resistors being measured with an uncertainty of 10^{-11} sec, including random and systematic errors.

2. *Principle of the method*

The principle of the new method is essentially that of the determination of two unknown quantities by the separate measurement of their difference and their sum. Two resistors are used, the time constants of which are to be found.

Firstly, the *difference* between the time constants of the resistors under test R_1, R_2 (*fig. 1*) is determined, using either an equal-arm bridge or a ratio bridge depending upon the question whether R_1 and R_2 are equal in value or not.

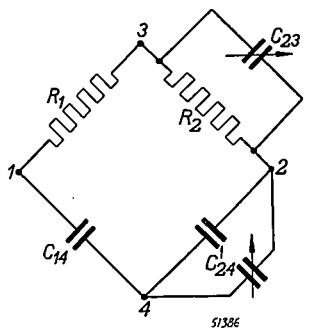


Fig. 1. Equal-arm or ratio bridge.

C_{14} and C_{24} are two of the standard condensers mentioned above, one of which (C_{24} in fig. 1) is shunted by a small variable condenser. The bridge is balanced by adjustment of this small condenser and of a condenser in parallel with R_1 or R_2 (in fig. 1 the latter case has been chosen). As the power factors of the standard condensers are so small that they may be ignored, the value of R_2 C_{23} is the required difference of time constants.

Then a second measurement, carried out with a Maxwell bridge, furnishes the *sum* of the time constants. This bridge network, shown in *fig. 2*, is ordinarily used for the determination of the effective self-inductance and resistance of coils. In the present case a coil is used of which the self-inductance and the effective resistance are determined with a resonance bridge, as will be described below.

It is assumed provisionally that the time constants of the resistors R_3 and R_4 are zero, and that R_1 and R_2 have the time constants T_1 and T_2 respectively. The condition of balance is as follows:

$$\frac{R_l + R_3 + j\omega L}{G_4 + G_c + j\omega C_4} = R_1(1 + j\omega T_1) \,.\, R_2(1 + j\omega T_2)$$

where R_l denotes the effective series-resistance of L, G_c the effective parallel conductance of C_4, and $G_4 = 1/R_4$. It is assumed that ωT_1 and ωT_2 are so small that their squares and products may be ignored. In this case separation of reals and imaginaries leads easily to:

$$R_3 + R_l = R_1R_2(G_4 + G_c) - \omega^2L(T_1 + T_2)\,. \qquad (1)$$

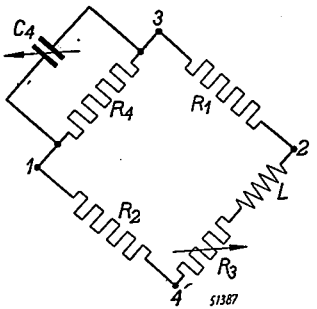


Fig. 2. Maxwell bridge.

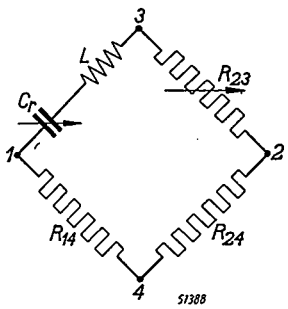


Fig. 3. Resonance bridge.

Evidently, this equation remains practically unchanged when R_3 and R_4 have small time constants, as in this case C_4 and L only are affected, and this leads to a negligible correction. When the other quantities involved are known, $T_1 + T_2$ may be calculated from (1). Apparently R_l and G_c have to be determined first.

G_c is measured with a ratio bridge [1]), while for the determination of R_l a resonance method is used as shown in *fig. 3*.

When balance has been obtained, L and C_r are in resonance, and consequently:

$$R_l + R_c = R_{23} \,.\, \frac{R_{14}}{R_{24}}\,. \qquad (2)$$

R_c denotes the effective series resistance of the condenser, and is also measured with a ratio bridge. R_{23}, R_{14}, and R_{24} may have small time constants as these affect the adjustment of C_r only.

3. *Further details*

In the chain of measurements described above there are some points where errors are likely to occur. These points will now be considered more closely.

a. *Resonance measurement*

The bridges used for the work described in this paper were provided with a Wagner earth-connection which, as is well known, renders harmless all stray admittances between branch points of the bridge and earth. Consequently, great care should always be taken when a bridge-network contains a branch consisting of two or more impedances in series, as in this case some stray capacitances still may cause errors. In *fig. 4a* the branch 1-3 of fig. 3 is drawn in detail.

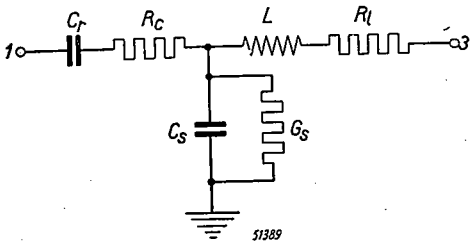


Fig. 4*a*. Series connection of coil and condenser in the resonance bridge.

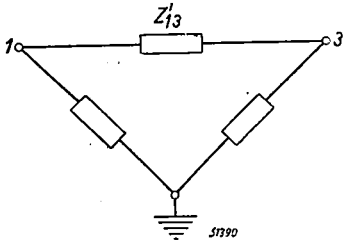


Fig. 4*b*. Transformed network of fig. 4*a*.

C_r and L have the same meaning as in fig. 3, while the losses of C_r are represented by the series resistance R_c and those of L by the series resistance R_l. A stray capacitance C_s between earth and the point joining coil and condenser has been added, its losses being represented by the parallel conductance G_s.

As the Wagner earth-connection deals with the admittances to earth from branch points, the star in fig. 4*a* has to be transformed into the equivalent mesh as shown in fig. 4*b*. From this mesh the impedance Z'_{13} is the only important quantity, the other two being impedances to earth from branch points. Z'_{13} may be written in the form:

$$Z'_{13} = R_c + \frac{1}{j\omega C_r} + R_l + j\omega L + \Delta Z_{13}.$$

The transformation now leads to:

$$\Delta Z_{13} = \left(R_l R_c + \frac{L}{C_r}\right) G_s - \left(R_c \omega L - \frac{R_l}{\omega C_r}\right) \omega C_s + \\ + j\left[\left(R_c \omega L - \frac{R_l}{\omega C_r}\right) G_s + \left(R_l R_c + \frac{L}{C_r}\right) \omega C_s\right].$$

It is assumed that C_s will be small compared with C_r, in which case $\omega^2 L = 1/C_r$ in close approximation. The real part of ΔZ_{13}, the only important one, may now be written as:

$$\Delta R_{13} = \left(R_l R_c + \frac{L}{C_r}\right) G_s + \frac{C_s}{C_r}(R_l - R_c).$$

$R_l R_c$ will be very small compared with L/C_r and may be ignored. Finally, G_s may be replaced by $\omega C_s \tan\delta_s$ and R_l by $\omega L \tan\delta_l$, whence:

$$\Delta R_{13} \approx \frac{C_s}{C_r}(R_l - R_c + \omega L \tan\delta_s) = \frac{C_s}{C_r}\{\omega L (\tan\delta_l + \tan\delta_s) - R_c\}. \quad (3)$$

Evidently it is essential to make C_s as small as possible. This can be achieved by surrounding the coil and the condenser each by a screen connected to one branch point (*fig. 5*), and by making the distance between these screens very small.

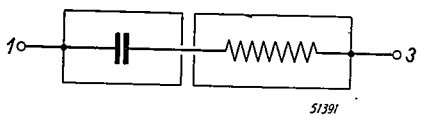


Fig. 5. Coil and condenser with screens connected to branch points.

The capacitance introduced directly between 1 and 3 is harmless as it shunts the whole branch which has a low impedance ($R_l + R_c$); the only effect is a somewhat different adjustment of the condenser.

The screen enveloping the coil will inevitably increase its losses in consequence of an increased parallel capacitance (*fig. 6*).

It is easily derived that the increase of R_l caused by C_p is given by

$$\Delta R_l = 2 R_l \frac{C_p}{C_r} + G_p \omega^2 L^2 = \frac{C_p}{C_r} \cdot \omega L (2 \tan\delta_l + \tan\delta_p). \quad (4)$$

It is important to make C_p as small as possible although it need not be zero. This point must be kept in mind when designing the coil.

The effective resistance R_l of the coil consists of two parts, *viz.* the D.C. resistance R_0 and the additional resistance R_a due to A.C. losses. Of these, R_0 is by far the greater and, moreover, it has a large temperature coefficient. As inevitably some time will elapse between the measurements

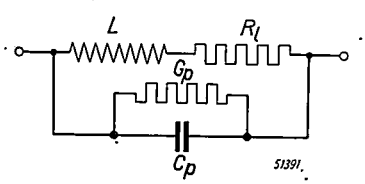


Fig. 6. Equivalent network of coil with parallel capacitance.

with the two bridges, a change of room temperature may thus cause a considerable error. If the D.C. resistance of the coil is eliminated from the measurements this error is avoided. For this reason R_0 is determined separately immediately after the resonance measurement and also immediately after the Maxwell-bridge measurement. In the case of the resonance bridge this is carried out by short-circuiting the condenser and connecting a D.C. source and a galvanometer to the bridge. For the calculation of R_a from the result of the resonance measurement R_c must be known. This quantity is measured with the same arrangement where now the coil is short-circuited while in branch 2-3 a variable air-condenser, the losses of which are known, is inserted in series with the resistance box. For this measurement a bridge ratio is chosen in accordance with the range of the variable air-condenser (100 to 1500 pF), while for the sake of convenience this ratio is also used for the resonance measurement. Unlike the resonance measurement, the measurement of R_c requires the resistors R_{24} and R_{14} to have equal phase angles.

Finally, the effective resistance of the leads connecting the coil and the condenser to the branch points has to be taken into account. This resistance also consists of a D.C. resistance r_0 and an additional A.C. resistance r_a. The resonance measurement furnishes the value of

$$R_0 + R_a + R_c + r_0 + r_a = r_1,$$

the D.C. measurement that of

$$R_0 + r_0 = r_2,$$

while the measurement of the condenser gives

$$R_c + r_0 + r_a = r_3.$$

Subtracting the second and the third equation from the first leads to

$$R_a - r_0 = r_1 - r_2 - r_3.$$

Evidently r_0 must be measured separately and then added to the result of the subtraction.

At high frequencies a small correction has also to be made for the resistance value of the resistance box used in branch 2-3 as this value deviates somewhat from the D.C. resistance.

b. *Measurement with the Maxwell bridge*

Here the branch 2-4 (fig. 2) containing the resistance box R_0 and the coil L is apt to give trouble. This branch is shown separately in *fig.* 7 where again R_l denotes the resistance of the coil and C_s a capacitance between earth and the point where the coil and the resistance box are

connected. In the same way as above, the star is transformed into the equivalent mesh, and an extra resistance ΔR_{24} is found given by:

$$\Delta R_{24} = -R_3 \omega^2 L C_s . \tag{5}$$

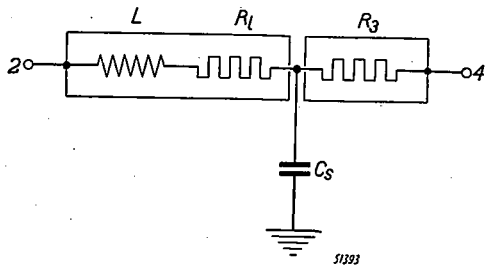


Fig. 7. Screened coil and resistance box in Maxwell bridge.

Evidently, making C_s small is as important in this case as it was in the case of the resonance measurement. Consequently here, too, the coil and the resistance box must be surrounded by screens connected to the branch points. In one respect this case is more difficult than that of the resonance measurement, as now a stray capacitance between the two screens is not harmless. Such a capacitance would shunt the whole branch as shown in *fig. 8* where C_p is the parallel capacitance.

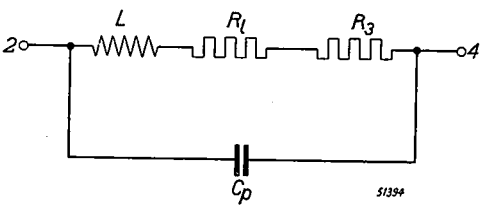


Fig. 8. Series connection of coil and resistance box with parallel capacitance.

C_p now causes on extra resistance $\Delta' R_{24}$ given by:

$$\Delta' R_{24} = 2(R_3 + R_l) . C_p . \omega^2 L . \tag{6}$$

This capacitance is avoided by placing an earthed screen between the screens of the coil and the resistance box. This screen might have a small capacitance towards the connection point but this can be limited to some tenths of a pF, and, moreover, its effect can be reduced by choosing for R_0 such a value as will make R_3 low (see eq. (5)).

The D.C. resistance of the coil is measured immediately after the A.C. measurement has been performed. It is very simple here, as the bridge

network remains unchanged and merely a D.C. source and a galvanometer have to be connected.

In case the A.C. resistance values of R_3 and R_4 deviate from the D.C. values, this error is traced and corrected in the following way. The coil is replaced by a short piece of resistance wire having approximately the same resistance, the A.C. value of which (at the frequency of the measurement) may safely be assumed to be equal to the D.C. value. Now an A.C. and a D.C. adjustment of the bridge are made and the difference between the values of R_0 in both cases is the error to be found.

The power factor of the condenser C_4 is determined separately. This condenser is connected to the Maxwell bridge by a pair of fairly long leads, the resistance of which is measured and allowed for. At high frequency this resistance limits the value of C_4, lest its influence on the power factor of C_4 becomes too great.

c. *Design of the coil*

As may be clear from section 2, the design of the coil must be such as to satisfy the essential requirement that the coil will have exactly equal losses when both measurements are made. This implies

a) that the losses of the coil must be very constant in the course of time;
b) that the losses of the coil must be influenced neither by the way the coil is connected in both cases, nor by its surroundings.

These requirements are satisfied by the following design. For the sake of constancy in the course of time an air coil was chosen while the demand b) can only be satisfied by making the external magnetic field of the coil extremely weak. Therefore a core of toroidal form was chosen and the whole coil was mounted in a copper housing of 5 mm thickness. These precautions resulted in the coil being highly independent of its surroundings; even a lump of copper placed upon the housing of the coil had no perceptible effect either on its losses or on its inductance. For the winding a bank winding was taken, this type of coil having the lowest possible self-capacitance, while by the use of finely stranded wire eddy-current loss is avoided. The toroidal core consists of porcellain, and it proved necessary to provide it with an air gap in the space between beginning and end of the winding in order to reduce the residual self-capacitance, since for the bank winding the direct capacity between beginning and end is the main component of the self-capacitance. After completion, the coil was fixed solidly into the housing and, in high vacuum, the whole housing was filled with compound. The housing serves as a screen, and consequently one of the terminals of the winding is connected to it. The other terminal is insulated from the housing by polystyrene.

Fig. 9 shows in detail how the coil is connected to the condenser box

for the resonance measurement. Here 1 and 2 are the terminals of the condenser. The condenser box is enveloped by two shields, the inner shield being connected to terminal 2, the outer one earthed. The coil, in its housing, is placed on the earthed shield of the condenser, separated from it by an insulating sheet. Point 1′ is the terminal of the coil insulated from the housing. It is of a very solid construction so as to minimize variations of capacitance here. This connection is screened off by a cover

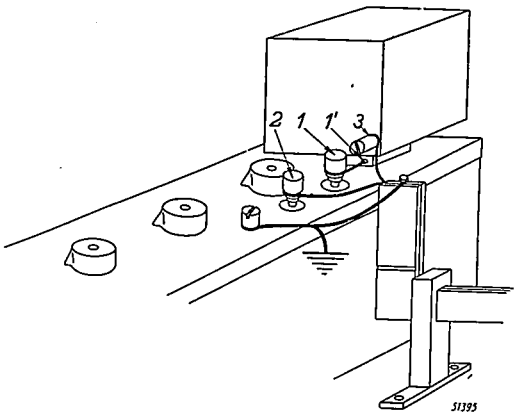


Fig. 9. Connection of coil and condenser box for resonance measurement.

connected to the inner shield of the condenser, and consequently has no earth capacitance. Moreover, this cover prevents any extra capacitance coming in parallel with the coil by screening off the terminal 1 from the housing of the coil.

The coil and the condenser are connected to the bridge by means of a special link ending in two flexible leads fixed to the terminals 2 and 3, the latter being screwed in the housing of the coil. For the measurement of the condenser losses these leads are connected to the terminals 1 and 2 and for the measurement of the D.C. resistance of the coil to the terminals 1 and 3. Thus, there are no complications arising from the use of short-circuiting straps.

Fig. 10 shows in detail how the coil is connected to the Maxwell bridge. The coil is placed in one of the compartments of a copper case containing the whole bridge network. This case is grounded, and the housing of the coil is insulated from it by a sheet of paper.

The connection point 4 is enclosed by two small boxes, *b* and *c*. Of these, *b* is connected to the screen of the resistance R_3 while *c* is earthed. Point 4 is connected to the coil terminal insulated from the housing *a*

while point 3 is connected to this housing. The connection leads pass through slots in the boxes, and after the connections have been made the boxes are closed so that the shielding is complete.

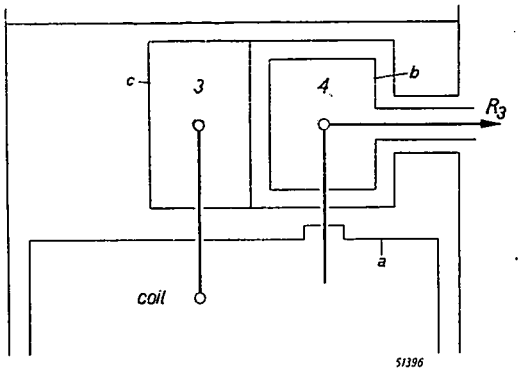


Fig. 10. Connection points of coil in Maxwell bridge.

Evidently the capacitance between a and b is very small because of box c. There is a small capacitance between the connection lead of 4 and the grounded box c, but this capacitance certainly will not exceed 1 pF.

4. *Accuracy of measurement and choice of impedances*

Equation (1) may be written in the form:

$$R_1R_2(G_4 + G_c) - (R_3 + R_l) = \Delta R = \omega^2 L(T_1 + T_2).$$

In the case that T_1 and T_2 have been made equal, this reduces to:

$$\Delta R = 2\omega^2 LT \tag{7}$$

where $T_1 = T_2 = T$.

This means that, independently of frequency, a certain time constant corresponds to a certain resistance difference if the self-inductances of the coils used for different frequencies be chosen in such a way that $\omega^2 L$ is a constant. On the other hand, when expressed in terms of phase difference, the accuracy attainable with the bridges used for this work is roughly independent of frequency, at least for the lower frequency range. Consequently, accurate determinations of time constants are best carried out at rather high frequencies.

Provided all necessary precautions are taken, an inaccuracy could be attained of 10^{-7} radian between 1 and 16 kc/sec. This corresponds to an inaccuracy of about 2.10^{-12} sec for the time constant. Between 16 and 100 kc/sec the accuracy in terms of phase angle decreases by about a factor of 6 so that at 100 kc/sec the uncertainty of bridge balance still

corresponds to about 2.10^{-12} sec in the time constant. It is impossible, however, to carry out the absolute determination of the time constant with the same accuracy, as a consequence of the great number of measurements involved. Therefore an absolute inaccuracy of $\pm 10^{-11}$ sec in time constant was aimed at. Evidently, variations in the time constants of the resistors must be well below this level if such an accuracy is to have any significance. Therefore both the resonance bridge and the Maxwell bridge were designed very carefully, containing a special "heart" as described in a previous paper [2]). When this arrangement is used, the variation of the time constant of a resistor when being connected several times and to different connection points, will certainly not exceed 10^{-12} sec.

Three coils were available to be used at three different frequencies. A number of numerical data concerning these coils are to be found in table I.

A further explanation of this table may follow here.

At 8 and 17·5 kc/sec the resistors used in the Maxwell bridge have the resistance values (see fig. 2) $R_1 = R_2 = 1000\ \Omega$, at 100 kc/sec $R_1 = 100\ \Omega$ and $R_2 = 1000\ \Omega$.

After the coil of 73 mH had been completed, it became evident that a convenient procedure is attained when for the resonance measurement and for the Maxwell measurement the same value of capacitance is used, as in this case one condenser-loss measurement is spared. It may be seen in table I that this condition is fulfilled for 17·5 and 100 kc/sec.

Evidently, the relative inaccuracy must be very low at 8 kc/sec (0·01%) whilst at 17·5 kc/sec the absolute inaccuracy of 2·2 mΩ can only be obtained with some difficulty. It may be clear that the variation of the D.C.

TABLE I

Coil	I	II	III	
Frequency	8	17·5	100	kc/sec
Self-inductance of coil	73·4	9·09	0·52	mH
Resonance capacitance	5375	9095	5200	pF
Permissible inaccuracy in difference of resistance measurements	3.7	2.2	3.8	mΩ
D.C. resistance of coil	33	4·2	0·34	Ω
Variation of D.C. resistance per °C	130	18	1·3	mΩ
Additional A.C. resistance of coil	3·440	0·483	0·457	Ω
Variation of additional resistance per °C	75	6	−0·5	mΩ
Necessary capacitance in Maxwell bridge	73000	9090	5200	pF
Correction for condenser losses in resonance measurement	284	60	30	mΩ
Correction for condenser losses in Maxwell measurement	259	60	33	mΩ
Permissible inaccuracy in measurement of power factor of the condenser	2·5	5	32	10^{-7} rad

resistance with temperature makes it absolutely necessary to eliminate this resistance by measuring consecutively with alternating current, direct current, and alternating current. When the room temperature is reasonably constant, the two A.C. measurements will not differ by more than 6 mΩ. The variation of the additional resistance with temperature makes it equally necessary to apply this procedure to the measurements with both bridges, so that the succession becomes: resonance bridge, Maxwell bridge, resonance bridge. In those cases where the room temperature is not very constant, a good accuracy is attained in the following way. In a series of preliminary measurements both the D.C. resistance R_0 and the additional A.C. resistance R_a were determined for a number of temperatures, and R_a was plotted against R_0. The two sets of values for R_0 and R_a obtained in the two resonance measurements must be in accordance with this curve, thus furnishing a first check. Then the value of R_0 obtained in the Maxwell-bridge measurement is used to determine from the curve the exact value of R_a at the moment of this measurement.

When considering the conditions of balance of both bridge networks it is easily seen that in both cases the condenser losses cause corrections of opposite sign in the measured resistance. Consequently an error in the loss measurement of the condenser is felt twice. As it was thought reasonable to require that not more than half of the total inaccuracy be caused by the inaccuracy of the measurement of the condenser losses, the permissible error in this measurement corresponds to one quarter of the required inaccuracy of the final result.

As a matter of fact, the values given in table I are well attainable if corrections for the temperature coefficient of the condenser losses are applied. This temperature coefficient has been determined separately and amounted to 2% per °C.

The value of R_4 in the Maxwell bridge (see fig. 2) was always chosen such that R_3 would not exceed 1 Ω. Since the earth capacitance of the point 4 in fig. 10 certainly will be under 1 pF, the error in the resistance caused by this capacitance remains below 2×10^{-4} Ω.

For the resonance measurement at 8 kc/sec the earth capacitance of the dangerous point (between coil and condenser) was measured and found to amount to 0·02 pF, causing an error of $1{\cdot}2 \times 10^{-4}$ Ω.

Another source of error lies in the fact that the additional resistance of the coil depends upon frequency. For the coils used at 8 and 17·5 kc/sec the losses are mainly caused by the self-capacitance of the coil, and consequently the additional resistance is roughly proportional to the cube of frequency. This implies that at 8 kc/sec the frequency must be constant within one part in 10^4, and at 17·5 kc/sec within 3 parts in 10^4. This requirement can very well be satisfied. At 100 kc/sec the situation is still easier,

since at this frequency the losses are mainly due to eddy currents and consequently the additional resistance is roughly proportional to the square of the frequency.

In conclusion, it may be stated that an inaccuracy in time constant of 10^{-11} sec as mentioned in the beginning of this section, may well be attained, provided the measurements are carried out carefully with all necessary precautions.

5. *Results*

As was mentioned in the foregoing section, the measurements were carried out with resistors of 1000 Ω and 100 Ω. These resistors belong to a series of standard resistors of 10000, 1000, 100 and 10 Ω. The differences varied less than 10^{-12} sec between 1 and 8 kc/sec, while in some cases this variation amounted to 7×10^{-12} sec between 8 and 100 kc/sec.

As stray inductances and stray capacitances have entirely different influences for different values of the resistance and, moreover, the construction shows great differences for different values, it may be considered very probable that the absolute values of T for all resistance values do not vary more than 10^{-11} sec for the whole range between 1 and 100 kc/sec. In order to obtain a further confirmation of this, a resistor was made consisting of a single bifilar loop of manganin wire 20 cm long and 0·04 mm thick. The resistance value amounted to 100 Ω, and since there was no insulating or magnetic material in the vicinity, the time constant may be considered independent of frequency. The time constant of this resistor was compared with that of a 1000-Ω standard resistor at 8 and 100 kc/sec.

The difference between the two time constants varied less than 5×10^{-12} sec for these frequencies. This result was in good agreement with the outcome of the absolute determination of the time constant of three of the standard resistors, the time constants of which had been made equal at the beginning of the experiment. This absolute determination, the results of which are to be found in table II, was carried out as a test of the new method.

The figures in parantheses were obtained with the coil of 9 mH giving

TABLE II

f(kc/sec)	8	17·5	100
$T(10^{-11}\text{sec})$	12·2 (12·2) (13·1)	12·3 11·6	11·5 11·5

at 8 kc/sec rather a low accuracy, as an error of 0·5 mΩ corresponds to an error in the time constant of 10^{-11} sec.

It may be seen that the deviations between the various figures are within the limits previously set, while there is no reason to regard as real the slight dependence on frequency.

It is difficult to tell at which of the three frequencies the most reliable results were obtained. At 100 kc/sec the reproducibility was excellent, deviations not exceeding 1 mΩ, corresponding to a time constant of $2{\cdot}5 \times 10^{-12}$ sec. On the other hand, corrections due to eddy currents, *etc.*, are greatest for this frequency, and consequently a small systematic error might be present. At 17·5 kc/sec an irreproducibility of 1 mΩ sometimes occurred, corresponding to a time constant of 5×10^{-12} sec, but there are no corrections.

At 8 kc/sec the relative inaccuracy must be very small, causing some uncertainty as a consequence of the temperature coefficient of the additional resistance. Nevertheless, the results of a great number of preparatory measurements were very constant, deviations of time constant not exceeding 10^{-11} sec.

It was concluded that all results can best be regarded as equivalent (the figures in parantheses of table II excepted), and the mean value taken as final result, giving

$$T = 11{\cdot}8 \times 10^{-11} \text{ sec.}$$

It is thus possible to measure the time constants of standard resistors of 100 Ω and 1000 Ω, and, indirectly, by using the standard condensers, those of standard resistors of 10 Ω and 10000 Ω. If it is thought advisable, it is also possible to adjust the time constants to any previously chosen value, for instance zero. The time constants of the whole set of standard resistors having been fixed, it is possible to use these results for the measurement of the time constant of any resistance value at any frequency between 1 and 100 kc/sec. The accuracy will then depend upon the resistance value and the frequency. These measurements may be carried out at other frequencies than those used for the absolute measurement, thanks to the fact that the time constants of the standard resistors depend so extremely little on frequency.

Eindhoven, March 1947

REFERENCES

[1]) C. G. Koops, Philips Tech. Rev. **5**, 300-308, 1940, The measurement of very small phase displacements.

[2]) J. W. L. Köhler and C. G. Koops, Philips Res. Rep. **1**, 419-446, 1946, A new alternating-current bridge for precision measurements.

Philips Res. Rep. 2, 468-474, 1947

ON THE CALCULATION OF IMPULSE-NOISE TRANSIENTS IN FREQUENCY-MODULATION RECEIVERS

by F. L. H. M. STUMPERS

621.3.015.33:621.396.62:621.396.822

Summary

The effect of impulse-noise transients is calculated by means of a series expansion of the phase, the general term of which contains $[A(t)]^n$ when the amplitude $A(t)$ of the disturbance is smaller than the amplitude of the signal, and $[A(t)]^{-n}$ in the inverse case. The Laplace transform is used to calculate the effect in the filters. The large effect of phase-opposition during the capture time is accounted for.

1. *Introduction*

Impulse noise is characterized by disturbances of very short duration, very high in amplitude and relatively widely separated. Automobile ignition noise belongs to this type [1]). In this paper we shall calculate the effect of an idealized impulse disturbance on the output of a frequency-modulation receiver tuned to a frequency-modulated transmitter. This is a slight extension of calculations given in my dissertation [2]) in which the same methods were applied.

When a frequency spectrum of these disturbances is made all components are found to be in phase at one moment. The situation is thus different from the disturbance by random noise — as caused, for instance, by shot effect — where the components have arbitrary and independent phases. We shall idealize an impulse disturbance by a δ-function (in operational calculus $\delta(x) = 0$, if $x \neq 0$; $\delta(0) = \infty$; $\int_{-\infty}^{\infty} \delta(x)\, dx = 1$).

At the moment the impulse occurs we receive a frequency-modulated signal of instantaneous frequency $\omega_0 + \Delta\omega \cos(qt + \psi)$. If we introduce the function $\mu(t) = m \sin(qt + \psi) + \varphi$, where $m = \Delta\omega/q$, then the input signal of our receiver is

$$\cos\{\omega_0 t + \mu(t)\} + a\,\delta(t), \qquad (1)$$

where a is an arbitrary constant defining the relative energy level.

For our calculations it is unadvisable to deform the amplitude characteristic of the filters so as to make it rectangular and at the same time to leave out of account the phase characteristic. The choice of the amplitude characteristic largely defines the phase curve, which is so important for the

computation of a transient. For this reason we have chosen as intermediate-frequency filter a single resonant circuit with the parameter $r/2L = a$ and an RC-circuit with the parameter $RC = \beta$ as audio-frequency filter (pre-emphasis included).

In this way the calculation is not too involved and the important points come out more clearly. The method can easily be applied to other filters.

We have to split the problem into two parts, dependent on the ratio of the instantaneous amplitudes of signal and disturbance. When the amplitude of the disturbance at the entrance of the ideal frequency detector exceeds the amplitude of the signal, the disturbance controls the instantaneous frequency during a short time, whereas for smaller disturbances the influence is much more restricted. If the filter a is placed in the anode circuit of a valve to which the voltage (1) is applied, the voltage on the filter will be proportional to

$$v(t) = \cos \{\omega_0 t + \mu(t)\} + 2\, aae^{-at} \cos \omega_0 t\; U(t), \tag{2}$$

$$U(t) = 0, \text{ if } t < 0;\; U(0) = \tfrac{1}{2};\; U(t) = 1, \text{ if } t > 0.$$

This is easily verified by the methods of operational calculus. It is assumed that a is small with respect to ω_0, but large enough not to introduce any appreciable distortion in the frequency-modulated signal. The voltage $v(t)$ is applied to the frequency detector. Its output is proportional to the derivative of the phase of $v(t)$ with respect to time. $2aa$ is the maximum amplitude of the impulse transient when it is has passed the intermediate-frequency filter (or at the entrance of the frequency detector), and the important point is whether or not this amplitude exceeds the signal amplitude.

2. *Small impulse disturbances* $(2aa < 1)$

For $t > 0$ the phase of $v(t)$ is given by

$$\omega_0 t + \mu(t) - \arctan \left\{ \frac{2aae^{-at} \sin \mu(t)}{1 + 2aae^{-at} \cos \mu(t)} \right\}, \tag{3}$$

in which the arc tan is taken between $-\pi/2$ and $\pi/2$. The discontinuity of the phase at $t = 0$ is represented by

$$-\left[\arctan \left\{ \frac{2aa \sin \mu(0)}{1 + 2aa \cos \mu(0)} \right\}\right] U(t). \tag{4}$$

This discontinuity in the phase gives a discontinuity in the frequency (replace $U(t)$ in (4) by $\delta(t)$). The effect of this discontinuity in the output of the audio filter is

$$-\left[\arctan \left\{ \frac{2aa \sin \mu(0)}{1 + 2aa \cos \mu(0)} \right\}\right] \beta e^{-\beta t}\, U(t). \tag{5}$$

For $t > 0$ we shall use the series development of (3),

$$\omega_0 t + \mu(t) + \sum_1^\infty{}^n \frac{(-1)^n}{n} (2a\alpha e^{-\alpha t})^n \sin n\mu(t) . \tag{6}$$

(The derivation of this formula was given in [3]).

The instantaneous frequency corresponding to (6) is given by

$$\omega_0 + \Delta\omega \cos(pt + \psi) + \sum_1^\infty{}^n (-2a\alpha)^n \Big\} - \alpha e^{-n\alpha t} \sin n\mu(t) + \Delta\omega \cos(qt + \psi) \cos n\mu(t) . \tag{7}$$

The first two terms give the undisturbed signal, which we shall omit from now on.

The Laplace transformation of the disturbance term is

$$\mathrm{Im} \sum_{-\infty}^{+\infty}{}^s \sum_1^\infty{}^n \frac{(-1)^n}{n} (2a\alpha)^n \frac{p}{(p + n\alpha - isq)} J_s(nm)\, e^{i\varphi + is\psi},$$

where Im stands for "the imaginary part of"; $J_s(nm)$ is the Bessel function of order s and argument nm.

To obtain the result after the low-frequency filter, this Laplace transform has to be multiplied by $\beta/(p + \beta)$, and, using the reciprocal transform, we get for the output voltage:

$$\mathrm{Im} \Big\{ \Sigma\Sigma (-1)^n \frac{(2a\alpha)^n}{n} \frac{\beta(-n\alpha + isq)}{(\beta - n\alpha + isq)} J_s(nm) \left(e^{-n\alpha t + isqt} - e^{-\beta t}\right) . e^{is\psi + i\varphi} \Big\} .$$

Now if $\alpha - \beta$ is so large that a filter of the bandwidth $\alpha - \beta$ does not introduce an appreciable distortion in the frequency-modulated signal, it is permitted to use the quasi-stationary approximation

$$\mathrm{Im} \Big\{ \Sigma (-1)^n \beta (2a\alpha)^n \frac{(-\alpha + i\Delta\omega \cos qt)}{\beta - n\alpha + in\,\Delta\omega \cos qt} e^{-n\alpha t + imn \sin(qt+\psi) + i\varphi} + \\ + \Sigma (-1)^{n+1} \beta (2a\alpha)^n \frac{(-\alpha + i\Delta\omega \cos \psi)}{\beta - n\alpha + in\,\Delta\omega \cos \psi} e^{-\beta t + imn \sin \psi + i\varphi} \Big\} . \tag{8}$$

As β is small with respect to α, we take the first approximation of the first term, and, because the first approximation of the second term is cancelled by (5), also the second approximation of the second term. We thus get the total effect of (8) and (5):

$$-\beta \arctan \frac{2a\alpha e^{-\alpha t} \sin \mu(t)}{1 + 2a\alpha e^{-\alpha t} \cos \mu(t)} + \\ + f(2a\alpha) \frac{\beta^2 e^{-\beta t}}{\alpha^2 + \Delta\omega^2 \cos^2 \psi} \Big\{ \alpha \sin \mu(0) + \Delta\omega \cos \psi \cos \mu(0) \Big\} , \tag{9}$$

where $f(x) = \sum_1^\infty{}^n (-1)^{n+1} x^n n^{-2}$. This function is nearly linear for $0 \leqq x \leqq 1$:

$f(0{\cdot}1) = 0{\cdot}09760$; $f(0{\cdot}2) = 0{\cdot}19080$; $f(0{\cdot}5) = 0{\cdot}44841$; $f(1) = 0{\cdot}8225$.

For amplitude modulation the disturbance term in the detected signal would be equal to

$$2aae^{-at} \cos \varphi \; U(t) . \tag{10}$$

Non-linear effects, to be expected for large values of $2a\alpha$, are left out of account. After the low-pass filter this would give

$$\frac{2aa\beta}{\alpha-\beta} (e^{-\beta t}-e^{-\alpha t}) \cos \varphi \; U(t) . \tag{11}$$

The energies of the disturbances are measured by the square-means of the expressions (9) and (11), which for $\Delta\omega = 0$ and $2a\alpha$ small are

$$F_1 = \frac{(a\beta)^2}{\alpha+\beta} (\alpha-\beta)^2 \tag{11a}$$

and

$$E_2 = \frac{a^2\alpha\beta}{\alpha+\beta}, \tag{11b}$$

respectively.

These energies should be compared with the maximum modulation energies $a^2/2$ and $\frac{1}{2}$ respectively. In the case of a sweep of 15 kc/s and a pre-emphasis time of 75.10^{-6} sec, the ratio of α to β will be 7, and the gain for frequency modulation is 9·8 db. This is compared to an amplitude-modulated signal of 15 kc/s bandwidth with the same pre-emphasis. If we go to 75 kc/s sweep, an extra 5·5 db is obtained.

Since the ear is less sensitive at high frequencies the short-duration term with $e^{-\alpha t}$ will have less influence, and the result will be still more favourable for frequency modulation.

3. *Large impulse disturbances* $(2a\alpha > 1)$.

Now it is useful to introduce the function $\mu_0(t)$, defined by the relations

$$\mu_0(t) = \mu(t) \pmod{2\pi}, \; -\pi \leqq \mu_0(t) < \pi .$$

Define t_0 by $e^{\alpha t_0} = 2a\alpha$; then, during the time interval $0 < t < t_0$, the phase can be expanded into the series

$$\omega_0 t + \sum_1^\infty \frac{(-1)^{n+1}}{n} \left(\frac{e^{\alpha t}}{2a\alpha}\right)^n \sin n \, \mu(t) ; \tag{12}$$

for $t_0 < t$ we have to use the series

$$\omega_0 t + \mu(t) + \sum_1^\infty \frac{(-1)^n}{n} (2aae^{-\alpha t})^n \sin n \, \mu(t) . \tag{13}$$

For $t = 0$ the discontinuity in the phase amounts to

$$-\mu_0(0) + \arctan\left\{\frac{\frac{1}{2a\alpha}\sin\mu(0)}{1+\frac{1}{2a\alpha}\cos\mu(0)}\right\}. \tag{14}$$

The phase is continuous at $t = t_0$, as from either side the limit is $\omega_0 t + \frac{1}{2}\mu(t)$. The instantaneous frequency is easily derived from (12) and (13) by one differentiation. With the help of the Laplace transform we can derive the transients after the low-pass filter in nearly the same way as in section 2. The results can be divided into four main parts.

First we have the terms for which α is the time constant:

$$\beta \arctan\left\{\frac{(2a\alpha)^{-1}e^{\alpha t}\sin\mu(t)}{1+(2a\alpha)^{-1}e^{\alpha t}\cos\mu(t)}\right\}\left\{U(t)-U(t-t_0)\right\} -$$
$$-\beta\arctan\left\{\frac{2a\alpha e^{-\alpha t}\sin\mu(t)}{1+2a\alpha e^{-\alpha t}\cos\mu(t)}\right\}U(t-t_0). \tag{15a}$$

These transients are quite analogous to those we have met for small values of $2a\alpha$. In the same way there are transients with the time constant β, which are the continuations of those in formula (8):

$$\beta^2\left\{\Sigma\frac{(-1)^{n+1}}{n^2}\right\}\frac{2\Delta\omega\cos(qt_0+\psi)}{\alpha^2+\Delta\omega^2\cos^2(qt_0+\psi)}\cos\mu(t_0)\,e^{-\beta(t-t_0)}\,U(t-t_0) +$$
$$+\beta^2\left\{\Sigma\frac{(-1)^{n+1}}{n^2(2a\alpha)^n}\right\}\frac{\alpha\sin\mu(0)-\Delta\omega\cos\psi\cos\mu(0)}{\alpha^2+\Delta\omega^2\cos^2\psi}e^{-\beta t}\,U(t). \tag{15b}$$

Then we have the transients occurring in the low-pass filter because the current corresponding to $\Delta\omega\cos(pt+\psi)$ is interrupted during the time interval from 0 tot t_0. These transients are

$$\left\{\frac{\beta^2}{\beta^2+q^2}\Delta\omega\cos\psi+\frac{\beta q}{\beta^2+q^2}\Delta\omega\sin\psi\right\}e^{-\beta t}\left\{U(t)-U(t-t_0)\right\} +$$
$$+\left[\frac{\beta^2}{\beta^2+q^2}\Delta\omega\left\{\cos\psi-e^{\beta t_0}\cos(qt_0+\psi)\right\}+\frac{\beta q}{\beta^2+q^2}\Delta\omega\left\{\sin\psi-e^{\beta t_0}\sin(qt_0+\psi)\right\}\right]\cdot$$
$$\cdot e^{-\beta t}\,U(t-t_0)\approx-\beta\Delta\omega t_0\cos\psi\,e^{-\beta t}\,U(t-t_0). \tag{15c}$$

And finally we have the terms:

$$-\mu_0(0)\,\beta e^{-\beta t}\,U(t)+\mu_0(t_0)\,\beta e^{-\beta(t-t_0)}\,U(t-t_0). \tag{15d}$$

If qt_0 is small with regard to 2π, as will be the case for all but exceptionally high peaks, and $\mu(t)$ does not pass an odd multiple of π in the time interval $(0, t_0)$, this term is approximated by:

$$\{\Delta\omega t_0 \cos\psi + \beta t_0\,\mu(0)\}\,\beta e^{-\beta t}\,U(t-t_0)\,.$$

This contribution is nearly cancelled by (15c). However, if there is an odd multiple of π between $\mu(0)$ and $\mu(t_0)$, this will give rise to an extra

$$\pm\, 2\pi\beta e^{-\beta(t-t_0)}\,U(t-t_0)\,. \tag{16}$$

If the same odd multiple of π is passed twice in opposite directions then there is no effect.

Of the terms (15) only (15a) and (15b) give a contribution when the modulation is absent and the receiver is tuned exactly. These expressions indicate a diminution of the audio disturbance for high peaks, but no large effect is present when $2a\alpha$ passes 1. Of the terms (15c) and (15d) the latter is the more important, because it gives the chance of a large disturbance (16). Here the large effects occur when signal and impulse transient pass through phase-opposition in the time interval $0 \leq t \leq t_0$ that the impulse transient captures the frequency detector. This was discussed graphically by Smith and Bradley [4]) and results directly from our formulae.

We shall now make an estimate of the energy. We assume again that $qt_0 = q/\alpha \ln 2a\alpha$ is small compared with 2π. During one low-frequency period the chance of passing an odd multiple of π is $2\Delta\omega/\pi q$. To each such multiple of π an angle qt_0 is attached in which the time $t = 0$ (the beginning of the impulse) must fall, if $t = t_0$ has to come on the other side of π. So the chance of the impulse coming in time for a large response is $\Delta\omega t_0/\pi^2$. The average energy corresponding to (16) is therefore

$$\frac{\Delta\omega t_0}{\pi^2}\cdot\frac{4\pi^2\beta^2}{2\beta} = 2\beta\,\Delta\omega\,t_0 = \frac{2\beta\Delta\omega}{\alpha}\ln(2a\alpha)\,. \tag{17}$$

If at full modulation the amplitude ratio after the i.f. filter $2a\alpha = 2$, then the contribution of this term (17) is already larger than that of all the others. It is the reason for the appearance of the impulse-noise threshold. If no modulation is present the same effect can be simulated by mistuning the receiver. If the transmitter has the frequency $\omega_0 \pm \omega_1$ the phase difference between 0 and t_0 is $\omega_1 t_0$. The chance that a peak will occur during this time is $\omega_1 t_0/2\pi$, and the corresponding energy is $(\pi\beta\omega_1/\alpha)\ln(2a\alpha)$. In this way the impulse-noise threshold was shown by Guy and Morris [5]).

As the larger effects occur only when $2a\alpha > 1$, in an area where much impulse noise is prevalent, a low deviation ratio has to be preferred.

The method of calculation can be applied to other filters. First one has to calculate the transient voltage at the entrance of the frequency detector. Then, if the amplitude of the impulse transient, $A(t)$, is smaller than the amplitude of the signal, one makes a series expansion of the phase with

reference to $\{A(t)\}^n$ and calculates the effect of each term in the low-pass filter with the Laplace transform. If the amplitude of the impulse transient exceeds the amplitude of the signal from t_0 to t_1, then expand the phase into a series with $\{A(t)\}^{-n}$ during this interval and continue as before.

Eindhoven, August 1947

REFERENCES

1) C. C. Eaglesfield, Motor-car ignition interference, Wireless Engr **23**, 265-272, 1946.
2) F. L. H. M. Stumpers, Eenige onderzoekingen over trillingen met frequentie-modulatie (diss. Delft, Netherlands, May 1946)
3) F. L. H. M. Stumpers, Interference problems in frequency modulation. Philips Res. Rep. **2**, 136-160, 1947
4) D. B. Smith and W. E. Bradley, The theory of impulse noise in ideal frequency-modulation receivers, Proc. Inst. Radio Engrs, W & E **34**, 743-751, 1946
5) R. F. Guy and R. M. Morris, NBC frequency modulation field test, RCA Rev. **5**, 190-225, 1940

ABSTRACTS

(Continued from page 453)

1723: P. Cornelius: Eén eenhedenstelsel in de electriciteitsleer (Faraday **16**, 57-67, 1947). (One system of units in electromagnetic theory).

A survey is given of the basic formulae of electromagnetism in rationalized Giorgi units (M.K.S. units). The didactic value of using these units is stressed.

1724: J. L. Snoek: Zeitabhängige Erscheinungen in Eisen enthaltenden Stoffen unter dem Einfluss mechanischer und magnetischer Kräfte (Schweizer Arch. angew. Wiss. Tech. **13**, 9-14, 1947).

For the contents of this paper, see J. L. Snoek, New developments in ferromagnetic materials, Amsterdam 1947 (see abstract **1729**).

1725: J. H. van Santen and G. H. Jonker: Effect of temperature on the permittivity of barium titanate (Nature, London, **159**, 333, 1947).

For crystals of barium titanate ($BaTiO_3$) and related compounds, such as $SrTiO_3$, $(Ba,Sr)TiO_3$, in the cubic region above a certain transition temperature ϑ, the permittivity has a sharp maximum. It decreases monotonically with increasing temperature. This decrease is explained from the Clausius-Mosotti formula. When the thermal expansion is taken into account, and the temperature dependence of the polarisability ignored, as proves to be legitimate in the case of barium titanate and related compounds and also for TiO_2 (tetragonal), it follows for larges values of ε that

$$1/\varepsilon = \beta(T-C)\,,$$

where β is the coefficient of thermal expansion. The absence of permanent dipoles in the cubic region ($T > \vartheta$) is suggested, whereas in the tetragonal region ($T < \vartheta$) the assumption of dipoles seems plausible.

1726: M. Gevers: The relation between the power factor and the temperature coefficient of the dielectric constant of solid dielectrics (Dissertation, Delft, 1947).

The relation between the power factor ($\tan \delta$) and the temperature coefficient of the dielectric constant of solid amorphous dielectrics is treated, both theoretically and experimentally. The methods of measurement are described, and various cases are discussed. Details on the properties of mixtures of dielectrics. (Also published in Philips Res. Rep. **1**, 197, 279, 361, 447, 1946; reports *R 15*, *20*, *25*, *30*).

1727: J. F. H. Custers: On the relation between deformation and recrystallization texture of nickel-iron with cubic orientation (Physica, 's-Grav. **13**, 97-116, 1947).

Polycrystalline nickel-iron ($\approx$ 50 weight % Ni) that has been severely cold rolled, exhibits on recrystallization at about 1000 °C a so-called cubic orientation. Aluminium does not show this tecture upon recrystallization. To trace any difference in slip mechanism between Ni-Fe and Al, the author investigates the deformation and recrystallization textures of polycrystalline Ni-Fe with cubic orientation. The findings are compared with observations of Burgers and Louwerse on single crystals of Al. On the whole, the deformation textures are found to be the same. Yet there are marked differences owing to the Ni-Fe specimen not being a monocrystal. The recrystallization textures are strongly different. The reason is to be found in the difference between the deformation textures, as can be explained by Burgers's theory. Barrett's criticism of this theory is discussed.

1728: J. M. Stevels: The effective permittivity of compressed and sintered samples of TiO_2 (Rec. Trav. chim. Pays-Bas **66**, 71-74, 1947).

The effective permeability of compressed and sintered samples of TiO_2 is measured, and the results are discussed in terms of the theory developed by Polder and Van Santen (abstract **1698**). The experimental curves show that in loose powders the holes are more or less disc-shaped. The more the samples are sintered the more nearly spherical the holes become. As a whole, the results confirm the theory.

SUBJECT INDEX

AUTHOR INDEX

INDEX ACCORDING TO THE UNIVERSAL DECIMAL CLASSIFICATION SYSTEM