VOL. 2 No. 1 JANUARY 1937

Philips Technical Review

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

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

COLOUR REPRODUCTION IN THE USE OF DIFFERENT SOURCES OF "WHITE" LIGHT

by P. J. BOUMA.

Summary. The requirements which have to be satisfied as regards colour reproduction differ considerably for different practical applications. An investigation is made as to what conditions the spectral distribution and the point in the colour triangle have to meet:

a) In order to obtain a "pleasing and agreeable" reproduction of colour (i.e. suitable for indoor illumination); and

b) To obtain a "correct and natural" colour reproduction (i.e. coinciding with that obtained in daylight).

A method is developed with the aid of Ostwald colour cards for the comparison and estimation of colour reproduction of different light sources. This is followed by a discussion of a series of examples of light sources which simulate daylight.

In a previous article [1]) it has already been shown that the colours which we perceive in surrounding objects are determined to a marked degree by the nature of the light which is incident on these objects. It thus follows that certain types of light are quite unsuitable for some purposes of illumination. Thus sodium lighting cannot be used in living rooms, where a differentiation between colours must be possible and the colours of different objects must appear pleasing and possess more or less their natural hues, requirements which sodium lighting is by no means able to meet. A further example is mercury light which cannot be indiscriminately used in stores where coloured materials are sold, for these colours not only appear different under mercury lighting than in daylight, and a purchase may easily lead to keen disappointment when seen later in daylight, but it may even occur that under mercury lighting two materials will produce exactly the same colour sensation, while appearing entirely different in daylight [2]).

These examples indicate that the problem of providing a satisfactory source of "white" light is rendered increasingly difficult owing to the variety of requirements which have to be met in individual cases. It is obvious that a single source of light cannot be evolved which will give a satisfactory illumination in all circumstances.

For use in living rooms the principal requirement is that the light must be pleasing and not irritating, while in stores and wherever colours have to be differentiated, judged or compared, the reproduction of colours must be as close as possible to their daylight hues. It will be found below that these requirements are in general incompatible.

We shall at the outset discuss the first requirement and consider the question:

What types of light give a pleasing reproduction of colour?

We are dealing here with an essentially practical problem which cannot be disposed with by mere theoretical analysis. To obtain an appropriate answer to our question we would have to exhibit a wide variety of light sources to a large number of observers, asking them to select that giving the most pleasing and agreeable illumination. But it is doubtful whether an investigation on even these comprehensive lines could lead to a satisfactory solution. Selection of the most pleasing light source would be governed, *inter alia*, by the following factors:

1) Philips techn. Rev., 1, 283, 1936.

2) This state of affairs is reached already when e.g. the reflecting powers of the two substances in the red band differ considerably and are the same for other spectral colours.

1. The question of habit: We have become so accustomed to seeing by daylight (with high intensities) during the day and by electric light (with lower intensities) during the evening, that we are inclined to regard all sources of light markedly different to them as "unnatural".

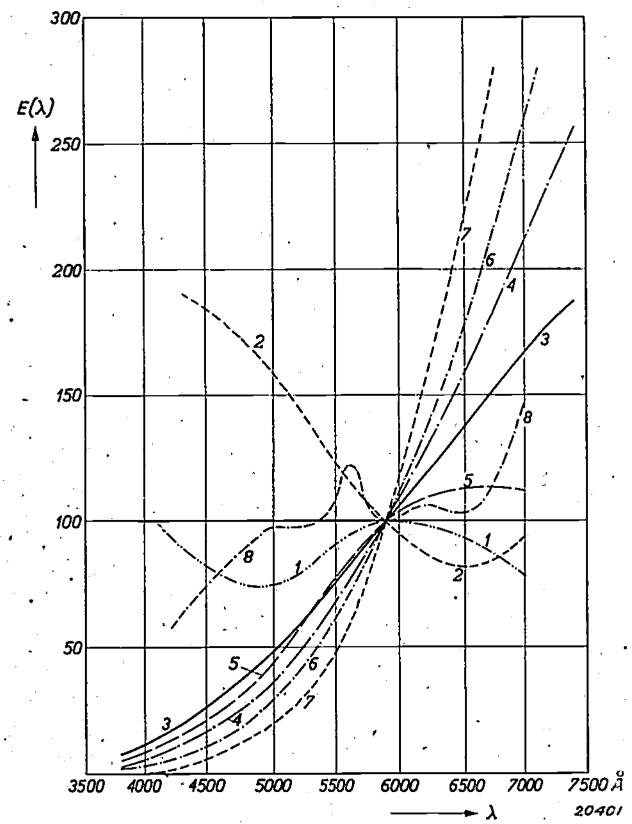


Fig. 1. Spectral distribution of various sources of "white" light. *1)* Daylight, overcast. *2)* The light from a clear blue sky. *3)* Gasfilled incandescent electric lamp. *4)* Vacuum incandescent electric lamp. *5)* Incandescent gas lamp. *6)* Carbon filament electric lamp. *7)* Paraffin lamp. *8)* Daylight lamp. (The intensity at 5900 Å is taken equal to 100).

2. The method of comparison: A source of light may be intrinsically quite pleasing and yet appear unsatisfactory if its suitability is tested at the same time as or immediately after a test on an entirely different type of light.
3. The question of intensity: It is evident that preference will be given to a different type of light at higher intensities than at lower intensities.

As a first step towards the solution of this problem, we shall investigate those sources of light which have hitherto been found suitable and satisfactory in practice.

For a series of light sources, the spectral distribution (the energy E at $\lambda = 5900$ Å being put equal to 100 for all light sources) is shown in *fig. 1*, while in the colour triangle in *fig. 2* the points are marked corresponding to the light emanating from the light source [3]). We have represented the different light sources in both ways here, because the spectral distribution alone does not give us directly an idea of the nature of the colour sensation produced by the light source itself, while on the other hand the colour triangle gives little information how each particular type of light reproduces the surrounding colours. In this respect, there is, for example, little similarity between daylight and a mixture of two complementary spectral colours (yellow and blue). If the mixture has been made in the correct proportions, then the light emanating directly from the two light sources will give the same colour sensation and is therefore represented by the same point in the colour triangle. But the reproduction of colour is entirely different: the second light source makes all objects appear white, grey, blue or yellow (the last-named colours with different degrees of saturation).

The following types of light are represented in the figures:

1) Daylight (results of measurements made at Utrecht with an overcast sky). The composition of the light varies fairly considerably; in fig. 2 a few results are given, while in fig. 1 only one spectral curve is given as an example.

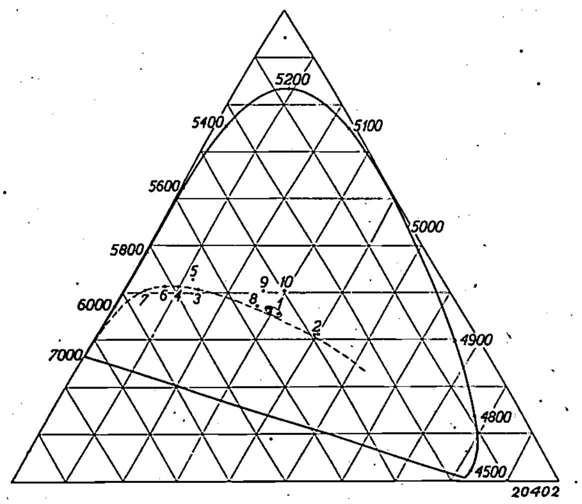


Fig. 2. Colour triangle with spectral colours (——), black bodies (----), the light sources in fig. 1 (Nos. *1* to *8*), high-pressure mercury lamp *(9)* and super-high-pressure mercury lamp *(10)*.

2) The light from a clear blue sky (also from measurements at Utrecht). This type of light is

3) Regarding the representation of colours in a triangle see Philips techn. Rev. 1, 283, 1936. In addition to the various sources of "white" light the spectral colours (——) and the black bodies of various temperatures (----) are also included in the diagram.

obtained in the main when on sunny days we are at any point in the shade. These lights *(1)* and *(2)* at the brightnesses occurring during the day are generally found to be satisfactory.

3) The light from a gasfilled incandescent electric lamp.

4) The light from a vacuum incandescent lamp.

5) Incandescent gas lamp.
Both as regards spectral distribution and the point in the colour triangle, these three light sources differ very considerably from daylight. Yet on the whole they are regarded as satisfactory, and we find them pleasing. The divergence in colour reproduction (compared to that obtained with daylight) only tends to become disturbing when an attempt is made to differentiate colours with these lights. In other respects we are hardly aware of how very different objects appear when seen in these lights than when viewed in daylight.

6) The oldest type of electric lamp: the carbon filament lamp.

7) The paraffin lamp.
These two types of light differ still more from daylight, and have a fairly saturated orange-yellow colour. The paraffin lamp is also accepted as giving an agreeable light, although the divergence in colour reproduction is so pronounced that we are quite well aware of this divergence in everyday life.

8) A socalled daylight lamp, i.e. a gasfilled incandescent electric lamp with a special blue bulb which serves to adapt the light to daylight.

The following general remarks may be added to supplement the above:

In the colour triangle all points for the types of light under consideration lie either on or close to the line for the black body (-----). All have spectral distributions without sharp maxima or minima; daylight *(1)* and *(2)* give very high colour temperatures (4500 to 8000 °K), while those sources used for interior illumination have much lower colour temperatures (1750 to 2850 °K).

Why do we prefer for interior illumination such sources of light which have a much lower colour temperature, i.e. are more yellowish than sunlight?

It might be assumed that we have chosen them just because they are the only sources of light which have been evolved on a commercial scale and that we have gradually grown accustomed to them. But this would not be correct. We can quite readily modify the colour of our light with lamp shades and coloured transparent materials, and the striking observation is then made that the great majority of these materials are y e l l o w, i.e. by shading our electric lamps we are still further reducing the colour temperature. To obtain a pleasing appearance we depart still further from daylight, electric lamps *per se* already marking a divergence therefrom. The conclusion to be drawn is therefore that during the day we are very well satisfied with daylight but . . . at the small intensities which we have in living rooms during the evening, we show a preference for a more yellowish light, i.e. light with the lower colour temperature. Light of equivalent spectral composition to daylight appears cold and lacks cheerfulness at low intensities.

These assertions are confirmed by a wealth of practical experience. During eclipses of the sun our surroundings appear pratically the same as in ordinary daylight, but are illuminated with a much lower intensity. The sensation obtained is by no means cheerful since the colour temperature is too high for this intensity. For the self-same reason daylight lamps used in a living room do not as a rule give a pleasing effect. If on dull days the daylight entering our rooms appears insufficient, supplementing it with electric light produces a depressing result: The combination gives us a spectral distribution to which we are not accustomed under these circumstances; we therefore don't like it.

An important factor in judging the suitability of a particular source of light for indoor illumination is whether colour distortion may suggest to us disagreeable things, for instance whether the face looks unhealthy and food assumes unnatural colours, etc. One of the factors which lead us to lower colour temperatures is certainly that with such light sources the skin appears to have a fresher and healthier colour.

In general, we are inclined to make up for the lack of illumination provided by artificial light as compared with daylight by the use of "warmer" colours (i.e. usually those containing more red).

The most acceptable sources of light for indoor illumination are therefore those whose spectral distribution closely resembles that of a black body with a low colour temperature (2800 °K or less). Yet to provide a comprehensive answer to the question posed above extensive investigations are still necessary.

We now come to the second question:

---

[4]) The colour temperature of a light source is the temperature which a black body ought to have to give the same relative spectral distribution.

What light sources offer a "correct" colour reproduction?

In other words: What requirements must a light source satisfy in order to reproduce colours in the same way as in daylight?

As we have already seen above, a consideration of the points in the colour triangle here again cannot provide us with an answer. We must moreover also take the spectral distribution into consideration. Colour reproduction will only be absolutely identical with that in daylight, when the spectral composition is fully equivalent to that of daylight.

But just as the eye experiences difficulty in differentiating between small differences in brightness [5]), so also is it unable to distinguish very slight differences in colour. There is in fact a certain minimum difference — which for the sake of brevity we shall term a step — which is necessary before differentiation becomes possible on comparing two colours.

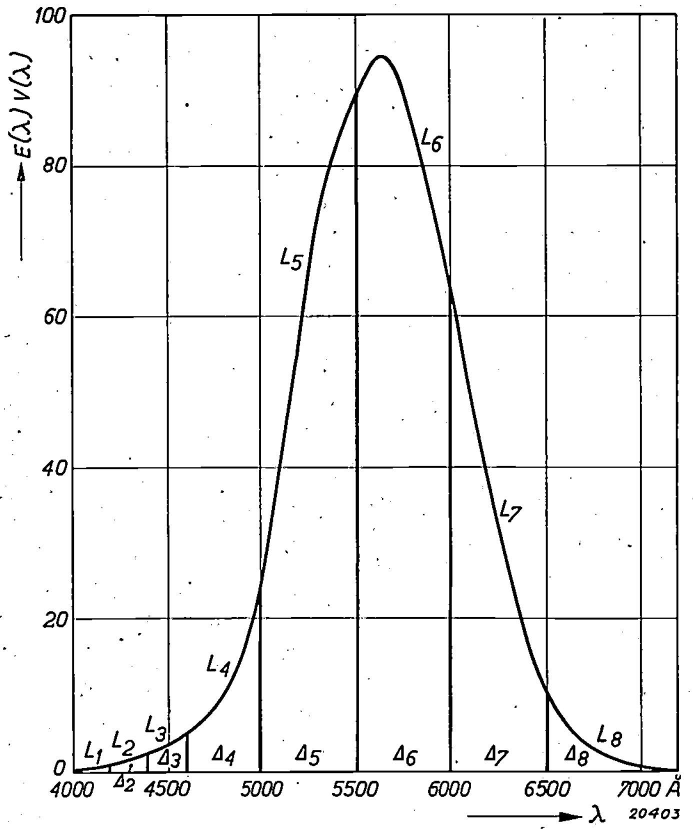


Fig. 3. Spectral flux distribution of daylight (energy distribution times ocular sensitivity curve) with subdivision into sections Δ_1, Δ_2, . . . each giving a total luminous flux of L_1, L_2, . . . respectively.

It thus follows that the relative spectral distribution of the lamp which is to simulate daylight need not be absolutely equivalent to that of daylight. Not only can the relative proportions of a specific wave length to the total intensity in the two spectra be slightly different, but the light with a particular wave length can also be replaced by that of a neighbouring wave length. Hence it is even possible to simulate daylight with a source of light which has a line spectrum, provided the lines (if necessary supplemented by a continuous band) sufficiently fill the whole of the visible region. It is found here also that mercury light (see also fig. 2; *9* applied to a high-pressure mercure lamp, and *10* to a super-high-pressure variety) even if suficient red light is mixed with it, is again not suitable for simulating daylight: The spectrum contains gaps of still too great an extent, particularly in the blue-green and the blue.

How can we arrive at a quantitative standard which the spectral distribution must satisfy? In *fig. 3* the spectral flux distribution of daylight is shown [6]). The wave-length band is divided into sections Δ_1, Δ_2, . . . , each section making a flux contribution of L_1, L_2, . . . (proportional to the area under the curve of each sectional component). When using the same subdivision for another source of light we get corresponding values L_1', L_2', . . . If these sections have been taken sufficiently small, the substitutional light source will simulate daylight with sufficient accuracy, when the components L_1', L_2' . . . differ by less than p_1, p_2, . . . per cent from the values L_1, L_2, ...

Investigations were made to determine to what extent the colour sensation due to an arbitrary pigment could fluctuate when illuminated by different light sources with components L_1, L_2, L_3 . . . It was found that the following subdivision of the wave-length band into eight sections was sufficient: 4000-4200-4400-4600-5000-5500-6000-6500-7000 Ångstrom units. It was desirable — as adopted here — to make a finer subdivision in the blue than in the rest of the band. The method of subdivision employed was found more satisfactory than a subdivision into ten equal components.

With this subdivision how large are the permissible tolerances p_1, p_2, p_3, . . . i.e. how far can the the flux components of the lamp and daylight differ in each of the selected divisions without the quality of colour reproduction being adversely affected? To investigate this question we shall assume that colour reproduction may be regarded as satisfactory when an arbitrary object, illuminated partly with daylight and partly from another source

[5]) See Philips techn. Rev. 1, 102, 1936.

[6]) This curve is obtained by multiplying the spectral energy distribution E (λ) (fig. 1, curve *1*) by the relative ocular sensitivity V (λ).

of light, exhibits a colour difference not exceeding 4 steps. On this basis, it can be calculated that the tolerances in the eight wave-length sub-bands chosen are roughly the same (confirming that the method of resolution was correct); they were all 20 to 25 per cent.

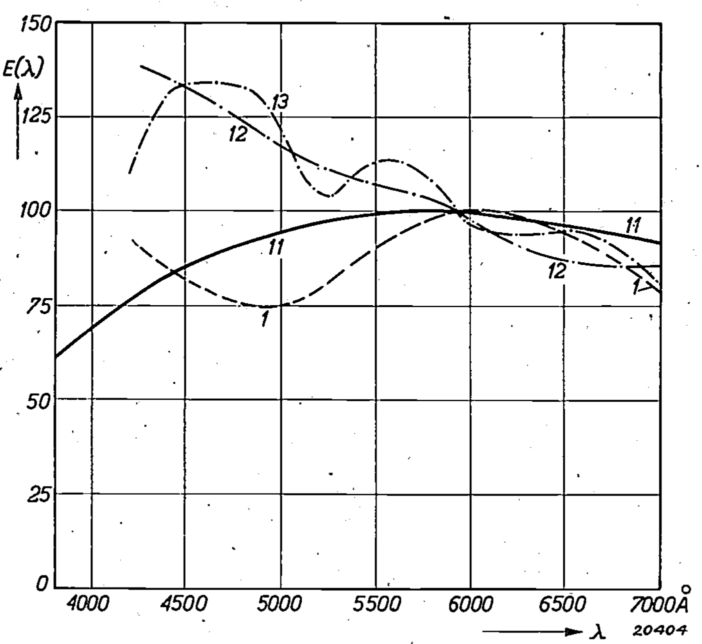


Fig. 4. Spectral distribution of various sources of "white light.
1) Daylight (overcast),
11) Black body (5000 °K.),
12) Mean of curves *1* and *2* in fig. 1,
13) Standard light source C (I.C.I., 1931).
(The energy at 5900 Å is taken as 100).

This conclusion will be discussed on the basis of a number of examples.

It is seen from fig. 1 that the energy distributions in the gas-filled incandescent lamp (*3*) and the vacuum lamp *(4)* differ by more than 25 per cent only in the blue. These two light sources thus differ slightly too much from each other to permit us to regard them as giving an identical colour reproduction. Also the difference between the daylight lamp *(8)* and daylight *(1)* is still too great.

Fig. 4 gives the curve *1* for the daylight measurements reproduced in fig. 1 and also the curve *11* for the radiation of a black body at 5000 °K. Within the tolerances stated above the two curves are in satisfactory agreement. In regard to colour reproduction it is thus permissible to replace the light obtained with a clouded sky by that afforded by a black body at 5000°K.

Curves *12* and *13* in fig. 4 also coincide within the limits laid down. These represent the mean of the curves for an overcast and a clear sky (curves *1* and *2* of fig. 1) and the standard light source (C) used as specified by the International Commission on Illumination (1931), in order to determine the appearance of colours under average daylight illumination.

Finally, it may be concluded from fig. 1 that the paraffin lamp *(7)*, the incandescent electric lamp *(3)* and daylight *(1)* differ far too much from each other to ensure equivalent colour reproduction. Colour reproduction with gaslight *(5)* is practically the same as with electric light *(3)* (the differences being rather large only in the red).

How can we compare the colour reproduction of two light sources?

Up to the present we have studied what conditions the spectral distribution of a light source must satisfy in order that its colour reproduction shall satisfactorily approximate to that of another light source. To ascertain under practical conditions how far the agreement between the two light sources extends, the best way lies in observing a large number of coloured objects under the illumination of each of the sources in question.

A series of objects most suitable for this purpose are the colour cards of the well-known Ostwald

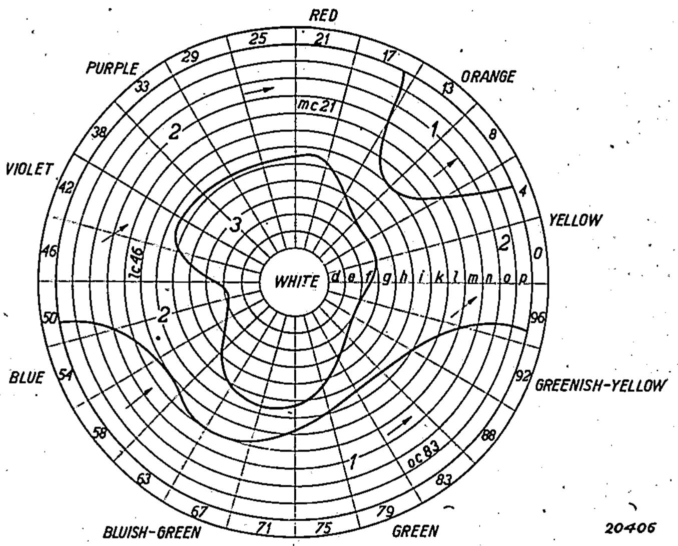


Fig. 5. Comparison of colour reproduction between a daylight lamp and daylight.
3) Region of unsatisfactory agreement,
2) Region of moderate agreement,
1) Region of good agreement, and
0) Region of full agreement.
The components represent the colour cards of the Ostwald colour atlas. The saturated colours lie at the outside and the less saturated colours at the middle.

colour atlas. This collection contains about 2000 coloured cards, each of which is denoted by two letters and a number (e.g. *oc* 83). The first letter denotes the saturation of the colour, the second the coefficient of total reflection, and the number the dominant wave-length [7]). Let us select from this series the groups *dc*, *ec*, *fc* . . . *pc* and of these the numbers *0*, *4*, *8*, . . . *96*, arranging the cards in concentric circles as shown in *fig. 5*. The complete notation is marked on several of the cards in the illustration. In broad outline this arrangement corresponds to that in the colour triangle; in the middle are all the very unsaturated colours (groups *dc* and *ec*), while towards the edges the colours become progressively more saturated (groups *oc* and *pc*).

If now each of the two halves of each card is illuminated with a different type of light, the degree of equivalence can be expressed numerically, viz.;

0: No difference;
1: Just perceptible difference;
2: Distinct difference;
3: Very marked difference (the sensation of an entirely different colour is created).

If the numbers ascribed to the different cards are plotted in fig. 5, we get a specific region *3* of very slight agreement, a region *2* of moderate agreement, a region *1* of good agreement and possibly also a region *0* of perfect equivalence. Fig. 5 show the results of a comparison between a daylight lamp (incandescent electric lamp with special blue bulb, see also fig. 1, curve *8*) and ordinary daylight. The unsaturated colours are still badly reproduced, and of the saturated colours purple is the least satisfactory. The direction of colour displacement is shown in the diagram by arrows. Purple is reproduced with a little too much red, the green is too much akin to yellowish-green, and the blue

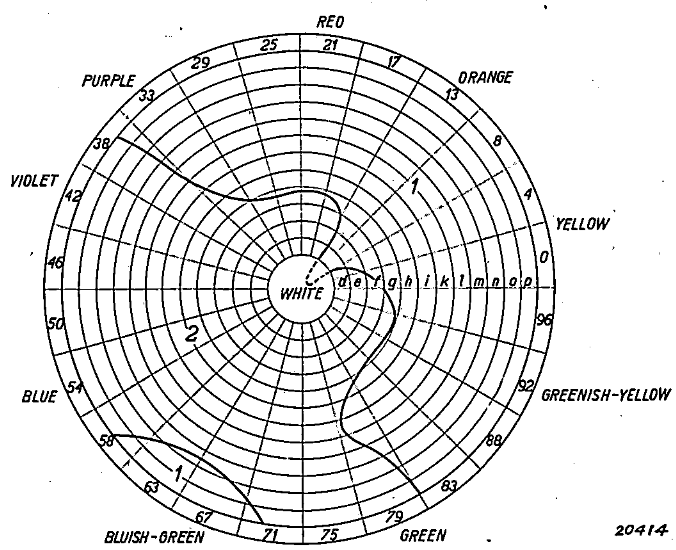


Fig. 7. Comparison of colour reproduction with a mercury-neon-fluorescent compound source with that obtained with daylight.

is a little too unsaturated. These deviations are mainly due to the fact that the daylight lamp still gives too little light at the shortest visible wave-lengths (see fig. 1).

Fig. 6 shows a comparison between a mercury lamp to which red has been added in the form of neon light, and daylight. Here there is a very large region of unsatisfactory colour reproduction, the reason for this being, as already indicated, the large gaps in the spectrum.

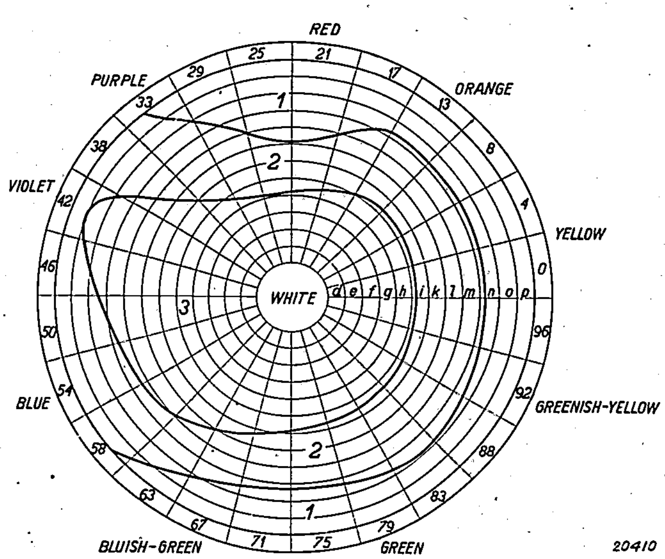


Fig. 6. Comparison of colour reproduction with a mercury-neon compound source with that obtained with daylight.

Fig. 7 compares daylight with a similar mixture of mercury and neon, but which has been improved by the use of a carefully-selected fluorescent bulb and accurate adjustment of the quantity of neon light. Region *3* has here just completely disappeared, and colour reproduction is throughout satisfactory. In fact with light from this source the surroundings appeared perfectly "natural". The gaps in the mercury spectrum were adequately filled by the neon light and the fluorescent light.

Finally, *fig. 8* compares daylight with a "blended-light" lamp (70-watt super-high-pressure mercury vapour supplemented with a 150-watt incandescent electric lamp). The dearth of red rays in the mercury lamp has here been made good by the excess of red light afforded by the electric lamp as compared with daylight (fig. 1, curves *1* and *3*). Here again there is a definite region of unsatisfactory colour reproduction, although this lack is of a different character to that found in the daylight lamp in fig. 5.

[7]) Regarding the definition of the terms "degree of saturation" and "dominant wave-length", see Philips tech. Rev. 1, 283, 1936. The degree of saturation and the dominant wave length of the colours here intended are those which the cards possess in daylight.

What is the relationship between the coefficients *0*, *1*, *2*, *3* and the number of "steps" in colour divergence? This relationship was found by a comparison of incandescent electric lamps which were fed with different voltages (i.e. black bodies with different temperatures). It was established that the coefficient *0* corresponds roughly to a difference of 0—1 steps, coefficient *1* to from 1 to 3 steps, coefficient *2* to from 3 to 5 steps and coefficient *3* to from 5 steps onwards.

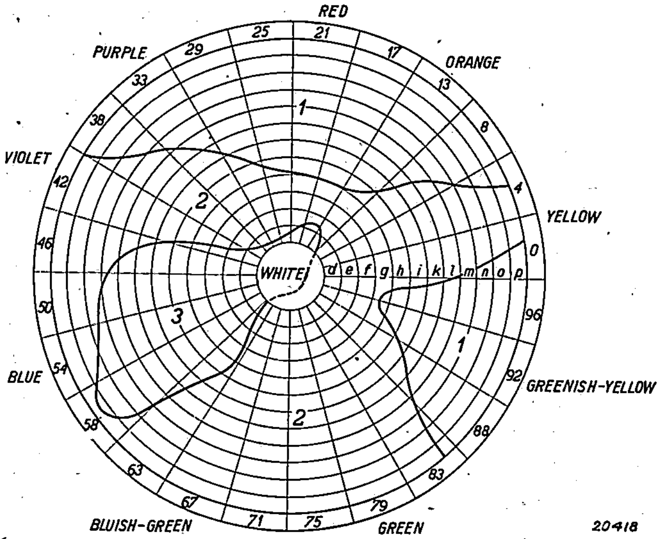


Fig. 8. Comparison of colour reproduction with a compound source made up of mercury light and an incandescent electric lamp, with that obtained with daylight.

The necessity for the complete suppression of region *3* thus coincides with the requirement stated above that the colour difference must not exceed 4 steps.

We thus see that it is indeed possible to obtain such high quality of colour reproduction with gas discharge lamps and fluorescent light (fig. 7). The light output of such a "daylight combination" may be higher than that of an ordinary incandescent lamp, while that of the earlier types of daylight lamps (incandescent lamps with coloured bulbs) was much smaller than the output of the ordinary incandescent lamp.

In many cases where the need for a correct colour reproduction is not so pressing a combination of mercury light and incandescent light (fig. 8) will already be found quite satisfactory.

ILLUMINATION AND ARCHITECTURE

By L. C. KALFF.

Summary. The intended appearance of the artistic forms employed in architecture is in many instances only brought out when they are suitably illuminated. The corresponding results obtained with daylight have been understood and utilised by architects from the earliest times. The innumerable applications and types of artificial lighting available at the present day open up a host of avenues for arriving at similar results with artificial sources of light, some of which have already been fully explored and the results applied in practice, while in many others only the threshold has been crossed and full development is reserved for the future.

Fig. 1. (Left) A building in the Doric style as seen in sunlight. All forms and profiles stand out with their true values. It is distinctly seen how e.g. the flutings and the cornice have been designed entirely on the marked contrast between light and shadow.

Fig. 2. (Right) The same building when seen in diffuse daylight loses all plasticity in its finer profiles and hence the characteristic appearance which the architect had intended.

An analysis of the relationship between illumination and architecture brings us to the conclusion that these two concepts are inseparably linked with each other. To state this fact may appear superfluous, although a number of interesting consequences accrue from it. Long before technological developments enabled electricity to be used for the illumination of architectural works, architects have been building for thousands of years and from the very first had functioned as what might be termed "illumination architects", i.e. daylight or sunlight architects. Some explanation of this aspect of the subject is perhaps necessary.

Every artistic form, which has been evolved on aesthetic lines from practical or other considerations, must be illuminated in a perfectly definite manner in order to reveal to the spectator its specific details of beauty, in other words the forms and plastic motifs of a building can be suitably enhanced in accordance with the aims of the artist by employing an appropriate illumination. That from the very outset every architect based his designs on a specific daylight illumination may be readily demonstrated with the aid of a few examples.

Examination of the different architectonic forms, profiles and ornaments as found in different parts of the world under different climatic conditions, will show that architects in different countries in evolving their designs have been guided strictly by the distribution of daylight prevailing in each particular country and associated with the climatic conditions obtaining there.

Thus it is seen in Egyptian and Assyrian architectural and plastic motifs *(fig. 3)* that in these countries with their abnormally dry climate and very bright and even harsh sunshine preference has been given to particularly pronounced forms and plastic designs ,while ornamentation has been limited to extremely shallow bas-reliefs which when illuminated by bright sunlight reveal a very striking and sharp delineation, even when cut only a few millimetres deep. The same principle is found to have been followed in the artistic productions of Ancient Greece, and it is interesting to note that bas-reliefs of this type are not found in countries further north where the same harsh sunlight is absent and where the sky is comparatively cloudy or overcast and the incident light would be too weak to bring out the finest architectural details.

A typical example of highly-refined "illuminated" architecture is offered by the Greek temples. In the profile of a frieze shown in *fig. 1* it may be seen how the angle of incidence of the sun's rays has been taken into account so that certain parts of the profiles lie in shadow, while again others are brightly illuminated; a modillion for example must stand out in the light against a shaded background, so that each individual modillion constitutes an illuminated unit in the long dark border of the frieze. It is noteworthy that the same profiles which are employed in the frieze behind the front

Fig. 3. Assyrian architecture with its bas-reliefs in stone or terra-cotta also brings out the characteristic features suited to a sunny climate.

row of columns in the peristyle exhibit entirely different forms. It is evident that there no part of the frieze receives the sun's rays, so that the only illumination consists of the sunlight reflected from the ground and the steps of the temple, which although powerful is yet markedly diffused. For this reason the profiles of these interior friezes have been modelled much deeper and with more emphasis, so that even in their location in a half-light their plasticity is yet prominent. On an examination of *fig. 2* it becomes still more clearly apparent that the architecture of the temple has been based on sunlight as the medium of illumination. In this photograph taken with an overcast sky the plasticity of the lightly modelled profiles is absent, and as a result one of the characteristic beauties which the architect had undoubtly intended has become lost.

If we pass to the more northern countries where an overcast sky, fogs and diffuse daylight are of common occurrence, it may be seen, for instance, on Gothic architectural monuments that extremely deep profiles and heavy ornamentation have been employed in order to obtain the desired effects of light and shade with the diffuse illumination there available. An excellent example of this fact is the Hotel de Ville (Town Hall) of Louvain *(fig. 4)* with its living silhouettes, its carved stones and balustrades and deeply chiselled relief. If an architectural masterpiece of this kind were transplanted to the harsh sunlight of Egypt, the greater part of its intrinsic beauty would be lost owing to the pronounced and heavy contrasts obtained between the excessively emphasised light and shadow effects.

A further characteristic consequence of the prevalence of diffuse lighting in northern countries lies in the preference shown for colour contrasts rather than contrasts produced by plastic modelling. An apt example of this is the typical bright exteriors of the burnt-clay and sandstone structures which in the Dutch Renaissance styles gave a much more characteristic appearance to the facades than mere sculptural ornamentation. Also in Germany, Sweden and Great Britain many examples are met where colour contrasts have been employed in place of plastic effects. On the other hand, examination of an Italian palace, such as the Palazzo Pitti, which has been built of a grey stone, reveals that here in the south the mere contrast between light and dark areas already produces a very fine effect.

On comparing a variety of modern applications

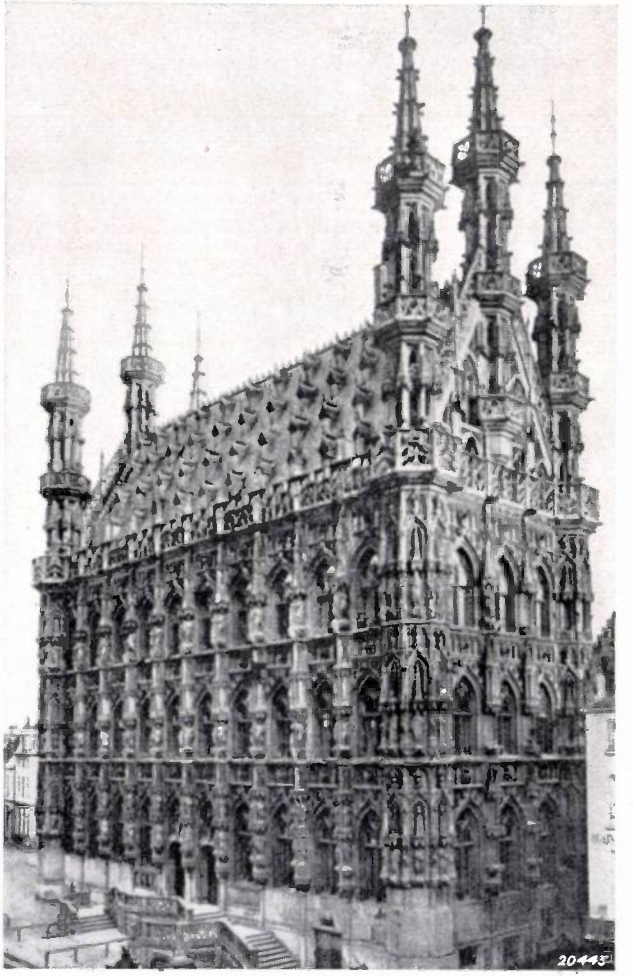


Fig. 4. North European Gothic structures, of which the Hotel de Ville at Louvain is an excellent example, reveals a living silhouette effect, stone balustrades and very deeply engraved reliefs and profiles, since the prevalence of dull weather makes these pronounced contrasts essential in order to obtain the desired effect.

Fig. 5. The Festival Hall of the Brussels World Exhibition, architect Van Neck, reveals a design determined to a large extent by the system of artificial illumination adopted.

interior whose forms have also been based on illumination by sunlight. An examination of such interior decoration when illuminated by an electric lamp suspended in the centre of the room reveals no shadows at all, and the only means of seeing the plastic design is offered by heavy deposits of dust collecting in the deeper hollows, the dust becoming visible owing to its difference in colour.

As long as the available light source remained weak and also evolved a large amount of heat, its most practical position was in the middle of the room, where everyone could conveniently gather round the light, so that full use could be made of the expensive and weak illumination and at the same time the danger of fire was reduced to a minimum owing to the light being at its maximum distance from the walls and ceiling. But the net result of this central position for illumination is, as has

of artificial illumination with the experienced use which has been made for years with natural light, it is realised that hitherto artificial illumination and architectural considerations have only very rarely been considered in the light of their mutual relationships. When floodlighting the profiles and friezes of a classical work of architecture, the floodlights are nearly always directed from below upwards, thus destroying all surfaces in shadow as worked out and intended by the architect. The frieze on the building acquires a flattened effect as a result of this illumination, the shadows disappear under the window sills and along the sides of the windows, and the greater part of the shadows produced by artificial illumination is almost invisible to the spectator who nearly always is looking in the direction of the light itself. Another example is offered by the classical

Fig. 6. The bar in the Capitol Theatre, Madrid, has also been exclusively designed from considerations of illumination technology.

already been indicated in the case of one of the classical examples referred to above, that only few shadows are obtained in the interior ornamentation. For this reason only a limited amount of plastic modelling was employed for these interiors, and the principal effect was obtained by a suitable choice of materials and colours. Daylight illumination has always been of unreliable quality and every consideration had to be given to the location of the room in regard to the direction of the wind and the ever fluctuating outdoor illumination, whose intensity depends on the time of the day and can therefore not be calculated *a priori.*

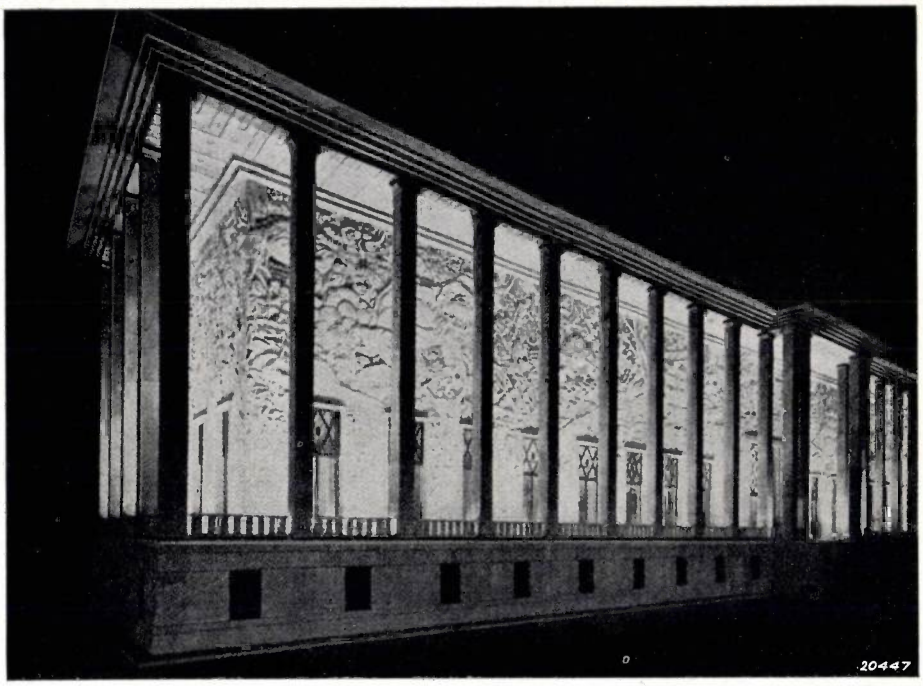


Fig. 7. Colonial Museum, Paris, a work of the architects Jaussely and Laprade, has a facade which has in fact a better appearance during the night than in daylight. The lights mounted behind the breastwork not only afford a very fine illumination of the bas-reliefs of the walls, but also the stepped profiles under the cornice have been designed for this form of illumination. Particularly successfull is the sihouette effect of the columns.

Fig. 8. The giant candelabrum of the Roxy Theatre in New York is not only an interesting lighting ornament, but also reveals shadow edges and decorative bas-relief which are only brought out by the grazing light rays.

In this direction there are innumerable potential applications for modern methods of artificial lighting. At no excessive cost incandescent electric lamps afford a high light output, while their wide ranges of shape and intensity with almost a complete absence of the danger of fire permit them to be placed in their optimum interior positions, and the direction and intensity of the light radiated to be accurately estimated *a priori*. In modern interior decoration we can therefore take as a basis the same general principles as adopted in previous centuries by architects and designers who only had to make provision for daylight illumination.

Naturally for daylight illumination we have almost unconsciously learnt how to give due consideration to the interplay of light and shadow so as to give life to plastic ornamentation and surface design, while to achieve the same natural result with artificial light remains a problem of the future.

Electric incandescent lamps, arc lamps, discharge tubes, etc., are now marketed with so many different shapes and output ratings that a suitable lighting unit can be selected for practically any purpose. But in this connection the appearance of the building under illumination by each light source and the impression made on the spectator by the resulting effect are points which require consideration. What is the most suitable illumination for each particular case? What impression and psychological effect have to be produced? Is the whole of the light to be reflected from the ceiling downwards; must the walls be lit up or is the room to be divided up by rich ornaments in order to form a silhouette? In what direction shall the gaze be attracted and shall this be done by adopting a colour contrast or by more intensive illumination? How do our moods react to coloured light, and how can we create a specific psychological effect? These are only a few of the innumerable problems which have to be solved.

The only true medium for plastic art has hitherto been exteriors where the interplay of light and shadow brings out to the full the many nuances associated with profiles, ornaments, bas-relief, niches, joints and entablatures. In past years these forms were introduced also into interior decorative schemes quite indiscriminately, with the result that the weak and usually centrally-situated light sources to a great extent neutralised the desired effects incorporated in design. Contrasts in colour and materials had then to be adopted as supplementary motifs. But with artificial light it has now become possible to create these effects in interiors exactly as realised with exteriors and even with more pronounced results. We have here a new mode of artistic expression with which we can in the first place gain practical experience and then incorporate the results in our architectural designs. Naturally the adoption of this system of lighting requires entirely new methods of construction, which at the present time have been followed in only isolated instances. These new principles imply the adoption of entirely different cross-sections for our buildings as compared with those adopted hitherto, and it will take many years before we are sufficiently skilled to apply this new method of interior decoration faultlessly and with as much facility and understanding as for exteriors.

The object of the present article is to indicate the new direction of architectural evolution and to convince the architect of the urgent necessity of a close study of its potentialities and applications in the light of modern technology. Examples of some modern designs are shown in *figs. 5* to *8*, in which the spacial forms and the lighting equipment have been rationally adapted to each other. Developments in this direction will continue unhaltingly; those who do not keep pace with them will be regarded as retrograde and their work soon forgotten, while those realising their import and significance will be abreast with the times and produce works of art infused with life.

THE PHOTO-ELECTRIC EFFECT AND ITS APPLICATION IN PHOTO-ELECTRIC CELLS

by M. C. TEVES.

Summary. This article discusses the principles of photo-electric phenomena. The external photo-electric effect is dealt with in greater detail, being followed by a description of technical developments culminating in modern photo-electric cathodes.

Introduction

About a century ago (1839) E. Becquerel discovered that an electromotive force is generated between two metal plates immersed in an electrolyte when these plates are illuminated. W. Smith in 1873 observed that the electrical resistance of the semi-conductor selenium varied when light was directed on it. In 1876 W. G. Adams and A. E. Day investigated the occurrence of electric currents when selenium connected in an electric circuit was exposed to light, while finally in 1887 W. Hallwachs prompted by an observation of H. Hertz on the diminution in disruptive voltage on illuminating the electrodes with light from another spark gap, observed that a negatively-charged plate loses its charge when it is exposed to ultra-violet light. If the plate is positively charged no discharge takes place. P. Lenard and J. J. Thomson demonstrated that the light induced the ejection of electrons from the plate. This group of phenomena is termed "photo-electric".

Those phenomena observed to take place with selenium in which the electrical process is located within the illuminated body are grouped together under the general term of the internal photo-electric effect. But if electrons are emitted from the illuminated surface, as in the last-mentioned experiments of Lenard and Thomson, the behaviour in question is classed as an external photo-electric effect.

In the present article we shall limit ourselves to a discussion of the latter effect. The ejected electrons issue from the plate with a specific velocity and can therefore overcome a potential difference. It is, however, surprising that the potential difference V_{max}, which is just able to retard and arrest the fastest electrons, is not determined by the intensity but exclusively by the frequency of the light, according to the equation:

$$V_{max} = \text{constant}\ (\nu - \nu_0).$$

Light with a frequency lower than ν_0 is no longer able to induce the ejection of electrons. ν_0 is termed the "limiting frequency" or also the "red limit" of the photo-electric effect. It is a constant for each substance.

This fact is difficult to interpret with the aid of the electromagnetic theory of light, for in some way or other the electric-field intensity of the light must be responsible for the emission of an electron from the cathode. But this field intensity is determined by the intensity of the light rays and is independent of the frequency.

A similar difficulty is encountered, moreover, in regard to other photo-effects. In the action of light on a photographic plate there is also a frequency threshold value below which the light produces no effect at all. Again in relation to its physiological effects the frequency of light is frequently of greater moment than its intensity. It is evident that the electromagnetic theory of light is incomplete in these directions.

This shortcoming is overcome by the assumption that light energy can be radiated or absorbed only in specific amounts, "quanta", of the value of $h\nu$ (i.e. a constant times the frequency). h is the socalled Planck's constant and is 6.55 10^{-27} erg per second. To liberate electrons from a metal a specific amount of energy E is required which can only be derived from the radiation when these quanta $h\nu$ are $> E$, i.e. when ν exceeds a definite value of ν_0. If $h\,\nu > E$ the excess energy is converted into kinetic energy of the emitted electrons.

From these considerations Einstein in 1905 deduced the famous equation:

$$^1/_2\ m\ v_{max}^2 = h\,\nu - E = h\ (\nu - \nu_0) \quad . \quad (1)$$

where e is the charge, m the mass and v_{max} the maximum velocity of the emitted electrons. The energy $E = h\nu_0$ can be expressed with the aid of equation $h\,\nu_0 = e\,V_0$ in the form of a potential fall V_0, which the electrons have to overcome on being emitted from the metal.

In Table I the value of the potential fall V_0 is given in volts for a number of metals, as well as the "red limit" in Ångstrom units (1 Å $= 10^{-8}$ cm).

Table I

Metal	Work-function in volts	Red limit in Å
Silver	4.61	2680
Gold	4.90	2520
Cadmium	4.00	3100
Mercury	4.53	2735
Tungsten	4.50	2700
Molybdenum	4.15	3000
Platinum	6.30	1960
Lithium	2.28	5400
Sodium	2.46	5000
Potassium	2.24	5500
Rubidium	2.15	5700
Caesium	1.90	6500

For a given spectral composition of the light the photo-electric current is proportional to the intensity. The response to the light takes place without lag, i.e. the photo-electric effect follows all fluctuations in intensity without measurable delay, even when fluctuations in time take place of the order of 10^{-8} sec.

If all the electrons emitted from the photo-electric cathode are diverted to the anode by a sufficiently powerful field, a "saturation current" is produced which is proportional to the number of light quanta incident per second, i.e. to the intensity of the light, provided of course that the frequency is above the value ν_0 defined by Einstein's equation. Furthermore, the current intensity is determined by the probability of absorption of light quanta resulting in the

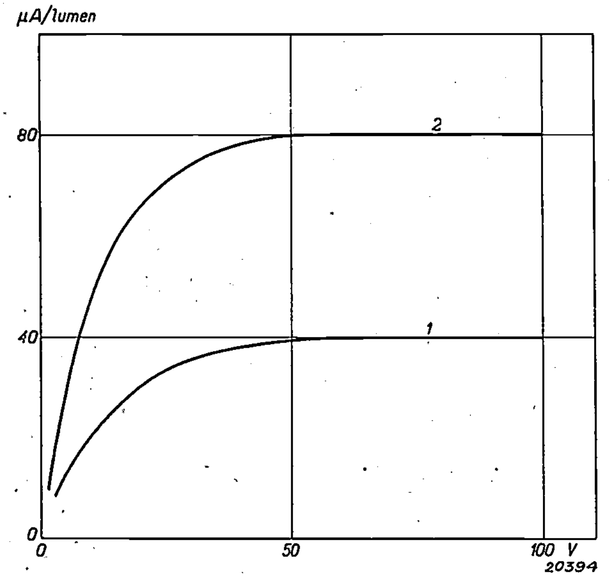


Fig. 1. Photo-electric current per lumen plotted as a function of the anode voltage for two different modern photo-electric cathodes of the type with caesium oxide deposited on silver in a vacuum. *1* Standard base; *2* special base. The curves apply for incident light from an electric lamp burning at a temperature of 2600° K.

emission of an electron, such probability being governed by the wave length of the light.

In *fig. 1* the variation of photo-electric current per lumen is plotted against the anode voltage for two different modern types of photo-electric cathodes. It is seen quite clearly that also with these (non-metallic) cathodes a saturation current is reached as the voltage is raised.

The following simple calculation will give an idea of the possible order of magnitude of the saturation current. We take as wave length λ = 7000 Å or ν = velocity of light/wave-length = $0.43 \cdot 10^{15}$ sec^{-1}, i.e. the long-wave red, since modern photo-electric cells usually exhibit their maximum sensitivity in this frequency range. A quantum $h\nu$ of this radiation has an energy of $6.55 \cdot 10^{-27} \times 0.43 \cdot 10^{15} = 2.81 \cdot 10^{-12}$ erg. Thus with 1 watt, $10^7 \div 2.81 \cdot 10^{-12} = 3.56 \cdot 10^{18}$ electrons can be liberated. In terms of current this is equal to $1.59 \cdot 10^{-19} \times 3.56 \cdot 10^{18}$ = 0.566 A. Actually in practice a sensitivity of not more than a good 1/500 of this sensitivity has hitherto been attained, as is shown for instance in the curves reproduced in *fig. 2*. Similar to those in fig. 1 these curves also apply to modern photo-electric cells, the characteristics of which will be discussed later in this article. The photo-electric currents of pure metals are much smaller still.

With the illumination from an electric lamp curve *1* in this figure corresponds to an output of 80 μA per lumen. For practical purposes, however, the currents obtained with this sensitivity will frequently prove inconveniently small (see e.g. the estimate made in the article "A surveillance system using infra-red rays" [1])). It is therefore extremely important as regards their technical applications that the efficiency of photo-electric cells is made as high as possible.

The curve which gives the variation in sensitivity as a function of the frequency of the incident light, is termed the spectral sensitivity curve, examples of which are reproduced in fig. 2. Along the ordinate the photo-electric current divided by the incident energy in mA per watt is plotted, and along the abscissa the wave length in microns, the frequency and a scale in "electron volts", the latter giving the potentials in volts corresponding to the frequencies in accordance with the equation $eV = h\nu$. It is seen that the red limit for both cathodes is situated at a wave length of approximately 1.5 μ, or at a frequency of 2.10^{14} cycles per second or at 0.83 electron volt. In the same

[1]) A. L. Timmer and A. H. van Assum, Philips techn. Rev., 1, 306, 1936.

way the maximum sensitivity for both cathodes can also be stated in each of these three scales.

Technical Developments

When employing the photo-electric effect the chief aim is to obtain the maximum current with the specific light source available. The light source is in nearly all cases an electric lamp of which the bulk of the energy radiated lies in the infra-red region of the spectrum.

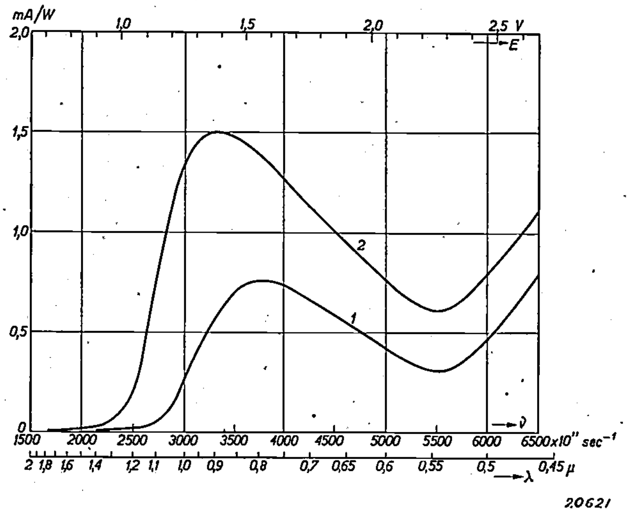


Fig. 2. "Colour sensitivity curve". Photo-electric current plotted against the frequency of the incident light for the same types of photo-electric cells for which the photo-electric current is given in fig. 1.

It is seen from Table I that only the alkali-metals lithium, sodium, potassium, rubidium and caesium, have their red limit located in the visible part of the spectrum, that of caesium being displaced furthest towards the red. These elements are insensitive to infra-red radiation. The maximum sensitivity for potassium is at 4350 Å, for rubidium at 4800 Å and for caesium at 5400 Å. Even when using a caesium cell, only 3 to 4 per cent of the light radiated from an electric lamp has a sufficiently high frequency to cause photo-electric emission. According to the table other metals are still less suitable.

Elster and Geitel found that, as a result of passing an electric discharge through hydrogen in contact with potassium deposited on the walls of a bulb, a photo-sensitive layer was formed whose red limit was displaced towards the red by several thousand Ångstrom units as compared with that for untreated potassium. These cells are termed "hydrogenated cells". Since the red limits of caesium and rubidium correspond to still longer wave lengths it was to be expected that hydrogenated cells of these metals would exhibit a more satisfactory behaviour than those of potassium. The stability of hydrogenated rubidium and caesium cathodes was, however, found to be unsatisfactory at ordinary temperatures.

In the development of the modern photo-electric cell another method was therefore adopted. Langmuir and Kingdon had found that the emission of electrons from incandescent filaments, e.g. those of tungsten, could be considerably increased by coating their surfaces with electro-positive metals, such as the alkaline earths (Ba) or alkali-metals (K, Rb or Cs).

If a tungsten filament on which a thin coating of caesium has been deposited is heated in a vacuum, electrons commence to be emitted already at 300 °C. If the temperature of the wire is gradually raised this emission increases, reaches a maximum and then begins to drop again when a temperature of about 700 °C is attained, at which the caesium is volatilised from the wire.

That the caesium atoms remain on the wire up to a high temperature is demonstrated by the fact that they are attached to their base with considerable cohesion. It is assumed that they are ionised and that an electron is given off to the underlying tungsten. The ionized caesium atoms build up a positive surface charge and so decrease the work done in the emission of the electron. This is shown in *fig. 3*. The electron enters an accelerating field as soon as the last tungsten atoms are passed, which facilitates the emission. This reduction in work done, which was originally observed on the emission of electrons from incandescent filaments, also applies to the photo-electric effect.

If the tungsten is coated with a layer of negative ions, for instance by absorption of oxygen atoms, the reverse process takes place: emission is rendered more difficult, and the voltage drop V_0 increases.

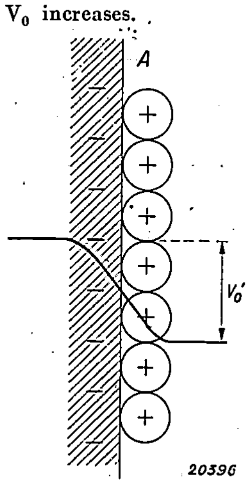


Fig. 3. Effect on work of emission of tungsten due to adsorbed caesium atoms. The caesium atoms deposited on the tungsten give up their electrons to the tungsten, and as a result an electrical double layer is formed. The electrons issuing from the tungsten encounter an accelerating field V_0' which lowers V_0 and so facilitates their emission.

Ives and other investigators found that surface coatings of electro-positive metals are also very useful for inducing photo-emission, and the red limit is displaced far into the infra-red region. Koller described the combinations: silver—monatomic oxygen layer—monatomic caesium layer, whose red limit has still a greater wave length.

Still another method was adopted by De Boer and Teves who based their work on the following considerations:

To obtain a high photo-electric current it is desirable:

1. To reduce as far as possible the work required for emission.
2. To increase as far as possible the efficiency, i.e. the percentage of absorbed light quanta which result in effective photo-electric emission.

The first condition leads, as already indicated, to the use of strongly electro-positive metals. But from the point of view of the second condition metals are not very suitable, for their electrons are extremely mobile, so much so that the greater part of the light energy promotes motion of the electrons and is lost in the form of heat. The photo-electric effect in insulators gives a much better efficiency. When light is absorbed by free atoms (in a gas) each absorbed light quantum actually liberates an electron, if the energy of the light quantum is greater than the ionisation energy.

The purpose of the investigations carried out by De Boer and Teves was to influence the photo-electric effect on solid surfaces in this direction, viz., by the adsorption of caesium on deposited coatings of salts. In the first instance barium fluoride coatings were used for this purpose, being deposited by sublimation *in vacuo* so as to obtain a laminated structure with a large surface (about 100 times greater than with a compact structure). This surface thus had a large number of areas exhibiting a very marked adsorption. Later oxides were used in place of the salts and these proved more efficient still.

The mechanism here responsible for the emission of photo-electrons differs fundamentally from that occurring with metals. The alkali-metal atom adsorbed at the insulating oxide layer behaves with respect to light similar to a free atom in a gas, except that the ionisation energy has been altered by adsorption. A light quantum can ionise this atom and the electron can as a result pass out into the vacuum. The photo-electric effect in this case depends on this process.

When an adsorbed atom has become ionised and an electron has been emitted, a positively-charged metallic ion remains. This must be neutralised before it can again participate in the emission process. But a neutralising electron cannot be withdrawn without further ado from the insulating coating, since this would merely result in a displacement of the charge. The electron must therefore be extracted from the metal substratum. This is indeed readily possible, if the thickness of the insulating coating does not exceed 100 to 1000 molecules, viz., by the field generated by a positive ion becoming sufficiently powerful to draw an electron directly from the underlying metal.

If the oxide layer still contains free metal, these particles can act as intermediate carriers, while if the conduction mechanism has been inadequately developed or has become overloaded by an excessive withdrawal of current, fatigue phenomena may set in since unneutralised positive charges remain extant in the layer. By inserting an additional metal particle the conduction mechanism is improved and the yield can be considerably increased.

A decisive factor in the choice of the insulating coating is the desire to reduce as far as possible the ionisation energy of the adsorbed alkali-metal atom. The best results in this direction are obtained with caesium oxide.

The optimum composition determined as a result of prolonged experiments consists of a silver mirror which is oxidised to silver oxide by a glow discharge in oxygen and then exposed to the action of caesium vapour. The caesium is converted to a caesium oxide, while the liberated silver remains in a very finely divided state and as already indicated increases the conductivity.

Table II gives the sensitivities of various photo-electric caesium cells, as well as the wavelength range corresponding to maximum sensitivity, and the red limit. The sensitivities have been expressed in μA per lumen, using light from a tungsten lamp operating at a temperature of 2600 ° K.

Table II

Sensitive layer	Max. Sensitivity approx.	Wave-length of max. sensitivity	Red limit
	μA/Lumen*	Å	Å
Pure caesium	0.15		6300
Ag with monatomic coating of O_2 and Cs	1.5	3500	8000
Ag with pure Cs_2O and Cs (monoatomic). . .	12	6100	11500
Same with Ag in Cs_2O	20	7000—8000	12000
Same with additional Ag in Cs_2O.	30	7500—8000	12000
Same with additional Cs in Cs_2O.	40	7500—8000	14000
Same with additional Ag and additional Cs in Cs_2O	55	7500—8500	17000

[1]) For light from a tungsten lamp burning at 2600 °K.

The efficiency i.e. the ratio of the number of photo-electrons emitted to the number of light quanta incident on the cell, is about 1 : 100 for the most sensitive cell using the most suitable light source. This ratio is still very much less than 1 : 1. Thus only a small fraction of the incident

a) b)

Fig. 4. Photo-electric cell with caesium cathode:
a) vacuum cell, type 3512;
b) with rare gas filling, type 3530.

light quanta is absorbed by the adsorbed caesium atoms; this is partly due to the fact that the bulk of the light is absorbed by the coloured covering surface.

In many cases the photo-electric current can be further intensified by filling the cell with a suitable rare gas: ionisation of the electrons by collision with the atoms of the rare gas then ensures that under favourable conditions an average of about 20 electrons reach the anode for each electron emitted from the photo-electric cathode. In this way 100 to 200 μA per lumen can be attained with low currents for long periods.

Disadvantages of the gas filling are its inertia, a result of the time required for ionisation to develop, also the fact that there is no linear relationship between the photo-current and the electron current, as well as the greater fortuitous fluctuations in the electron current with constant light intensity, this latter resulting in a disturbing noise.

Fig. 4 shows two photo-electric cells with caesium cathodes: a) a vacuum cell, and b) a gas-filled cell. *Fig. 5* gives the mean photo-electric current in μA per lumen for standard gas-filled caesium cells. No saturation current occurs with these cells. The current, which increases with the voltage eventually causes an arc owing to the steady increase in ionisation, i.e. to a passage of current which is maintained also in the absence of any incident light.

For both types of cells the permissible load is 5 μA per 100 sq. cm. of cathode surface. After 1000 hours the sensitivity is then still 60 per cent of the initial value. The arcing voltage of the gas-filled cell is 150 volts in the dark, and the working voltage 100 volts. In the dark the current spontaneously emitted from the cathode (the socalled dark current) is 10^{-10}A per sq. cm. of cathode surface. At temperatures from 15 to 30 ° C this dark current rises by 10 per cent per degree.

For ultra-violet light, with for instance wave lengths below 4000 or 3000 Å, other cells are employed, viz., sodium cells for below 4000 Å and cadmium cells for below 3000 Å. Below 3500 Å a jacket of glass permeable to ultra-violet rays must be used, while below 2800 Å quartz is necesary.

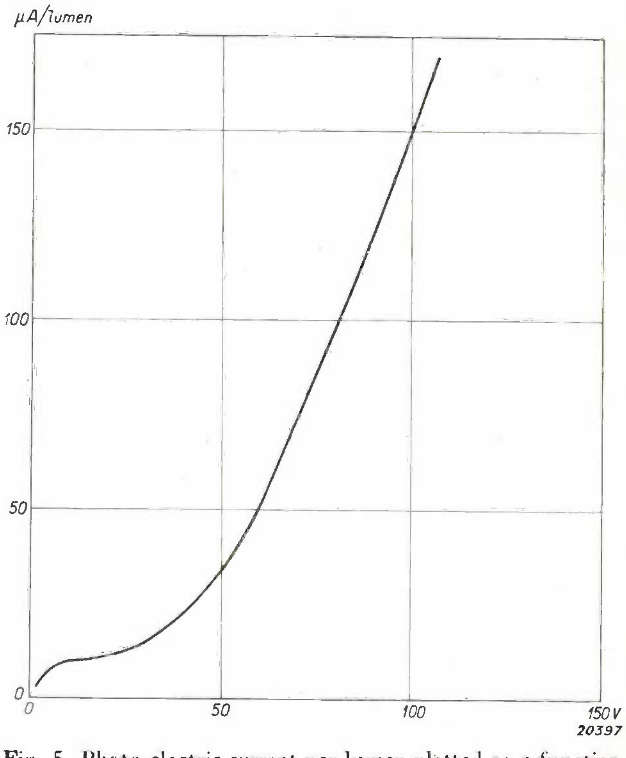


Fig. 5. Photo-electric current per lumen plotted as a function of the anode volts for a gas-filled cell. Contrary to the curves for vacuum cells (fig. 1) this curve exhibits no saturation as the anode voltage increases.

SPECTRAL DISTRIBUTION OF THE RADIATION FROM THE "BIOSOL"

By A. VAN WIJK.

Summary. Short description of the "Biosol" and supply arrangements, followed by a discussion of measurements of the spectral intensity distribution. The spectral sensitivity curves for various biological and therapeutic applications of ultra-violet radiation are also discussed and as far as permissible conclusions are drawn regarding the efficacy of the "Biosol".

Fig. 1. Philips "Biosol" Irradiation Unit.

The "Biosol" and Supply Arrangements

The "Biosol", which was primarily evolved for medical purposes, is a tubular quartz lamp in which an arc discharge through mercury vapour is generated between electrodes. The density of the mercury vapour is maintained constant, during the operation of the lamp, while the vapour pressure is of the order of 1 atmosphere. Two types of lamp are made: Type A rated for 250 watts and Type B rated for 475 watts. In general the "Biosol" is connected directly to an A.C. supply through a choke coil (for voltages from 200 to 260 volts). In the case of an A.C. supply below 200 volts the lamp is connected up over a transformer. The choke coil or transformer is accomodated in the base of the lamp stand which supports the lamp and its reflector. The latter is chromium plated on the inside and can be adjusted both in height and direction *(fig. 1)*. It is made up of six facets arranged at such angles to each other that the radiation reflected from each facet is focussed on the same strip, 40 cm wide, situated 50 cm in front of the lamp *(fig. 2)*. In this way a very uniform distribution throughout the irradiation field is obtained, also at a greater distance from the lamp *(fig. 3)*.

The lamp in most cases lights directly the current is switched on; if it does not doso, a push is operated so that a higher potential is applied to the lamp for a short time. The circuit is shown in *fig. 4*. To obtain the higher voltage required for ignition, operation of the push connects a small transformer in series with the mains supply. The current is then limited by a high resistance. A glass filter partially transmitting ultra-violet rays can also be fixed over the lamp; it absorbs the greater part of the shorter ultra-violet rays (below approximately 0.28 μ) but allows the longer rays to pass through, thus adapting the lamp to other biological purposes. The spectrum of the "Biosol" with and without the filter is shown in *fig. 5*, from which it may be seen that a line spectrum is obtained.

Dosage Control of Ultra-violet Radiation

As soon as we commence to deal with a compound radiation made up of a variety of wave-lengths,

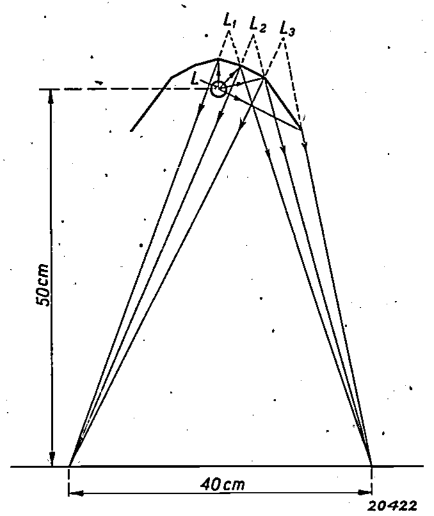


Fig. 2. Concentration of rays with the "Biosol" reflector. The virtual images L_1, L_2, L_3, etc., of the light source L are all projected on to the same strip, 40 cm wide, at a distance of 50 cm.

instead of a simple monochromatic radiation, the estimation of the radiation intensity becomes complex and cannot be determined by the application of any general rules. The evaluation of intensity depends in fact on the nature of the effects and results which are to be achieved with the radiation. Thus, e.g. different compositions of visible radiation, producing an equivalent impression on the eye, are found to have entirely different values when compared as regards their photographic activity, or their influence on the assimilation processus in the green portions of plants. The reason for the different effects produced by

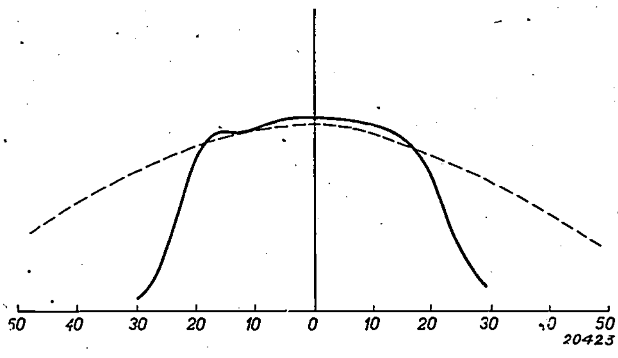


Fig. 3. Relative irradiation intensity in a plane at 50 cm distance from the "Biosol" lamp in the reflector as a function of the distance from the centre of the irradiation field. The curves relate to measurements parallel to the axis of the lamp (broken lines) and perpendicular to it (full lines).

radiations having different spectral compositions is that the action of the radiation on the processes above-mentioned depends on the various wave lengths to different extents. To calculate the action produced by a compound radiation it is necessary to multiply the radiation energy in each wave-length interval by a factor expressing the sensitivity of the particular process under consideration with regard to that wave length, and then add or integrate the result over the whole wave-length range entering into question.

But since the different radiation effects may correspond to different sensitivity curves, the results obtained by such calculation for a specific compound radiation and a specific action cannot be employed *pari passu* for drawing any conclusions of the effect which the same radiation might produce in other processes.

In reference to our introductory remarks regarding the action of a visible compound radiation it must be emphasised that a statement of the intensity of illumination, expressed in lux units, produced by a given radiation on a given surface is no indication of the potential effect on the same

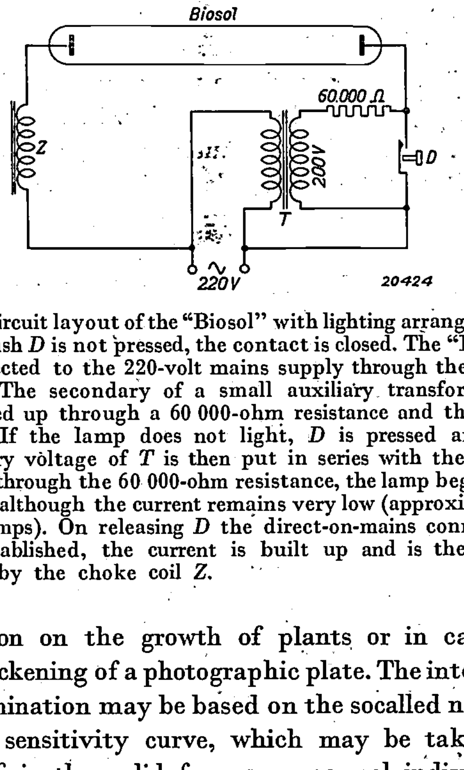


Fig. 4. Circuit layout of the "Biosol" with lighting arrangement. If the push D is not pressed, the contact is closed. The "Biosol" is connected to the 220-volt mains supply through the choke coil Z. The secondary of a small auxiliary transformer is connected up through a 60 000-ohm resistance and the push switch. If the lamp does not light, D is pressed and the secondary voltage of T is then put in series with the mains voltage through the 60 000-ohm resistance, the lamp beginning to burn although the current remains very low (approximately 7 milliamps). On releasing D the direct-on-mains connection is re-established, the current is built up and is then only limited by the choke coil Z.

radiation on the growth of plants or in causing the blackening of a photographic plate. The intensity of illumination may be based on the socalled normal ocular sensitivity curve, which may be taken to be sufficiently valid for every normal individual. For a given photographic plate a similar sensitivity curve can also be deduced, such curve then serving as a basis for calculating the "photographic illumination-intensity" for the particular type of plate under consideration. For other kinds of plates different curves must then be employed.

A similar state of affairs is encountered in estimating the action of ultra-violet radiation when used for biological and therapeutic purposes. The efficacy of the radiation depends here entirely on the action which it is desired to produce. But in comparison matters are less favourable here, since it has hitherto not been possible to deduce a sensitivity curve for every biological therapeutic process entering into consideration. It is in fact doubtful whether standard sensitivity curves, whose validity is reliable as regards normal individuals, at all exist for the various processes in question. But such curves are necessary if we are to evolve a biological unit of irradiation intensity for any specific process, although a knowledge of the corresponding sensitivity curve is not absolutely indispensable for this purpose.

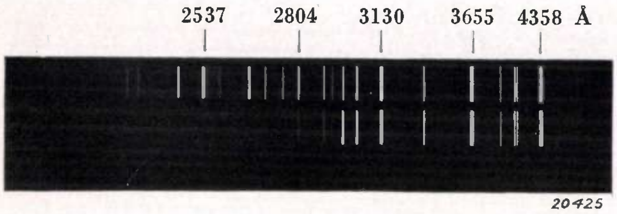


Fig. 5. Spectrum of the "Biosol" with and without filter. Equivalent exposure times.

It thus follows that there can be no question at all of introducing a single unit for expressing the intensity of ultra-violet radiation in reference to its biological and therapeutic effects, and that at most it might be possible to introduce a whole series of these units although at the present time no basis for this development has been evolved.

That the biological and therapeutic sensitivity curves for ultra-violet radiation have not been adequately studied in the past is due mainly to the serious experimental difficulties encountered, *inter alia*, in producing sufficiently powerful monochromatic radiations. In addition, there are the ever-present and serious difficulties of carrying out a really quantitative biological experiment. Actually the only relative sensitivity curve which has been sufficiently accurately investigated up to the present is that relating to the socalled erythema effect. It may be recalled here that the term erythema is applied to the redness of the skin exhibited after a few hours' irradiation with a sufficiently powerful dosage of ultra-violet light. If the intensity of the radiation is sufficiently high, the redness usually becomes converted to a tan which can persist for a fairly long time and is termed pigmentation. If the radiation is not powerful enough, the redness disappears, for instance after about 20 to 40 hours, without leaving any visible trace. The most accurate erythema curve which has been evolved to date is that due to Coblentz and Stair [1]). In how far this erythema curve may be regarded as a standard applicable to all (normal) individuals is still a matter of controversy.

Some knowledge has also been gathered as regards the influence of wave length on a variety of other processes, but not sufficient for the data to furnish the requisite unit of dosage measurement. Thus it is known that pigmentation (tanning, see above) develops to some extent parallel to an erythema, although it has frequently been assumed that there are differences towards both the short and the long waves. Regarding the treatment of rickets which is based on the formation of vitamin D in the skin, it has been established that, just as in the production of an erythema, only wave lengths below approx. 0.31 μ are efficacious, although in other respects the curve is definitely different [2]). The bactericidal (bacteria-destroying) action is most powerful at approximately 0.25 to 0.26 μ and drops off in both directions. In the treatment of lupus (tuberculosis of the skin), the wave length interval between 0.32 and 0.35 μ appears to have the greatest activity. If the connective tissues of the eyes become inflamed by exposure to ultra-violet radiation, the eyes will ache after a few hours and will feel as if sand has entered them. Those who have ascended a glacier omitting to protect the eyes with a suitable pair of goggles will remember these symptoms. The protecting forehead affords a fairly satisfactory protection for the eyes against the solar rays incident from above, but the radiation reflected from snow, ice, still water surfaces and bright sand is incident on the eyes from a direction in which they are quite unprotected. This effect which is termed conjunctivitis (since is consists of an inflammation of the conjunctiva) soon disappears, for instance after the elapse of ten hours. The effect on the conjunctiva is still comparatively weak with rays of about 0.3 μ, but increases considerably in intensity with shorter waves. *Fig. 6* depicts, in addition to the erythema curve, also the conjunctivitis curve calculated from the data of Fischer, Eymers and Vermeulen [3]).

In view of the circumstances outlined above, the only satisfactory method for determining the dosage of ultra-violet radiation necessary in a specific case is to ascertain the physical composition of the radiation as accurately as possible, i.e. the intensity of the various wave lengths (in absolute

[1]) Res. Paper No. 631, Bur. Stand., J. Research, **12**, 13, 1934.

[2]) Cf. e.g. Bunker and Harris, Science, **83**, 487, 1936 E. Gorter, J. Pediatrics, **4**, 1, 1934.

[3]) Cf. the 1935 report of the Commission for ultra-violet rays of the International Illumination Committee.

units, e.g. ergs per sq. cm and sec) and to give the time of exposure. All other data, such as those expressed in "erythema units" — which one might be tempted to use owing to the ease with which they can be determined — provide no information

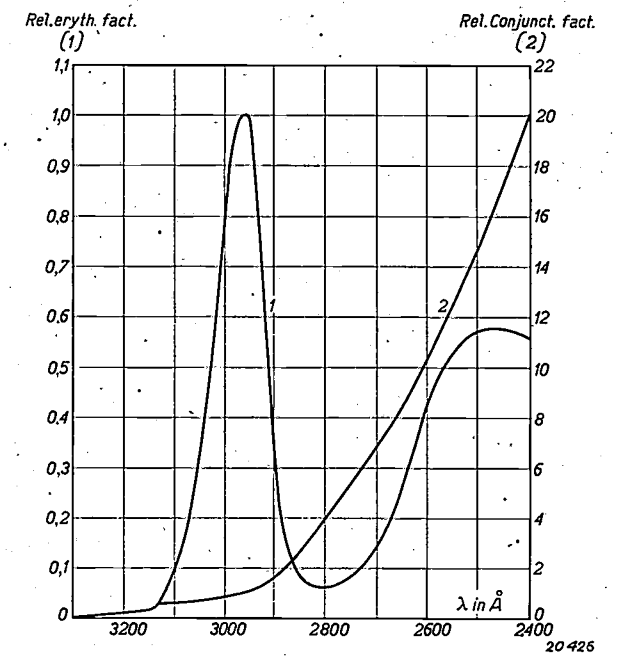


Fig. 6. *1)* Erythema curve of Coblentz and Stair.
2) Conjunctivitis curve of Fischer, Eymers and Vermeulen.

In both cases the reciprocal of the dosage required to obtain a just perceptible effect is plotted as a function of the wave length. The maximum values in the range of wave lengths shown have been arbitrarily put equal to unity.

of practical utility and are incomplete. The more knowledge is gained regarding the curves of biological action, the better equipped will be the investigator to calculate in advance the potential effect of a specific irradiation, i.e. as obtained with a radiation having a known spectral distribution curve.

The questions of principal current interest as regards dosage measurement and control with ultra-violet radiation are the following:

Determination of the spectral intensity distribution in the radiation from ultra-violet light sources and the plotting of the biological sensitivity curves [3]).

Spectral Intensity Measurements with the "Biosol"

The intensity of the spectral lines emitted by the "Biosol" between 3130 and 2480 Å was determined in absolute units. *Table I* gives the results of measurements for both types of "Biosol" without a filter or reflector. To permit comparison, the relative values obtained by Krefft and Pirani with a similar mercury discharge at a vapour pressure equal to that in the "Biosol" are also included.

Table I. Radiation Intensity of the "Biosol" at 50 cm distance.

Wave-length	Absolute intensity of "Biosol" A	Absolute intensity of "Biosol" B	Relative intensity (Kr. & P.)	Absolute : Relative A	Absolute : Relative B
Å	1000 ergs per sq. cm/sec.	1000 ergs per sq. cm/sec			
3130	2.01	4.76	66.4	30	72
3022	0.864	2.10	31.5	27	66.5
2967	0.493	1.05	14.8	33	71
2894	0.180	0.425	5.9	30^{5}	72
2804	0.325	0.855	11.3	29	75.5
2753	0.115	0.270	3.8	30	71
2699	0.150	0.350	4.9	30^{5}	71.5
2652	0.687	1.640	22.4	30^{5}	73
2537	1.17	2.370	29.1	40	81.5
2483	0.370	0.880	12.3	30	71.5
			Average	30	71.5

With the exception of the line with a wave length of 2537 Å, whose intensity in particular is closely dependent on the pressure, our values are in satisfactory agreement with the relative values of Krefft and Pirani. Since these investigators also made measurements at longer wave lengths, including those in the visible spectrum, we can calculate from their observations the intensity of the "Biosol" also in this range, using the abovementioned conversion factor. As a check the intensity of the "Biosol" was calculated (on the basis of the relative ocular sensitivity curve and the value of the mechanical equivalent of light) from the intensity of the visible lines obtained in this way, and compared with the values obtained by direct measurement (Compare *Table II*).

Table II. Candle Power (Intern. Candles) of "Biosol" perpendicular to axis of tube.

Type	Calculated	Measured	Calc. : Measd.
"Biosol" A	675	625	1.08
"Biosol" B	1630	1570	1.04

Furthermore, by measuring the irradiation intensity with different wave lengths on a surface at a distance of 50 cm from the lamp both with and without the reflector, the intensification factor of the reflector was determined as a function of the wave length. *Table III* gives the results obtained for both types of "Biosol" with and without the filter.

Table III. Radiation Intensity of "Biosol" with reflector at 50 cm distance.

Wavelength	Intensity of "Biosol" A		Intensity of "Biosol" B	
	Without filter	With filter	Without filter	With filter
Å	1000 ergs per sq. cm/sec.	1000 ergs per sq. cm/sec.	1000 ergs per sq. cm/sec.	1000 ergs per sq. cm/sec.
5770/91	6.55	5.9	16.3	14.7
5461	5.9^5	5.35	14.7	13.2
4358	5.2	4.7	12.5	11.3
4078	0.48	0.43	1.15	1.04
4047	2.9	2.6	6.95	6.25
3655	9.4	8.45	22.4	20.2
3342	7.6	6.1	1.82	1.45
3130	5.75	3.8	13.6	9.0
3022	2.43	1.27	5.9	3.08
2967	1.35	0.59	2.85	1.25
2894	0.46	0.15	1.1	0.35
2804	0.82	0.15	2.15	0.41
2753	0.27	0.035	0.67	0.086
2699	0.35	0.028	0.84	0.067
2652	1.60	0.084	3.85	0.20
2537	2.55	0.032	5.15	0.064
2483	0.78	0.003	1.85	0.007
Illumination in lux	$7.85 \cdot 10^3$	$7.05 \cdot 10^3$	$1.97 \cdot 10^3$	$17.7 \cdot 10^3$

Erythema Action of the "Biosol" With and Without Filter

For each of the emitted lines the fractional contributions to erythema can now be calculated by multiplying the intensity by the erythema factor. *Table IV* gives the results of this calculation for "Biosol" B (with reflector, and at a distance of 50 cm) with and without the filter.

Summation gives a measure for the total erythema effect. It is found that with the "Biosol" B without filter the total effect is equivalent to that produced by a radiation exceeding 13 000 ergs per sq. cm and sec with a wave length of 2967 Å (at which, according to Coblentz and Stair, the erythema factor has its maximum value and is taken as equal to unity). The lamp with filter gives a radiation which as regards erythema effect is equivalent to that obtained with an intensity of 3560 ergs per sq. cm and sec at 2967 Å. It was ascertained experimentally that to obtain a specific erythema the time of irradiation required with the filter was three times that required without the filter. The agreement between calculation and direct measurement is thus satisfactory. According to Coblentz (IIe Congrès International de la Lumière, Copenhague, 1932) a just perceptible erythema on the inside of the forearm requires a radiation dosage of approximately 200 000 ergs per sq. cm at $\lambda = 2967$ Å. We found in good agreement with this figure that the limiting dose at a distance of 50 cm from the "Biosol" B with filter was about 45 secs with the majority of the individuals examined ($45 \times 3560 = 160\,000$). When using the lamp without a filter a weak erythema was observed in most cases after only 15 secs irradiation. However considerable differences existed between individual subjects.

As may be seen from Table IV, when using the "Biosol" without a filter about 45 per cent of the erythema is due to radiation with a wave length less than 0.280 μ, i.e. to a radiation which is not present in the solar spectrum. If the filter is fixed in front of the lamp, only a negligible proportion (about 3 per cent) of the erythema is due to this short-wave ultra-violet radiation. Corresponding differences naturally also apply to other biological processes. It is to this effect that the value of the filter is due, viz., that it adapts the action of the "Biosol" to make it comparable to that of sunlight.

Table IV. Intensity I, erythema factor f and erythema ratio I.f of "Biosol" B at 50 cm distance for different wave-lengths in absolute measure and as a percentage of the total erythema effect.

Wave-length	Erythema factor	Without filter			With filter		
		Intensity	Erythema ratio		Intensity	Erythema ratio	
Å		1000 ergs per sq. cm/sec	Abs.	%	1000 ergs per sq. cm/sec	Abs.	%
3130	0.03	13 6	408	3.1	9.00	270	7.6
3022	0.55	5.90	3240	24.6	3.08	1700	47.8
2967	1.00	2.85	2850	21.6	1.255	1255	35.2
(2925)	0.70	0.435	304	2.3	0.161	113	3.2
2894	0.25	1.10	275	2.1	0.352	88	2.5
2804	0.06	2.16	130	1.0	0.410	25	0.7
2753	0.07	0.67	47	0.4	0.086	6	0.2
2699	0.14	0.84	118	0.9	0.067	9	0.3
2652	0.25	3.84	960	7.3	0.201	50	1.4
(2576)	0.49	0.755	370	2.8	0.013	6	0.2
2537	0.55	5.16	2840	21.6	0.064	35	1.0
2483	0.57	1.85	1055	8.0	0.007	4	0.1
(2464)	0.57	0.27	154	1.2	—		
(2400)	0.56	0.74	415	3.2	—		
			13166	100.1		3561	100.2
Proportion of radiation with wave-lengths below 2800 Å in total erythema effect			45.4 per cent			3.2 per cent	

Conjunctivitis Effect of the "Biosol" B with and without Filter

With the aid of the curve for the conjunctivitis effect reproduced in fig. 6, the corresponding curve for the radiation of the "Biosol" B was determined.

for an intensity equivalent as regards this effect, at the wave length 2967 Å *(Table V)*.

Table V. Intensity I, conjunctivitis factor c and conjunctivitis ratio I.c of "Biosol" B at 50 cm distance for different wavelengths.

Wave-length	Conjunctivitis factor	Without filter		With filter	
		Intensity	Conjunctivitis ratio	Intensity	Conjunctivitis ratio
Å		1000 ergs per sq. cm/sec,		1000 ergs per sq. cm/sec.	
3130	0.75	13.60	10	9.00	6.8
3022	0.85	5.90	5	3.08	2.6
2967	1.0	2.85	2.8	1.255	1.3
2925	1.5	0.435	0.6	0.161	0.2
2894	2	1.10	2.2	0.352	0.7
2804	4	2.16	8.6	0.410	1.6
2753	6	0.67	4.0	0.086	0.5
2699	7	0.84	6.3	0.067	0.5
2652	8	3.84	32.6	0.201	1.7
2576	10	0.755	7.5	0.013	0.1
2537	12	5.16	62	0.064	0.8
2483	15	1.85	27.8	0.007	0.1
2464	17	0.27	44.6	—	—
2400	20	0.74	14.8	—	—
			$188.8 \cdot 10^3$		$16.9 \cdot 10^3$

According to Fischer, Eymers and Vermeulen, the limiting dose at 2967 Å for conjunctivitis is approximately $12 \cdot 10^5$ ergs per sq. cm. This dosage can be obtained with approximately 7.5 secs irradiation with the "Biosol" B without filter (perhaps even a shorter exposure would be sufficient owing to the proportion of wave lengths below 0.240 μ which could not be taken into account). On the other hand, when using the filter on the lamp approximately 80 secs irradiation would be required to reach the limiting dosage. Experiments to determine this have not yet been made; but the results of calculation agree with general experience that between the lamps with and without the filter there is a marked difference as regards the conjunctivitis effect, which difference is much more pronounced than that found for erythema. The calculation given here is however very unreliable owing to the considerable uncertainty attaching to the conjunctivitis curve.

Treatment of Rickets

The requisite daily dosage found by Gorter and Soer [4]) for the treatment of rickets in children was as follows: $4.2 \cdot 10^6$ ergs (0.1 cal) on irradiation with a wave length of 2967 Å. This radiation was concentrated on a skin surface of 200 sq cm, i.e. $21 \cdot 10^3$ ergs per sq cm. In agreement with the dosage values given above for producing an erythema, no erythema results with this irradiation. Neglecting all consideration of other wave lengths, the requisite exposure to irradiation for the treatmant of rickets with the "Biosol" B with filter at a distance of 50 cm is found to be $21 \cdot 10^3 : 1255 =$ approximately 17 secs, if an area of 200 sq. cm is to be irradiated. On increasing the irradiated surface the requisite exposure time diminishes roughly in inverse proportion.

These examples will suffice to demonstrate how with an adequate knowledge of the spectral biological-activity curves the potential effects and the requisite irradiation times can be deduced from the known spectral distribution of the radiation from the "Biosol".

4) E. Gorter and J. J. Soer, Ned. T. Geneesk., 74, 4310, 1930.

THE "PHILORA" SODIUM LAMP AND ITS IMPORTANCE TO PHOTOGRAPHY

by J. A. M. VAN LIEMPT.

Summary. In photographs taken with sodium light illumination the relative brightnesses of coloured objects are rendered very well. The reproduction becomes almost perfect if the sodium light is mixed with a suitable quantity of incandescent or mercury light. Furthermore sodium light offers unexpected advantages for the illumination of dark rooms.

Introduction

It may be stipulated as a requirement for non-coloured photographs that the colours naturally occurring in practically every group of persons or things photographed shall be reproduced in a greyish tint whose subjective brightness-impression is in agreement with the corresponding colours.

The old type of photographic plate, sensitive only to blue, did not meet this requirement at all. It was therefore an important step in the development of the photographic negative when H. Vogel discovered, in 1873, that by adding certain colouring matters to the photographic emulsion one could greatly increase the sensitivity of the emulsion to red and yellow rays. This gave rise to orthochromatic and panchromatic plates, now in general use. Meanwhile it has been found that although these plates considerably improved the photographic reproduction of colour, they have not yet proved capable of reproducing colours in an ideal manner when working with any of the artificial light-sources hitherto available in great variety. The only way to come to some extent near the desired result was to select, with a given light-source, a suitable filter for each class of plate.

As a matter of fact it is a laborious matter to select such a filter combination; besides, in such filters much light is lost by absorption, so that the required exposure time becomes longer than is desirable.

In collaboration with the court photographer F. Ziegler of The Hague we carried out several experiments on sodium lamps as a light-source for photography. The result, surprising though it first seemed, revealed the fact that this nearly monochromatic light-source solves the problem of photographic reproduction in a most satisfactory manner without making use of filters [1]).

By way of further illustration we are setting forth below a series of measurements carried out with various types of plates on the market with sodium light, which may be combined with other auxiliary light-sources.

For these measurements use was made of the Colour Test Chart (Stufenfarbentafel) [2]) issued by Agfa, a photographic reproduction of which is given in *fig. 1* (taken in daylight on panchromatic

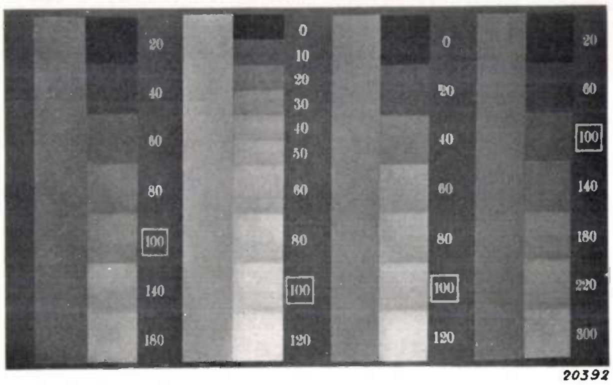


Fig. 1. Reproduction of the Agfa Colour Test Chart on panchromatic plate with daylight.

plates. This chart consists of 4 strips coloured red, yellow, green and blue respectively, each colour having beside it a numbered shade of grey. These four colours are the main colours of the spectrum and represent average colours as met with in practice (i.e. not pure colours of the spectrum).

As a result of scientific measurements in the Agfa laboratory it has now been definitely established that every colour, when correctly reproduced photographically, can be represented by a certain shade of grey which is indicated by the figure 100, which figure means that this particular shade of grey makes the same brightness-impression on our eye as the original colour.

A shade of grey that is indicated, for instance, by the figure 80 is one that reflects 80% white light compared with the shade of grey indicated as 100, and so on.

[1]) Several articles by F. Ziegler on this subject have appeared in „Bedrijfsphotographie" 17, 247 1935); 18 331 (1936).

[2]) See H. Arens and J. Eggert, Z. wiss. Phot. 27 279 (1930) and M. Biltz, Wiss. Ver. Agfa 4, 42 (1935).

If, therefore, a photograph is taken of the Colour Test Chart in the correct exposure time, using for illumination the light-source to be examined, we obtain a negative in which can be seen what shade of grey in the scale corresponds to the colour concerned.

If, for instance, red is reproduced by grey-shade 40, this means that the reproduction of red is too dark. In this case, therefore, the light-source used did not furnish sufficient red rays, or possibly the plate used was not sufficiently sensitive to red.

The correctness of the exposure time can easily be verified by the reproduction in the scale of grey. With correct exposure times all the stages of the scale of grey will be seen quite distinct from each other on the negative.

All the measurements were carried out with the following types of lamps:

1) "Philora" Sodium Lamp type SO 650, a 100 watt lamp with an efficiency of approx. 62 lm/watt [3]).
2) "Philora" mercury vapour lamp type HP 300, a 75-watt quartz lamp with an efficiency of approx. 36 lm/watt [3]).
3) "Philora" mercury vapour lamp type HO 1000, a 250-watt glass lamp with an efficiency of approx. 36 lm/watt [3]).

Along with these lamps some ordinary incandescent lamps are also used.

The spectrum of the sodium lamp is practically of a monochromatic yellow colour, in addition to which the lamp gives a little red neon radiation; the spectrum of the mercury lamps is a line spectrum in various colours, whilst the H.P. lamp emits a faint continuous spectrum as well.

The photographs were taken with the light of these lamps (or combinations thereof) on negative material as used in photographic practice. By way of comparison some photographs were also exposed in daylight in a room with a window facing northward, the sky being practically cloudless.

The results were as follows:

If, in the above *table I*, we first examine the photographs taken in daylight we find that the colour reproduction is far from perfect even when using good panchromatic plates. Moreover, mercury light is unsuitable, as it contains too much blue and too little red. Sodium light, on the other hand, gives the almost ideal colour reproduction in yellow and red — somewhat less good in blue — whereas the reproduction of green is comparatively poor with practically every kind of photographic material, because (for reasons we need not go into here) the negative material on the market is by its nature poorly sensitive to green.

Tabel I. Measurements in unmixed "Philora" light

Plate	Light-source	Colour reproduction red	yellow	green	blue
Non-chromatic plates					
Gevaert Ultra Rapid	Sodium	100	100	50	60
	Daylight	30	10	25	180
Orthochromatic plates					
Gevaert Super Press	Sodium	90	110	50	60
	H.P.lamp	50	10	15	220
	H.O.lamp	40	20	25	190
	Daylight	30	20	30	180
Gevaert Ortho Process	Sodium	90	100	50	60
	Daylight	30	15	30	200
Panchromatic plates					
Gevaert Ultra Panchro	Sodium	120	120	50	40
	H.P.lamp	50	40	35	180
	H.O.lamp	45	40	35	160
	Daylight	60	40	40	140
Agfa Superpan	Sodium	140	120	60	40
	Daylight	90	40	30	140
Ilford Hypersensitive Panchromatic	Sodium	120	110	50	60
	H.P.lamp	55	30	30	180
	H.O.lamp	50	60	35	140
	Daylight	75	40	35	150

Furthermore, it is a noticeable fact that colour reproduction in sodium light is just as good on non-chromatic material as on orthochromatic material, so that — if the duration of the exposure is no object — negative material that is sensitive only to blue will do just as well as the more expensive panchromatic material; moreover, to a certain extent with orthochromatic material the the technique of development is simpler than with panchromatic.

We thus come to the remarkable paradox that it is possible to obtain good colour reproduction with the monochromatic light of sodium in conjunction with non-chromatic material.

From daylight photos we have already seen that the imperfect reproduction of colour obtained even with the most up-to-date panchromatic material, is due to the fact that this material is still insufficiently sensitive to red and yellow. Whilst the

[3]) In the determination of the efficiencies the losses in the series resistances and transformers are taken into account.

use of incandescent lamps as light-source does considerably improve the reproduction of red, that of yellow is still at a disadvantage. With sodium light, thanks to the large proportion of yellow it contains, the deficiency of yellow-sensitivity is entirely made good. As regards the reproduction of red it should be noted that the sodium lamp contains in the first place a small proportion of red neon rays. Moreover, the red occurring in practice is not a spectrally pure colour but contains, in the sense of Ostwald's colour theory, an appreciable proportion of white which reflects all other colours and especially the yellow adjacent to red. The ability of sodium light to give very fair, though not perfect, reproduction of blue, is also due to a white content in the blue colour met with in practice.

Since, further, the light reflected from sodium-lit objects of different colours is, unlike incandescent lamp light, nearly monochromatic, it is obvious that the same reproduction of colour will be produced on a nonchromatic as on a panchromatic plate. The exposure time will, however, be considerably longer for the non-chromatic plate on account of its low sensitivity to the yellow sodium line (0.59 μ). For sodium light photography it would indeed be sufficient to have a plate sensitive to yellow only, but this sensitivity to yellow should be as high as possible.

Measurements in mixed light

The reproduction of blue sodium light may be brought to perfection by adding the light of other light-sources to the sodium light. We merely state the final result.

On non-chromatic material, thanks to its high sensitivity to blue, a very slight admixture of incandescent lamp light proves sufficient.

Examples:
Emulsion: Gevaert Ultra Rapid (non-chromatic)
Light-source: Sodium lamp in conjunction with vacuum spiralised-filament lamp.

Light-source	Colour reproduction			
	red	yellow	green	blue
Sodium lamp	100	100	50	60
1w sodium lamp-power to 0.01w incand. light	100	100	50	100

In the case of orthochromatic plates which are not so sensitive to blue it is necessary to add somewhat more incandescent lamp light.

Emulsion: Gevaert Super Press (orthochromatic)
Light-source: Sodium lamp with vacuum spiralised-filament lamp.

Light-source	Colour reproduction			
	red	yellow	green	blue
sodium lamp	90	110	50	60
1 w sodium lamp-power to 0.30 w incandescent light	100	100	50	100

For panchromatic material — which is generally to be preferred on account of the shorter exposure time — the best results are obtainable with a combination of sodium and mercury light.

Emulsion: Ilford Hypersensitive Panchromatic
Light-source: Sodium light with HO mercury lamp.

Light-source	Colour reproduction			
	red	yellow	green	blue
Sodium lamp	120	110	50	60
1 w sodium lamp-power to 0.5 w mercury lamp-power	90	90	40	90

With these same light-sources we also found, in a similar manner:

Emulsion: Gevaert Ultra Panchro 8000 H and D
Light source: Sodiumlamp with HO mercury lamp

Light-source	Colour reproduction			
	red	yellow	green	blue
Sodium lamp	120	120	50	40
1 w sodium lamp-power to 0.5 w mercury lamp-power	120	100	50	80

Instead of adding mercury light we can also obtain good results by combining the sodium light with an "Argaphoto" lamp with blue filter.

The foregoing examples are merely intended as a guide to practical photographers, each of whom can, with the aid of Colour Test Charts, select for himself the right combination to use in conjunction with the negative material he prefers.

We would furthermore draw attention to the high efficiency of "Philora" lamps to which we referred at the commencement, this efficiency being such that, for the same current consumption, the lamps give two to three times as much light as ordinary incandescent lamps.

Another incidental advantage afforded by the sodium lamp when used for portrait photography is that, for a given illumination on the eye, the pupil is rendered larger by sodium light than by

Photo F. Ziegler, Den Haag

Photograph taken on the betrothal of H.R.H. Princess Juliana and H.S.H. Prince Bernhard.
The exposure was made by sodium light.

white light [4]) — a fact to which F. Ziegler has already drawn attention.

The accompanying photos *(fig. 2)* plainly indicate this. The same person has been photographed — in one instance after the eyes (while accomodated to infinate distance) had been exposed for 15 minutes to the light of a 300-lux "Argaphoto" lamp, and in another instance after such an exposure of the eyes to 300 lux of sodium light (luxmeter calibrated by means of the flicker photometer). In the first case the ratio $\frac{\text{iris diam.}}{\text{pupil diam.}}$ is 3.7, in the second case 2.0.

For the time of exposure in seconds the following formula may be used:

$$t = \frac{13500\, F^2}{L \times HD},$$

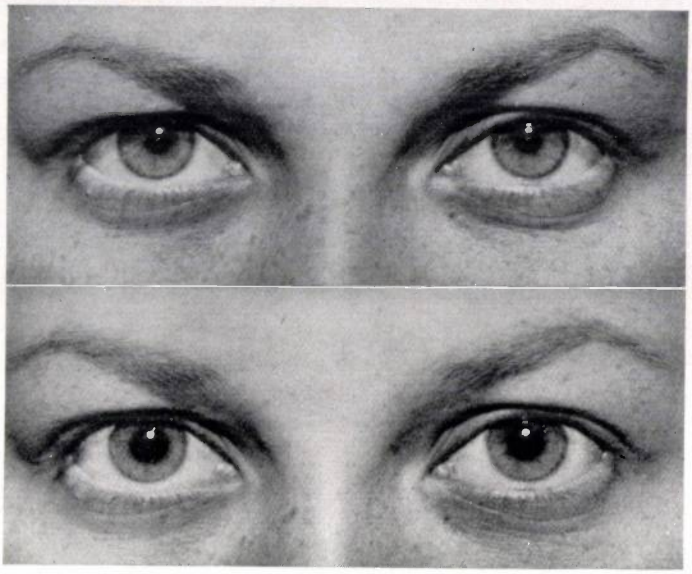
in which: F is the opening of the lens

L the illumination in lux on the place of the object to be photographed

HD the sensitivity of the plate for artificial light in Hurter and Driffield units.

This relation may be used for the mixed light for ortho- as well as panchromatic plates. Using pure sodium light it is valid only for panchromatic material.

[4]) A similar effect is also obtainable with the "Photoflux" lamp. See Z. wiss. Phot. 33, 287, 1935.



20593

Fig. 2. Photo's taken after accomodation during 15 min. with 300 lux incandescent light (gasfilled) and with 300 lux sodium light respectively.

Illumination of the darkroom

Finally, we should like to draw the attention to an important application of the "Philora" Sodium Lamp, viz, as a dark-room lamp for use in handling gaslight papers when making prints and enlargements. Since the radiation of the sodium lamp is almost purely monochromatic and ordinary gaslight paper is insensitive to yellow, a photographer working by a 50-watt sodium lamp can flood his "dark" room with a sea of light — a boon formerly undreamed of! As, however, the electrodes of the lamp still give a faint continuous spectrum it is advisable that the lamp be fitted not only with an ordinary vacuum globe but also with an effective yellow filter whereby this white light is cut off and only a slight total loss of light occurs. Gevaert Ridax paper immersed in the developer can, for instance, be exposed for hours to this light without showing the slightest blur. However, when using bromide paper, which is more sensitive, greater caution is required and one cannot indiscriminately use this lamp.

The foregoing explanations will no doubt be sufficient to show the great advantages which are offered by sodium light both for photography and for cinematography. At the same time it furnishes us with a striking example of the manner in which a technical product devised for a different purpose may sometimes unexpectedly prove to be of far-reaching importance in allied branches of technique.

PRACTICAL APPLICATIONS OF X-RAYS FOR THE EXAMINATION OF MATERIALS IX

By W. G. BURGERS.

To decide whether in a particular case of the examination of a material the X-ray method might prove of use, it must always be remembered that this method gives in the first instance only information of the crystalline structure of a material. If a specific difference between two materials is due to a divergence in their crystalline texture, it may be assumed at least in principle that such difference can be detected by X-ray examination. In many instances the difference in texture is a result of differences in composition of the substances under consideration, and in a number of examples already cited, e.g. No. **9** (metallic deposition in the formation of metallic carbides), No. **10** (detection of thorium oxide and metallic thorium in a tungsten wire), and No. **15** (excessive tantalum content of tantalum carbide) it has been observed that the radiographs in cases of this type differ from each other in the relative positions or number of the interference lines.

Also in a material of homogeneous composition, differences in crystalline configuration can also occur, for instance as regards size and complete formation, or also the mutual orientation, of the crystals making up the material.

Differences of this type correspond in general to very pronounced differences in the physical and mechanical properties of the substance. Also in these cases, it may be expected that X-ray examination will give information of the characteristic state of the body. Some appropriate applications of the X-ray method in this direction have already been discussed in examples No. **3** (production of fissures in soapstone on firing) and No. **5** (texture of electrolytically-deposited nickel coatings), where it was found that the observed differences in properties were related to difference in the orientation of the crystals). Attention should also be called to examples Nos. **17** and **19** where in a homogeneous substance differences in the conditions of internal stress were revealed by X-rays.

In the present and subsequent articles, a selection of further examples will be discussed in which different "states" of one and the same material can be differentiated by X-ray examination.

20. Recrystallisation of Drawn Tungsten Wire

An instructive example of the potential application of the X-ray method to the detection of changes in crystalline texture is obtained in the heat treatment of a machined metal. It is well known that plastic deformation of a metal causes the marked deformation of the original crystallites, a change which may be detected by the loss of definition in the crystals on etching the metal and the production of a more or less fibrous texture. If the deformed metal is maintained at a suitable high temperature for a sufficient time, recrystallisation will take place, in other words the fibrous texture will become reconverted to a normal crystalline texture made up of crystals with sharply-defined boundary planes. Recrystallisation takes place as a result of the displacement of atoms at specific highly-deformed and hence heavily stressed points in the machined metal, these atoms forming nuclei for new normal crystals which then grow at the expense of their distorted surroundings.

The etched figures reproduced in figs. 1*a* to *c* show this process in regard to a tungsten filament about 90 μ thick, using a magnification of over × 300. Pieces of this wire were subjected to heat treatment at an elevated temperature, and it is seen how as a result the original fibrous texture becomes altered: At the lowest reheating temperature employed (fig. 1*b*) there is little change to be seen; at the very most an incipient "coarsening" of the fibrous texture has developed. In the next stage illustrated (fig. 1*a*) this "coarsening" is more pronounced and the fibres are seen to disappear gradually; the formation of new crystallites however cannot yet be definitely deduced from the etched figure (the point marked with an arrow is probably the first beginning of this formation). In the section of the wire reheated to a still higher temperature (fig. 1*d*) this can already be distinctly observed: The fine striations at the top of the figure are due to a crystal which has become formed from a nucleus in this part of the wire and has already consumed part of the original fibrous texture. In the last figure (fig. *c*) recrystallisation has proceeded so far that the whole section of the wire has become occupied by a single crystal.

Figs. 1*f* to *k* reproduce the corresponding radio-

[1]) The difference between real and cultered pearls (example 4) is also due to this.

graphs of the wire sections [2]), and these also clearly reveal the changes which have taken place in the wire as a result of reheating. Fig. 1*f* and 1*g* both contain continuous lines, although the doublet, which is "sensitive" to a small alteration in the intra-atomic distance (for the explanation of this see article VIII of this series), is definitely separated in g, as compared to the original state in f, where it consists of a single ill-defined line. The latter is due to the fact that in the drawn wire the lattice

[2]) To obtain the radiographs a length of wire of about 1 mm was irradiated, i.e. about four to five times the length shown in the polished sections. The points in the radiographs are therefore due to a larger number of crystals than shown in the polished sections.

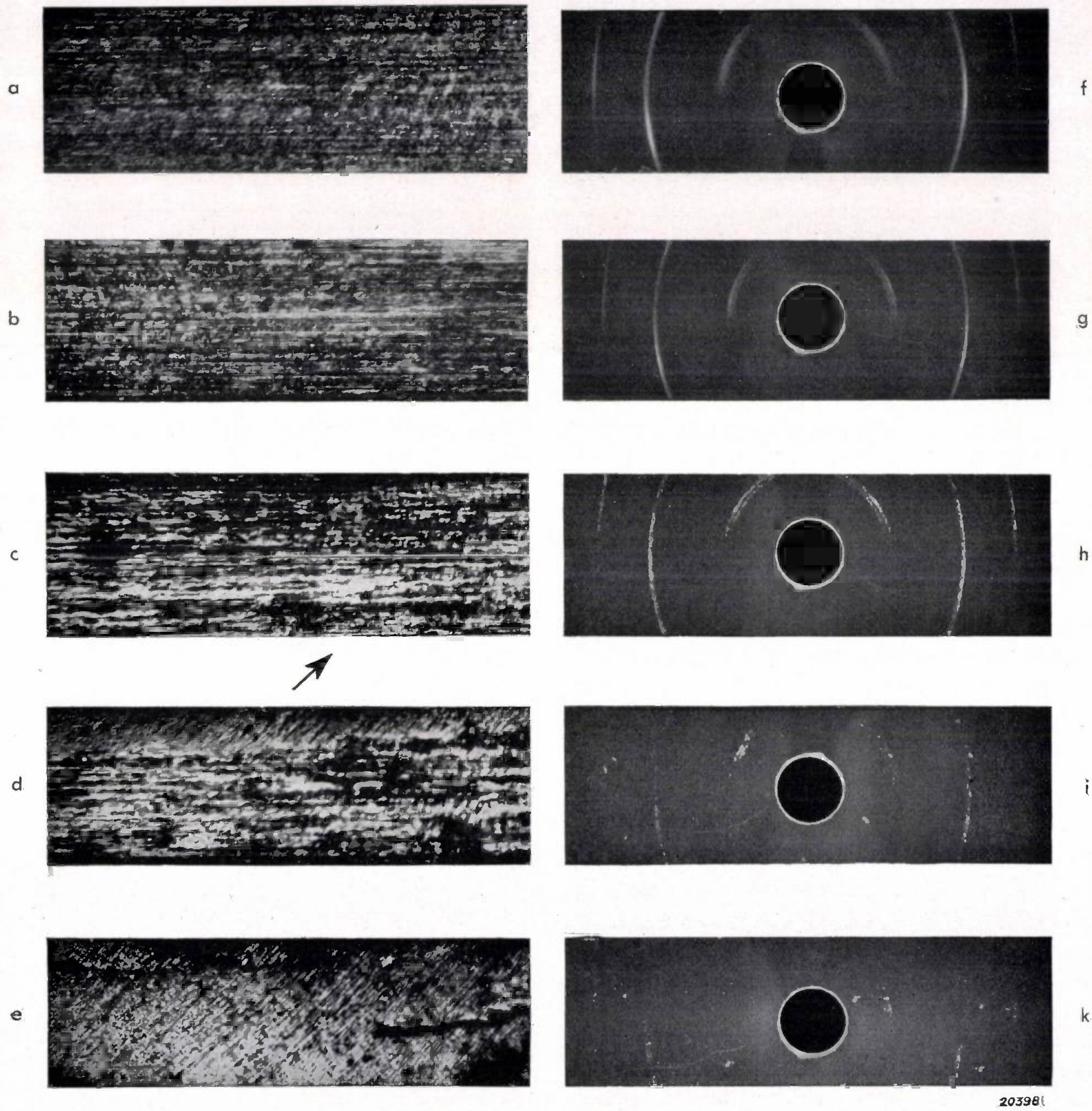
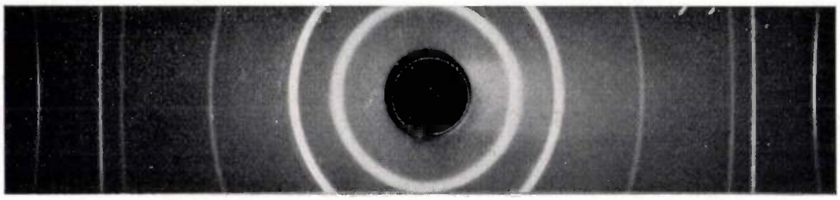


Fig. 1. Alteration in the crystalline texture of a drawn tungsten wire on heating to different temperatures in sequence.
a to *e*: Etched figures (magnification 325).
f to *k*: Radiographs.
The recrystallation taking place in the wire is indicated in the etched figures by the fibrous structure (figs. *a* and *b*) becoming converted to newly-formed crystals (fig. *d* above and fig. *e*) as revealed by fine parallel striations. In the radiographs recrystallisation is revealed by the initially continuous interference lines becoming resolved into separated spots. At the lowest reheating temperatures at which the fibrous structure is still practically unaltered (cf. fig. *a* with fig. *g*), the radiographs show a sharpened definition of the interference lines (cf. fig. *f* with fig. *g*) from which a diminution of the internal stresses may be concluded.

of the crystals has become somewhat distorted by the drawing process and thus gives rise to internal micro-stresses (see example **18** in article VIII). The first effect of comparatively small temperature rise is thus seen to consist of a decrease in these deformations of the lattice and hence of the internal stresses. Crystal growth susceptible to detection by X-rays has however not yet taken barely observable in the polished sections (cf. *a* with *b*), the radiographs reveal (cf. *f* with *g*) that the stresses created by the drawing process have already diminished. Furthermore, it may be deduced from the definition of the interference spots in the radiographs of the crystallising wires that the freshly-formed crystals are to a large extent free from stresses, which also cannot be directly con-

a



b

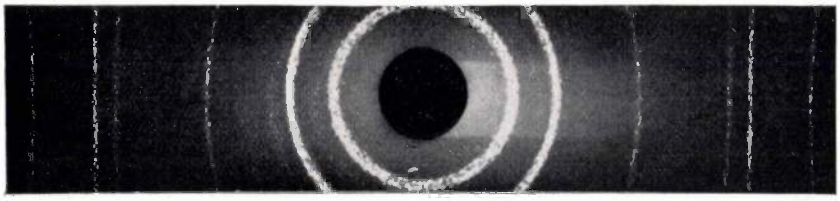


Fig. 2. Occurrence of recrystallisation on heat treatment of nickel cathode tubes.
a: Radiograph of drawn tube.
b: Radiograph of recrystallised tube.

place. This is brought out by the fact that the lines in the radiograph have retained their continuous character.

In the piece of wire which was reheated to the next highest temperature (fig. 1*h*), the lines begin to be resolved into individual points. This phenomenon indicates that in the wire section irradiated by the X-rays crystals were already present of at least about 10μ in size, i.e. nuclear formation had been initiated, and the wire at least partially was in a state of incipient recrystallisation. With the progressive development of the recrystallisation process and the simultaneous disappearance of the original fibrous texture, the continuous character of the interference lines disappears completely (figs. 1*i* and 1*k*), and only a few spots remain which emanate from the relatively small number of newly-formed crystallites.

The value of radiographic examination in regard to polished sections is that each method of investigation supplements the other. While, for instance, in reheating the wire to the lowest temperature employed structural changes were not at all or cluded from the polished sections. On the other hand the polished sections aptly bring out the shape and size of the crystals finally formed.

21. Nickel Tube for Oxide Cathodes

In indirectly-heated receiving valves with oxide cathodes, the oxide coating is deposited on short drawn nickel tubes which have been subjected to a special heat treatment, Recrystallisation may take place during their heat treatment, and is revealed in the radiograph, as in the previous example, by the continuous lines obtained with the drawn tube becoming resolved into individual spots.

On the examination by X-rays of two series of nickel tubes of different origin the difference was found that one series (cf. fig. 2*a*) has not yet recrystallised while in the other series recrystallisation has occurred (cf. fig. 2*b*). Moreover, from the density of the spots in radiographs of different recrystallised tubes certain conclusions could be drawn as regards the more or less advanced stage of the recrystallisation process.

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

No. 1122: M. J. Druyvesteyn: Die Brennspannung eines Niedervoltbogens (Physica, 3, 724 to 726, July, 1936).

The arc voltage of a low-voltage arc is measured as a function of the current intensity and the gas pressure, in the course of which distinct differences are observed between a tungsten and an oxide cathode. Partly these differences are probably due to differences in temperature of the gas just in front of the cathode, as a result of which, at the same gas pressure, the density in front of the oxide cathode is made much greater than in front of the tungsten cathode.

No. 1123: W. de Groot and J. H. Gisolf: De Geiger-Müller teller (Ned. T. Natuurk. 3, 161-169, June, 1936).

This paper which was read before the Nederl. Natuurk. Ver. presents a survey of the literature dealing with the Geiger-Muller counter, followed by a report of observations made by the authors themselves on the discharge of the counter by means of a cathode ray oscillograph. The different forms of the wire potential as a function of the time which were found may be traced to the different values possessed by unavoidable capacities and resistances, as was illustrated by means of a substitutional circuit. It was shown in particular that the occasional instantaneous values of the wire potential, which are greater than to be expected from the corona characteristic, are in general not due to the potential between the electrodes falling below the "initial potential".

No. 1124: A. Bouwers and W. J. Oosterkamp: Die Unschärfe einer Röntgenaufnahme (Fortschr. Röntgenstr. 54, 81-91, July, 1936).

The best definition is obtained in radiographs when the kinematic want-of-sharpness U_b, the screen want-of-sharpness U_f and the geometrical want-of-sharpness U_g are almost equal to each other, i.e. $U_b = U_f = U_g$. It is just as incorrect to use too short an exposure as to work with too small a focal point or with intensifying screens with an excessive high-definition.

No. 1125: E. J. W. Verwey and J. H. de Boer. Surface oxide films. (Rec. Trav. chim. Pays-Bas, 55, 675-687, July, 1936).

The electrokinetic potential of rare metals and of carbon may be positive or negative according to the conditions ruling. The adsorption of oxygen at the surface of a pure metal imparts a positive potential to the latter with reference to the solution. In most cases the potential in the outermost part of the double layer is however of opposite sign, owing to the existence of a thin oxide film at the surface, i.e. a monatomic layer of oxygen which reacts chemically with the metal or the carbon, causing the production of a compound with a pronounced dipole moment. With carbon the activation energy for this reaction is so great that the surface oxide is formed only very slowly at room temperature, so that the carbon still has an electrokinetical positive charge. The nature of this activation energy and the manner in which the potential in the double layer varies are discussed in detail.

No. 1126: T. Jurriaanse and M. J. Druyvesteyn: The transition from a glow discharge to an arc discharge (Physica, 3, 825-840, August, 1936).

The probability of the transition of a glow discharge to an arc discharge on contact between the cathode and a conductor is investigated. The probability of this transition is determined by a large number of factors, *inter alia* by the capacity and the potential difference between the conductor and the cathode; by the current density and the cathode fall in the glow discharge. Transition may occur equally on making or breaking contact between the conductor and the cathode. A determining factor in the probability of this transition to an arc is the energy of discharge between the conductor and the cathode.

No. 1127: K. F. Niessen: Zur Theorie der Brownschen Bewegung eines kritisch gedämpften Galvanometers (Physica 3, 841-854, August, 1936).

A theoretical method is evolved for determining the Boltzmann constant k from the Brownian movement of a critically damped galvanometer. The method described differs fundamentally from the standard method in that the result is obtained from the time function of the mean value under measurement.

VOL. 2 No. 2 FEBRUARY 1937

Philips Technical Review

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

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

A TELEVISION RECEIVER

By C. L. RICHARDS.

Summary. Description of a television receiver designed for two scanning systems, viz., for 240 lines standard scanning and for 405 lines interlaced scanning.

Introduction

The design of a television receiver is determined to a large extent by the principle of operation of the transmitter from which radiated pictures are to be picked up, particularly by the method of scanning used and the type and form of the synchronising signals employed at the transmitter.

The television receiver described here was designed for picking up the programmes of the B.B.C. transmitter in London, as well as the transmissions from the experimental television transmitter installed at the Philips Laboratory. The London station commenced transmitting television by two different scanning methods: one with 25 pictures per second, 240 lines per picture and an aspect or picture ratio of 3 : 4, and a second with interlaced scanning with 2 × 25 pictures per second, $202^1/_2$ lines per picture and an aspect ratio of 4 : 5 [1]). The Philips transmitter can be adjusted to various scanning standards from 25 pictures with 90 lines per picture to 2 × 25 pictures with $202^1/_2$ lines per picture (interlaced scanning).

Since, in these transmitters, the method of modulating the picture and synchronising signals on the carrier wave is substantially the same, the same receiver can be employed for picking up both transmitters. Two different scanning systems are in fact provided, and each can be regulated independently, so that by changing over from one system to the other a wide range of scanning speeds can be covered. In view of the very high modulation frequencies entailed in scanning a picture resolved into 405 lines, the receiver has been designed for a maximum modulation frequency of 2.5 megacycles.

[1]) Since this article was written the 240 lines standard transmission has been discontinued in London.

Type of Signals

Vision and associated sound are transmitted by modulating each on its own carrier wave. In the case of the London transmitter, vision is for instance transmitted on a carrier wave of 45 megacycles (6.67 metres) and sound modulated on a carrier wave of 41.5 megacycles (7.23 metres). The method of modulating vision and the associated synchronising signal is shown in *fig. 1*. The ordinate gives the amplitudes of the high-frequency oscillations as a percentage of the maximum amplitude. If no picture signal is received, i.e. the picture screen is perfectly dark, the modulation of the high-frequency oscillations is approximately 30 per cent. During scanning of a picture line along which the brightness fluctuates from bright to dark, the carrier wave is modulated to varying degrees between 30 and 100 per cent. At the end of each line the carrier wave is completely suppressed for a short interval (zero modulation), and this decrease in modulation from 30 per cent to zero

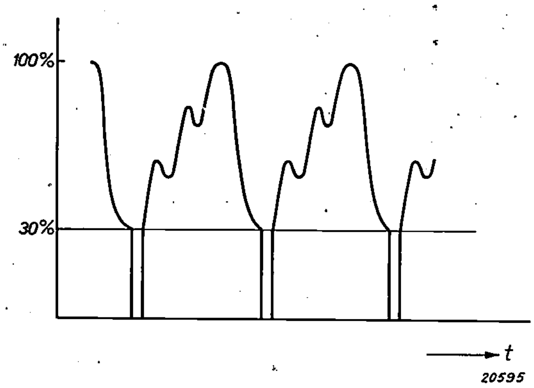


Fig. 1. Variation of amplitude of high-frequency oscillations in a wave modulated with vision signals.

constitutes the line synchronising signal. Similarly at the end of each picture, a picture synchronising signal is produced, which is not shown in fig. 1.

In interlaced scanning, the synchronising signals are not produced in the same simple manner, but lack of space precludes this point being discussed in detail here.

plification must not alter the mutual positions of the oscillations with different frequencies from which the picture signal is built up, since the phase of an oscillation has a marked effect on the distribution of the bright and dark elements produced on the screen. On the other hand, television reception is simpler than sound reception inasmuch as pro-

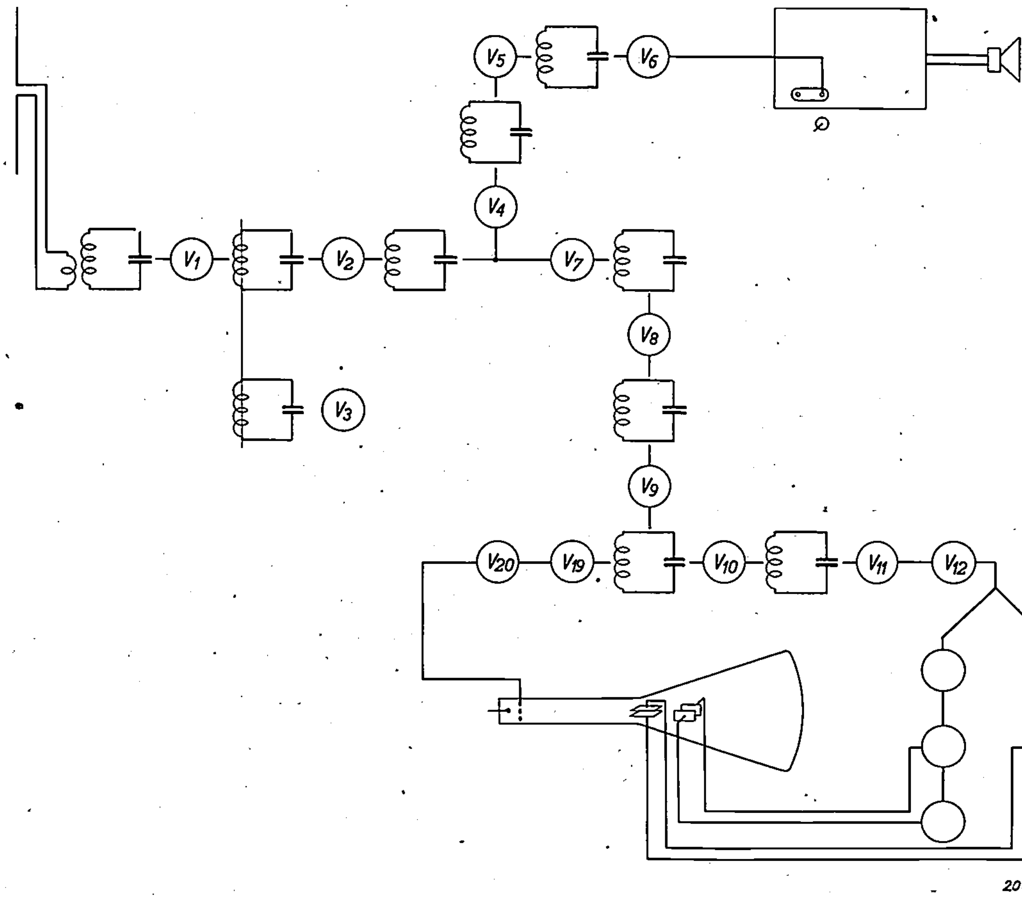


Fig. 2. Circuit diagram of a receiver set for television with associated sound.
V_1 Ultra-short wave amplifying valve for vision with sound.
V_2 Mixing valve with oscillator V_3.
V_4 to V_6 Amplification and rectification of sound signals.
V_7 to V_9 Amplification of vision and synchronising signals.
V_{19} and V_{20} Transmission of vision signals to the control electrode of the cathode-ray tube.
V_{10}, V_{11}, V_{12} Amplification of synchronising signals and transmission to the corresponding saw-tooth wave generators for deflecting the cathode ray.

Desiderata of the Television Receiver

A television receiver differs fundamentally from an ordinary radio receiver in that for the reception of television pictures a receiver must give a uniform amplification over a very wide frequency range. It is therefore not possible to use circuits with a high selectivity, so that there is a marked reduction in the amplification which can be obtained per stage. Furthermore in receiving vision — as opposed to the reception of sound — any phase displacement which may be obtained during amplification must not alter the mutual positions

nounced non-linear distortion is permissible. As a result, the carrier wave and one of the side bands are, for instance, quite sufficient for good reception. This offers the important advantage that the frequency band to be treated is only half as wide, so that the amplification in the individual stages is practically doubled.

Circuit Arrangements of the Receiver

In a television receiver the wave on which the picture is modulated must be picked up, amplified and rectified; the vision signal obtained in this way must then modulate the voltage at the control electrode of a cathode-ray tube in such a way that the scanning spot on the fluorescent screen always has the same intensity as that of the corresponding point in the original picture. In addition, a scanning arrangement consisting of two saw-tooth generators is required, so that the scanning spot will cover the whole surface of the screen; a device is also needed responding to the synchronising signals picked up and thus synchronise the operation of the scanning device with that of the transmitter. Finally, the sound which is modulated on another carrier wave must also be picked up, amplified, rectified and finally passed to a loudspeaker.

In the receiver described here, a television cathode-ray tube is used giving a picture measuring 25 × 18 cm, almost white in colour and sufficiently bright for viewing in a dimly-lit room. The tube operates on an anode tension of 5000 volts. The cathode ray is deflected in both directions by electrical means through the agency of two pairs of deflecting plates; the requisite deflection voltage being approximately 1000 volts. The modulation voltage at the control electrode of the tube is approximately 30 volts.

A simplified circuit of the receiver is shown in *fig. 2*. The two carrier waves for vision and sound are picked up with the same aerial and both amplified in the same high-frequency stage by an amplifying valve V_1; the frequency characteristic of this stage must naturally be sufficiently broad for this purpose. In the top right hand corner of *fig. 3* is shown the frequency range occupied by the carrier wave and the side bands carrying the vision signals, while in the left hand corner is shown the much narrower range of the other carrier wave and the sound signals modulated on it. Curve *1* in fig. 3 shows the frequency characteristic of the high-frequency amplifying stage. The high-frequency circuits in front of and following the amplifying valve V_1 can be finally adjusted by means of small trimming condensers, thus permitting these circuits to be tuned to both the B.B.C. transmitter and the Philips transmitter.

To separate the picture and sound signals the heterodyne principle employed in radio receivers has been introduced. The two signals are passed to the mixer valve V_2 together with an auxiliary frequency generated by the oscillator V_3.

Two new carrier waves are produced in the mixer valve, each with a lower frequency than the two primary carriers. In conformity with the nomenclature employed in heterodyne radio receivers, these new waves are termed intermediate-frequency carrier waves, although their frequencies are

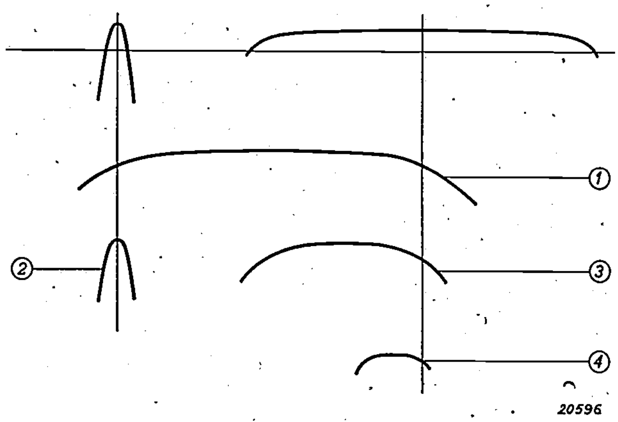


Fig. 3. Diagrammatic representation of frequency bands. Above: Radiated signals for vision and sound.
1 Characteristic of stages V_1 and V_2.
2 Characteristic of stages V_4 to V_6 for amplification of sound signals.
3 Characteristic of stages V_7 to V_9 for amplification of vision and synchronising signals.
4 Characteristic of stages V_{10} to V_{12} for amplification of synchronising signals.

situated in an entirely different band (e.g. 10 and 7·5 megacycles) than is the case in radio reception.

Vision is modulated on one intermediate-frequency carrier wave and sound on the other. The two carriers are passed together to the grids of two intermediate-frequency amplifying valves V_4 and V_7.

The anode circuit of V_4 is tuned to the sound intermediate frequency and has such a narrow frequency characteristic that the vision signals are no longer able to pass. In fig. 3 this characteristic is represented by curve *2*. In this connection it should be remembered that behind the mixer valve the signals are transmitted on an entirely different carrier; since fig. 3 is intended merely to give a general indication of the mutual positions of the frequency characteristics in the different stages, this displacement of the carriers has not been taken into consideration in this figure. The intermediate-frequency carrier modulated with vision is further amplified in V_5 in the usual way and rectified in the diode V_6; the low-frequency signal so obtained is applied to the gramophone pick-up sockets of a standard radio receiver equipped with loudspeaker and volume control.

As already indicated, the two intermediate frequency carriers are also passed to the grid of the amplifying valve V_7. The circuit following this valve, as well as the two succeeding amplifying stages with the valves V_8 and V_9, together furnish a frequency characteristic of the type shown in curve *3* in fig. 3, so that in the anode circuit of the valve V_9 there remain only the picture and synchronising signals. Behind V_9 a separation again takes place. Passing to the left the modulated intermediate-frequency waves reach the valve V_{19}, which functions as an anode rectifier. Following rectification picture signals are obtained which can also be termed picture-frequency signals; they are amplified by V_{20} and finally passed to the control electrode of the cathode-ray tube.

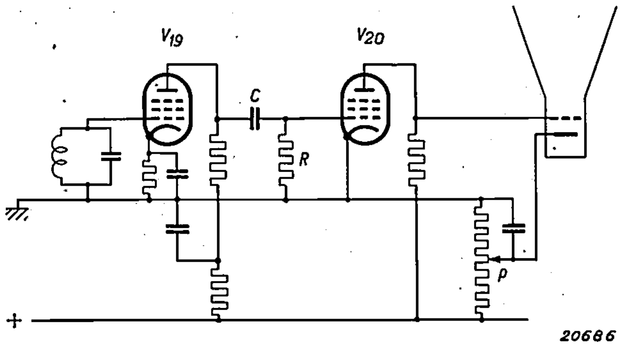


Fig. 4. Circuits of valves V_{19} and V_{20} which apply the vision signal to the control electrode of the cathode-ray tube.

Since this part of the circuit requires closer attention, it has been reproduced separately in *fig. 4.* The picture-frequency signals furnished by the anode rectifier V_{19} are passed through a condenser C to the grid of V_{20}; this grid is connected to the cathode through a high resistance R; as long as no signal is received the grid is at zero potential with respect to the cathode. If now the picture-frequency signal is applied to the grid, grid current will flow and produce a voltage drop in the resistance R, as a result of which the mean grid bias becomes negative. This is shown in *fig. 5a;* the grid voltage adjusts itself in such a way that the peaks on the right, which in this case correspond to the base of the synchronising signal, just fall within the region in which grid current flows. For each line scanned the vision signal is of different form and in general exhibits marked fluctuations in modulation depth, as shown e.g. for two lines in fig. 5*a*. On the other hand, the synchronising signals always have the same modulation depth, so that the broken lines marked *1* and *2* will always occupy the same positions. In consequence, the values V_1 and V_2 (fig. 5*b*), of the voltage at the anode of V_{20}, corresponding to the grid voltages *1* and *2* will also always be the same irrespective of the modulation of the vision signals. The control electrode of the cathode-ray tube is connected directly to the anode of V_{20}. By applying a suitable negative tension (adjusted with the potentiometer P) between the control electrode and the cathode, the picture-frequency signal can be applied to the control electrode of the cathode-ray tube in the manner shown in fig. 5*c*; it is seen that the line DE corresponding to the perfectly dark picture occupies such a position that current just fails to pass through the cathode-ray tube, while modulation to the right of DE allows a beam current to pass, which is approximately proportional to the modulation and thus increases and diminishes with the brightness of the radiated picture.

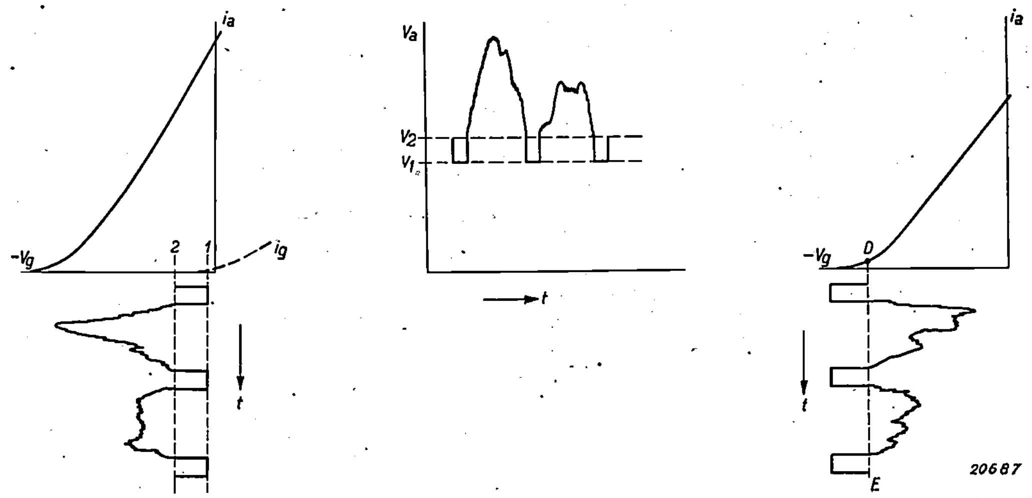


Fig 5 *a* Characteristic and applied grid voltage of valve V_{20}. *b* Anode volts of valve V_{20}. *c* Voltage at control electrode of cathode-ray tube.

Reference will be made to some further details of the receiver circuit. Correct tuning is obtained by adjusting the oscillating circuit of the oscillator, shown next to V_3 in the circuit diagram, in such a way that the loudspeaker reproduces the sound picked up, i.e. tuning is done on the sound and not on vision, since the circuits in the receiver have much sharper resonance curves than those of the television receiver. Automatic volume controll is provided in the sound intermediate-frequency amplifier to prevent overloading and also to minimise variations in the loudspeaker sound output, occurring as a result of a slight drift in the oscillator frequency.

It is evident from curves *1* and *3* in fig. 3 that only one of the two side bands of the carrier on which the vision signals are modulated is amplified. The suppression of the other side band does not adversely affect the quality of the picture, as has already been pointed out above.

Synchronising Signals

In discussing the receiver circuits in reference to fig. 2 it was already mentioned that an intermediate-frequency carrier on which the picture and synchronising signals are modulated is obtained in the circuit behind V_9. The carrier wave is amplified by the valve V_{10}, and the circuit following this valve has a transmission characteristic of the type shown in curve *4* in fig. 3, V_{11} is a triode with incorporated diode *(fig. 6)*. The incoming signal

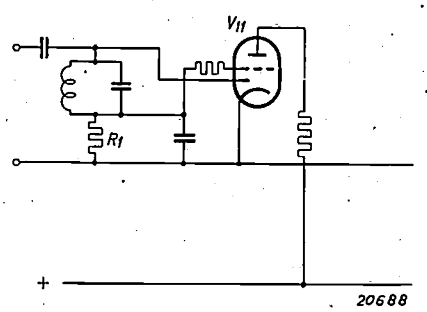


Fig. 6. Circuits of the diode-triode V_{11}. The short line represents the diode anode. In this stage the synchronising signals are separated from the vision signals.

is rectified in the diode giving a picture-frequency signal at the resistance R_1. As shown in the figure the signal is also applied at the same time to the grid of the triode. The characteristic of the triode is so chosen *(fig. 7)* that the amplitudes of the vision signals all fall within those grid voltages at which no anode current flows, while the opposite is the case for the synchronising signals. The anode current of the triode is thus of the form shown in fig. 7*b*; it is seen that the synchronising signals are almost completely separated from the vision signals proper.

The circuit shown in fig. 6 is followed by the stage V_{12}, which serves for the still more complete separation of the synchronising signals from the vision signals. Following V_{12} are filters in which the line and picture synchronising signals can be separated from each other on the basis of their

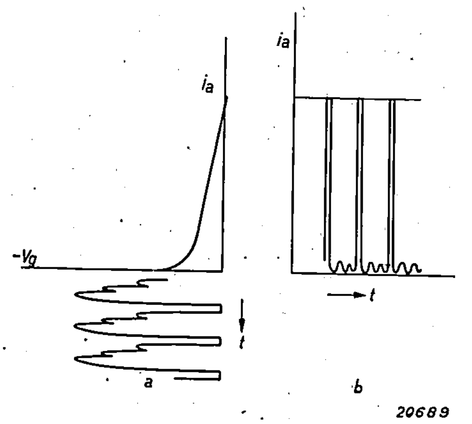


Fig. 7. *a* Characteristic and applied grid voltage of the diode-triode valve V_{11}.
b Anode current of triode section of V_{11}.

frequency difference; finally, each is applied to one of the saw-tooth wave generators serving for deflecting the cathode ray.

Saw-tooth Wave Generators

In the circuit diagram the saw-tooth wave generators are represented by two groups each composed of three circuits (not specifically indicated) on the right of the cathode-ray tube. Each of the two generators contains a condenser which is charged through a resistance and discharged by a valve (top circuit). The saw-tooth voltage of the condenser is amplified by the two amplifying valves following it, and which apply the scanning voltage to the deflection plates of the cathode-ray tube. The circuit containing the two amplifying valves is so arranged that the saw-tooth generators automatically continue in operation, even in the absence of synchronising signals. Means are also provided for improving the linearity of the saw-tooth waves. Each saw-tooth generator is actually duplicated in order to be able to employ different scanning methods, a point already indicated at the outset.

Assembly of Components

The cathode-ray tube *(f in fig. 8)* is mounted horizontally in a chassis *a*, on which the various amplifying stages and the saw-tooth generators are also fixed. The tube is inserted from the front, after removing a loose front panel on the cabinet. Behind this panel, those controls are also mounted which normally require a single adjustment only and need no further attention during normal operation of the apparatus.

Four control knobs serve for the adjustment of the television receiver (in the actual apparatus these are combined to a pair of twin knobs *h* and *l*), and with which the following adjustments can be made: 1. Adjustment of the amplification of the vision amplifier, permitting a variation of the amplitude of the picture-frequency signal at the control electrode of the cathode-ray tube (see fig. 5*c*) and thus increasing or reducing the brightness of the picture; 2. adjustment of the potentiometer *P* in fig. 4, permitting the line *DE* (fig. 5*c*) corresponding to a dark screen to be brought into the correct position, viz., just at the current cut-off point of the cathode-ray tube; 3: tuning of the oscillating circuit; 4: changing over from one scanning system to the other.

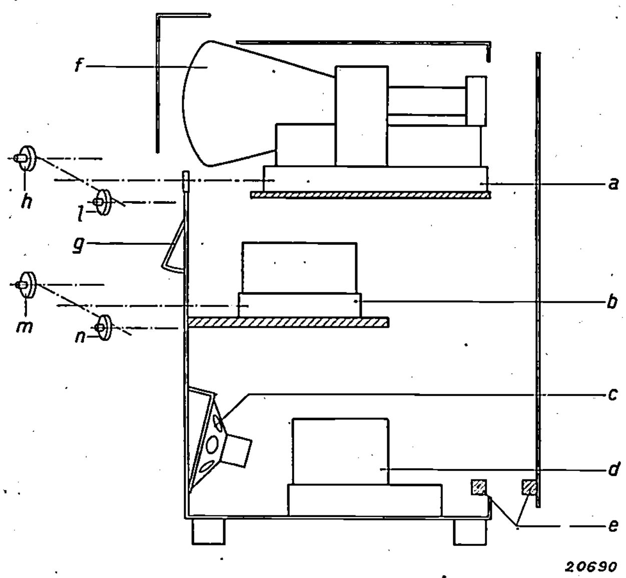


Fig. 8. Diagrammatic section through the receiving set. *a* Vision receiver. *b* Broadcast receiver. *c* Loudspeaker. *d* Anode current unit *e* Safety contact. *f* Cathode-ray tube. *g* Tuning scale of radio receiver.

Vision controls: *h* Picture adjustment. *l* Tuning and selecting of scanning method.

Radio controls: *m* Volume and band width. *n* Tuning and wave range; on and off switch; changeover for radio or television reception.

The chassis *b* for a broadcast receiver with the control knobs *m* and *n*, which can also be used for normal radio reception, is mounted lower in the cabinet. The television receiver is switched on in the fourth position of the wave switch, and the associated sound is then reproduced through the pick-up connections of the sound receiver.

The loudspeaker and the various anode voltage units are located is the bottom of the cabinet. All high-tension leads are provided with metallic screening throughout the set; protection against high-tension shocks is also provided, the mains voltage being cut off and the high-tension condensers discharged when the cabinet is opened.

THE REPRESENTATION OF COLOUR SENSATIONS IN A COLOUR SPACE-DIAGRAM OR COLOUR TRIANGLE

By P. J. BOUMA.

Summary. On the basis of the experimental results obtained on colour mixing, it is demonstrated how all colour sensations can be represented in a colour space-diagram and how various transformations can be applied to this space-diagram. In many cases such a space-diagram can be replaced by a suitable colour triangle. Various characteristics of these triangles, particularly the I.C.I. 1931 standard triangle (fig. 4), are discussed, and a series of formulae given for the computation of colour mixtures.

Introduction

Considering the practical importance which has been recently acquired by the problem of characterising colour sensations produced by modern sources of coloured light, it appeared to us desirable to expand our purely qualitative analysis of colour perception given in a previous paper [1]) to a more precise quantitative discussion of the method for characterising these colour sensations and to describe the experimental principles of this method.

A colour sensation is produced when a specific type of light stimulates the retina of the eye. Colour sensations produced by different means, (e.g. by mechanical agents) will not be discussed here. The nature of the colour sensation is determined by:

1) The physical composition (spectral distribution) of the incident light, and
2) The characteristics of the eye.

Construction of a Colour Space-Diagram

It is assumed that we are dealing with a "standard" and unstrained eye, which does not receive simultaneously widely-different colour stimuli. The effects of fatigue, after-images, simultaneous and successive contrast are considered to be absent. It is also assumed that the brightness level is neither so low that the Purkinje phenomenon [2]) is obtained, nor so high that a condition of glare results.

In these circumstances the well-known rule of mixture applies:

If three arbitrary relative spectral distributions (i.e. three types of light) are selected, each colour sensation can be created by suitably mixing these three "primary colours". The required result can be obtained in each case only by a specific ratio of the primary colours.

To ensure the general validity of this law we must conform with the two following new conditions:

1) The three primary colours must be selected in such a way that it is not possible to produce any one of the colours by a suitable mixture of the other two.
2) In each selection of the fundamental colours, there are always colours K present which cannot be produced in the manner indicated; in these cases it is however always possible — and actually in one way only — to select a mixture of K with one or two primary colours and to obtain an identical colour sensation by combining the two other primary colours (or by taking a specific amount of the third primary colour). The fact that a definite quantity B_1 of a certain primary colour is added to K may be expressed as follows: "We add the quantity $-B_1$ of this primary colour to the simulative mixture of the primary colours"; hence the rule of mixtures assumes general validity in that *negative* quantities of the primary colours can also be dealt with.

Since with a specific choice of the primary colours a certain colour sensation (brightness B) can be simulated in one and only one way by mixing the three primary colours with brightnesses B_1, B_2 and B_3, this colour sensation is completely characterised by the three quantities: B_1, B_2, B_3. It may be recalled that these three numerals determine merely the colour sensation and do not express the spectral composition of the light.

The totality of the colour sensations which can be produced is therefore three-dimensional, in other words if we employ a three-dimensional diagram [3]) with the co-ordinates x_1, x_2, x_3, each colour sensation can be represented by a point in

[1]) Philips techn. Rev. 1, 283, 1936.

[2]) Philips techn. Rev. 1, 102, 142, 166, 1936.

[3]) It is immaterial whether the systems of axes are rectangular or oblique; in figs. 1 and 2 they have been taken as oblique.

this diagram, when we take $x_1 = B_1$, $x_2 = B_2$, and $x_3 = B_3$. To each colour sensation there thus corresponds one point, and to each point no more than one colour sensation.

Characteristics of the Colour-Space-Diagram

To obtain some idea of the position of colour sensations in this space diagram, attention must be called to a few laws which with the restrictions outlined above (standard unstrained eye, brightness above the limiting value of approximately 3 candles per sq.m, etc.,) are of general validity:

1) On a proportional change in energy two "physiologically equivalent" colours remain physiologically equivalent, i.e. if two different illuminants produce the same colour sensation, this equivalence is retained when the energy of each illuminant (and hence also the brightnesses) are multiplied or reduced by the same factor (thus leaving the relative spectral distributions unaltered).
2) Physiologically equivalent colours behave similarly in any mixing process, and they can therefore be substituted for each other in any mixture [4]).
3) The brightness of a mixture is equal to the sum of the brightness values of the components.

It follows from the first law that on an alteration in the energy of an illuminant (with constant spectral composition) displacement takes place along a line through the origin in the (x_1, x_2, x_3) space diagram.

For if B (colour k) $\sim B_1$ (colour 1) $+ B_2$ (colour 2) $+ B_3$ (colour 3) [5]) we have according to this law:

nB (colour $k \sim n B_1$ (colour 1) $+ n B_2$ (colour 2) $+ n B_3$ (colour 3).

The brightness B (colour k) thus has the co-ordinates $x_1 = B_1$, $x_2 = B_2$, $x_3 = B_3$, and the brightness nB (colour k) the co-ordinates $x_1' = n B_1$, $x_2' = n B_2$, $x_3' = n B_3$. Displacement is thus along a straight line through the origin.

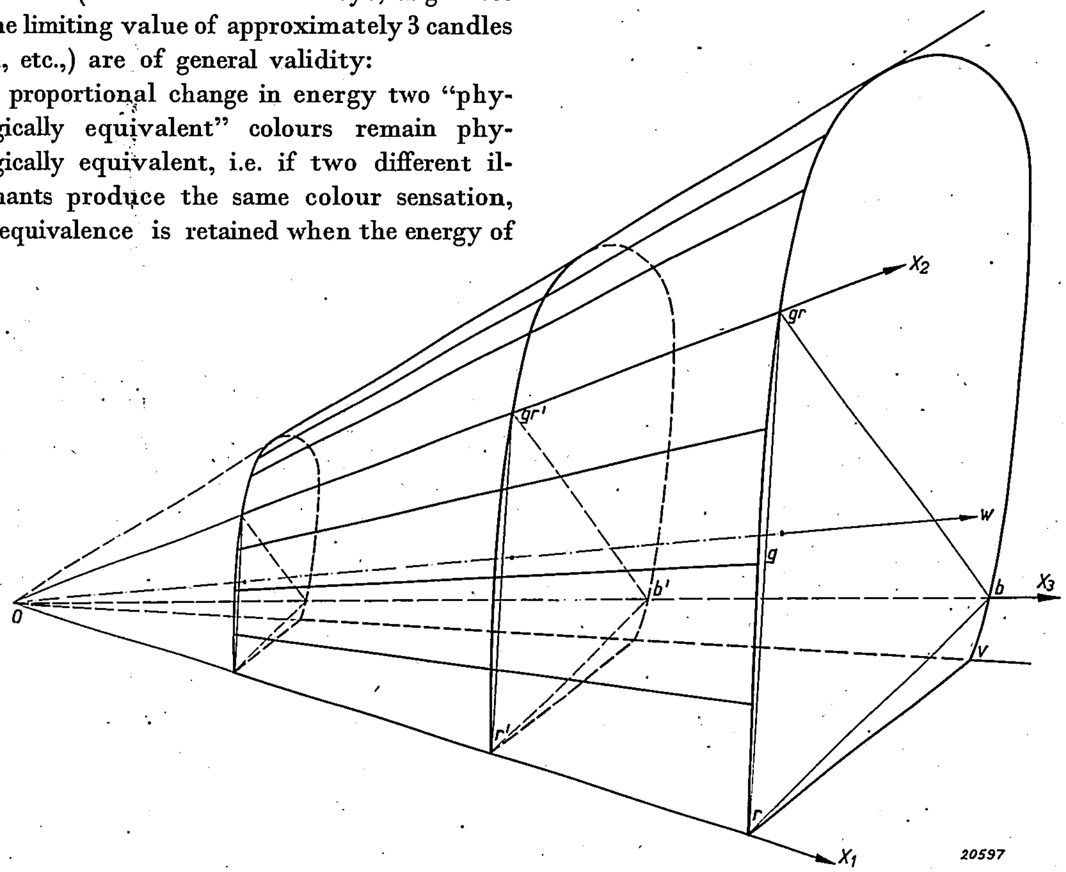


Fig. 1. Diagrammatic representation of a colour space-diagram with red, green and blue spectral colours as primary colours. The spectral colours lie on the surface of a cone, the saturated purple colours in a plane O-r-v, and all other colours within this cone. Colour triangles are obtained in the planes $x_1 + x_2 + x_3$. The line O-w represents white; the cone is truncated on the left for taking account of the Purkinje phenomenon.

In particular a line through the origin corresponds to every spectral colour. These lines lie on the surface of a cone which is shown diagrammatically in *fig. 1*, where the primary colours selected are at 4358, 5461 and 7000 Å. These are the three spectral colours which Guild and Wright employed for their calibration of the spectrum. The second is the blue mercury line and the third

[4]) It should be carefully noted that this rule does not hold in regard to the mixing of dyes.

[5]) The symbol $\sim$ is used to indicate "physiological equivalence".

the blue mercury line. The letters r, g, gr, b and v represent approximately the positions of the spectral red, yellow, green, blue and violet. r and v correspond to the ends of the spectrum. The cone is truncated to the left in accordance with the above restriction regarding the brightness level.

From the second law it follows that a mixture of colours in the space diagram can be represented by vectorial addition.

For if:

B (colour k) $\sim B_1$ (colour 1) $+ B_2$ (colour 2) $+ B_3$ (colour 3)

and:

B' (colour k') $\sim B_1'$ (colour 1) $+ B_2'$ (colour 2) $+ B_3'$ (colour 3)

then according to this law we have:

B (colour 1) $+ B'$ (colour k') $\sim (B_1 + B_1')$ (colour 1) $+ (B_2 + B_2')$ (colour 2) $+ (B_3 + B_3')$ (colour 3).

The following co-ordinates therefore apply to the colour sensations B (colour k) and B' (colour k'):

$x_1 = B_1, x_2 = B_2, x_3 = B_3$ and $x_1' = B_1', x_2', = B_2', x_3' = B_3'$

and to the mixture of these two colours:

$x_1'' = B_1 + B_1', x'' = B_2 + B_2', x_3'' = B_3 + B_3'$.

The vector (x_1'', x_2'', x_3'') is thus actually obtained by adding the vectors (x_1, x_2, x_3) and (x_1', x_2', x_3').

We also see why the primary colours may not be selected in such a way that one of them can be produced by a mixture of the other two: Such a combination of three primary colours would be shown in fig. 1 by three coplanar vectors, and with these primary colours only the colour sensations laying in this plane could be obtained. It follows furthermore from the rule of mixtures deduced above that:

1) The plane O-v-r contains the mixtures of spectral red and spectral violet (saturated purples).
2) All mixtures of spectral colours (i.e. all colour sensations) lie within the surface of the cone formed by the spectral colours and the plane O-v-r. In other words: Each point within this conical surface represents a specific colour sensation; a point outside this surface represents no colour sensation at all and has therefore only a geometrical significance. The line O-w represents white light.

From the third law we get the brightness equation:

$$B = B_1 + B_2 + B_3 \quad \text{or} \quad B = x_1 + x_2 + x_3. \quad (1)$$

In fig. 1, r-gr-b and r'-gr'-b' are planes which cut off equal lengths from the three axes; their equation is therefore $x_1 + x_2 + x_3 =$ constant, and these planes are thus planes of constant brightness.

Transformations of the Colour Space-Diagram

Having represented the various colours in this way we can pass in different ways from this system to other systems, in which a colour is determined by three other numerals, i.e. we can transform the colour space-diagram obtained in various ways.

The most important transformation is the homogeneous linear transformation, which consists in substituting the co-ordinates (x_1', x_2', x_3') for the co-ordinates (x_1, x_2, x_3), the relationships between the two sets of co-ordinates being as follows:

$$\left.\begin{aligned} x_1 &= a_1 x_1' + a_2 x_2' + a_3 x_3' \\ x_2 &= b_1 x_1' + b_2 x_2' + b_3 x_3' \\ x_3 &= c_1 x_1' + c_2 x_2' + c_3 x_3' \end{aligned}\right\} \quad . \; . \; . \quad (2)$$

This substitution corresponds to a geometrical rotation of the axes with simultaneous and different contraction or elongation in the directions of the three axes. In many cases this can be interpreted physically by a different choice of primary colours coupled with a choice of different brightness units for the three different primary colours. But this interpretation is only applicable if the new axes lie within the cone in fig. 1. If they are outside it (and transformations of this type are frequently employed in practice) transformation (2) loses its simple physical meaning. It is advisable therefore to regard this transformation always as a pure mathematical artifice [6]). The purpose of these transformations is purely of a practical nature, for by a suitable choice of the coefficients in (2) different purposes can be served, e.g.:

1) The white line O-w can be brought into a specific position with respect to the axes, e.g. so that it can be represented by the equation $x' = x_2' = x_3'$;
2) The brightness equation which after transformation is:
$$B = (a_1 + b_1 + c_1)\, x_1' + (a_2 + b_2 + c_2)\, x_2' + (a_3 + b_3 + c_3)\, x_3';$$
can be given a specific form, e.g.
$$B = x_2';$$
3) That all colour sensations shall have positive co-ordinates exclusively; this cannot be done if the three axes are all within the cone of fig. 1 or lie along its surface. These axes can therefore no longer be regarded as representing colours.

[6]) We prefer this interpretation to the introduction of the rather problematic and obscure concept of "imaginary primary colours". A line outside the cone does not represent an imaginary colour but it represents no colour at all.

Fig. 2 shows an example of a colour space-diagram which has been obtained by a transformation of this type and which has the three above-mentioned characteristics. The planes $x_1' + x_2' + x_3' =$ constant shown are now no longer planes of constant brightness, such planes then being all parallel to the plane O-x_1'-x_3' [7]).

Construction of Colour Triangles

Since all colours which lie in a straight line through the origin appear to be of practically the same character to the eye and only differ in brightness, it is frequently desirable to replace such a line by a point. This can be done by the following transformation:

$$\left.\begin{aligned} t_1 &= \frac{x_1}{x_1 + x_2 + x_3} \\ t_2 &= \frac{x_2}{x_1 + x_2 + x_3} \\ t_3 &= \frac{x_3}{x_1 + x_2 + x_3} \end{aligned}\right\} \quad \ldots \ldots \quad (3)$$

7) The truncation of the cone made necessary by the Purkinje phenomenon has not been shown in this figure. It would also be parallel to the plane O-x_1'-x_3'.

All points on a straight line through O have the same co-ordinates (t_1, t_2, t_3), and the equation $t_1 + t_2 + t_3 = 1$ is always satisfied. We can therefore say that if we are only interested in the character of a colour sensation and not in its brightness, it is sufficient to know the ratio of x_1, x_2 and x_3, i.e. the co-ordinates (t_1, t_2, t_3). The co-ordinates determined by equation (3) can be represented in different ways:

1) In an equilateral triangle. If we draw an equilateral triangle with height 1, the well known property, that the sum of the distances of any point from the three sides of the triangle is always equal to unity, allows us to represent the colour (t_1, t_2, t_3) by a point with distances t_1, t_2 and t_3 from the three sides. Geometrically we can visualise such a triangle being obtained by choosing the trihedral angle O-x_1-x_2-x_3 in the corresponding colour space-diagram equilateral and projecting the space from the centre on to a plane $x_1 + x_2 + x_3 =$ constant (which is generally **not** a plane of constant brightness).

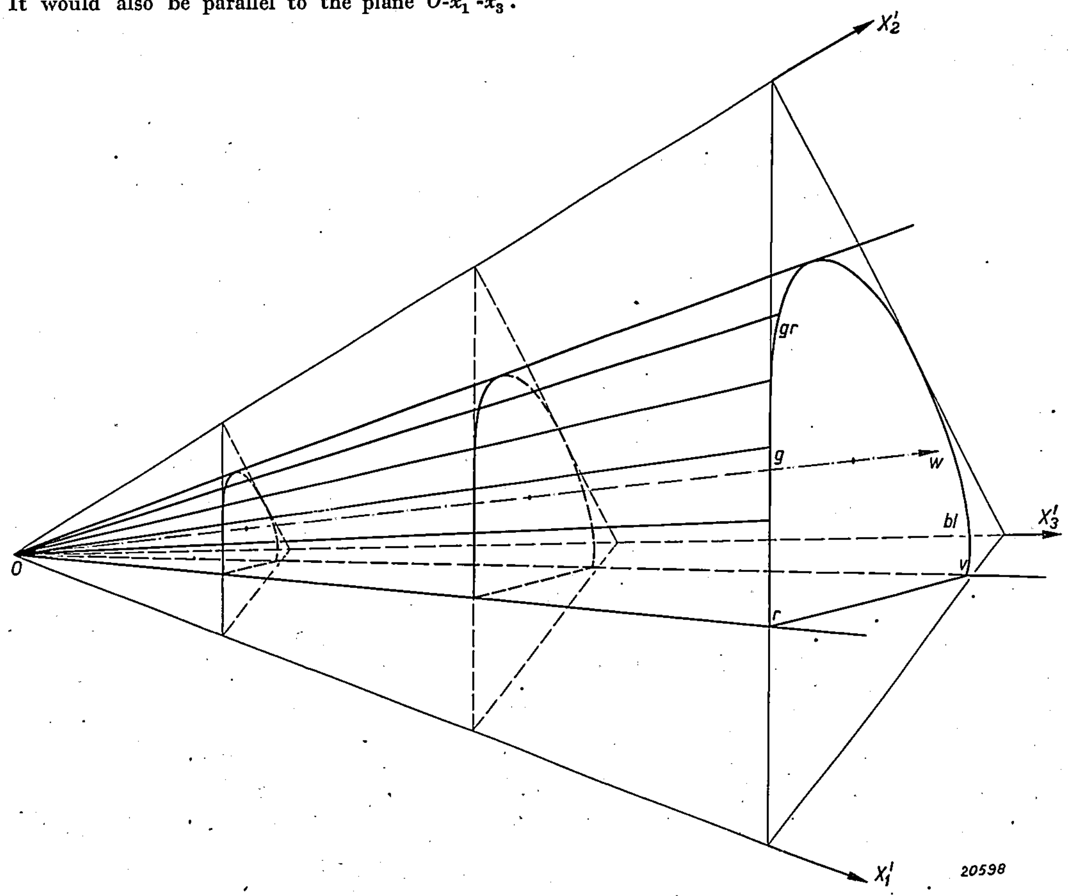


Fig. 2. Diagrammatic representation of a colour space-diagram obtained from fig. 1 by linear transformation. The axes of co-ordinates lie outside the colour cone. O-w (white) has equal distances to the three axes. Contrary to fig. 1 all colours have three positive co-ordinates.

Examples of this geometrical construction are shown in figs. 1 and 2.

2) In an isosceles right-angled triangle (the advantage being that it is easier to construct). Since $t_1 + t_2 + t_3 = 1$ it is sufficient to give two of the co-ordinates, for instance t_2 and t_3. We can therefore also plot the points in a rectangular system of co-ordinates, with t_3 and t_2 as abscissa and ordinate. It is readily seen that then $t_1 = \sqrt{2}$ times the distance of the point from the line $t_2 + t_3 = 1$ (the hypotenuse of the triangle).

The following illustrations give examples of colour triangles (drawn from the second method) with spectral colours and the white point W both inserted, and which have been derived from different space-diagrams:

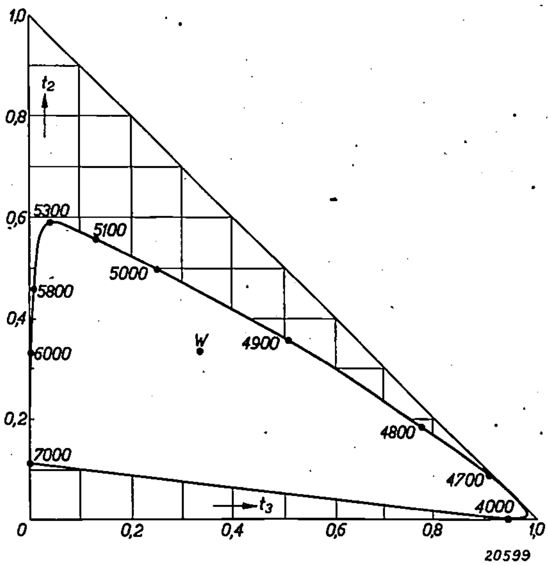


Fig. 3*a*. A colour triangle based on the early measurements of König. For white we have here $t_1 = t_2 = t_3$. The brightness equation for the corresponding space diagram is:

$$B = 0.568\ x_1 + 0.426\ x_2 + 0.006\ x_3$$

The axes are outside the spectral cone and are chosen in those directions in which we can move through space without certain types of dichromates perceiving any change. If we assume that the three dimensional nature of the totality of colour sensations is due to the existence of three photo-sensitive systems in the eye (red, green and blue systems), and that in those suffering from colour blindness one of the systems is lacking and that a white sensation is obtained by an equivalent stimulus of each of these three systems, the triangle 3*a* will represent approximately to what extent these systems are stimulated by specific types of light.

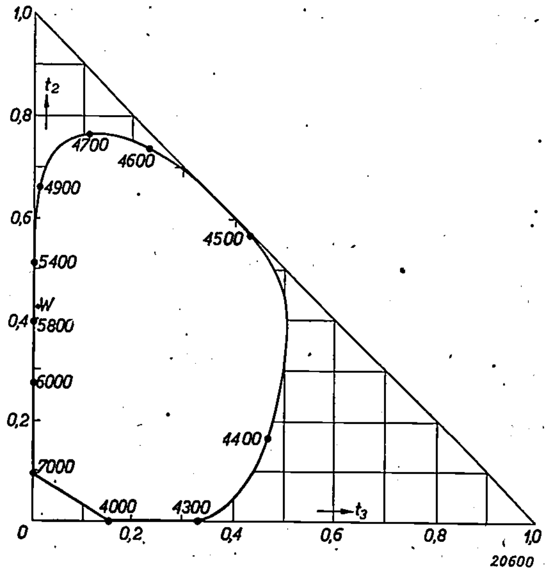


Fig. 3b. A colour triangle based on the same measurements as fig. 3a, and for whose space-diagram the brightness equation $B = x_1' + x_2' + x_3'$ applies. By the transformation $x_1' = 0.568\ x_1$, $x_2' = 0.426\ x_2$, $x_3' = 0.006\ x_3$ this space diagram becomes transformed to the previous diagram. The advantages of the simpler brightness equation and the simpler rule of mixing resulting therefrom (see below) are discounted by the serious disadvantage that the white point W is now situated very close to the spectral curve and an appreciable part of the colours is hence concentrated in a small space. For this reason space diagrams with $B = x_1' + x_2' + x_3'$ are never used in practice.

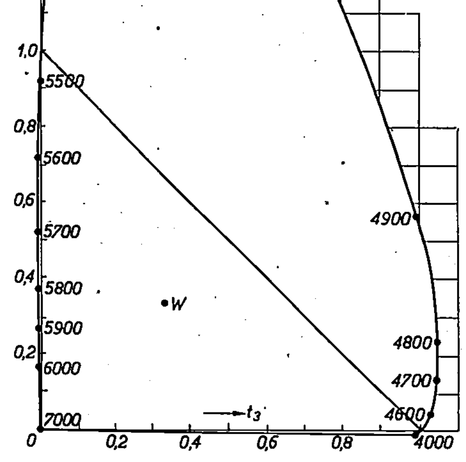


Fig. 3c. A colour triangle reproducing the recent measurements of Guild and Wright. The primary colours chosen here are the spectral colours 4358 Å,

5461 Å and 7000 Å; the brightness equation for the space-diagram is $B = x_1 + 4.39\ x_2 + 0.048\ x_3$. The white (the standard N.P.L. white fixed by the National Physical Laboratiory which very closely resembles sunlight) is located at the centroid of the triangle. As is always the case where the axes of the space diagram actually represent colours, some colours occur with negative co-ordinates (thus for λ = 5000 Å, t_1 is negative).

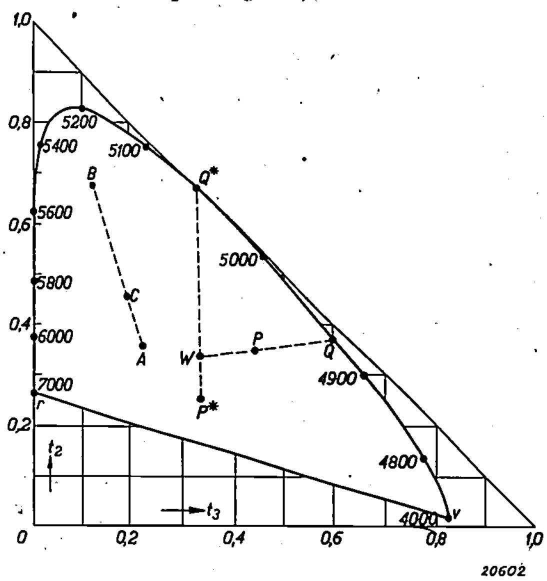


Fig. 3*d*. Another colour triangle also based on the measurements of Guild and Wright. This triangle was recommended in 1931 by the International Commission on Illumination (I. C. I.) and is now employed internationally. The white (here the equal energy spectrum, i.e. the spectrum which has the same energy $E\,d\,\lambda$ in all wave length regions $\lambda \rightarrow \lambda + d\,\lambda$) is located at the centroid; two of the sides touch the spectral curve [8]); the brightness equation in the corresponding space-diagram is: $B = x_2$. In a physical sense this equation has a some what surprising character as it does not contain the co-ordinates x_1 and x_2; but it must be remembered that this system has been evolved by a purely formal mathematical transformation from a system in which the physical interpretation of all magnitudes and formulae was directly apparent. The corresponding space diagram is shown in fig. 2.

Characteristics, Concepts and Formulae of Colour Triangles

Rules of Mixtures. From the space-diagram rule of mixing (vectorial addition) it follows immediately, in conjunction with the geometrical method of construction of the colour triangle, that in the triangle all mixtures of two colours lie on the line joining the colours [9]).

The position of the mixture on the join depends on the brightness ratio of the components. If A (t_1, t_2, t_3) and B $(t_1', t_2'\ t_3')$ represent the colours in the triangle which are mixed with brightness B and B' (fig. 3*d*) and if the brightness equation for the corresponding space-diagram is $B = ax_1 + bx_2 + cx_3$, then it can be shown mathematically that the position of the point C representing this mixture is given by the expression:

$$\frac{A\,C}{C\,B} = \frac{B'}{B}\,\frac{a\,t_1 + b\,t_2 + c\,t_3}{a\,t_1' + b\,t_2' + c\,t_3'}\,.$$

For colour triangles derived from space diagrams with $B = x_1 + x_2 + x_3$, this expression becomes:

$$\frac{A\,C}{C\,B} = \frac{B'}{B}.$$

The mixture of two colours having equal brightnesses is in this case always located at the midpoint of the line joining the colours. For the colour triangle in fig. 3*d* we have:

$$\frac{A\,C}{C\,B} = \frac{B'}{B}\,\frac{t_2}{t_2'} \quad \ldots\ldots \quad (4)$$

Dominant Wave-Length and Colorimetric Purity. It is seen from fig. 3*d* that a colour P (t_1'', t_2'', t_3'') brightness B'' can be obtained by mixing white light with brightness B' with a brightness B' of a specific spectral colour Q (t_1', t_2', t_3'). The wave length of Q is termed the dominant wave length, and the ratio $E = B'/B''$ the colorimetric purity of the colour P.

The colours P^* which lie in the triangle r-W-v (fig. 3*d*) cannot be formed from white light and a spectral colour; yet the colour P^* (t_1'', t_2'', t_3'') (with brightness B'') can be mixed with spectral colour Q^* (t_1, t_2, t_3) (with a brightness B') the result being white light (t_1, t_2, t_3) (with brightness B). In this case the wave-length of Q^* is termed the dominant wave length, and the ratio $\sigma = -\ B'/B''$ the colorimetric purity of the colour P^*: points in the r-W-v triangle (purples) therefore have a negative colorimetric purity.

---

[8]) By very careful interpolation it is found that this is not quite correct for the hypotenuse in fig. 3*d*; the difference is however extremely small.

[9]) This rule was introduced in the previous article as a convention. It is shown here that this convention was quite justified.

For all points in the colour triangle the equation applies:

$$\sigma = \frac{WP}{WQ} \cdot \frac{a\,t' \ + b\,t_2' + \ c\,t_3'}{a\,t_1'' + b\,t_2'' + c\,t_3''} =$$

$$= \frac{t_i - t_i''}{t_i - t_i'} \cdot \frac{a\,t_1' + \ b\,t_2' + \ c\,t_3'}{a\,t_1'' + b\,t_2'' + c\,t_3''}\,,$$

where *1*, *2* or *3* can be substituted for i as required, while WP/WQ must be taken negative, when W lies between P and Q, and positive when P lies between Q and W.

For colour triangles whose space diagram is given by $B = x_1 + x_2 + x_3$, the colorimetric purity is given by:

$$\sigma = \frac{WP}{WQ} = \frac{t_i - t_i''}{t_i - t_i'}\,.$$

For the triangle in fig. 3*d* we thus have:

$$\sigma = \frac{WP}{WQ} \cdot \frac{t_2'}{t_2''} = \frac{t_i - t_i''}{t_i - t_i'}\,\frac{t_2'}{t_2''} \ . \ . \ . \quad (5)$$

This triangle has been reproduced again in *fig. 4* with the curves of constant colorimetric purity inserted. Since equation (5) also contains the factor t_2'/t_2'', the scale of colorimetric purity along a straight line through W is generally not uniform (especially marked deviations in the blue).

Determination of the point in the colour triangle corresponding to a specific spectral distribution $E(\lambda)$.

Consider the various spectral colours are taken in turn each with the same energy. These different colour sensations will give points in the space diagram which form a curve on the spectral conical surface:

$$x_1 = \overline{x_1}\,(\lambda),\ x_2 = \overline{x_2}\,(\lambda),\ x_3 = \overline{x_3}\,(\lambda).$$

For the space diagram from which fig. 4 was derived the co-ordinates for this curve are given in *Table I*.

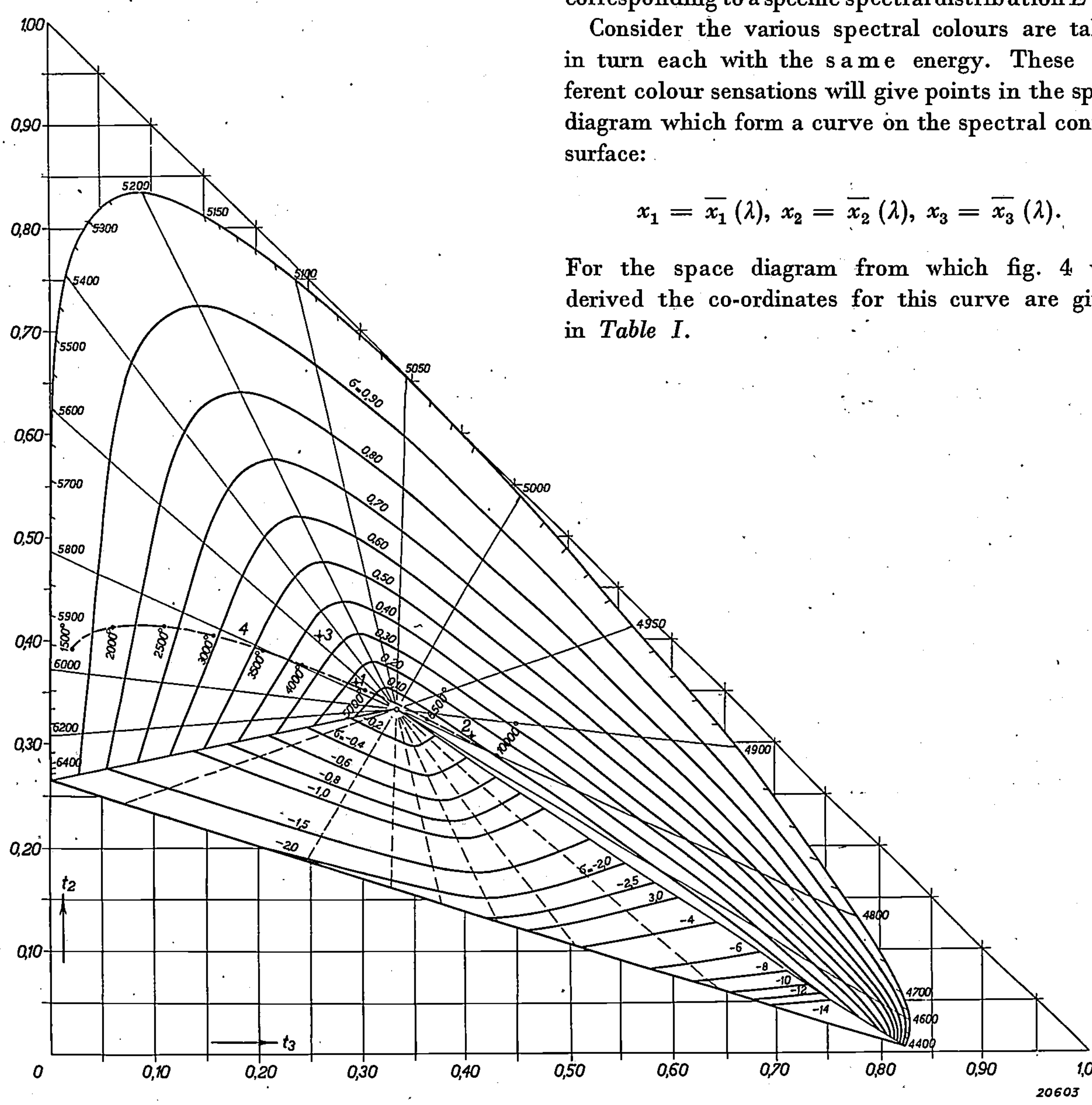


Fig. 4. The colour triangle specified by the I.C.I. (see also fig. 3*d*) with lines of constant colorimetric purity and a series of special points: 1) Daylight (sun and sky); 2) Blue sky; 3) Mercury light; 4) Black-body radiation.

Table I. Colour co-ordinates based on the I.C.I. System.

λ	$\overline{x_1}(\lambda)$	$\overline{x_2}(\lambda)$	$\overline{x_3}(\lambda)$	λ	$\overline{x_1}(\lambda)$	$\overline{x_2}(\lambda)$	$\overline{x_3}(\lambda)$
3800	0.0014	0.0000	0.0065	5800	0.9163	0.8700	0.0017
3900	0.0042	0.0001	0.0201	5900	1.0263	0.7570	0.0011
4000	0.0143	0.0004	0.0679	6000	1.0622	0.6310	0.0008
4100	0.0435	0.0012	0.2074	6100	1.0026	0.5030	0.0003
4200	0.1344	0.0040	0.6456	6200	0.8544	0.3810	0.0002
4300	0.2839	0.0116	1.3856	6300	0.6424	0.2650	0.0000
4400	0.3483	0.0230	1.7471	6400	0.4479	0.1750	0
4500	0.3362	0.0380	1.7721	6500	0.2835	0.1070	0
4600	0.2908	0.0600	1.6692	6600	0.1649	0.0610	0
4700	0.1954	0.0910	1.2876	6700	0.0874	0.0320	0
4800	0.0956	0.1390	0.8130	6800	0.0468	0.0170	0
4900	0.0320	0.2080	0.4652	6900	0.0227	0.0082	0
5000	0.0049	0.3230	0.2720	7000	0.0114	0.0041	0
5100	0.0093	0.5030	0.1582	7100	0.0058	0.0021	0
5200	0.0633	0.7100	0.0782	7200	0.0029	0.0010	0
5300	0.1655	0.8620	0.0422	7300	0.0014	0.0005	0
5400	0.2904	0.9540	0.0203	7400	0.0007	0.0003	0
5500	0.4334	0.9950	0.0087	7500	0.0003	0.0001	0
5600	0.5945	0.9950	0.0039	7600	0.0002	0.0001	0
5700	0.7621	0.9520	0.0021	7700	0.0001	0.0000	0

The following formula is employed to determine the point in the triangle corresponding to a given spectral distribution:

$$t_1 : t_2 : t_3 = \int E(\lambda)\overline{x_1}(\lambda)d\lambda : \int E(\lambda)\overline{x_2}(\lambda)d\lambda : \int E(\lambda)\overline{x_3}(\lambda)d\lambda, \quad (6)$$

or for a line spectrum, the expression:

$$t_1 : t_2 : t_3 = \Sigma E(\lambda)\,\overline{x_1}(\lambda) : \Sigma E(\lambda)\,\overline{x_2}(\lambda) : \Sigma E(\lambda)\,\overline{x_3}(\lambda). \quad (7)$$

In fig. 4 a number of these calculated spectral distributions have been inserted.

1) Daylight (mean result of measurements of Miss Eymers) consisting of a combination of direct sunlight and reflected sky radiation. Dominant wave length 5600 to 5770 Å; colorimetric purity 0.14 to 0.22.
2) Light of blue sky (Eymers): Dominant wave length 4835 to 4855 Å; colorimetric purity 0.09 to 0.13.
3) Mercury light (1 atmos. pressure).
4) The dash line represents black-body radiation at various temperatures. Close to $T = 2800$ °K is the colour of glowlamp light.

Fig. 4 can be used for the graphical determination of the dominating wave length (λ) and the colorimetric purity (see equation (5)) from the co-ordinates (t_1, t_2, t_3). As examples:

$t_1 = 0.1$, $t_2 = 0.5$, $t_3 = 0.4$ gives $\lambda = 5021$ Å, $\sigma = 0.81$ (fairly saturated blue-green)

$t_1 = 0.4$, $t_2 = 0.2$, $t_3 = 0.4$ gives $\lambda = 5215$ Å, $\sigma = -1.12$ (a mixture of purple and white with green complementary colour).

BIBLIOGRAPHY

Handbook of Colorimetry, Cambridge, Massachusetts, 1936.

Schrödinger, Müller Pouillet's Lehrb. d. Phys., II, Optik, first half, pp. 458 to 488.
Fundamental principles of the colour space-diagram and colour triangle.

Description of the I.C.I. specification:
Smith and Guild, Trans. Opt. Soc., **33**, 74, 1931/1932.
Judd, Journ, Opt. Soc. Am., **23**, 359, 1933.
Richter, Das Licht, **4**, 205 and 231, 1934.
Ribaud, Revue d'Optique [10]), **15**, 161, 1936.

Schrödinger Ann. Physik, **63**, 397, 481, 1920.
In addition to the construction of the colour space-diagram this paper also deals with the "Physiological interval" between two points in the space-diagram.

Bouma and Heller, Proc. Kon. Ak. Wet. A'dam. **38**, 34. 148, 258, 1935.
The analysis is expanded to the brightness levels where the Purkinje phenomenon takes place.

[10]) It should be noted that the formulae on pp. 168 and 169 of this article are not correct.

OCTAVES AND DECIBELS

By R. VERMEULEN.

Summary. With the aid of practical data, the relationships are discussed between the acoustic magnitudes: Pitch interval and loudness, and the corresponding physical magnitudes: frequency and sound intensity. The object of this article is to emphasise that the logarithmic units, octaves and decibels, should be given preference in acoustics for denoting physical magnitudes.

Why is a preference shown in acoustics and associated subjects for logarithmic scales and are even new units introduced to permit the employment of such scales? The use of logarithmic scales in acoustics is justified from a consideration of the characteristics of the human ear, apart from many other advantages offered in both this and other subjects. Reasons for this are that to a marked extent acoustic measurements, and particularly technical sound measurements, represent attempts to find a substitute for direct judgement by means of the ear. An instructive example of this is offered in the determination of the characteristics of a loudspeaker, in which not only must the reproduced music and speech be analysed critically but *inter alia* the loudness also determined, i.e. the sound output of the loudspeaker at different pitches for an equivalent electrical power input. Yet neither the loudness nor the pitch is a physical magnitude; these are physiological and even to some extent psychological factors, which cannot be evaluated by direct measurement, since they are purely subjective. To arrive at a numerical value it is therefore necessary to make do with an objective measurement of specific physical factors, and in the case under discussion, for instance, the sound output measured as a function of the frequency. To do this it is very desirable to select such units for expressing the results arrived at that the numerical values are to some extent proportional to the corresponding physiological sensations. The same applies to electrical magnitudes when these are used for sound transmission, as for instance in telephony.

Pitch and Frequency

It is well-known that the pitch is related to the frequency of the fundamental tone [1]). That this relationship, is, however not a linear function will be evident from the following considerations. In music, where without any close insight into the physics of sound the pitch has always been denoted in musical scores by notes, the concept "musical or pitch interval" has been introduced to represent a specific difference in pitch. The most important pitch interval is the o c t a v e: Two tones which are at an interval of an octave have such a marked natural similarity that musicians have not considered it necessary to introduce different names for them, but have merely distinguished them by means of indices. In addition to the octave, there are pitch intervals, such as the fifth, fourth and several others, which are similarly perfectly natural and self-evident, not only because two tones which differ from each other by such a pitch interval harmonise with each other, but also, and this is the more important for what follows, because such pitch intervals occur at all pitches with equivalent differences.

Musicians therefore indicate the pitch in musical notation in such a way that a pitch interval nearly always corresponds to the same difference in pitch [2]), independent of the absolute pitch value. The keyboard of the pianoforte is also based on the principle that equal pitch intervals are reproduced. at uniform intervals throughout the whole compass of the instrument.

A pitch interval regarded from the physical aspect is, however, not representative of a specific d i f f e r e n c e, but of a specific r a t i o of frequencies, an octave corresponding to a doubling of the frequency, and the fifth to a ratio of 2 : 3, etc. (see *Table I*). The l o g a r i t h m of the frequency shares in common with the pitch the property that on an alteration of the tone by the same pitch interval it always increases by the same amount, irrespective of the absolute frequency values:

$$log \frac{f_1}{f_2} = log\, f_1 - log\, f_2.$$

[1]) That this relationship is not strictly absolute, but is determined somewhat by the intensity and timbre, is left unconsidered here.

[2]) These intervals are not absolutely equal since the pitch interval between two successive tones in musical notation is usually equal to a whole tone but sometimes also equal to a semitone.

Hence in the light of our introductory remarks the logarithm is a suitable measure for the pitch.

Octaves

Owing to the fundamental importance of the octave to hearing it appears desirable to select not 10 the Briggsian base, but the number 2 as the base of logarithms. In this way we get for the magnitude I of a pitch interval with the octave as unit:

$$I = {}^{2}log\,\frac{f_1}{f_2} = 3.320\ {}^{10}log\,\frac{f_1}{f_2}\,.$$

The octave can be subdivided in various ways. The most common subdivision has been developed from the desire to arrive at a scale enabling the natural pitch intervals to be obtained with the minimum number of keys per octave in instruments with a fixed keyboard, such as the organ and pianoforte. The principal natural pitch intervals are shown on a logarithmic scale in *fig. 1* with a fundamental tone denoted by "C" as the tone of origin. If the same pitch intervals were furthermore based on one of the tones obtained in this way, new tones would be continually produced, some of which are shown in the figure by thin lines; it is thus seen that a finite number of scale divisions does not suffice. Nevertheless it is found that the most important pitch intervals can be obtained to a very close approximation by subdividing the octave into twelve equal parts, which are also shown in fig. 1 and by which additional new tones are obtained between C and D and between G

Table I.

1	2	3	4	5	6	7
Pitch interval	Tone	Frequency ratio			Milli-octaves	
		Natural scale		Tempered scale	Natural scale	Tempered scale
Unison	C	1	1.000	1.000	0	0
Comma		81/80	1.013	1.000	17.92	0
Semitone	C♯	25/24	1.042	1.059	58.89	83.33
Limma		16/15	1.067	1.059	93.11	83.33
Minor second	D♭	27/25	1.080	1.059	111.0	83.33
		10/9	1.111	1.122	152.0	166.6
Major second	D	9/8	1.125	1.122	169.9	166.6
Augmented second	D♯	75/64	1.172	1.189	228.8	250.0
Minor third	E♭	6/5	1.200	1.189	263.0	250.0
Major third	E	5/4	1.250	1.260	321.9	333.3
Diminished fourth	F♭	32/25	1.280	1.260	356.1	333.3
Augmented third	E♯	125/96	1.302	1.335	380.7	416.5
Perfect fourth	F	4/3	1.333	1.335	414.8	416.5
Augmented fourth	F♯	25/18	1.389	1.414	473.9	500.0
Diminished fifth	G♭	36/25	1.440	1.414	526.1	500.0
Perfect fifth	G	3/2	1.500	1.498	585.0	583.3
Augmented fifth	G♯	25/16	1.562	1.587	644.0	666.6
Minor sixth	A♭	8/5	1.600	1.587	678.1	666.6
Major sixth	A	5/3	1.667	1.682	737.0	750.0
Augmented sixth	A♯	125/72	1.736	1.782	795.8	833.3
Minor seventh	B♭	9/5	1.800	1.782	848.0	833.3
Major seventh	B	15/8	1.875	1.883	906.9	916.6
Diminished octave	C♭	48/25	1.920	1.883	941.1	916.6
Augmented seventh	B♯	125/64	1.953	2.000	965.7	1000.0
Octave	C′	2	2.000	2.000	1000.0	1000.0

In addition to the musical notation (2), the above table gives for the various pitch intervals, which are also indicated in fig. 1, the following data: (3) the numerical ratio determining the pitch interval and which for the less simple pitch intervals is always a simple ratio, e.g. $125/64 = (5/4)^3$ (i.e. the pitch interval is composed of three successive thirds). The next column (4) gives this fraction in decimals, while (5) gives the tempered scale approximating to the natural pitch interval. In the last two colums both the natural (6) and the tempered (7) scales are given in milli-octaves.

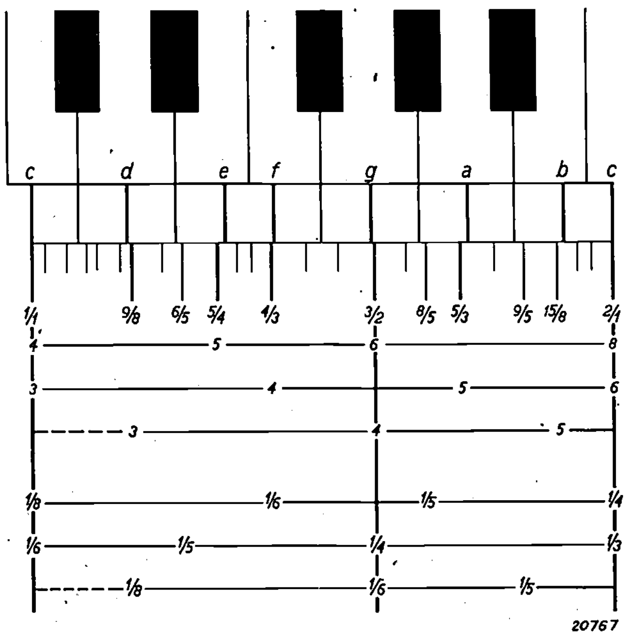


Fig. 1. Pitch intervals and scales. Below the keyboard of the pianoforte, the frequency ratios of the tones on the tempered scale are given logarithmically, and below these again the ratios for tones with natural pitch intervals. The musical notation of the notes is indicated by letters and the frequency ratios of the natural tones based on the fundamental tone C by vulgar fractions. The tones of the major scale are indicated by thick lines, and may be regarded as evolved by a threefold application of the major chord whose tones have a frequency ratio of 4 : 5 : 6, i.e. a major third and a fifth. The figure shows how with this chord the tones of the scale are obtained. It follows from the figure that this scale can be represented to a close approximation by those tones on the tempered scale which correspond to the white keys on the pianoforte.
The pitch intervals, which are shown by slightly thinner lines and which may partly be approximately represented by the black keys, have been arrived at by corresponding application of the minor chord. This minor chord may be regarded as a harmonic of the major chord; it is made up from the frequency ratio (1/4 : 1/5 : 1/6). The tempered scale gives an additional two tones between C and D and between F and G, to which no simple pitch intervals of C correspond. It will be seen, however, that these are required for arriving at the natural pitch intervals by taking one of the other tones of the scale as roots.

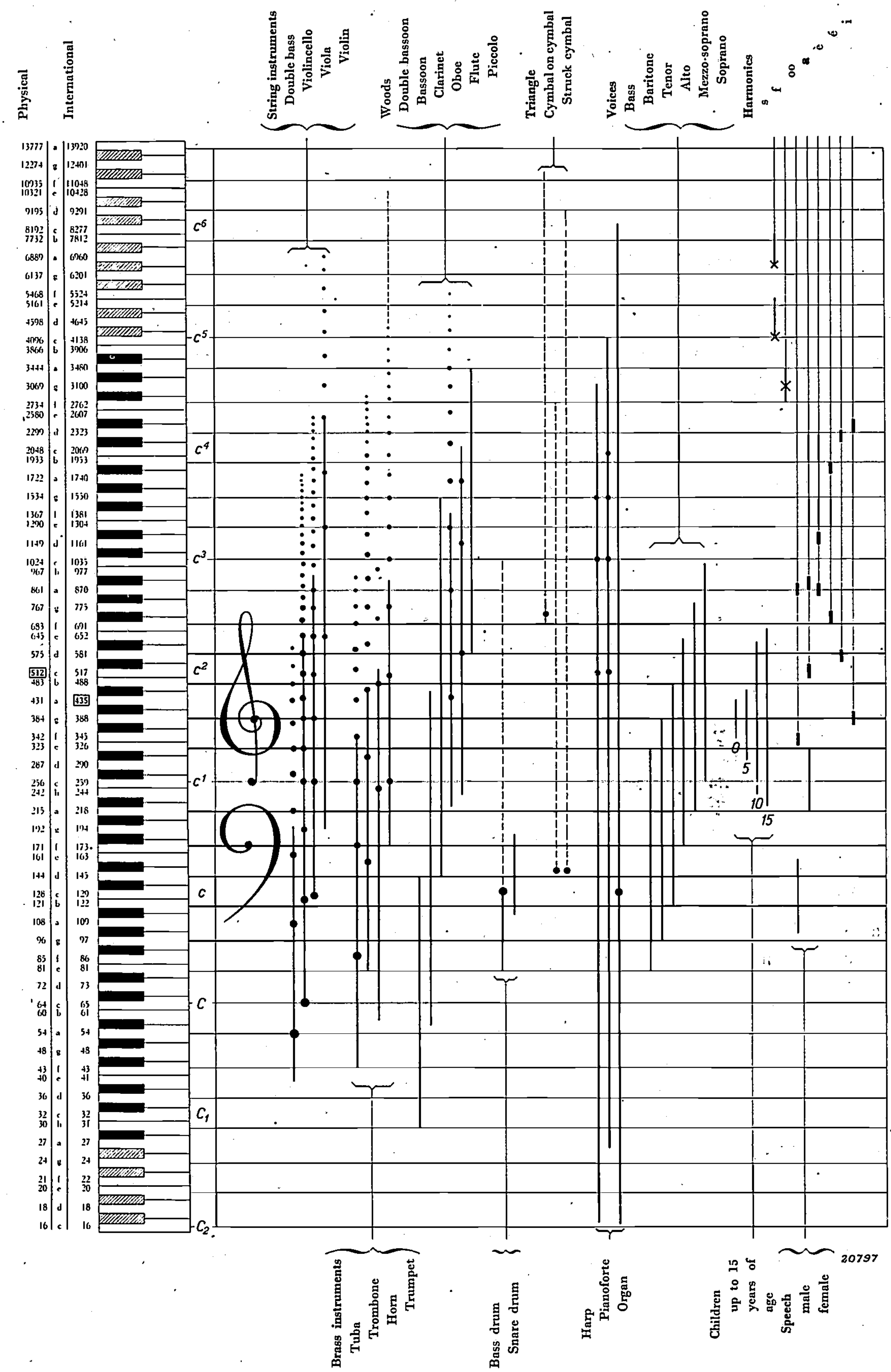


Fig. 2. Compass of musical instruments and voices. The frequency range of the fundamental tone for various musical instruments and voices is shown by a continuous line. For a number of tones (heavy dot) the corresponding harmonics have been included as small dots or a broken line. For speech the characteristic overtones (socalled harmonics) have been included in the figure.

and A. Such uniform subdivision, termed a scale of just temperament, has the advantage over scales based on natural pitch intervals that deviations in the pitch intervals on a keyboard from the natural intervals are always of the same magnitude, irrespective of the fundamental tone taken as basis.

Owing to the general application of the decimal system, a decimal subdivision is favoured in physical work, and pitch intervals of centi-octaves and milli-octaves are used; these are also shown in Table I.

The octavo provides a method for ascribing numerical values to differences in pitch. To be able to deduce the absolute pitch from the frequency, the frequency of a specific tone must be fixed as a basis. Various conventions have been adopted for this purpose. The standard of pitch of orchestral instruments is based on "A" whose frequency according to the socalled international scale is 435 cycles, while the socalled physical scale takes 430.5 cycles for this tone, for it is considered an advantage that C^3 then has a frequency of 1024, i.e. exactly 10 octaves above 1 cycle. In American scientific literature a fundamental tone of 1000 cycles is employed in conjunction with decimal subdivision of the octave.

Fig. 2 shows the compass of a number of common musical instruments and brings out the importance of the various frequency bands as well as the frequency of the tones in the two aforementioned scales.

Loudness and Sound Intensity

Just as the logarithm of the frequency has been employed as a measure of the pitch, the logarithm of the sound intensity [3]) is used to express the intensity of an auditory sensation. The justification for this is here however not equally valid since the auditory sensation is not a complete parallel to the logarithm of the sound intensity. The true relationship between these two magnitudes is difficult to establish, since the loudness is determined not only by the sound intensity but also largely by the frequencies composing the sound.

Furthermore, the estimation of equivalent differences in loudness is much less accurate and far more difficult than is possible with differences in pitch. The results of experiments designed to determine the relationship between loudness and sound intensity are far from satisfactory, and their discussion is beyond the scope of the present article. It is nevertheless possible to determine with a fair degree of accuracy the equality of loudness of two sounds, even if they differ considerably in character. At a given frequency and sound intensity, the difference in intensity which produces a just perceptible difference in loudness can moreover be established fairly accurately. It is found that for a tone of approximately 1000 cycles the ratio I_2/I_1 of the sound intensities between which the ear can still just differentiate is constant over a wide range and is approximately 1.2. If a logarithmic scale is used two sound intensities just at the threshold of differentiation will be roughly the same distance apart over a greater part of the graph. This is very convenient, where, for instance, it is desirable for the accuracy of a graph to conform to the acuity or sensitivity of differentiation of the ear. Moreover, with this scale the very wide compass of intensities of $1 : 10^{13}$ to which the ear reacts (see *fig. 3* and *4*), can be covered, which

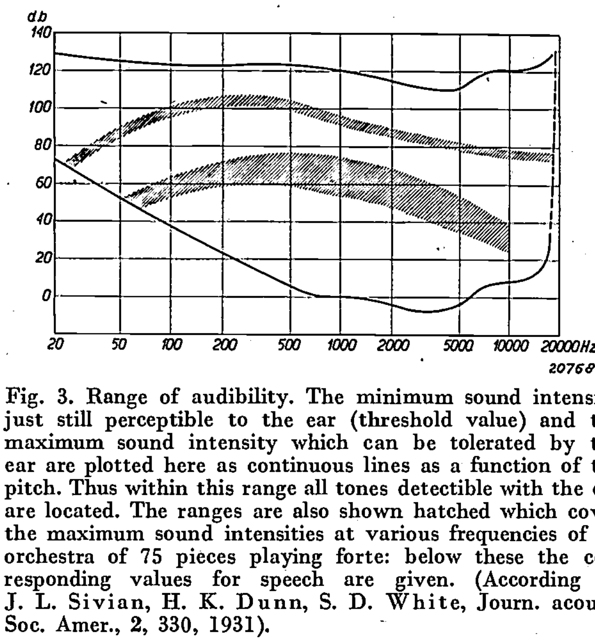


Fig. 3. Range of audibility. The minimum sound intensity just still perceptible to the ear (threshold value) and the maximum sound intensity which can be tolerated by the ear are plotted here as continuous lines as a function of the pitch. Thus within this range all tones detectible with the ear are located. The ranges are also shown hatched which cover the maximum sound intensities at various frequencies of an orchestra of 75 pieces playing forte: below these the corresponding values for speech are given. (According to J. L. Sivian, H. K. Dunn, S. D. White, Journ. acoust. Soc. Amer., 2, 330, 1931).

would be quite impossible with a linear scale. The logarithmic scale offers other practical advantages, which will be discussed at the conclusion of this article.

Bel and Neper

To ascribe a numerical value to the difference in intensity between two sounds we can therefore adopt the convention "The logarithm of the ratio of their sound intensities shall be N", although

[3]) The sound intensity of a progressive wave is the amount of energy which is conveyed by the sound wave in unit time and per unit of surface normal to the direction of flow. The unit used to express this intensity is the erg per sec per sq. cm, although frequently the intensity is given in watts per sq. cm = 10^7 erg sec^{-1} cm^{-2}.

it is evident that this cumbersome method is useless. For this reason the units "Bel" and "Neper" have been introduced.

The difference between these two units lies in the first place in the choice of the base of logarithms used [4]), the **Briggsian** base of 10 being selected for the "Bel" and the base of natural logarithms e = 2.718 for the "**Neper**". They also differ in the type of the physical magnitudes to which they are applied: The bel is always used for expressing energies or powers, while the neper is used for amplitudes, such as sound pressure, velocity of sound (sound-particle velocity), electric currents or voltages. To exemplify the fundamental difference between these two units and to bring out their relationship, the following discussion is instructive:

The intensity of a plane progressive sound wave at a certain point is determined by the sound intensity I (see footnote 3), by the intensity of the pressure fluctuations p or by the sound-particle velocity v. These factors are related to each other as follows:

$$I = p \cdot v = \frac{p^2}{\varrho c} = \varrho\, c\, v^2, \quad \ldots\ldots (1)$$

where ϱ is the density of the medium of propagation (for air ϱ is 0.0013 gr per cm^3) and c is the velocity of propagation through the medium (for air c = approximately 332 m per sec). This equation is valid only for plane progressive waves.

Similarly, in the transmission of electrical energy, either the power W transmitted, the voltage e or the current i can be stated, provided that in the latter case the resistance R of the consumer is also given, thus:

$$W = e\, i = \frac{e^2}{R} = i^2 R \quad \ldots\ldots\ldots (2)$$

It is seen that the product $\varrho\, c$ in equation (1) bears a close analogy to the resistance R in equation (2). $\varrho\, c$ is termed the "acoustic resistance". For air $\varrho\, c = 41$ gr cm^{-2} sec^{-1}. *Fig. 4* gives the relationship between the ratio of the amplitudes, the number of nepers or decibels and the ratio of the intensities or energy values.

It follows from equations (1) and (2) that the logarithm of the ratio of two powers or intensities is double the logarithm of the ratio of the corresponding voltage, current, pressure, or velocity amplitudes.

$$log \frac{W_2}{W_1} = 2\ log \frac{e_2}{e_1} = 2\ log \frac{i_2}{i_1},$$

$$log \frac{I_2}{I_1} = 2\ log \frac{P_2}{P_1} = 2\ log \frac{v_2}{v_1}.$$

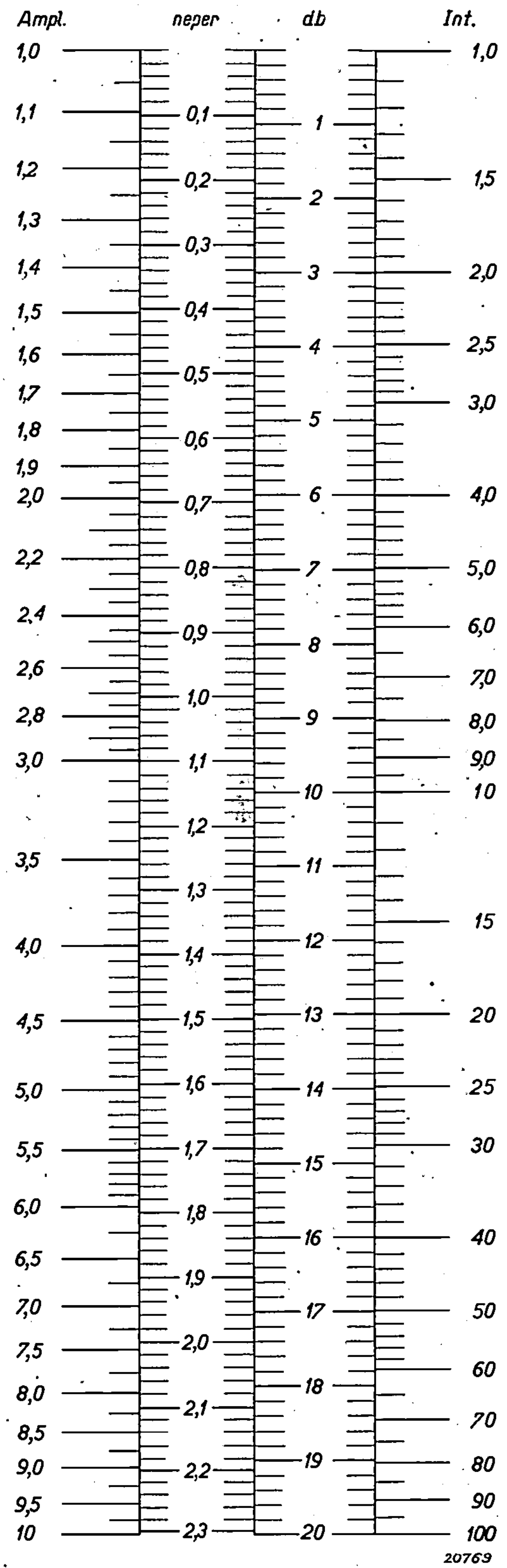


Fig. 4. Scale subdivision in amplitude ratios, nepers, decibels and intensity ratios.

---

[4]) We shall write $log_{10}\, a$ as $lg\, a$ and $log_e\, a$ as $ln\, a$, the simple relationships between these logarithms being:

$lg\, a = 0.434\, ln\, a$ and $ln\, a = 2.3\, lg\, a$.

In the following we shall term the logarithm of the ratio of two magnitudes a difference in level, thus arriving at the following definitions for the bel and neper:

Bel: The difference in level between two powers or intensities is N bels, when:

$$lg \frac{W_2}{W_1} = N \quad \text{or} \quad \frac{W_2}{W_1} = 10^N.$$

Decibel: The decibel is the tenth part of a bel, so that the difference in level of two powers or intensities is n decibels when:

$$10\ lg \frac{W_2}{W_1} = n \quad \text{or} \quad \frac{W_2}{W_1} = 10^{\frac{n}{10}}.$$

Neper: The difference in level between two amplitudes (of voltage, current, pressure, velocity, etc.) is M nepers when:

$$ln \frac{a_2}{a_1} = M \quad \text{or} \quad \frac{a_2}{a_1} = e^M.$$

In the light of these definitions the difference in level of two powers or intensities can only be expressed in bels or decibels, while the difference in level between two voltages or veliocities must not be expressed in bels but in nepers. To calculate the difference in level between two powers the measured voltages (e), currents (i), velocities (v) or pressures (p) are naturally useful, provided the resistance of the consumer is the same in both instances. The following equation can then be used:

$$n = 10\ lg \frac{W_2}{W_1} = 20\ lg \frac{e_2}{e_1}$$

If it is remembered that $log\ 2 = 0.30$ and that 1 decibel corresponds to an energy increase of approximately 25 per cent, the intensity ratio corresponding to a specific number of decibels can readily be calculated mentally. The following series of corresponding values is then found:

Decibels	0	1	2	3	4	5	6	7	8	9	10
Ratio	1	1.26	1.6	2.0	2.5	3.2	4.0	5.0	6.3	7.9	10

[5]) The following equivalents are employed for converting nepers into decibels, provided that the resistances are equivalent:
1 decibel = 0.115 neper; 1 neper = 8.7 decibels.
But it is incorrect to calculate the amplification of an amplifier with an input impedance of $R_1 = 1$ megohm and a resistance load of $R_2 = 1000$ ohms, in terms of decibels as $n = 20\ lg\ e_2/e_1$. The amplification is moreover given by the expression:

$$n = 10\ lg\left(\frac{e_2^2}{R_2} \cdot \frac{R_1}{e_1^2}\right) = 20\ lg \frac{e_2}{e_1} + 10\ lg \frac{R_1}{R_2} = 20\ lg \frac{e_2}{e_1} + 30.$$

To determine the ratio of the amplitudes the number is sought corresponding to half the number of decibels.

Sound Intensity Level and Loudness Level

Decibels and nepers are units for expressing differences in level. With pure tones the loudness itself can be established by taking a specific sound intensity as zero level. The difference in decibels compared with zero level is termed the sound intensity level.

Tabel II. Relationships between different acoustic magnitudes for plane progressive waves in air.

Decibels	Phons	Intensity	Sound pressure	Sound-particle velocity	Amplitude of vibrations in air at
Above 10^{-16} wattsper sq. cm	db above $2.5 \cdot 10^{-16}$ watts per sq. cm	10^{-9} watts per sq. cm	Dynes per sq. cm	in air, 10^{-3} cm per sec.	1000 cycles, 10^{-6} cm
64	60	0.25	0.32	8	1.8
65	61	0.32	0.36	9	2.0
66	62	0.40	0.40	10	2.2
67	63	0.50	0.45	11	2.5
68	64	0.63	0.50	12	2.8
69	65	0.8	0.56	14	3.2
70	66	**1.0**	0.63	16	3.6
71	67	1.25	0.71	18	4.0
72	68	1.6	0.80	20	4.5
73	69	2.0	0.89	22	5.0
74	**70**	2.5	**1.0**	25	5.6
75	71	3.2	1.1	28	6.3
76	72	4.0	1.25	32	7
77	73	5.0	1.4	36	8
78	74	6.3	1.6	40	9
79	75	8	1.8	45	10
80	76	10	2.0	50	11
81	77	12.5	2.2	56	12
82	78	16	2.5	63	14
83	79	20	2.8	71	16
84	80	25	3.2	80	18

It is a matter of serious concern that not only is more than one zero level used but that moreover reference is frequently omitted to the actual zero level used as a basis of comparison [6]). In America

[6]) The whole statement of data may as a result become meaningless. "The intensity of a loudspeaker is 80 decibels" has no meaning, unless it is stated which intensity is taken as zero level and at what distance this intensity is measured, or, if the total power output is meant, which power value has been selected as zero level. The sentence above can therefore mean: "The sound intensity in the axis of the loudspeaker at a distance of 2 m is 80 decibels above 10^{-16} watts per sq. cm", or also e.g.: "The total output of the loudspeaker is 80 decibels above 1 microwatt."

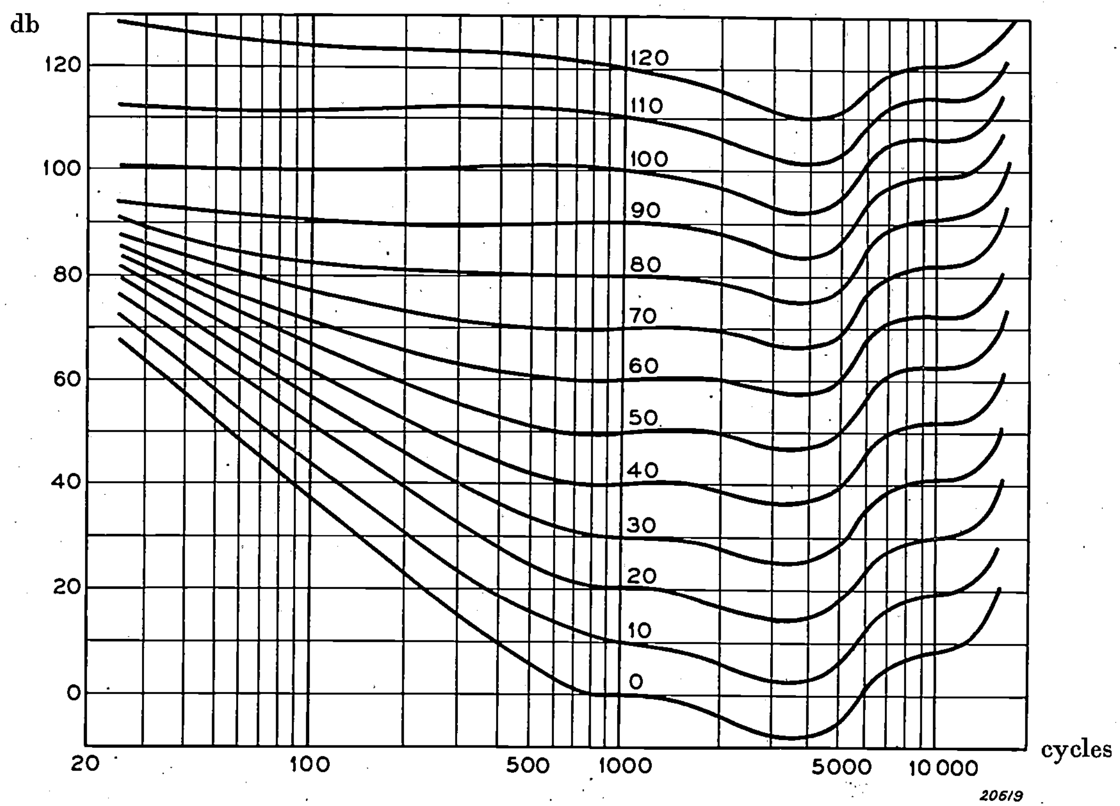


Fig. 5. Sound intensity level in decibels above 10^{-16} watts per sq.cm as a function of the frequency for constant loudness. The numbers next to the lines indicate the loudness level, i.e. the sound intensity level of the corresponding loudness for a 1000-cycle tone. The bottom line represents the threshold of audibility of the human ear. It is striking that the corresponding sound intensity increases considerably with diminishing frequency, while the curves for higher loudness values are much flatter.
The irregularity which occurs in all curves in the band between 2000 and 6000 cycles is probably due to the diffraction of the sound waves around the head. In measurements of the sound pressure in the ear canals the curves are smooth. (According to H. Fletcher and W. A. Munson, Journ. acoust. Soc. Amer., **5**, 82, 1933).

an intensity of 10^{-16} watts per sq.cm which lies approximately at the threshold of audibility of the ear is used as zero level; in Germany the standard intensity level corresponding to 1 dyne per sq.cm sound pressure in a plane wave is put equal to 70 decibels. Zero level then has a pressure of $0.32 \cdot 10^{-3}$ dyne per sq. cm and a sound intensity of $2.5 \cdot 10^{-16}$ watts per sq. cm. In general the use of the decibel is limited to differences in level only and the level itself is expressed in phons, 70 phons corresponding to a sound pressure of 1 dyne per sq. cm. In *Table II*, the comparison is given for a plane progressive wave between the intensity level (decibels above 10^{-16} watts per sq. cm and phons) and the intensity, as well as the corresponding effective value of the pressure fluctuations and the sound-particle velocity. As a guide the amplitudes of the air particles are given in the last column for a frequency of 1000 cycles [7]).

[7]) The amplitude a is calculated from the effective value of the sound-particle velocity for a specific frequency f using the expression:

$$a = \frac{v}{2\pi f}\sqrt{2}.$$

For sounds differing in character the loudness may differ considerably at the same intensity level. In particular pure tones with different frequencies and at the same level exhibit very marked differences in loudness. In *fig. 5* the intensity level at constant loudness is plotted as a function of the frequency. It is seen that a much higher sound intensity corresponds to an equivalent loudness at low frequencies than at medium frequencies. The difference may be as great as 60 decibels, i.e. by a factor of 1 million.

The sound intensity which is just at the limit of audibility is termed the threshold value. The line of constant loudness corresponding to this value is marked with the numeral 0 in the figure. The other lines are marked with the sound intensity level which is required to obtain the loudness in question at 1000 cycles. This numerical value is termed the loudness level. Also for compound sounds, such as street noises, the loudness level can be determined by comparison with a tone of 1000 cycles. A selection of loudness levels of various common sounds and noises is reproduced in *fig. 6*.

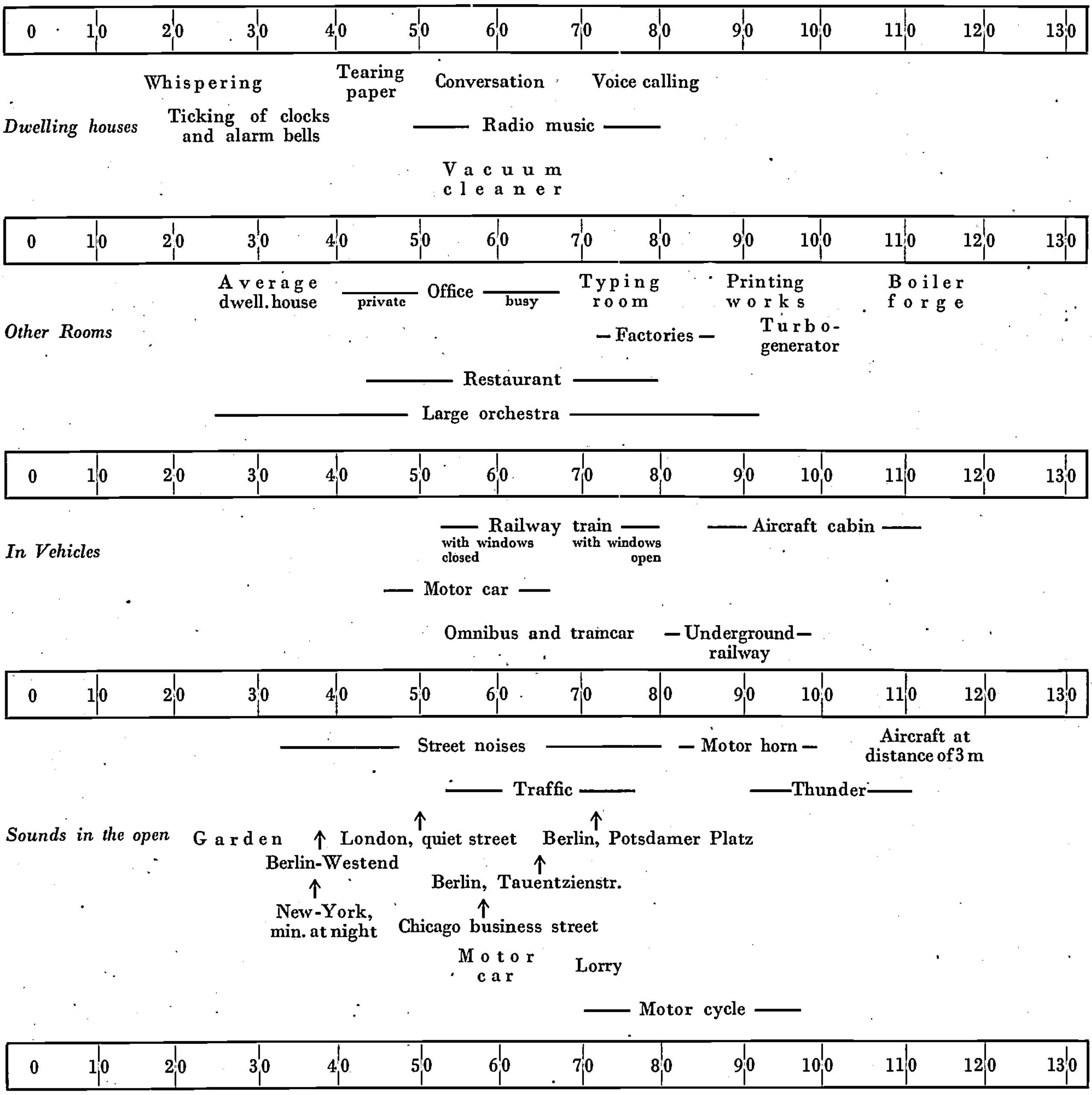


Fig. 6. Range of loudness level in phons for various common sounds and noises.

Comparison of Diagrams with Logarithmic and Linear Scales

To illustrate clearly the differences between the linear and logarithmic scales used for graphical representation, the same arbitrary characteristic showing, e.g. the variation of sound intensity of a loudspeaker, has been plotted in various ways in *fig. 7* as a function of the frequency.

1) The first point which becomes apparent, for instance, on a comparison of *a* and *b*, is that with the logarithmic frequency scale all octaves are shown to an equal extent, while with the linear scale the band below 1000 cycles, which is the most important as regards audibility, has been compressed to an insignificant width. On the other hand the band above 5000 cycles, which has a compass of only $1^1/_2$ octaves, takes up a disproportionately large part of the figure.

2) Secondly (cf. *a* and *b*) the widths of the various peaks and troughs of the curves are equal on the logarithmic scale where they are the same percentage fractions of the frequency. This gives a better insight into the damping of the resonant frequencies, which in certain

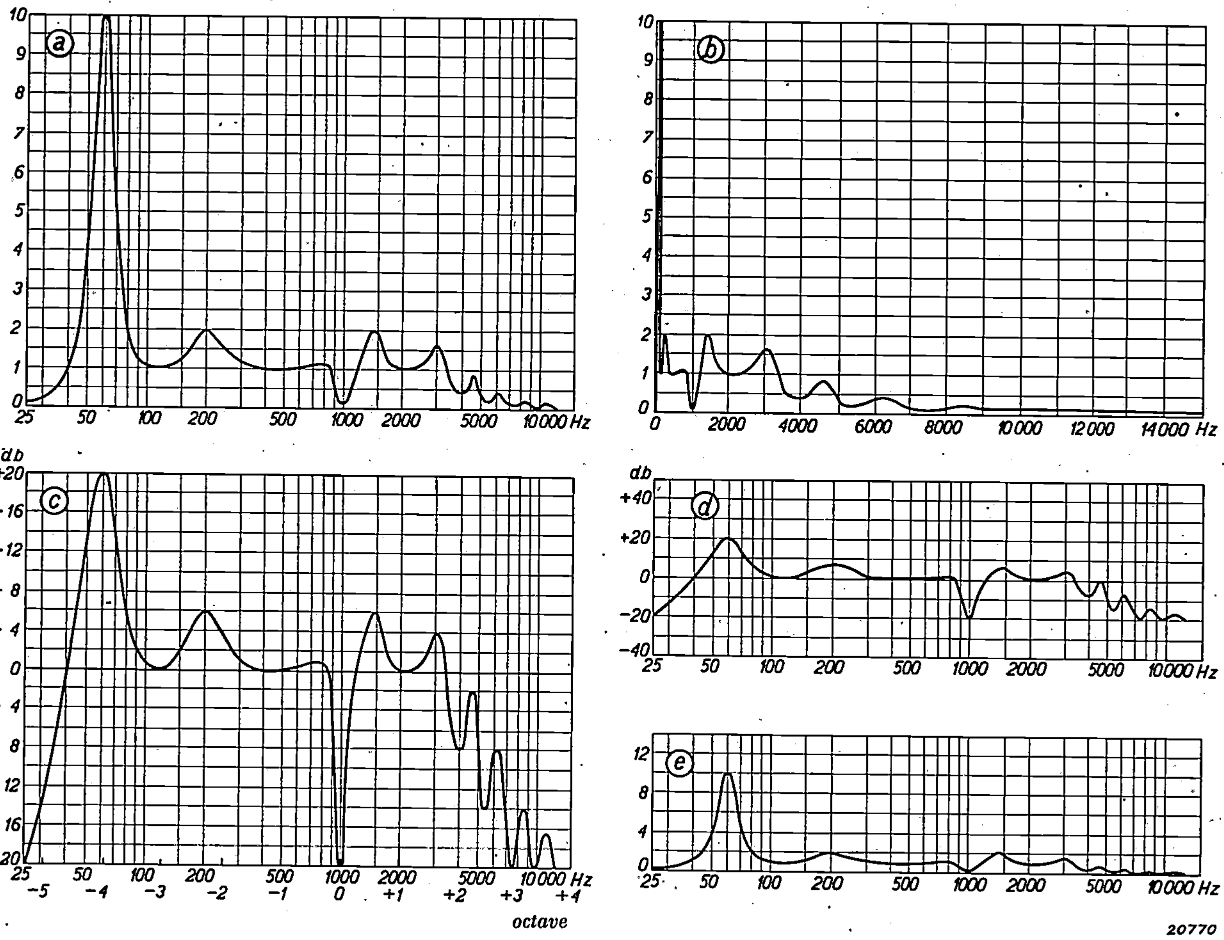


Fig. 7. Representation of one and the same frequency characteristic on linear and logarithmic scales along the abscissa and ordinate.

a abscissa with logarithmic scale, ordinate with linear scale;
b abscissa and ordinate both with linear scales;
c abscissa and ordinate both with logarithmic scales.
d as for *c* but with reduced scale along the ordinate.
e as for *a*, but with reduced scale along the ordinate.

circumstances are responsible for abnormalities.

3) Thirdly (cf. *a* and *c*) it is evident that the linear intensity scale brings out more clearly the peak at 60 cycles than the trough at 1000 cycles, while these are of the same size on the logarithmic scale.

4) The character of the curve above 5000 cycles (cf. *b* and *c*) is almost flat with the purely linear scale, but on the logarithmic scale still reveals marked irregularities which certainly affect the quality of sound reproduction.

5) It is interesting to note that on measurements with e.g. a quarter of the power the shape of the curve is not changed at all if a logarithmic scale is used. The only difference is a displacement of the whole curve downward by 6 decibels. When using a linear intensity scale (cf. *a* and *e*) the curve is however considerably distorted and in case *e* appears to be much smoother.

6) The objection is frequently advanced against a logarithmic scale that the irregularities of a characteristic are smoothed out. In general this is not the case, for although the peaks are reduced in size as compared with the troughs (cf. *a* and *c*) or regarded from the acoustical aspect they are reduced to their true proportions, the magnitude of the deviations yet depends entirely on the choice of unit for the scale (cf. *c* with *e*, where the logarithmic scale brings out greater irregularities, and *a* with *d* where this is the case with the linear scale).

7) The frequency and intensity range which can be conveniently covered in a single diagram is practically unlimited when using a logarithmic scale and with a linear scale is limited to a ratio of 1 : 10.

8) A disadvantage of the logarithmic scale for many applications is that the zero is located at $-\infty$ and also that negative values cannot be plotted. But since the ear itself has a specific

threshold value no objection can be raised to this from acoustical considerations.

9) With a linear scale the numerical value of the intensity of a compound sound is equal to the sum of the numerical values of its components. With a logarithmic scale this is not the case, although on this scale the total amplification or damping (attenuation) can be found directly by expressing the corresponding components in decibels and adding them. In many instances the ability to do this is alone sufficiently important to justify the use of decibels or nepers.

THE PLAYING SPEED OF GRAMOPHONE RECORDS

By J. DE BOER.

If a gramophone record is not played at the correct speed [1]), both the pitch and time of the music will be reproduced incorrectly. At Philips Laboratory a series of experiments have been carried out with the object of establishing the reaction of listeners to this effect. A disc of Beethoven's Fifth Symphony was played at different speeds, and each member of the audience of about 30 persons was then asked whether the speed was correct, too fast, or too slow. The results of these tests are shown graphically in *fig. 1,* where the percentage of the listeners who judged that the record had been played at the right speed is given for different playing speeds.

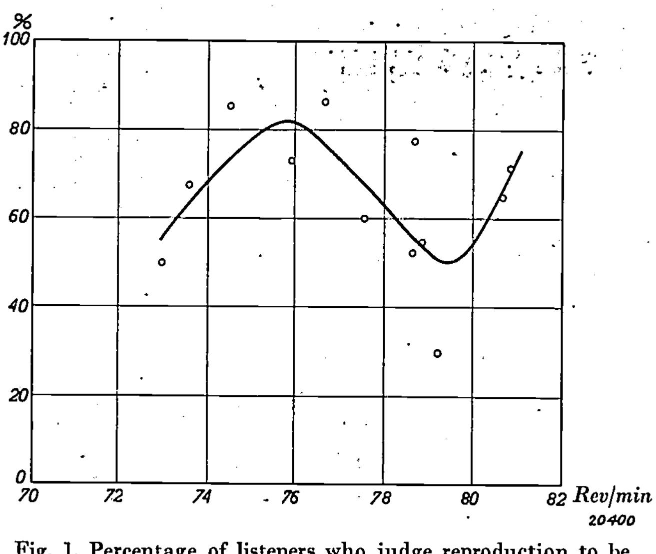


Fig. 1. Percentage of listeners who judge reproduction to be good at different playing speeds.

Owing to the small number of listeners entailed the accuracy of the points plotted is somewhat uncertain, although the shape of the curve was later confirmed in a second series of tests. It may be deduced from this curve that the correct playing speed for this record is 76 revolutions per minute, instead of 78 revs. per min. as printed on the record.

This divergence was confirmed by an observer with a perfect ear. According to this observer also, the pitch (C minor) of the music reproduced was correct when the playing speed was 76 rev. per minute. In addition, the frequencies of two tones on the record, e′ and g″, were also determined objectively. With an equal tempered scale with a′ = 435 cycles, these frequencies should be 308 and 775 cycles respectively. These tones were recorded on a strip of film by the Philips-Miller system of sound recording. The film strip was passed through the reproducing machine in the form of an endless band, so that the pitch could be readily compared by the beat method with a calibrated tone generator. The frequencies were found to have the specified values when the playing speed of the records was 76 ($\pm$ 0.3) revolutions per min. This value is in agreement with the results obtained in the tests with an audience and with those made by an observer with a perfect ear.

The difference between the optimum playing speed and the specified speed is probably not due to the speed of the recording apparatus, making the discs, being incorrect. It is more feasible that the observers took as a basis an equal tempered scale with a′ = 435 cycles, while the orchestral instruments were tuned to a higher pitch.

The result of this investigation demonstrates that the public is able to distinguish whether the pitch and time of the music reproduced are correct. The curve in fig. 1 again shows an increase at a speed of approximately 80.5 revolutions per min., which may be associated with the fact that at this speed the notes are just transposed by half a tone and that the music is then reproduced at a pitch more familiar to listeners than at intermediate playing speeds.

[1]) This speed is defined as the speed at which the record was originally recorded.

threshold value no objection can be raised to this from acoustical considerations.

9) With a linear scale the numerical value of the intensity of a compound sound is equal to the sum of the numerical values of its components. With a logarithmic scale this is not the case, although on this scale the total amplification or damping (attenuation) can be found directly by expressing the corresponding components in decibels and adding them. In many instances the ability to do this is alone sufficiently important to justify the use of decibels or nepers.

THE PLAYING SPEED OF GRAMOPHONE RECORDS

By J. DE BOER.

If a gramophone record is not played at the correct speed [1]), both the pitch and time of the music will be reproduced incorrectly. At Philips Laboratory a series of experiments have been carried out with the object of establishing the reaction of listeners to this effect. A disc of Beethoven's Fifth Symphony was played at different speeds, and each member of the audience of about 30 persons was then asked whether the speed was correct, too fast, or too slow. The results of these tests are shown graphically in *fig. 1*, where the percentage of the listeners who judged that the record had been played at the right speed is given for different playing speeds.

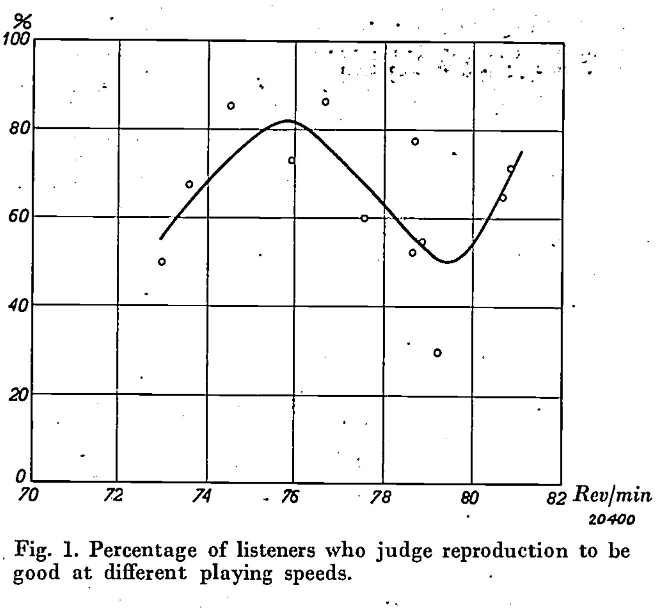


Fig. 1. Percentage of listeners who judge reproduction to be good at different playing speeds.

Owing to the small number of listeners entailed the accuracy of the points plotted is somewhat uncertain, although the shape of the curve was later confirmed in a second series of tests. It may be deduced from this curve that the correct playing speed for this record is 76 revolutions per minute, instead of 78 revs. per min. as printed on the record.

This divergence was confirmed by an observer with a perfect ear. According to this observer also, the pitch (C minor) of the music reproduced was correct when the playing speed was 76 rev. per minute. In addition, the frequencies of two tones on the record, e′ and g″, were also determined objectively. With an equal tempered scale with a′ = 435 cycles, these frequencies should be 308 and 775 cycles respectively. These tones were recorded on a strip of film by the Philips-Miller system of sound recording. The film strip was passed through the reproducing machine in the form of an endless band, so that the pitch could be readily compared by the beat method with a calibrated tone generator. The frequencies were found to have the specified values when the playing speed of the records was 76 (± 0.3) revolutions per min. This value is in agreement with the results obtained in the tests with an audience and with those made by an observer with a perfect ear.

The difference between the optimum playing speed and the specified speed is probably not due to the speed of the recording apparatus, making the discs, being incorrect. It is more feasible that the observers took as a basis an equal tempered scale with a′ = 435 cycles, while the orchestral instruments were tuned to a higher pitch.

The result of this investigation demonstrates that the public is able to distinguish whether the pitch and time of the music reproduced are correct. The curve in fig. 1 again shows an increase at a speed of approximately 80.5 revolutions per min., which may be associated with the fact that at this speed the notes are just transposed by half a tone and that the music is then reproduced at a pitch more familiar to listeners than at intermediate playing speeds.

[1]) This speed is defined as the speed at which the record was originally recorded.

A UNIVERSAL TESTING SET FOR RADIO VALVES

By D. ERINGA.

Summary. A testing set is described which combines the functions of a receiving-valve test unit and a universal measuring unit.

All switching operations are performed automatically by the closing of a contact bridge with 140 contacts initiated after inserting a selected code card into the bridge. The efficiency of a receiving valve is indicated by the deflection of the pointer into either the red or blue area of the scale of the milliammeter. Measurements of voltages, currents, resistances and capacities and the output voltage of wireless receivers can also be carried out with this apparatus. Resistances are measured on direct current and condensers with 50-cycle alternating current.

Introduction

For a long time the need has been felt for a test set for radio valves, which was easy to manipulate and with which the efficiency of radio valves in service could be tested without extensive laboratory equipment. An apparatus of this type is required where a radio dealer wishes to demonstrate to a customer the efficient or defective functioning of a particular valve being used by the latter. The test units hitherto evolved for this purpose were either too complicated for unskilled users or did not afford a satisfactory and adequate test.

The testing set designed by us offers a very practical solution of the problem, since it fulfils the following requirements:

1) It is suitable for all standard types of valves;
2) Manipulation has been made extremely simple;
3) Mistakes in adjustment cannot damage either the valves or the testing circuits;
4) Schedules giving the limits of efficiency and serviceability for different currents and voltages can be dispensed with;
5) Incorrect manipulation of the apparatus does not make the valves appear more or less efficient than they are actually.

General Design and Manipulation of the Apparatus

The general appearance of the apparatus may be gathered from *fig. 1.* It contains the following components:

1) Twelve different valve holders to take almost all standard commercial types of valve bases (European, British and American).
2) A milliammeter with a scale divided in two portions red and blue. If the pointer reads in the red scale the valve is defective, while if it gives a reading on the blue scale the valve is in good condition.
3) A neon lamp for detecting short circuits between the individual electrodes.

Fig. 1. General view of complete testing set.

4) An electric glowlamp for detecting broken filaments.
5) A switch with eight pushes for testing the electrodes for shorts and continuity; this unit is so designed that only one push can be depressed at a time, hence not more than one switching operation can be performed at a time.
6) A potentiometer for adjusting the mains voltage to the correct value.
7) Two connector plugs for a test cord permitting measurements of voltages, currents, resistances, capacities or the output voltage of a radio receiver.
8) A slot for inserting the code cards in the socalled "contact bridge". *All switching operations required for testing a specific type of valve are performed*

automatically when the corresponding code card is inserted in the contact bridge. One or more cards are provided for each type of valve.

A view of the contact bridge is shown in *fig. 2*. One half of the bridge (on the right) consists of a stationary plate with 140 contact pins. The other half can be displaced by means of the handle on the right hand side-wall of the apparatus and has 10 contact bars. When the code card is inserted and the bridge is closed, the contact pins in front of the perforations in the code card make contact with the bars. The card is made of an insulating material, so that where there is no perforation the contact pin is insulated from the corresponding contact bar behind the card.

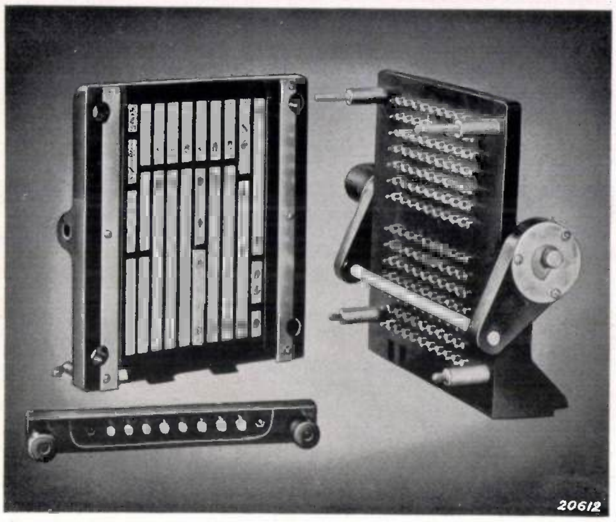


Fig. 2. View of contact bridge.

Fig. 3 gives as an example the code card for the AL 4 receiving valve. At the top of the card the circuit of the valve under test is shown diagrammatically, and indicates with which connector sockets the various electrodes are connected. The distinguishing numbers of the electrodes correspond to the numbers on the valve holders and on the eight-way switch.

The connections set up when the contact bridge is closed after inserting a code card serve the following purposes:

a) Connection to the requisite filament voltage.

b) Choice of anode voltage, and if necessary adjustment of the test voltage for condensers or resistances.

c) Adjustment of the sceen grid voltage, if necessary (with rectifying valves) selection of alternating voltage applied to the anodes.

d) Adjustment of negative grid bias.

e) Fixing of correct loading resistances for measurements on rectifying valves.

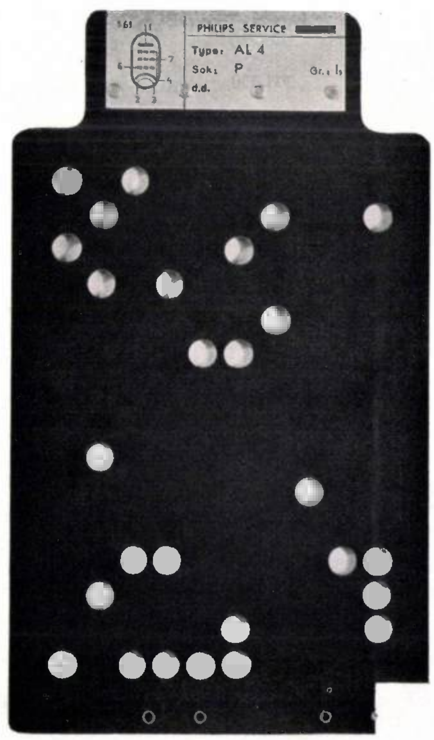
f) Connecting parallel and shunt resistances to the measuring circuit, so that always the same red-blue scale can always be used for various types of valves.

g) Connection to the correct valve holder. In this way each electrode of the valve holder receives the correct voltage through the perforations in the card.

By employing the special circuit under g) above the use of a number of valve holders of the same type is superfluous.

A test is carried out as follows:

The valve is fixed in the holder, and the card inserted in the contact bridge. The glowlamp and the neon lamp L_6, which indicate the results of the test made, are not connected up through the contact bridge, so that for the time being the latter can remain open. The bridge is only closed after the tests with the open bridge have indicated no defect in the valve.



20613

Fig. 3. Code card for an AL 4 receiving valve.

1) Immediately after inserting the valve, test the continuity of the filament;

2) Depress the pushes of the eight-way switch in succession to test for shorts between the electrodes.

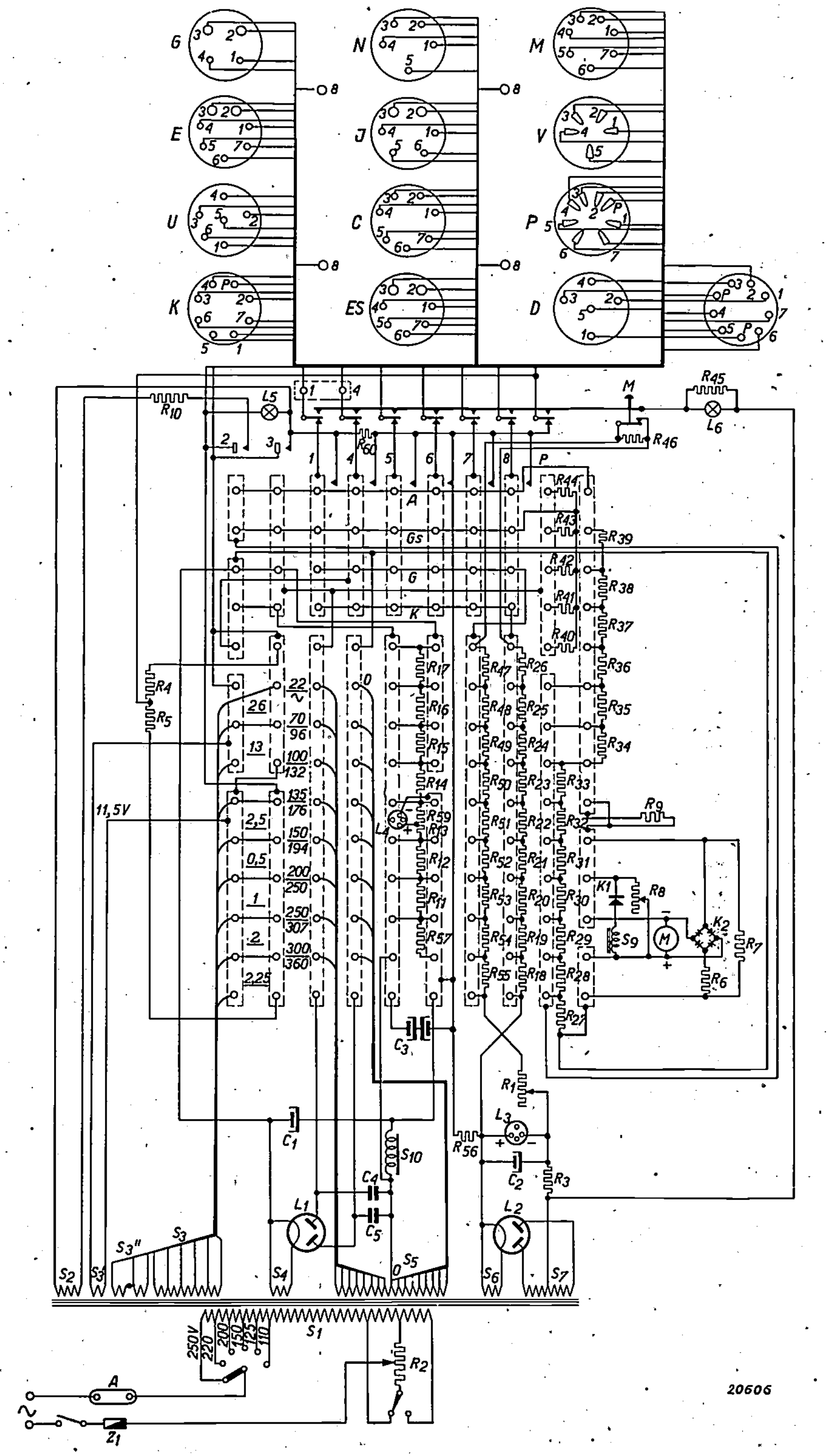


Fig. 4. Circuit diagram for testing set.

If test 1 and 2 give satisfactory results, the contact bridge is closed, and the following tests are then carried out:

3) A test for anode or auxiliary grid current;

4) A test for slope;

5) A test for electrode disconnections or possible bad insulation.

After closing the bridge, manipulation of the apparatus is again limited to pressing in succession the pushes on the eight-way switch.

Explanation of Circuits

The circuit arrangements of the testing set are shown in *fig. 4;* the principal circuits are described below. In the circuit diagram the permanent circuits are indicated by dots, while connections which are set up by means of the contact pins are indicated by a circle.

Filament voltage. The filament voltage is furnished by eight windings marked S_3, S_3', S_3'' in the circuit diagram. These windings are able to furnish all filament voltages from 1 to 56 volts in stages of $^1/_2$ volt, and partly in stages of $^1/_4$ volt. The filament voltage is taken from the first two bars on the left.

Anode Voltage. The third and fourth bars from the left serve for setting up the requisite circuits. The rectifying valve L_1 to whose anodes different alternating voltages can be applied according to requirements, furnishes the anode voltage. The anode-voltage unit has been so designed that up to about 30 mA the voltage is independent of the current tapped.

Auxiliary Grid Voltage. The auxiliary grid voltage is tapped from a potentiometer which is connected to the fifth and sixth bars from the left.

The potentiometer resistance also acts as a load resistance for the aforementioned anode voltage unit. The potentiometer is composed of the resistances R_{11} to R_{17}. The duty of the neon lamp L_4 is to maintain the auxiliary grid voltages constant at 60, 80 and 100 volts.

Loading Resistances for Rectifiers. For testing rectifying valves, resistances R_{11} to R_{17} are used as load resistances. In this case the A.C. supply from the third and fourth bars is not applied to the anodes of the rectifying valve L_1, but to the anodes of the rectifying valve under test.

Negative Grid Bias. The negative grid bias is tapped from the potentiometer in the usual way, the latter being fed from a separate rectifying valve L_2. The seventh and eighth bars from the left are provided for selecting the correct tappings.

To measure the slope of the characteristic, the potentiometer resistance can be adjusted so that that the negative grid bias is increased by 2 volts on pressing the push M. Thus from the difference in the reading on the milliammeter the difference in anode current for a change of 2 volts in grid voltage can be measured, this measurement being sufficient for most purposes.

Milliammeter with Different Shunts. The two last bars serve for connecting up a number of shunts in parallel with the milliammeter. The combinations possible with the different contacts are so numerous (about 70) that the testing ranges "bad" and "good" as marked on the meter scale of the testing set can be retained for all valves and measurements.

Protecting of Testing Set

Since the testing set is intended for the use of unskilled operators, it is naturally possible that a valve may be tested which already has a short-circuit, for instance between the grid and the cathode, or that during test a short of this type develops in the valve. It becomes imperative, therefore, to protect the milliammeter from damage. Since the method adopted by us is not generally employed, it will be described in detail below.

The circuit employed is shown in *fig. 5.* A metal

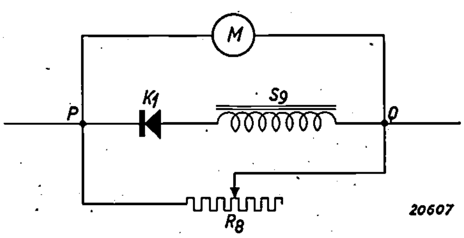


Fig. 5. Circuit diagram for instrument with oxide cell.

rectifier K_1 is connected in parallel with the ammeter as well as a correction resistance R_8. A "blocking layer" rectifier — (cuprous oxide —) cell has the property that its internal resistance depends on the voltage across its terminals. *Fig. 6* shows

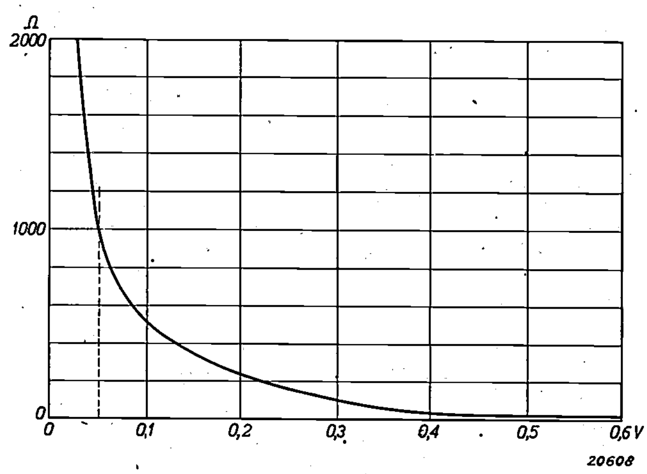


Fig. 6. Characteristics of "blocking-layer" cell.

the resistance in ohms plotted as a function of the applied voltage in volts. If the voltage between P and Q is only of the order of 0.05 volt, the resistance of the cell is about 1000 ohms, and the bulk of the current applied externally flows through the measuring instrument whose resistance is

200 ohms. If, however, the applied current rises considerably the voltage across P and Q does not increase in the same ratio, since as the terminal voltage at the cell increases its resistance diminishes considerably. If the terminal voltage is 0.5 volt, the resistance is only a few ohms. A higher current applied to the terminal P thus does not overload the instrument, for the greater part of the current flows through the cell shunt, whose resistance then has a reduced value.

This solution of the problem introduces, however, a very undesirable secondary factor. The milliammeter is not exclusively used for testing receiving valves, but is also employed for measurements on rectifying valves in which the measuringcurrent is a rectified alternating current, i.e. a pulsating direct current.

As a rule the circuits of rectifying valves are so arranged that the anode current exists only a small part of the cycle. The pulsating direct current is then approximately of the type shown in *fig.* 7. In this diagram the broken line represents

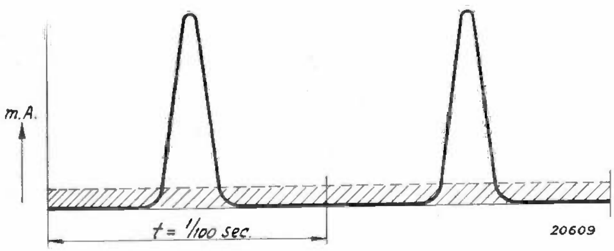


Fig. 7. Current curve for rectifying valves. The broken line represents the mean value indicated by a moving-coil ammeter.

the mean value as indicated by a moving-coil ammeter. The instantaneous value of the current during part of a period is, however, about 10 times greater, and during this interval the voltage also rises between the points P and Q. It follows from the characteristic of the cell under discussion that these peak currents will select the path through the cell instead of through the coil of the measuring instrument with its comparatively greater self-induction. Unless suitable precautions are taken the reading obtained on the instrument when testing rectifying valves will be inaccurate. To eliminate this source of error a small choke of several henries (S_9 in fig. 5) is connected behind the cell K_1. This gives the circuit K_1-S_9 so great a reactance that the resultant error is not too high. It is unfortunate that the coil S_9 is necessary since it has a certain non-reactive resistance, of about 20 ohms. This consequently reduces to some effect the shunting effect of the cell.

Testing a Receiving Valve

A simplified circuit reproduced in *fig.* 8 illustrates how any receiving valve, e.g. one with three

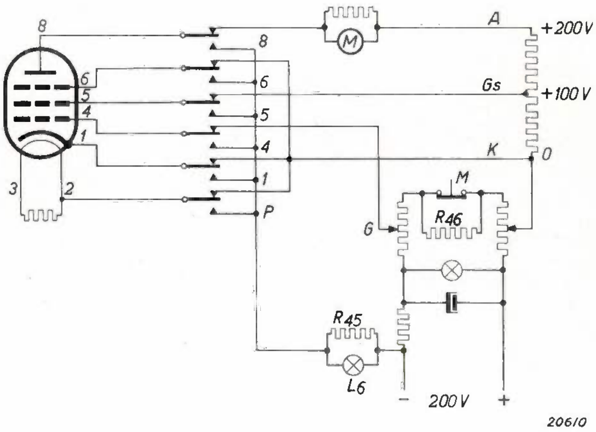


Fig. 8. Circuit for testing a pentode.

grids, has the various voltages supplied to it on closing the bridge. The feed circuit of each electrode passes through a single-pole change-over switch. The seven single-pole switches are the seven pushes of the eight-way switch shown in *fig.* 9. The tests made by means of these switches are together with the slope and anode currents of the greatest importance.

If a push is not depressed the corresponding electrode is connected to its current supply. On depressing the push, the feed circuit of the electrode is opened and the electrode is connected to the communal bar situated under the pushes, this bar having a negative voltage of approximately 200 volts. In series with the communal bar is the neon lamp L_6 with the resistance R_{45} as shunt.

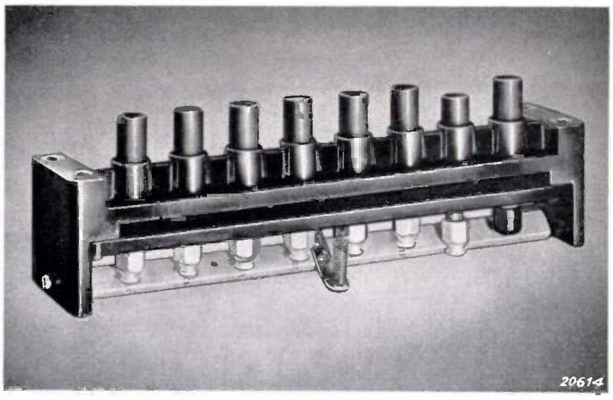


Fig. 9. Eight-way push-button switch.

When no pushes are depressed, the requisite voltages are applied to all of the electrodes and the measuring instrument M is in the anode or auxiliary grid circuit so that the current to either of these electrodes may be tested. What happens when the various pushes are depressed?

Assume that, according to fig. 8, the filament is connected to the contacts *2* and *3*, the cathode to contact *1*, the control grid to contact *4*, the screen grid to contact *5*, the suppressor grid to *6* and the anode to *8*, and that all pushes are out of circuit. With a good valve the instrument M will then give a certain reading. If the pointer lies over the blue part of the scale the anode current of the valve is sufficient. The anode current is then above the specified limit and which is the end of the red measuring range. If now the push marked M is depressed the resistance R_{46} in the potentiometer is added to the negative grid bias circuit, and the bias is increased by 2 volts. The reading of the milliammeter will then be a few divisions lower, and the half number of these divisions for a specific valve and with a specific code card will be a measure of the slope of that valve at the testing or working point.

If push *1* is pressed, the cathode feed circuit is opened and the cathode connected through the neon lamp L_6 to 200 volts negative with respect to the common neutral point of the whole circuit. With a hot cathode the neon lamp L_6 will burn with its maximum brightness. This test is very important since by its means it can be established whether or not the cathode is emitting, also when a disconnection in the anode circuit causes no anode current reading to be obtained. On pressing push *4* (push *1* will then automatically spring back into the off position), a negative voltage of 200 volts is applied to the control grid *4*. If during heating of the valve a short has developed between this grid and the cathode, the neon lamp will burn brightly. If however the grid insulation is good and the grid connexion is not broken, the pointer of the measuring instrument will return to zero. On the other hand, if the grid lead is broken in the interior of the valve, there would be no change in the meter-reading on pressing the push. The break in the circuit may therefore be directly established. Since in this operation the neon lamp L_6 does not light and is therefore non conducting, metallic contact between the grid and the 200 volts terminal has been made for by connecting the resistance R_{45} in parallel with the lamp. Similar contact is made when pressing pushes *5* and *6*. When push *8* is pressed the meter drops back to zero, since the circuit containing the instrument is opened, while the anode is connected to the 200 volts terminal and a test thus made for insulation.

Finally, if, push P is pressed, then with certain types of valves, according to the arrangement of the perforations in the corresponding code card, the cathode heater insulation may be tested on 200 V with L_6 in circuit. If the insulation is inadequate lamp L_6 will light.

Detection of Broken Filament

The lamp L_5 is in parallel (cf. fig. 4) with the filament pins *2* and *3* and is fed from winding S_2 of the transformer through the resistance R_{10}. The voltage drop in R_{10} is of such magnitude that the voltage across lamp L_5 is just correct. On inserting a valve in the holder the filament is connected in parallel with lamp L_5, as a result of which the voltage drop across R_{10} increases to such an extent that the brightness of the lamp is considerably reduced. Immediately on inserting the valve to be tested the lamp L_5 will indicate whether or not the filament is broken.

Universal Measuring Apparatus

It is evident that an apparatus as comprehensive as that described above can also be made to carry out other measurements in addition to valve testing. In designing the apparatus this possibility was given due consideration. Leads for connecting up with current sources which it is desirable to measure can be inserted in the pushes *1* and *4* (fig. 4). To enable alternating currents also to be measured with the instrument incorporated in the test set, provision has been made for connecting a rectifying bridge K_2 in series.

By inserting suitable code cards the testing set can also be employed for the following:

a) Voltage measurements: Alternating current and direct current up to a maximum of 500 volts.

b) Current measurements: Alternating current and direct current up to a maximum of 1 A.

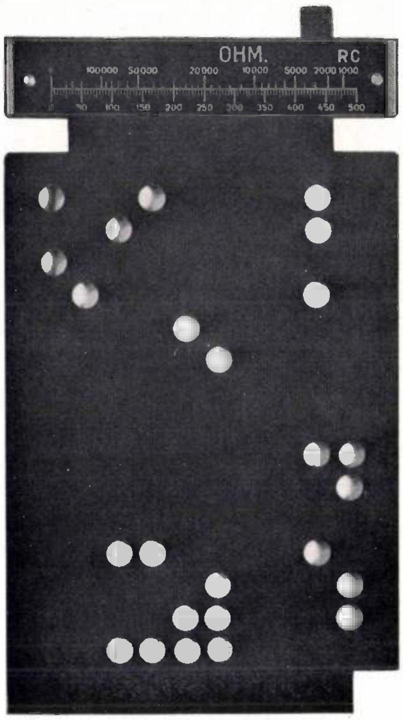
c) Measurements of output voltages of radio receivers.

d) Resistance measurements from 1 ohm to 5 megohms.

e) Capacity measurements from 1000 μμF to 200 μF.

Fig. 10 illustrates a code card for resistance measurements in the range from 1000 to 100 000 ohms.

The apparatus can also detect short-circuits. If the contact bridge is opened and one of the pushes *1* or *4* is pressed, a short in a circuit connected to the terminals *1* and *4* will be indicated by the neon lamp L_6 (see fig. 8). This is a very sensitive method, for the rapid discovery of shorts in a receiver. It is not necessary to use a special code card for this purpose.

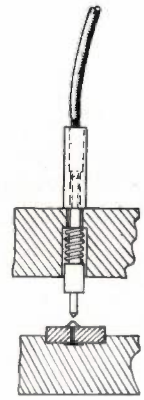


20615

Fig. 10. Code card for resistance measurements from 1000 to 100 000 ohms.

Constructional Details

Contact pins. One of the principal factors in ensuring the efficient operation of the testing set described here is naturally the design of the 140 contact pins. A cross-section through a contact pin is shown in *fig. 11*. The point of the pin is made of silvered brass. If when being inserted exactly vertically, the point comes in contact with one of the flat bars below it, it is quite possible that good contact will not be made should the contact surface and the pin be separated by a microscopically-thin layer of oxide or dirt. The greatest danger of this occurring is in the case of contacts which carry



20611

Fig. 11. Section through one of the 140 contact pins.

no current, such as the leads to the control grid of a receiving valve. To guard against this a conical contra-contact of solid silver is mounted on the bar opposite to the conical apex of the contact pin. On closing the contact bridge these two cones slide over each other during part of the rotation, so that the contact surfaces are self-cleaning.

Contact Bridge. The contact bridge is constructed on such lines that the movable plate can be removed in a few minutes by undoing four nuts and withdrawing a spindle. The contacts are therefore readily accessible, which considerably facilitates maintenance.

Mechanical Lock. Under the contact bridge is a mechanical contact. The main transformer of the measuring instrument is only switched on after a card has been inserted in the contact bridge. This contact is so designed that the apparatus cannot be switched on if a card is inadvertently inserted incorrectly into the contact bridge.

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

No. 1128: J. L. Snoek: Messung der Koerzitivkraft an kleinen Proben (Physica, 3, 855 - 858, August, 1936).

A method is described for the simple and rapid aproximate estimation of the coercivity on specimens of almost arbitrary shape.

No. 1129: W. Elenbaas: Die Intensitätsverteilung und die Gesamtstrahlung der Super-Hochdruck-Quecksilberentladung (Physica, 3, 859 - 871, August, 1936).

The spectral intensity distribution is measured between 0.4 and 3 μ for several types of super-high-pressure mercury discharges. As the pressures increase the spectral lines and the continuous background broaden out, a result also obtained with increase in the current density. The total radiation was measured, a correction being made for the absorption of mercury radiation by the cooling water. Thus, e.g. the total radiation of a discharge in a tube with an internal diameter of 1 mm constitutes 75 per cent of the energy input at 1.1 amp and 800 volts per cm. In conclusion, some reasons are advanced for the fact that the ratio of the intensity of the yellow line λ (5770/5790 Å) to that of the green line (5461 Å) is here smaller than in discharges with a vapour pressure of the order of 1 atmosphere.

No. 1130: K. H. Klaassens and R. Houwink: Viskosität in Lösung und Kondensationsgeschwindigkeit von Phenolformaldehydharzen (Koll. Z., 76, 217 - 223, August, 1936).

Viscosity measurements were made on solutions of resins with a concentration exceeding 5 per cent (socalled Staudinger limiting concentration). From the results obtained conclusions can be drawn regarding the form of the molecules. The viscosity increases with the degree of condensation at constant concentration, a behaviour which can be accounted for the transition from the small molecules in the resin into larger ones.

No. 1131: E. M. H. Lips and J. Sack: A hardness tester for microscopical objects (Nature, 138, 328 - 329, August, 1936).

An apparatus is described by means of which the hardness of microscopical objects (such as inclusions or structural components of metal alloys) can be determined with a Vickers diamond.

No. 1132:* R. Houwink: Technisch-wetenschappelijke beschouwingen over de elasticiteit van verven en lakken (Verfkroniek, 9, 221 - 228, August, 1936).

Report of a paper read before a meeting of the Association for Materials Technology.

No. 1133: J. L. Snoek: Kristalstructuur en magnetisme (Ingenieur, 51, E 81 - 84, July, 1936).

A review is given of theories of ferromagnetism. Magnetic hysteresis only assumes high values when the displacement of the boundaries of the magnetic elementary regions is retarded or prevented. This state may be achieved by the artificial production of a pronounced non-homogeneous voltage distribution in the material. In modern magnetic steels an attempt is made to arrive at this result by quenching to supercool a supersaturated solid solution, after which partial precipitation is induced by heating to medium temperatures. In the case of soft materials, on the other hand, the endeavour is made to obtain a maximum homogeneity; to do this a pure material is chosen which exhibits a low magnetic-striction or in which the latter's disturbing effect has been eliminated (permalloy treatment; cooling in a magnetic field). At the same time the crystal anisotropy must also be made as low as possible.

No. 1134: J. L. Snoek: Ferromagnetische materialen (Ingenieur, 51, E 87 - 89, July, 1936).

The subjects considered in the previous article are here discussed in greater detail with reference to various ferromagnetic materials.

No. 1135: M. J. O. Strutt and A. van der Ziel: Einfache Schaltmassnahmen zur Verbesserung der Eigenschaften von Hochfrequenz-Verstärkerröhren im Kurzwellengebiet (El. Nachr. Techn., 13, 260 - 268, August, 1936).

The principal characteristics of receiving valves in the short-wave range are correlated. An investigation is made of the effect of introducing suitable impedances, particularly in the cathode lead, on these characteristics.

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

VOL. 2 No. 3 MARCH 1937

Philips Technical Review

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

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

ELECTROLYTIC CONDENSERS

By W. CH. VAN GEEL and A. CLAASSEN.

Summary. This article discusses the characteristics of electrolytic condensers and outlines the principles of their operation. Different types of electrolytic condenser for a variety of duties are described.

Introduction

It has been known for several decades that various metals, such as aluminium, tantalum, niobium, zirconium and titanium, can be coated with an oxide film by electrolytic means. This may be done by introducing the metal into a suitable electrolyte, for instance aluminium in a sodium phosphate solution, and passing a current through it, the metal forming the positive pole of the circuit *(fig. 1)*. Oxygen is evolved at this pole and oxidises the metal.

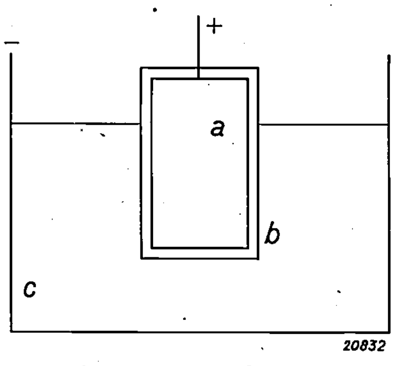


Fig. 1. Diagrammatic construction of an electrolytic condenser. *a* is the aluminium electrode, *b* the separating film of aluminium oxide, and *c* the electrolyte.

The thin film of oxide (Al_2O_3) produced on aluminium offers a very high resistance to the further passage of the current, whose intensity therefore diminishes. If the applied voltage is kept constant the current through the system after a time tends towards a certain minimum value, termed the dissipating current. A cell of this type with aluminium as the positive electrode and an electrolyte as negative electrode can be used as a condenser, the oxide film separating them acting as a very thin dielectric and thus making the resulting capacity comparatively high.

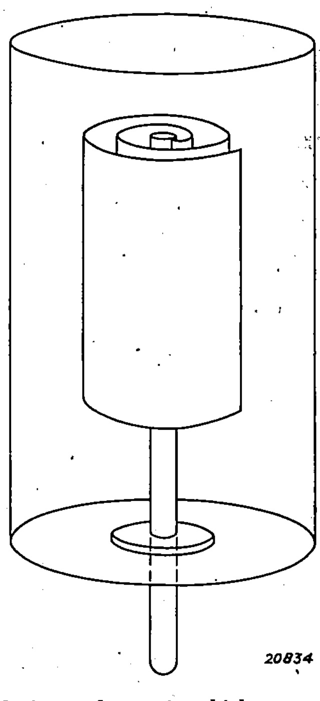


Fig. 2. Electrolytic condenser in which one of the electrodes consists of a spirally-coiled aluminium strip.

In the past it has not been found possible to reduce the dissipating current to such a low value that the heat evolved in the electrolyte could be neglected. The heat evolved was for instance 2 watts with a condenser containing approximately 50 c.c. of electrolyte, and through which a dissipating current of 0.01 A flowed at a potential difference of 200 volts. This signified the evolution of approximately 0.5 cal of heat per sec with a thermal capacity of 50 cals per degree, so that the temperature of a condenser of this construction was raised by about $^1/_2$ degree per minute.

Only after the development of radio technology and as a result of researches in various industrial laboratories were suitable means discovered for sufficiently reducing the above-mentioned dissipating current, e.g. to below 1 mA.

About 1930 electrolytic condensers were introduced for radio purposes. These consisted of an aluminium plate of suitable shape which was covered with a very thin film of aluminium oxide. The aluminium electrode was surrounded by an electrolyte, usually boric acid with some added borate, and the complete condenser enclosed in a small container *(fig. 2)*.

Characteristics of the Electrolytic Condenser

The different characteristics of the electrolytic condenser will first be reviewed.

The electrolytic condenser has a high capacity. The thickness d of the oxide film coating the aluminium is very thin (approximately 10^{-5} cm), and the dielectric constant k of the Al_2O_3 produced is high (about 10). Calculating the capacity c per sq. cm from these data, we get:

$$c = \frac{k}{4\pi d} = \frac{10}{12.6 \cdot 10^{-5}} = 8 \cdot 10^4 \text{ cm.}$$

For an aluminium electrode of 100 sq. cm surface a capacity of already approximately 8 μF can thus be obtained. More detailed reference will be made below to the extent the thickness d of the oxide film depends on the potential employed when forming the condenser, i.e. when producing the oxide film.

The electrolytic condenser can only be employed with a flow of current in one direction. The aluminium electrode must always be connected to the positive side of the applied voltage, and the body of the condenser, i.e. the electrolyte, must therefore always be negative. With current flowing through the condenser in this direction, the current intensity is small; it appears as if the condenser has a slight leak and as if a high resistance were connected in parallel with the condenser. This resistance is determined by the potential. If the direction of current flow is reversed, a strong current will flow through the condenser and the latter becomes useless as such.

It is thus seen that the system exhibits the characteristics of a rectifier, and the condenser does not then differ in any way from the well-known electrolytic rectifier.

As already indicated the condenser coatings consist of aluminium and electrolyte. The specific resistance of an electrolyte is, however, high as compared with that of metals (of the order of 100 ohms per cm). The condenser coating composed of electrolyte has therefore an appreciable resistance which is in series with the condenser.

If the potential difference applied to an ordinary type of condenser, for instance a mica condenser, is progressively raised a puncturing of the dielectric will result. What happens with an electrolytic condenser when the potential difference is similarly increased? If this difference between the aluminium and the electrolyte is steadily raised, a sparkover will be initiated between them at a certain voltage. The dissipating current, which already increased during the stepping up of the voltage, commences to rise much more rapidly when sparking begins. This sparking is analogous to the breakdown of the dielectric in an ordinary condenser, but with the fundamental difference that the flashover does not have destructive results in the case of the electrolytic condenser. The resulting disturbance of the dielectric is in fact instantaneously eliminated by the oxygen which is simultaneously evolved by the dissipating current.

Fig. 3 shows the diagrammatic circuits of an electrolytic condenser. C is a condenser of the same capacity as the electrolytic condenser. The condenser plate a represents the aluminium electrode, the dielectric O the film of aluminium oxide. b is the electrolyte whose resistance R is in series with the condenser. Between the condenser plates a and b there is a resistance r corresponding to the dissipating current and which diminishes with rising voltage.

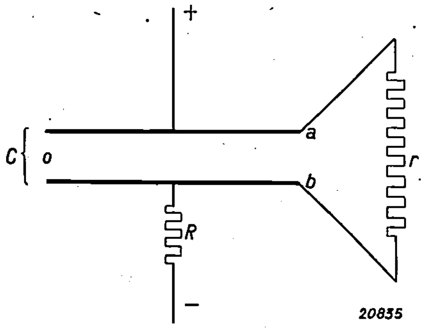


Fig. 3. Diagrammatic circuit of the electrolytic condenser. a is the aluminium electrode, o the oxide-film dielectric of the condenser C, and b the other electrode composed of electrolyte. r is the resistance of the separating layer in parallel with the capacity C; R denotes the resistance of the electrolyte in series with the condenser C.

As a rule the electrolytic condenser is employed in circuits in which a comparatively low alternating current is superimposed on the unidirectional current.

When considering this alternating current the resistance r can be neglected. The series resistance R is composed of the resistance of the electrolyte which is governed by the frequency, and a resistance which we may term the loss resistance in the oxide film. The losses in the oxide film increase with the frequency, the tangent of the phase difference of the oxide film (the resistance of the electrolyte thus being neglected) varying from approximately 2 per cent at 50 cycles to approximately 6 per cent at 1000 cycles. The capacity itself is practically independent of the frequency, and in the band from 50 to 1000 cycles it does not alter by more than 5 per cent.

Explanation of Electrolytic Rectification

The difference in current transmissibility through the system Al — Al oxide film — electrolyte in the two directions may be accounted for as follows: We have already seen that the oxide film is very thin, being of the order of 10^{-5} cm, so that if a potential difference of 100 volts is applied between the aluminium and the electrolyte, the field strength in the dielectric will be about 10^7 volts per cm. With such high field intensities, a "cold emission" always takes place, i.e. the negative electrode emits electrons.

Consider now two plane metal electrodes between which the field intensity F is so high that the negative electrode emits electrons; it is then found that the electron current i can be represented by an equation of the form:

$$i = AF^2 e^{-\frac{B}{F}}$$

where A and B are constants of the materials.

Assume that the plate a emits electrons more easily than plate b. This signifies that when an alternating voltage is applied the current passes through with greater facility in one half wave than in the other half, the greater current flowing when the plate more susceptible to electronic emission constitutes the negative pole.

To rectify a current a thin layer of insulation is therefore necessary and must be bounded by two substances capable of emitting electrons to very different degrees. If the substance which emits electrons the more easily is made negative, a higher current will flow than when it is positive.

Substances are indeed known which emit electrons with different facilities and intensities; metals emit electrons easily, and semi-conductors and electrolytes emit them with difficulty. The electrons in the electrolyte are in fact not free but are linked to ions, although the powerful field obtaining can detach the electrons from the ions and transfer them to the separating layer.

It is thus seen why the electrolytic condenser must always be connected up in such a way that the electrolyte is the negative pole, for then only will a weak dissipating current flow through the condenser. It would be most desirable to have no dissipating current at all, but as we shall see later this is unavoidable. It also follows from the above that the dissipating current will be the lower, the smaller the number of ions present in the electrolyte, i.e. the less current it conducts.

Electrolytes with a high specific resistance actually exhibit low dissipating currents, but these electrolytes have the drawback that the resistance R *(fig. 4)* in series with the condenser is made very great.

A compromise therefore becomes imperative, and consists in fixing the permissible series resistance R and tolerating the corresponding dissipating current. Fortunately in this way very low dissipating currents (less than 1 mA for 450 volts at several μF) have been realised.

It is now evident why it is impossible to use a metal as a second electrode in place of the electrolyte. In such case the separating layer (aluminium oxide film) would be bounded by two substances which emit electrons with almost the same facility, so that a high current would flow in both directions and the condenser be of no practical value. Moreover a breakdown of the dielectric would occur already at low voltages. The specified polarity and the use of an electrolyte are therefore indispensable requisites.

Condensers for Different Voltage Ratings

We have already seen that the dissipating current is generated because the electrolyte is also able to emit electrons when a powerful electric field is applied to it, such electrons migrating through the oxide film to the aluminium. This current is determined by the field strength. If in condensers made of the same materials the dissipating currents are the same at equal potential differences, it may be concluded that the oxide films are of the same thickness. The field intensity is then equal to:

$$F = \frac{\text{applied voltage}}{\text{thickness of separating layer}} = \frac{V}{d}.$$

If an aluminium electrode is oxidised in an electrolyte and a specific potential difference V_1 is applied, then as already indicated at the outset the current through the electrolyte will steadily diminish. At a certain small terminal value i_1 of this current, the oxidation process may be regarded as having terminated. If now a second aluminium electrode with the same dimensions is placed in the same electrolyte and a potential difference V_2, which is double V_1, is applied, and we wait until the dissipating current has attained the same final value i_1, it may then be assumed that in the two condensers the same field strength prevails at the separating film of the electrolyte. Since however $V_2 = 2\ V_1$, d_2 must be 2 d_1 and hence also the capacity of the second condenser half as great as that of the first. With an anode rated for 500 volts the thickness of the oxide layer is 0.6 μ.

Thus with the same area of aluminium we can make a condenser with e.g. a rating of 10 μF at 500 volts, 20μF at 250 volts, 50 μF at 100 volts, etc. One of the principal advantages of the electrolytic condenser is that the thickness of its insulating layer is automatically matched to the potential difference.

Disruptive Voltage of Electrolytic Condensers

Mention has already been made of the fact that sparking occurs with an electrolytic condenser when a certain critical potential difference is applied to it, such sparking being analogous to the breakdown of the dielectric in a standard condenser.

It is observed that this sparking voltage V is determined by the specific resistance (ϱ) of the electrolyte, and for a particular thickness of film is given by the expression:

$$V = a \log \varrho + b$$

where a and b are constants.

The increase of V with ϱ may be interpreted as follows: The greater the concentration of the ions in the electrolyte, the greater will be the current emanating from the electrolyte, and hence the greater the number of electrons migrating to the dielectric, and the easier will breakdown (sparking) develop.

Theoretically therefore condensers with a very high disruptive voltage can be produced by making the specific resistance ϱ of the electrolyte sufficiently great. But the need for none too high a series resistance compels a compromise to be met, as already indicated when considering the dissipating current.

That the condenser does not suffer permanent deterioration from sparking has already been indicated, and this is still the case when the applied over-voltage is very high. While in this case the dissipating current is indeed higher, it yet assumes its normal value again with a normal working voltage soon after being overloaded.

Effect of Temperature on an Electrolytic Condenser

The temperature primarily affects the conductivity of the electrolyte, which rises with the temperature so that the series resistance R diminishes. With temperature increase the density of the free ions and their mobility also increase, resulting in an increase in the dissipating current. Between 20 and 60 °C this current becomes about three times greater, but is then still very small.

Electrolytic Condenser in Use

It has already been stated that the dissipating current continuously generates oxygen at the anode, the gas producing oxidation. The oxide film hence becomes thicker during the use of the condenser and the capacity diminishes. This decrease is however extremely small, since the dissipating current is itself low. Condensers having a liberal rating of 10 μF, after being in use for a year at 450 volts, have exhibited an average reduction in capacity of only 0.8 μF. The dissipating current and the resistance also remain practically constant.

When an electrolytic condenser has been taken out of service for a period of a few months, a fairly high current will be found to flow through the condenser during the first few seconds a potential difference is applied to it. The dissipating current drops very rapidly however, becoming small after a few minutes, and shortly after regains the value which it had when the condenser was taken off load.

Construction of Electrolytic Condensers

The electrolytic condensers made by Philips may be conveniently divided into three main groups differing from each other in construction as well as in their working voltage rating.

The most common type in use, which is termed the star type from the shape of the anode, is made for working voltages of 320, 350 and 450 volts. For working voltages of 500 and 550 volts a special high-tension type of condenser is made, while the socalled low-tension type is designed for working voltages of 25 and 12.5 volts. Details of these different types are given below, as well as some considerations which determined their specific designs.

1) Star Type *(Fig. 4)*

The anode in this condenser consists of a star-shaped aluminium electrode combining a large

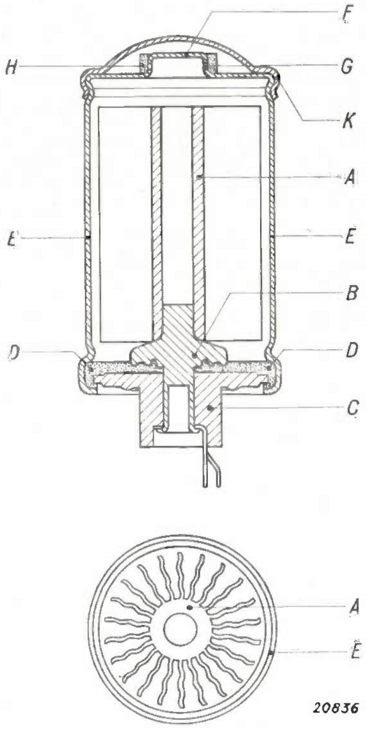


Fig. 4. Construction of a star-type electrolytic condenser. The anode *A* consists of a star-shaped aluminium rod. By means of the pin *B* the anode is rigidly connected with the "Philite" plug *C* which is threaded. The rubber gasket *D* provides a liquid-tight seal between the anode and the "Philite" plug and between the plug and the aluminium case (cathode). The aluminium sleeve *E* contains a valve at the top consisting of the bell-shaped extension *F* in the projecting wall of which holes *G* have been drilled. These holes are closed by the surrounding rubber band *H*. If the pressure in the condenser rises too far, pressure relief is obtained through the holes *G* along the rubber band *H*. The whole assembly is closed with the aluminium cap *K* which contains an absorbent material to take up any liquid trickling out of the valve. The sleeve is filled with electrolyte up to the edge of the anode.

surface with a small volume. A method was developed in the laboratory for increasing the surface of the aluminium by chemical means (etching), as shown in *fig. 5*. By this etching process the initially smooth surface was roughened to such an extent that the active surface was enlarged several times. If aluminium treated in this way is prepared in the usual way, a capacity is obtained which on the average is seven times greater than that with a smooth surface. To this the compact dimensions of these condensers are due.

The condenser case which also acts as the cathode consists of an aluminium cylinder closed at the bottom by a threaded "Philite" plug. This plug also serves for fixing the condenser and as an insulator for passing through the anode to the anode terminal. A rubber gasket giving a perfectly liquid-proof seal is inserted between the aluminium

Fig. 5. Photograph of an etched electrode surface.

case and the "Philite" plug. The whole is filled with electrolyte of specific composition which differs according to the voltage rating. A valve is situated in the top of the aluminium case to allow the escape of gases evolved during service. It opens with an excess pressure less than 1 atmosphere. An absorbent mass is plugged round the valve to absorb any liquid trickling out through the valve, e.g. on high overloads; the whole assembly is closed by a cap.

The following table gives the principal data relating to this type of condenser.

Table I

Capacity	8	16	32	μF
Max. working voltage	450	450	320	volts
Peak voltage	480	480	350	volts
Series resistance at 50 cycles and 20 °C	ab. 70	ab. 45	ab. 10	Ω
Max. permissible alternating voltage at 50 cycles	23	23	16	volts
Max. dissipating current	0.8	1.6	2	mA
Height	49	69	69	mm
Diameter	40	40	40	mm
Weight	90	125	125	g
Max. working temperature	60	60	60	°C

For this type of condenser the dissipating current is plotted against the potential difference in *fig. 6*. Since electrolytic condensers are mainly used to eliminate any A.C. component present in a unidirectional voltage (voltage smoothing in radio apparatus) the maximum permissible alternating voltage is also given in the above table. Care must be taken that in every case the direct voltage simultaneously applied is always greater than the amplitude of the alternating voltage, in other words the anode must never become negative with respect to the cathode. The working voltage in the table represents the sum of the applied direct

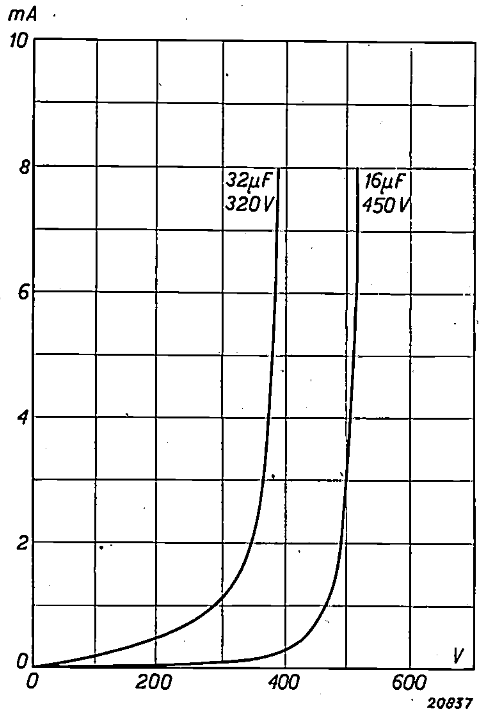


Fig. 6. Dissipating current of an electrolytic condenser plotted as a function of the voltage.

voltage and the amplitude of the alternating voltage. The peak voltage is the voltage which can be safely applied to the condenser for short periods. It should be noted why an electrolytic condenser cannot sustain alternating voltages of unlimited magnitude. The alternating current flowing through the condenser makes the cathode, which is also of aluminium, alternately positive and negative with respect to the electrolyte. As already indicated above the aluminium becomes covered with an oxide film when it is positive with respect to the electrolyte. As a result this alternating current produces a capacity also at the cathode; this capacity is in fact very high, but as it is in series with the anode capacity it may reduce the total capacity. Experiments have shown that with an alternating current density of 0.5 mA per sq. cm (at 50 to 100 cycles) the cathode capacity remains so high after several thousand hours' service that no appreciable diminution of capacity takes place. By submitting the aluminium used for the cathode to special treatment it has been found possible to raise the permissible alternating current to from 4 to 5 mA per sq. cm (at 50 to 100 cycles). It may be noted that oxidation of the cathode ceases at frequencies above 500 cycles.

For this condenser type *fig.* 7 shows the tangent of the phase difference (tan δ) at 50 cycles as a function of the temperature from —10 °C to 60 °C. The decrease in tan δ, or what is the same thing the drop in the series resistance, is due to the diminution of the specific resistance of the electrolyte with rising temperature. The dielectric losses in the separating layer are very small and can therefore be neglected in this connection.

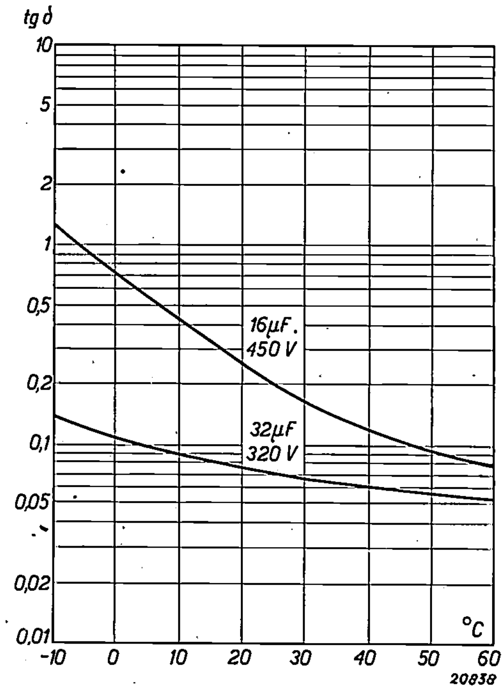


Fig. 7. Tangent of the phase difference (tan δ) at 50 cycles as a function of the temperature.

2) High-Tension Type

For working voltages of 500 volts and over, electrolytes must be used with a fairly high specific resistance (approximately 10 000 ohms per cm at 20 °C). An 8 μF condenser of the star-type would thus give a series resistance of 300 ohms and over, which is far too high. The only means available for reducing the series resistance is to make the distance between the anode and the cathode as small as possible, as has been done in the high-tension type of condenser (see *fig. 9*). The anode, which is a die casting, here consists of an aluminium cylinder which is surrounded internally and externally by the cathode placed at a distance of approximately 1 mm. The characteristics of this type of condenser are given in the following table.

Tabel II.

Capacity	8	8	μF
Max. working voltage	500	500	volts
Peak voltage	550	600	volts
Series resistance at 50 cycles, 20 °C	ab. 40	ab. 60	Ω
Max. A. C. Rating, 50 cycles	30	30	volt
Max. dissipating current	2	2	mA
Height	112	112	mm
Diameter	40	40	mm
Weight	115	117	g
Max. working temperature	50	50	°C

3) Low-Tension Type

These condensers are of similar construction to paper condensers; a cathode and an anode of thin aluminium foil are coiled together, being separated by a layer of paper impregnated with electrolyte, and enclosed in a compressed paper container; the two ends of the coil are compound sealed. The characteristics of this condenser are given in the following table:

Table III.

Capacity	25	50	μF
Working voltage	25	12.5	volts
Series resistance, 20 °C	ab. 6	ab. 6	Ω
Max. A.C. rating, 50 cycles	6	3	volts
Max. dissipating current	50	50	μA
Diameter	17	17	mm
Length	53	53	mm
Weight	14	14	g

Comparison between Electrolytic and Paper Condensers

The advantages offered by the electrolytic condenser over the paper condenser are as follows:

1) Higher capacity per unit of volume, especially at low working voltages; the oxide film can be made very thin by employing lower voltages during surface oxidation, while in the paper condensers a minimum thickness of paper is imperative.

2) While in the paper condenser surges may result in a breakdown of the dielectric and hence the complete destruction of the condenser, the electrolytic condenser can sustain overloads without damage, any disturbance being automatically made good.

The advantages of the paper condenser over the electrolytic condenser are the following:

1) No continuous e.m.f. for polarisation is required, so that there is no dissipating current.

2) Only the paper condenser can be used for a pure alternating current load.

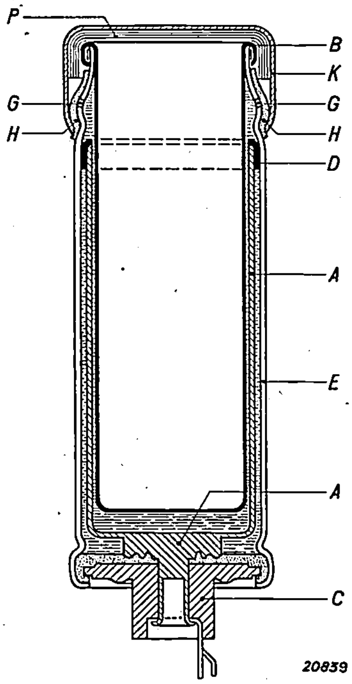


Fig. 8. Construction of a high-tension type of electrolytic condenser. The anode is a die casting and is shaped like an aluminium beaker closed at the bottom with the "Philite" plug C similar to that used in the star type of condenser. The cathode is made in two parts: one part constitutes the outer case of the condenser E, and the other an inner sleeve lapped at B to make liquid-proof contact with E. By adopting this construction the distance between the anode and the cathode is made very small. A strip of insulating material is fixed at D to facilitate the centering of the anode in the outer case during assembly. The valve again consists of a series of holes G in the outer container which are covered by a rubber band H. The valve is separated from the cap K by an intermediate layer of an absorbent material P.

3) The losses ($\tan \delta$) in the paper condenser are smaller than the electrolytic condenser, particularly at high frequencies.

4) For low capacity ratings (below 2 μF) the paper condenser is usually much cheaper than the electrolytic condenser.

In selecting a condenser for a specific duty these relative advantages and disadvantages must naturally be given careful consideration.

TELEVISION SYSTEM WITH NIPKOW DISC

By H. RINIA and C. DORSMAN.

Summary. A television transmitter suitable for transmitting films is described. By using a Nipkow disc, a secondary electron multiplier and a high-pressure mercury lamp a very simple layout of apparatus is achieved, as well as a high line frequency and excellent results in transmission. The electrical components of the system consist of two multipliers, a photo-electric cell and three amplifying valves.

Introduction

The possibility of using a revolving apertured disc for scanning television pictures was described by Nipkow already in 1884, and for a time a disc of this type was fairly extensively used for reproduction purposes. As the number of picture elements is however increased the difficulties of reproducing a satisfactory picture become serious so that at the present day the cathode ray tube is employed almost universally for reproduction.

It has nevertheless been found that the Nipkow disc can be used in a transmitting system with very good results in certain cases, as for instance for transmitting films up to a very high line frequency; its adoption also results in a marked simplification of the components required in the system. This applies in particular when a socalled multiplier tube is used as an amplifier and the high-pressure mercury lamp serves as a light source.

and is projected on the film by the objective lens O. When the disc revolves the light spot sweeps out nearly a horizontal line on the film. The film is moved in a downward direction at a uniform velocity, so that the luminous line produced by the next hole on the film lies above the first line.

The objective lens O consists of two standard film-projection lenses with the sides, which in normal use are directed towards the screen, placed opposite to each other.

The use of two lenses requires in the first place the provision of a wide beam angle. The lenses are designed for use where the rays of the issuing pencil are roughly parallel and also accordingly corrected; such conditions in fact obtain in the Philips experimental transmitter. The disc and film are each situated at the focus of one of the lenses, and the scale of reproduction is determined by the ratio of the two focal lengths.

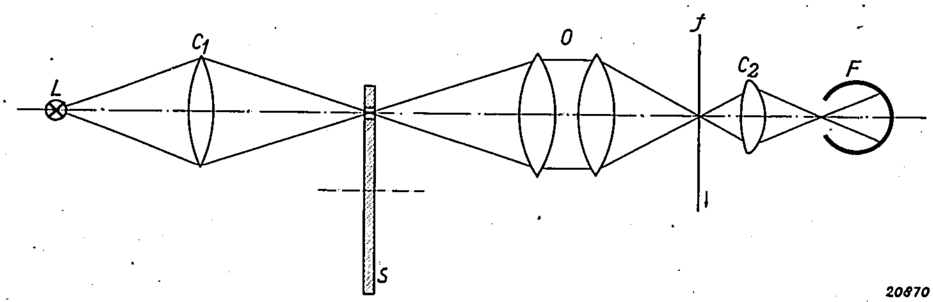


Fig. 1. Path of rays in film-scanning system using the Nipkow disc; L light source (high-pressure mercury lamp); C_1, C_2 condenser lenses; S Nipkow disc; O lens; f film; F photo-electric cell.

Description of the System

The path of the rays in the experimental television unit constructed at the Philips Laboratory is shown diagrammatically in *fig. 1.*

Close to the edge of the disc small holes are located at uniform intervals and at equal distances from the centre. The light from the elongated light source L is thrown on the perforated disc by the condenser lens C_1, in such a way that the line through the centre of the image touches the circle passing through the centres of the apertures. Part of the light thus passes through one of the holes

Finally, the transmitted light is condensed by C_2 and projected on to the photo-electric cell, where an electrical signal is produced, which after amplification is passed to the transmitter. It is apparent that the disc and film can be interchanged without any modification in the principle of the system or any diminution resulting in the maximum light intensity.

Potentialities of the disc

We propose to investigate the maximum amount of light which can be obtained for controlling the photo-electric cell.

Neglecting consideration of the absorption of the film, it is evident that the amount of light falling on the photo-electric cell is determined by the brightness of the lamp, the width of the aperture, the solid angle of the beam included at the disc, and the various losses sustained in the whole optical system.

If the brightness of the lamp is H, the solid angle of the beam passing through the hole Φ, the absorbed fraction of light a and the area of the hole o, the luminous flux incident on the cell is given by the expression:

$$H \cdot \Phi \cdot a \cdot o \quad . \; . \; . \; . \; . \; . \; . \quad (1)$$

Brightness H

An elongated source of light with a high brightness value is required in the system under consideration; a very suitable illuminant for this purpose is the water-cooled high-pressure mercury lamp, which has a brightness value of 30 000 candles per sq. cm, or about three times the brightness of a standard carbon arc.

Solid Angle Φ

The solid angle Φ is limited by the capacity of the optical system, in fact by that lens of the two forming the compound objective O which has the shortest focal length. In our present system this is the lens on the side nearest the disc, so that an enlarged image of the disc is thrown on the film. We have made use of a lens with an aperture of 1 : 2, i.e. $\Phi = 0.2$.

Size of Aperture o

For a given number of picture elements, the size of the aperture is related to the peripheral velocity of the disc at the location of the holes. The width of the holes d_s bears the same ratio to the width d_b of the picture element determined by the scanning method employed, as the velocity of the disc V_s to the scanning speed V_b which is similarly determined:

$$d_s = d_b \cdot \frac{V_s}{V_b} \quad . \; . \; . \; . \; . \; . \quad (2)$$

The peripheral velocity of the disc is however limited by the maximum stresses which the material can sustain and in the apparatus constructed at this Laboratory is approximately 130 m per sec at the periphery of the circle swept out by the holes. The size of the holes is then arrived at as follows:

In *fig. 2* the picture area to be scanned is reproduced; the ratio b/h will be termed β. Assuming that during scanning n horizontal lines are successively traversed and that each hole is a square of side h/n, i.e. with a width equal to the interval between two consecutive lines, we then have:

$$d_b = \frac{h}{n} = \frac{b}{\beta\, n} \, .$$

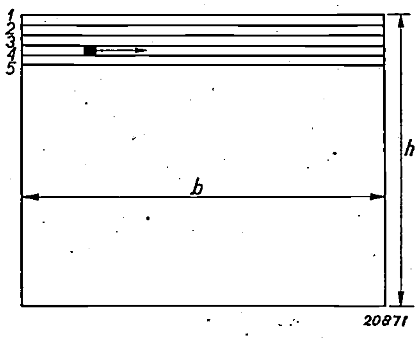


Fig. 2. Traversing of the exploring spot over the surface of the picture.

The total distance to be traversed by the aperture in scanning the picture is $n \cdot b$, i.e. the velocity V_b with N pictures per second is $V_b = N.n.b$.

Since $b = V_b/nN$, we have:

$$d_b = \frac{V_b}{\beta n^2 N} \, .$$

From equation (2) we thus get for the size of the hole the expression:

$$o = d_s^2 = \left(\frac{V_s}{\beta\; n^2\; N}\right)^2 \quad . \; . \; . \; . \quad (3)$$

This formula indicates that the size of the hole diminishes in inverse proportion to the fourth power of the line frequency, and that a high disc speed is desirable.

On inserting the following numerical values:

V_b = 13 000 cm per sec,
n = 405 lines per picture,
N = 25 pictures per sec, and
β = 1.2

we get:

$$o = \left(\frac{13\,000}{1.2 \cdot 25 \cdot 164\,000}\right)^2 = (2.64 \cdot 10^{-3}\ \text{cm})^2 =$$
$$= 7 \cdot 10^{-6}\ \text{cm}^2 \, ;$$

The length and width of the hole are therefore 26.4 μ. The total luminous flux according to equation (1), assuming that $\varrho = 0.5$, is therefore:

$$30\,000 \cdot 0.2 \cdot 0.5 \cdot 7{\cdot}10^{-6} = 2.1 \cdot 10^{-2}\ \text{lumen.}$$

If we assume that the sensitivity of the photo-electric cell for the light-source employed is 20 μA per lumen, the signal strength is found to be 0.42 μA, which as will be shown below is quite adequate.

If the holes in the disc were placed a greater distance apart than the width of the film, and hence the lens throws a reduced image on the film, the cone of light would have a greater solid angle on the side towards the film than on the side towards the disc. When therefore the maximum aperture obtainable with the optical system has been achieved on both sides, it is useless to increase the speed of the disc still further. In practice this limit is usually not reached when using standard film as well as a high line frequency.

Requisite Signal Current Intensity

The permissible minimum intensity of the signal current is determined by the fact that the signal during scanning a picture element must generate such a large number of electrons that the statistical irregularities of the photo-electric effect cannot cause any distortion. In the present case with $\beta . n^2 = 197\,000$ picture elements the duration of scanning an element is:

$$\frac{1}{25 \cdot 197\,000} = 2 \cdot 10^{-7} \text{ sec.}$$

Since the charge of an electron is $1.6 \cdot 10^{-19}$ coulomb, the number of electrons liberated per picture element is:

$$0.42 \cdot 10^{-6} \cdot 2 \cdot 10^{-7} \div 1.6 \cdot 10^{-19} = 500\,000.$$

For the dark portions of the picture the number of electrons may be 40 times less. Yet in the case under consideration the number is always so large that the statistical deviations can be neglected.

Where, for instance, it is desirable to increase the number of picture elements, it is important to know how far the intensity of the signal current can be reduced. If the number of electrons emitted during scanning a single picture element with constant illumination is on an average a, it follows from the theory of least squares that this number varies on the average by $\sqrt{a}$; this places a lower limit to the number of electrons still of practical value. Assuming that in an area of the size of a picture element in the dark sections of the picture, fluctuations of approximately 20 per cent are still imperceptible, the minimum number of electrons is found to be 25. Where a maximum ratio of light to dark of 40 : 1 has still to be transmitted, this gives a total of 1000 electrons at maximum illumination. In the film-reproduction example worked out above the actual number is 500 times greater, so that in this respect a very liberal reserve is still available.

This favourable state of affairs is largely due to the exceptionally high brightness of the light source. If a system of this type were employed for direct scanning, the brightness values in the scanning room would have to be at least 10 000 times smaller, which would introduce considerable difficulties as regards the signal strength.

Amplification of the Signals

To operate a television transmitter the signal obtained in the manner described above must be amplified. In view of the still small strength of the signal currents, care must be taken that the method of amplification does not introduce its own specific distortion. For instance, the signal current could be passed through a resistance and the potential difference at the resistance terminals applied to the control grid of an amplifying valve. But in this case the thermal motion of the electrons in the resistance would produce fresh voltage fluctuations at the resistance terminals, which would be much greater than those due to current fluctuations in the photo-electric cell [1]). Moreover, the amplifying valve itself would introduce its own distorting effect. In spite of these difficulties amplification with receiving valves has yet proved quite feasible in the present system.

An amplifier designed on entirely different principles, viz., the secondary electron multiplier, offers in every respect a very great improvement as well as a simplification in design. This multiplier is briefly described below (*fig. 3*).

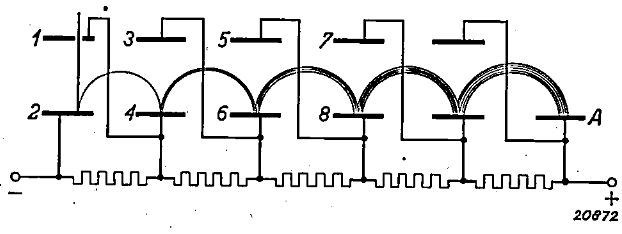


Fig. 3. Diagrammatic sketch of the secondary electron multiplier. A beam of light passes through electrode *1* on to *2* and at the latter causes the emission of photo-electrons. Under the action of the magnetic fields perpendicular to the plane of the paper and the electric fields between the different electrodes, the electrons are forced to move in the manner indicated. Each electron impinging on the electrodes *4*, *6*, *8*, etc., liberates further secondary electrons from these electrodes, so that *in toto* each photo-electron causes about 100 000 electrons to reach the anode A.

---

1) The voltage fluctuations due to thermal motion increase with the square root of the resistance, while the voltages generated by the signal current increase in direct proportion to the resistance. By increasing the resistance, the temperature effects can be reduced, although there is a limit to this reduction due to the adverse grid-cathode capacity of the amplifying valve, which is in parallel with the resistance. With a capacity of 12 and a frequency of $3 \cdot 10^6$ cycles, the total impedance can never exceed e.g. approximately 4500 ohms.

A series of flat electrodes *1*, *2*, *3*, etc., are mounted in an evacuated bulb. Electrode *2* is photo-sensitive and is illuminated through an aperture in electrode *1*. The surfaces of electrodes *4*, *6*, *8*, etc., have been pre-treated in such a way that an electron impinging on them with an energy of, say, 100 volts causes the emission of, for instance, 3 secondary electrons. The electrodes are connected up in the manner shown in the figure. A suitable potentiometer ensures an equal potential difference between all pairs of electrodes, e.g. 100 volts.

A magnetic field is applied perpendicularly to the plane of the figure. The operation of the multiplier is as follows:

An electron emitted by the action of the incident light is attracted upwards by the electric field between *1* and *2*. The magnetic field at the same time draws it towards the right so that it impinges on the electrode *4* with a velocity corresponding to the potential difference between *2* and *4*. At this electrode three electrons are liberated, which are similarly directed to strike *6* and each again causes the liberation of *3* electrons at this electrode. This cumulative action continues, such that with a fairly low number of electrons a current amplification of e.g. 10^5 can be obtained. If a multiplier of this type is used the whole of the photo-electric amplifier can be dispensed with and the root cause of thermal fluctuations is completely eliminated. The whole process takes place with practically no lag, so that even the very high frequencies are transmitted without any attenuation. Since the number of secondary electrons is affected only slightly by the acceleration voltages, the multiplier is practically insensitive to small voltage fluctuations and hence not subject to disturbances. Reaction, which for instance in the case of amplifiers may result in an undesirable oscillation, can also be entirely neglected. Since the multiplier can equally well amplify direct current and alternating current, it is evident that this method of amplification is practically ideal for the present purpose.

Specific Constructional Details

Measured to the circle of holes the Nipkow disc is 35 cm in diameter; it revolves at a speed of 125 r.p.s. Thus for a line frequency of 405, 81 holes are required round the periphery of the disc. The interval between the holes is 13.5 mm, while the width of the film picture to be scanned can be a maximum of 23 mm. The disc is thus reproduced with an enlargement of 1 : 1.7.

The above calculation has been made on the basis of square holes. Thus in scanning, as opposed to reproduction, the shape of the holes is not of paramount importance. In practice round holes are quite satisfactory, and have a diameter roughly equal to the distance between the lines or slightly larger than this interval. The greater ease with which this type of disc can be made is more than compensated by the slight loss in light. In our system the diameter of the holes is 27 μ. It is absolutely essential for the holes to be all exactly the same size, as otherwise disturbing dark and bright lines will be produced in the picture.

To obtain a uniform definition over the whole area of the picture, the holes must lie in the same plane. The high centrifugal force (here 12 000 times the force of gravity) ensures that even if the disc is initially slightly bent it will always rotate perfectly flat. Furthermore to obtain a satisfactory picture, all holes, both in a tangential and in a radial direction, must be in their exact theoretical positions (permissible error 5 μ) as well as retain these positions during rotation of the disc. To reduce the effects of air friction, the disc is enclosed in a iron housing which is evacuated (a residual pressure of about 1 cm of Hg is sufficiently low).

The mechanism for transporting the film is very simple, as the film motion must be continuous. The film is drawn across the film gate by means of a sprocket rotated at a speed of 1500 r.p.m. through two gears by a small synchronous motor. Contrary to a standard projector, it is unnecessary here to provide a drive for the sprocket feeding the film to the gate. Sound scanning with sound films requires no special arrangements for maintaining a perfectly uniform film speed, in view of the continous film motion employed.

Further Amplification of the Signals

To pass the picture signals obtained through a cable to the television transmitter without distortion, they must have a voltage of the order of several volts. The end of the cable is terminated with its impedance (here approximately 100 ohms), in order to avoid reflections.

The maximum signal which the secondary electron multiplier used can furnish is about 2 to 3 mA. The circuit used for amplifying these signals is shown in *fig. 4*. The current from the multiplier flows through a resistance R_1 connected across the grid and cathode of the valve L_1. The signal voltage at the grid of the valve can therefore be raised by making R_1 large. For the reasons given in footnote [1]) the magnitude of the resistance which can be used is however limited. It was found feasible in the present system to use a resistance of $R_1 = 2500$ ohms, when a maximum signal of 5 to 7 volts is obtained at the grid and — with a cable with an impedance of 100 ohms and an amplifying valve slope of 10 mA per volt — the same

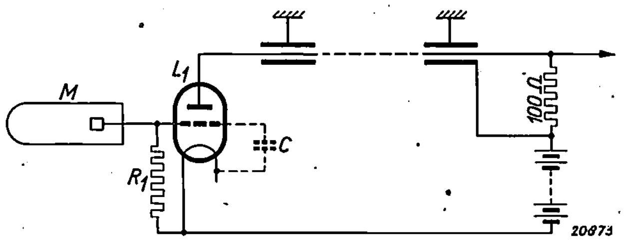


Fig. 4. Transmission of the picture signal generated by the multiplier to the television transmitter. The multiplier current generates voltages at the resistance R_1 which are ampified by the valve L_1; c is the adverse capacity of the valve and limits the frequency which can be transmitted.

voltage at the terminating resistance of the transmission line. As shown in the circuit diagram, the valve is connected up as a D.C. amplifier.

Synchronising Signals

In addition to the picture signals, picture and line synchronising signals must also be transmitted (as to the form of these signals, see the articles by Van der Mark and Richards in this Review [2]).

The synchronising signals are generated as follows:

A disc is mounted on the shaft of the synchronous motor which drives the film; it interrupts the beam of light between a lamp and a photo-electric cell, so that only during the time between two pictures does the light pass through a hole in the disc on to the cell. The cell current furnishes a signal which is amplified and which is applied to the line in the same way as the picture signal. The line synchronising signal is generated in the same way, using however the apertures in the Nipkow disc. Images of the holes are thrown by means of a second optical system on a fixed aperture behind which a secondary electron multiplier is again set up. It is apparent that in this way a short rectangular signal is produced. The circuit is in all other respects exactly the same as that for picture signal amplification. To suppress the line synchronising signals during picture synchronising, a revolving diaphragm is arranged in the path of the beam used for the line synchronising singals.

Apart from any supervisory apparatus required, two secondary electron multipliers, a photo-electric cell and three amplifying valves are employed to raise both the picture signals and the synchronising signals to the level required for transmission through the cable to the transmitter.

[2]) J. van der Mark, Philips techn. Rev. 1, 325, 1936. C. L. Richards, Philips techn. Rev. 2, 33, 1937.

MAGNETIC CORES FOR LOADING COILS

By J. L. SNOEK.

Summary. Following a brief discussion of the technical reasons for the use of loading coils with magnetic cores, the characteristics of solid and subdivided cores are analysed with special reference to their use in loading coils. To avoid eddy current losses a subdivision by gaps in the direction of the field is necessary. Hysteresis losses and instability are considerably reduced by gaps perpendicular to the lines of force, but can also be brought down to the desired levels in the case of a nickel-iron alloy by suitable mechanical and thermal pre-treatment. In conclusion the magnetic state of this alloy is analysed more closely and the possible causes of its characteristic properties are discussed.

Magnetic Cores are Necessary for Loading Coils

The function of loading coils is to reduce the attenuation in telephone cables and at the same time make the attenuation independent of the frequency within the band of speech frequencies. As has already been shown in a previous article [1]), this end may be attained by raising the selfinduction of the line, provided the line resistance is not increased too much as a result thereof. The following equation was deduced in the previous article for the attenuation constant:

$$\alpha = \sqrt{{}^1/_2 \sqrt{(R^2+\omega^2L^2)(G^2+\omega^2C^2)} + {}^1/_2 (RG-\omega^2LC)}, \quad (1)$$

where L is the self-induction, R the resistance, C the capacity and G the dielectric conductivity, all expressed for unit length of the cable. This equation also applies for moderate frequencies when a part of the resistance and the self-induction are not uniformily distributed but are concentrated at loading coils situated at uniform intervals along the line.

An investigation will be made on the basis of equation (1) of the self-induction required in loading coils (and the permissible resistances) in order that these coils effectively reduce the attenuation at speech frequencies. To simplify this analysis G will be neglected compared to ωC (in practice G is approximately $1/10\,000\,\omega C$ at speech frequencies) so that we get the simplified expression:

$$\alpha = \sqrt{{}^1/_2 R\omega C} \cdot \sqrt{\sqrt{1 + \frac{\omega^2L^2}{R^2}} - \frac{\omega L}{R}} \, . \quad (2)$$

In a non-loaded line $\omega L \ll R$ for the speech frequencies, so that the second root of the equation (2) is nearly unity and hence $\alpha = \sqrt{{}^1/_2 R\omega C}$. If $\omega L/R$ is increased, α becomes smaller, while if $\omega L \gg R$ we get to a first approximation:

$$\alpha = \sqrt{{}^1/_2 R\omega C} \cdot \sqrt{\frac{R}{2\omega L}} = \frac{R}{2}\sqrt{\frac{C}{L}} \quad . \; . \; . \quad (3)$$

The attenuation is therefore reduced by the factor $\sqrt{2\omega L/R}$ and is independent of the frequency [2]).

To determine the requirements to be met by loading coils, the resistance is resolved into two components, viz., the component R_l due to the line and the component R_p due to the loading coil, as well as the self-induction neglected in the first case. We then have with $\omega L \gg R$:

$$\alpha = \sqrt{{}^1/_2 \omega (R_l + R_p) C} \cdot \sqrt{\frac{R_l + R_p}{2\,\omega\, L_p}}$$

$$\alpha = \alpha_l \left(1 + \frac{R_p}{R_l}\right) \cdot \sqrt{\frac{R_l}{2\,R_p \omega}} \cdot \sqrt{\frac{R_p}{L_p}}\, , \quad . \quad (4)$$

where α_l is the attenuation of the non-loaded line.

The general specification as regards loading coils is that they must not raise the resistance of the lines by more than 15 per cent and they must also reduce the attenuation at 800 cycles to at least a quarter. Taking $R_p/R_l = 0.15$ we then have from equation (4)

$$\alpha/\alpha_l = (1 + 0.15) \sqrt{\frac{1}{2 \cdot 0.15 \cdot 2\pi \cdot 800}} \cdot \sqrt{\frac{R_p}{L_p}}\, .$$

If furthermore we put $\alpha/\alpha_l < 1/4$ we have:
$R_p/L_p < 75$ ohms per henry.

[1]) W. Six, Philips techn. Rev. 1, 357, 1936.

[2]) As was shown earlier (footnote 1), the attenuation is made entirely independent of the frequency when $LG = RC$. Since G is very small, this equation can generally only be satisfied in the case of telephone lines by providing a very high self-induction. Nevertheless, it is seen here that it is unnecessary to satisfy this condition. The attenuation is made sufficiently independent of the frequency if $\omega L \gg R$, which is in fact already the case in the speech-frequency range with much lower self-induction values.

If the attempt is made to build self-inductions with such low resistance values without introducing magnetic cores, the space required for the requisite cross-section of copper of each separate coil is found to be of the order of several cubic decimetres. Apart from the unnecessary waste of copper entailed, there is also the difficulty of finding accommodation for a large number of such coils in a buried cable box. In addition to the greater space requirements, air-core coils cannot be piled as closely as coils with magnetic cores, since their mutual interference is so much greater owing to their high leakage that cross-talking between one cable circuit to another results. To reduce the space requirements and the leakage of the coil it is imperative to raise the induction flux with the aid of magnetic cores, although these latter possess a number of undesirable characteristics as will be indicated below.

Distortion and Energy Losses with Magnetic Cores

If the air core of a coil is replaced by a magnetic core, the induction B, which in the air core is equal to the field intensity H, becomes multiplied by a specific factor; the same applies to the self-induction of the coil. The factor is termed the permeability of the core.

If B and H were exactly proportional to each other, i.e. if μ were constant, the coil would have exactly the same properties as an air-core coil, apart from its higher self-induction value.

But with solid ferro-magnetic materials the relationship between B and H is by no means so simple. In most cases the B-H curve is curved which results in non-linear distortion being produced. Moreover the relationship between B and H is not single-valued, since the induction in an alternating field with diminishing field intensity is on the average higher than when the field intensity is the same but increasing. This phenomenon which is termed hysteresis causes energy losses and also augments the non-linear distortion effects. A third undesirable property of the solid magnetic core is its electrical conductivity. The eddy currents induced in the iron represent an additional energy loss, which is exhibited as a resistance increasing with the square of the frequency.

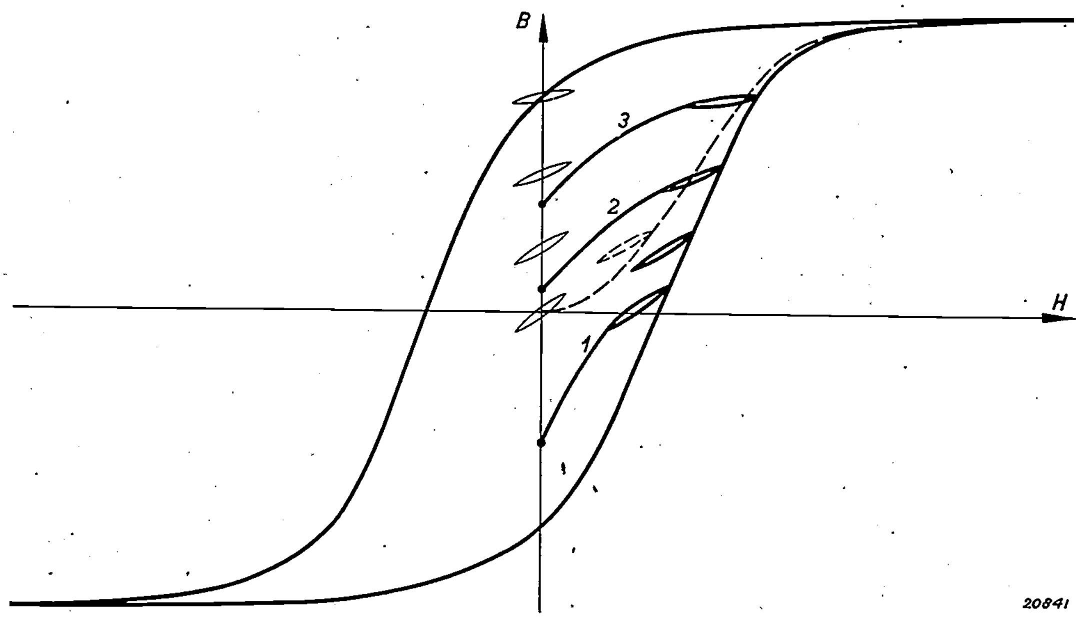


Fig. 1. Magnetisation curve of a hard magnetic iron. Depending on its magnetic history, every state can be realised corresponding to a point within the hysteresis loop. The small loops are traced out when the material is subjected to the same magnetic alternating field commencing from different points within the hysteresis loop.

Instability

The above discussion does not by any means exhaust the complex relationship between induction and field intensity shown in *fig. 1*. The first striking point indicated in the figure is that as the field intensity is raised the induction tends towards a saturation value. On reverse magnetisation different curves are traced out in the two directions, so that with a sufficiently high amplitude for the field intensity the lower loop in the figure is obtained.

If the field intensity is allowed to fall from smaller maximum values to zero, the induction values will follow the curves *1*, *2*, *3*, according to the initial values attained. Thus with a zero field a certain residual induction is retained.

The small loops at different parts of the diagram show the variation of induction when a weak alternating field is applied to the iron after submitting it to the afore-mentioned pre-treatment. The mean gradient of these loops is termed the *A.C. permeability*. This A. C. permeability differs in the different parts of the *B-H* diagram. R. Gans was the first to call attention to the remarkable fact that this permeability value is determined almost wholly by the magnitude of the induction and is only very loosely related to the field intensity. (The loops with the same ordinate in the figure are therefore parallel). In general the A.C. permeability is a maximum for $B = 0$, at first diminishing slowly with increasing induction and then more rapidly.

It follows from the above facts that the A.C. permeability in a zero field (which alone enters into consideration on practice) is determined by the cycles of pre-magnetisation, the effect being the greater the higher the residual induction. The resulting relationship between the self-induction of a coil and its previous magnetic history is termed its instability.

Improvement of Magnetic Cores

Efforts have been made in two directions to ameliorate the shortcomings of solid iron cores.

In the first place, an appreciable improvement in the characteristics of the core were achieved by suitably subdividing it by leaving gaps occupied by an insulating non-magnetic material. In this method marked differences are however observed between:

a) The effect of a gap parallel to the lines of force, and
b) The effect of a gap perpendicular to the lines of force.

As will be seen below, gaps running parallel to the lines of force cause a reduction in the eddy-current losses without the magnetic properties of the core being affected. The hysteresis losses and the instability can, however, be considerably reduced by putting the gaps perpendicular to the lines of force. No special requirements have to be met by the material of the core where this method is employed, except that it must have a maximum permeability.

It was furthermore found possible to reduce to negligible values both the hysteresis and the instability of the magnetic material by employing a suitable alloy and submitting it to suitable mechanical and thermal treatment. The gaps under b) above are then eliminated, while those under a) remain.

Gaps parallel to the Lines of Force

Fig. 2 shows a part of a core which is subdivided by parallel insulated gaps running parallel to the lines of force. The induction produced in the core at a specific field intensity is not affected by these

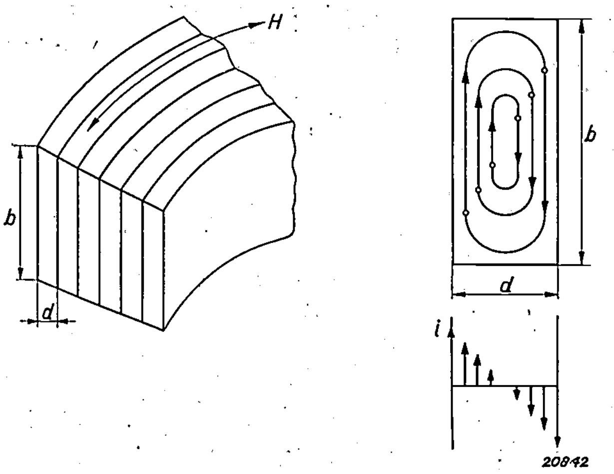


Fig. 2. Part of a coil core subdivided by gaps in the direction of the field *H*, which reduce the eddy currents. The diagram on the right illustrates the paths of the eddy current in a lamina.

gaps, so that they also do not influence the magnetic properties. On the other hand, it is readily seen that the eddy-current losses can be made as low as desirable by subdividing the core sufficiently. The path of the eddy currents is shown in the figure. When the induction B is sinusoidal relative to the frequency ω in radians, the peripheral voltage at the edge of each element of the subdivided core will be:

$$u_{eff} = \omega\, B_{eff} \cdot d \cdot B \cdot 10^{-8} \text{ volt.}$$

The energy q developed per second in unit volume is however equal to the square of the electric field intensity divided by the specific resistance ϱ. The circumference is roughly $2b$, and the effective peripheral field intensity is hence $u_{eff}/2b$. We thus have:

$$q = \left(\frac{u_{eff}}{2\,b}\right)^2 \cdot \frac{1}{\varrho} = B^2_{eff}\,\omega^2\, d^2/4\,\varrho \cdot 10^{-16} \text{ watts/cm}^3.$$

Towards the centre the eddy currents become weaker and weaker, while the heat evolved also diminishes. A very simple calculation based on the current distribution shown in the figure indicates that the mean eddy-current losses per cub. cm are $^1/_3$ of the value calculated above. The losses thus increase with the squares of the frequency, induction and thickness of strip. By employing a sufficiently fine subdivision the losses can be reduced to a negligible and harmless amount.

It should be noted that a subdivision of the core into wires instead of into strips does not make for any considerable improvement. The eddy-current losses of a core made of strip material are only twice as great as those of a core made from wires with the same thickness.

Gaps Perpendicular to the Lines of Force

If the lines of force are intersected by gaps perpendicular to the field (as shown in *fig. 3)*, the induction is reduced as a result thereof. But a greater decrease is obtained in the instability and the hysteresis, so that by maintaining the self-induction constant the hysteresis losses can be reduced. This may be shown from fig. 3 by investigating how the induction in the core is determined by the mean field intensity H, i.e. by the number of ampere turns.

The field intensity in a sufficiently small gap perpendicular to the lines of force is equal to the

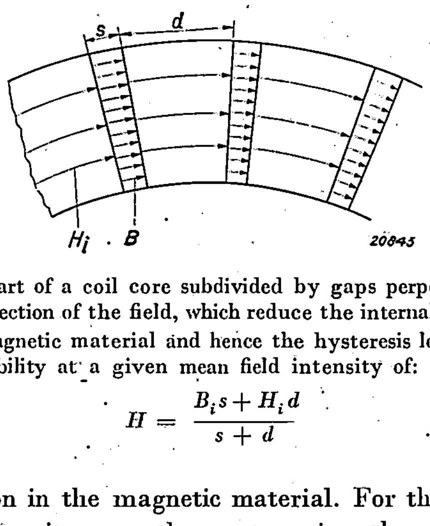


Fig. 3. Part of a coil core subdivided by gaps perpendicular to the direction of the field, which reduce the internal field H_i in the magnetic material and hence the hysteresis losses and the instability at a given mean field intensity of:

$$H = \frac{B_i s + H_i d}{s + d}$$

induction in the magnetic material. For the mean field intensity we thus get, using the notation adopted in fig. 3:

$$H = \frac{B s + H_i d}{s + d} \quad \ldots \ldots \quad (5)$$

If the permeability of the magnetic material is sufficiently high, the induction B will be a multiple of H_i so that with a gap which is not too wide we have:

$$H_i\, d \ll B s \quad \ldots \ldots \quad (6)$$

We can then neglect $H_i.d$ in equation (5) without making a serious error, which gives the equations:

$$B = \left(1 + \frac{d}{s}\right) H, \quad \ldots \ldots \quad (7)$$

i.e. a straight line relationship between B and H. The hysteresis and instability disappear and for the permeability we get independent of the properties of the material:

$$\mu = 1 + \frac{d}{s} \quad \ldots \ldots \quad (8)$$

Actually the internal H_i field never completely disappears, so that a certain instability and hysteresis always remain. It may be readily seen that these effects will be the smaller, the greater the inequality (6) is made. It is thus possible to improve the inequality by increasing the ratio s/d, yet by doing so the permeability is reduced as indicated by equation (8). As a result the total losses in a loading coil increase again when the gaps s are made too large, for with diminishing permeability and given dimensions for the core and coil the number of turns required to arrive at a specific self-induction increases and hence the copper losses become greater. The total losses are a minimum when the iron losses are equal to the copper losses.

Compressed Cores

The system of twofold gaps described above has been realised in the socalled compressed iron powder cores or briefly compressed cores. These cores are composed of powdered granules which are moulded under pressure to the requisite shape with a non-magnetic binder. To avoid eddy currents the grains must be satisfactorily insulated against each other. A high permeability is also desirable, and requires the greatest care as regards the density of packing.

Finally it is desirable for the grains to be well rounded and free from sharp edges and corners, for at these irregularities high local induction values can be created in the material, thus introducing additional hysteresis losses. Moreover, irregularities in contour make effective insulation more difficult.

The degree of subdivision is not only of importance as regards the eddy current losses, for the ratio of the iron path to the gap path above determines the intensity of the hysteresis effects and not their absolute values. Yet if the gaps are too large stray fields and an assymetrical distribution will result, and these must be avoided at all costs in loading coils. By making the core of a compressed mass of small grains in which the stray fields can be neglected, a simple solution of this difficulty has been arrived at.

Loading-Coil Cores without Gaps perpendicular to the Lines of Force

It has already been reported in this Review that the Philips Laboratory has succeeded in improving the magnetic properties of a certain nickel-iron alloy by means of a combined rolling and heat-treatment process to such an extent that gaps perpendicular to the lines of forces are superfluous. The material therefore does not require pulverising, and to avoid the production of eddy currents it is sufficient to adopt the subdivision shown in fig. 2. To arrive at this result the loading-coil core is wound of a strip material, the surface of which has been electrically insulated by a coat of varnish. The permeability of the loading-coil cores made in this way is approximately 90, and is much smaller than with other kinds of soft iron, such as those employed for transformer cores, but yet double the value obtained with the compressed cores available at the time, so that the discovery of the new material allowed an appreciable reduction in the

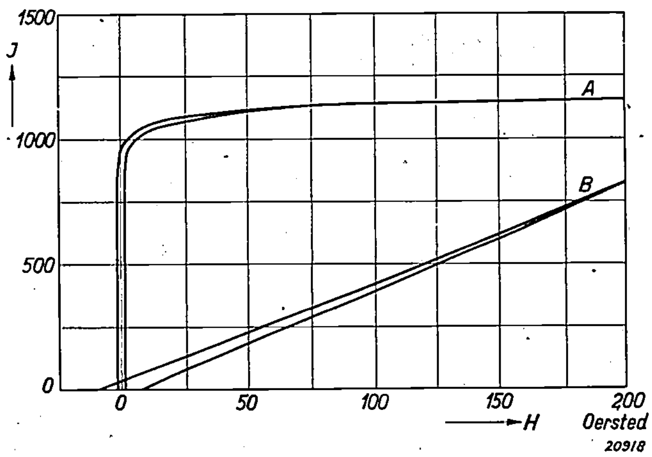


Fig. 4. Magnetisation curves for nickel-iron strip for loading coils.
A. Magnetic field perpendicular to the direction of rolling.
B. Magnetic field parallel to the direction of rolling.

dimensions of loading coils. *Fig. 4B* reproduces a magnetisation curve of the new material in the direction of rolling, i.e. in the direction of the lines of force passing through the loading coil. This curve shows a very small remanence, so that since only small remanent induction values can be obtained a very low instability may be expected in view of the above-mentioned relationships between the A.C. permeability and the induction. The curve also shows that the relationship between B and H is linear over a wide range.

An entirely different behaviour is shown by the curve when the direction of magnetisation is turned through 90 deg. in the plane of the strip. The magnetisation curve then obtained is of the type shown in fig. 4*A*, which, in contradistinction to the curve in fig. 4*B*, reveals an extremely high remanence (90 per cent of saturation) and a pronounced curvilinear relationship between B and H.

Measurements have confirmed the supposition that constant values for the A.C. permeability, instability and hysteresis have a much more undesirable effect on magnetisation in this direction that in the first case.

In the third principal direction, i.e. perpendicular to the plane of the strip, the magnetisation curve can only be measured if special precautions are taken, and is then found to be of the same type as the curve obtained in the direction of rolling (fig. 4*B*).

Magnetisation curves for directions other than in that of rolling have no practical significance, since magnetisation takes place exclusively in this direction. Particular interest attaches to them, however, for elucidating the magnetic conditions since they give an indication as to how the optimum properties are obtained in the direction of rolling.

We shall, therefore, make brief reference again to the general causes of remanent induction and hysteresis.

Causes of Hysteresis and Remanence

P. Weiss assumed that in ferromagnetic materials there is a force which compels the atomic magnets to take up the same direction in specific zones (*Fig. 5*). These Weiss zones are hence magnetically saturated; their individual directions of magnetisation are however in general quite haphazard, such that the body when viewed microscopically appears to be unmagnetised. According to Weiss the direction of magnetisation is capable

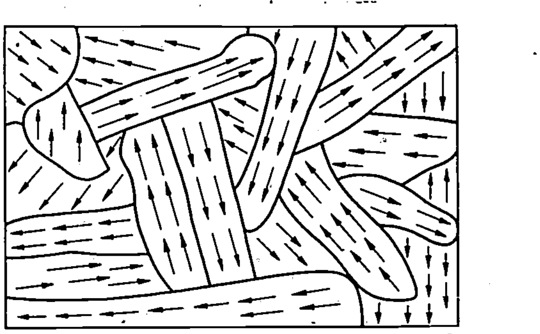


Fig. 5. Weiss zones. The unmagnetised iron is composed of submicroscopic zones which are magnetically saturated. The directions of magnetisation are however uniformily distributed such that the magnetised condition produces no external field.

of free rotation, the only restriction being that the prevailing direction must be the same for all atoms in each zone. If therefore a weak magnetic field is applied to the iron, the directions of magnetisation in all Weiss zones will be rotated in the direction of the field, with the result that the iron becomes immediately magnetically saturated.

There are indeed certain kinds of iron which exhibit this behaviour, while with other kinds powerful fields are required to obtain saturation. The reason for this was given by R. Becker, who demonstrated that the magnetisation in these materials is not devoid of directional control,

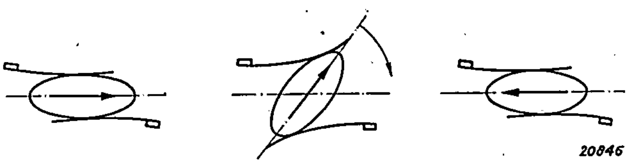


Fig. 6. Diagram of the forces acting on the arrow (which indicates the direction of magnetisation) where a preferred direction of magnetisation obtains. The magnetic induction can be directed either to the right or left and from any intermediate position tends to occupy either one of these directions.

but inseparately associated with certain preferred directions. This restriction to a preferred direction may be regarded as somewhat similar to the spring switch shown in *fig. 6;* the arrow on this switch has an equal tendency to point either to the left or right, and if moved into any inclined position will of its own accord jump back into one of these preferred positions.

The preferred directions as a rule will be entirely different for different zones *(fig. 5)*, yet the nature of hysteresis can be quite simply interpreted if it is assumed that there is only one preferred direction in a material. This state of affairs is shown in *fig. 7*, where the initial condition is unmagnetised. Half of the zonal magnetisation directions are towards the right, and those of the other half to the left. What will happen if a field directed to the right is slowly increased in intensity? As long as the field intensity is not high enough to overcome the elastic attraction to the preferred direction, the initial condition will be preserved. At a certain critical field intensity H_k the "switch" is reversed in direction from right to left, so that saturation is suddenly reached. If the field is again reduced to zero, the magnetic saturation suddenly changes sign, and thus gives the magnetisation curve *A* shown in fig. 7.

Consider the action of a magnetic field perpendicular to the preferred direction. In this case the directions of magnetisation will commence to rotate already with the weakest field strengths. But no reversals will take place. The directions of magnetisation of the different Weiss zones will rotate more and more in the direction of the field as the field intensity increases and will return in the same way to their initial positions on a reduction of the field, thus giving the magnetisation curve *B*.

With an arbitrary spacial distribution of the preferred directions, magnetisation will take place by the combination of rotations and sudden changes in orientation [3]). This aspect of the subject will not be discussed here, for the simplified model is indeed quite suitable for accounting for the magnetic properties of the strip material used for making loading coils, if it is assumed that the preferred direction of magnetisation in the plane of the strip is perpendicular to the direction of rolling. This follows from the close agreement between the diagrammatic curves in fig. 7*A* and *B* and the measured curves in fig. 4*A* and *B*.

In manufacture it is most important to know

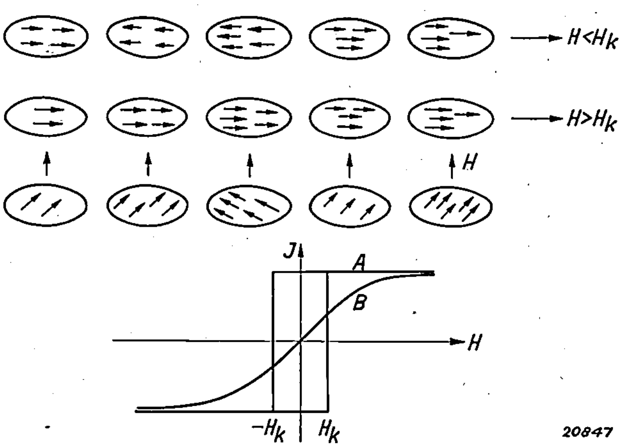


Fig. 7. Effect of a magnetic field on the induction, when in all Weiss zones the same preferred direction obtains:

A. Field in preferred direction; initial state unmagnetised, i.e. half of the magnetic arrows are opposite to the direction of the field. A magnetic field weaker than H_k has no effect. A magnetic field greater than H_k causes the opposing magnetic arrows to become suddenly reversed; the material becomes magnetically saturated and remains so also when the field is reduced (hysteresis).

B. Field perpendicular to the preferred direction. The magnetic arrows are rotated and no reversals take place. The material is therefore devoid of hysteresis and exhibits a slowly diminishing permeability with increasing field even when starting with weak intensities.

The curves *A* and *B* are the magnetisation curves obtained for cases *A* and *B* above. They are somewhat similar to curves *A* and *B* in fig. 5 and thus offer an explanation for the magnetic properties of the new loading-coil material.

---

3) Closer investigation of the sudden reversal process has demonstrated that the assumption that the Weiss zones are invariable in shape and size introduces too great a simplification of the actual conditions. Particularly with weak fields, magnetisation is mainly produced by the zones in a more favourable situation growing at the expense of the other zones.

how the preferred direction of magnetisation in question is produced, although much doubt still attaches to this point. Becker's theory advances two possible causes.

In the first place the majority of ferromagnetic substances have a preferred direction of magnetisation with reference to the axes of the crystals. In the cubic iron crystals, the three directions along the edges of the cube are preferred, and in nickel crystals which also have a cubic structure the four spacial diagonals.

Secondly, internal stresses may result in the appearance of new preferred directions; in the case of iron the direction of maximum tension is favoured, while in nickel that of maximum stress [4]. It is however difficult to determine whether one of these factors has a decisive influence on the magnetic properties of the new material or not.

At a first glance it appears as if the first factor has a direct bearing, since the preferred direction of magnetisation also has a crystallographic significance. With very limited scattering the crystals are so arranged that the axes of the cubes coincide with the longitudinal and transverse directions of the strip. On the other hand, the predominance of magnetisation in specific directions is too small in the crystals of this series of alloys (iron with 30 to 70 per cent nickel) for the order of magnitude of the observed phenomena to be determined. For a specific alloy of this series (65 per cent nickel) the different directions in the crystal are in fact completely equivalent from a magnetisation standpoint. Nevertheless this alloy after suitable pre-treatment also exhibits the same behaviour as the other alloys. The crystalline configuration is thus alone incapable of interpreting the magnetic behaviour, which in fact already follows from the fact that the extremes in magnetisation curves discussed above correspond to equivalent crystallographic directions.

The required magnetisation curve B is actually only obtained by a second rolling after arriving at the crystalline configuration by a first rolling and subsequent recrystallisation.The manufacture of this loading-coil material has been described and studied by means of X-rays, as outlined in an article published in this Review.

It should be noted that by a second rolling the position of the crystals is on the whole not altered, from which it follows that the conditions of shear are entirely different in such an "orientated" material to those in standard materials in which, after similar reduction by rolling, nothing remains of the original texture of the material.

The experience gathered in the manufacture of this material indicates that the peculiar distribution of stress occasioned by the specific shear characteristics is essential for obtaining the preferred direction of magnetisation perpendicular to the direction of rolling. The exact nature of these stresses has however not yet been fully established and it has not yet been possible to define a system of stresses which with the alloy in question here would result in the observed phenomena. It appears probable that entirely different and hitherto unsuspected conditions have a direct bearing on the magnetic properties of the material in question.

4) Actually the orientation of the stress with reference to the crystal axes is also an important factor. Cf. R. Becker, Phys. Z., **33**, 905, 1932.

AN APPARATUS FOR HYSTERESIS MEASUREMENTS ON SOFT MAGNETIC MATERIALS

By J. J. WENT.

Summary. To carry out hysteresis measurements on soft magnetic materials the specimen under investigation is preferably employed in the form of a closed ring which is placed in an annular magnetising coil. To insert the ring in the coil and to remove it from same, the coil is made in two parts which are joined in the plane of the ring and make contact while surrounding the ring specimen.

When plotting the hysteresis loop of a magnetic material, the relationship has to be found between the magnetic field strength H and the induction B. To determine this relationship it is desirable in every case to employ a method by which the field strength in the specimen under examination can be accurately calculated and kept absolutely constant, as otherwise it is not possible to obtain the value of the induction B corresponding to a specific value of H.

These two requirements are met by specimens in the form of an:

a) Ellipsoid, or a

b) Closed annular ring.

a) The ellipsoid is brought into a homogeneous field. The induction obtained can be determined by, for instance, measuring the deflection produced in a magnetic needle by the magnetised specimen under examination. An arrangement for this purpose has the advantage that it affords a homogeneous field of mathematical precision in the specimen and that magnetisation is easy to carry out. Against these points are the disadvantages that an ellipsoid is not easy to make; also difficulties accrue owing to part of the lines of force being through air. The latter is a particularly serious difficulty with soft materials, since these then exhibit a marked demagnetisation as a result of which the accurate determination of the field strength in the iron is rendered a very difficult operation. Moreover, measurements are very readily affected by external fields such as produced by the presence of iron or electric currents in the vicinity.

b) For this reason test specimens of soft materials are usually made in the form of a closed ring. The ring is then magnetised by a primary coil wound uniformly round it. The induction B is determined by measuring with a ballistic galvanometer the volt seconds (v sec) induced in a second coil on an alteration in the field strength. The following equation then applies:

$$B = \frac{\text{Vsec}}{f \cdot n} \cdot 10^8 \text{ Gauss},$$

where f is the cross-section of the iron and n the number of turns in the secondary coil.

The method of measurement described here has to meet the following requirements:

a b c

Fig. 1. View of two-part coil for carrying out hysteresis measurements. a) Top half, b) Lower half with iron ring, c) Two halves joined. In b) the small secondary coil wound round the iron ring may be seen.

1) The ring formed of the material under investigation must constitute a closed circuit; it must have no air gaps as otherwise demagnetisation may occur and hence the field strength not correspond entirely to the current through the primary coil and its number of turns.
2) The ring must have a uniform diameter throughout.
3) The internal circumference of the ring must be only slightly smaller than the outer diameter. If this condition is not met the field strength will not be adequately constant, since along every line of force $\int H\,ds$ = field strength $\times$ circumference has the same value.
4) The primary coil must be uniformly distributed over the whole ring.

In view of the first requirement it becomes impossible to make the core from several separate parts which are pushed into the coil individually, and hence each ring must be provided with a new set of windings. To avoid this tedious operation, a coil has been constructed in two halves and which can be fixed round the ring under examination (see *fig. 1*). Requirements 2 and 4 can be satisfied without difficulty. The ring is made of such dimensions that the maximum error under 3 above is 0.2 per cent, when the field strength is assumed to be that obtaining at the mean diameter of the ring.

Construction of Coil

Two ebonite discs A' and A'' (see *fig. 2*, right half) of 120 mm diameter and 20 mm thick are

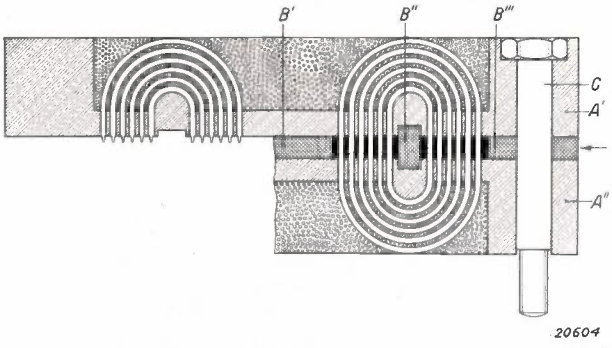


Fig. 2. Section through two-part coil, A' and A'' Ebonite discs, C Bolts, B', B'' and B''' Filler plates.

turned on a lathe leaving a wide flange at the sides by means of which the two parts can be rigidly joined together with four bolts C. Temporarily the fillers B', B'' and B''' are inserted between the two discs, B'' taking the place of the specimen which is later submitted to test. Both inside and outside B'' six series of 66, 60, 55, 45, 40 and 32 holes respectively of 1.5 mm diameter are now drilled along circles through the two plates. The holes are spaced at equal distances over the circles, so that on the inner circle a space of approximately 2 $\times$ 2 mm is available for each hole. A long enamelled copper wire (1.5 mm diameter) is now wound through all the holes, the windings nearest to B'' being made first. After completing each layer of windings the wire is again varnished and the surface then covered with a smooth coating of litharge and glycerine which very quickly hardens. When finally all six windings have been completed, the whole of the turned cavity of the disc is filled with this compound. The two discs are now closed with flat ebonite plates above and below, these being firmly screwed to the broad flange. The bolts are then removed and the whole assembly sawn through in the direction of the arrow. In the one disc all projecting copper points are smoothed down to conical points with a special hollow cutter (see fig. 1, left half), while in the other disc corresponding conical holes are drilled in the copper points. Naturally the fillers temporarily inserted are removed before these operations. If the discs are now again placed on each other, all 596 contacts can be simultaneously closed by tightening the nuts on the bolts. The contact resistance is so small that the resistance of the whole arrangement is about 1.5 ohms, this resistance moreover being found adequately constant. To ensure that the ebonite discs on tightening the nuts remain perfectly flat and the pressure is uniformly distributed over all contacts, two thick brass plates are inserted between the ebonite discs and the heads of the screws or nuts, as shown in fig. 1*c*.

The heat generated by the current in the primary coil is very small owing to the low contact resistance. With a current intensity of 5 A, corresponding to a field strength of approximately 120 oersted, the temperature rise in the ring is about 1 °C. per min. It is therefore possible to carry out measurements also at very high field intensities, should this be found desirable.

For a measurement one or more rings of the magnetic material under investigation are made with an overall thickness of 4 mm and with internal and external diameters of 45 and 55 mm respectively. Three to ten turns of wire 100 μ thick to act as a secondary coil are wound round the rings being insulated by a layer of paper 10 μ thick, the number of turns depending on the thickness of the rings. The whole arrangement is placed in the coil, which is closed and is then ready for making ballistic measurements in the usual way.

At first such a powerful field is generated by passing a strong current (e.g. 6 amps) through the primary coil, that a magnetic condition is attained which lies outside the hysteresis region (e.g. A in *fig. 3*). If the current is then allowed to drop to zero again, the field H also becomes zero, while the ring retains the induction B_r, which is termed the remanence or remanent induction. The magnitude of this remanence can be obtained by direct measurement of the induction currents in the secondary coil,

a) By again applying a current of 6 A and measuring the change in induction B_rA';
b) By applying a current of — 6 A and again measuring the change in induction B_rC'.

We get directly from the graph the remanence $\frac{1}{2}(B_rC' - B_rA')$ and for the magnetisation at a field strength of 6 A, $\frac{1}{2}(B_rC' + B_rA')$. The other points of the hysteresis loop are obtained by starting from A and reducing the field intensity or starting from B_r and applying negative field intensities. It is assumed here that the hysteresis loop has the same form for both negative and positive field strengths.

If finally the virgin or initial magnetisation curve is to be traced the ring must first be completely demagnetised, which can be done by means of an alternating current of gradually increasing strength; the corresponding induction for the virgin curve can then be determined by applying a field of known intensity.

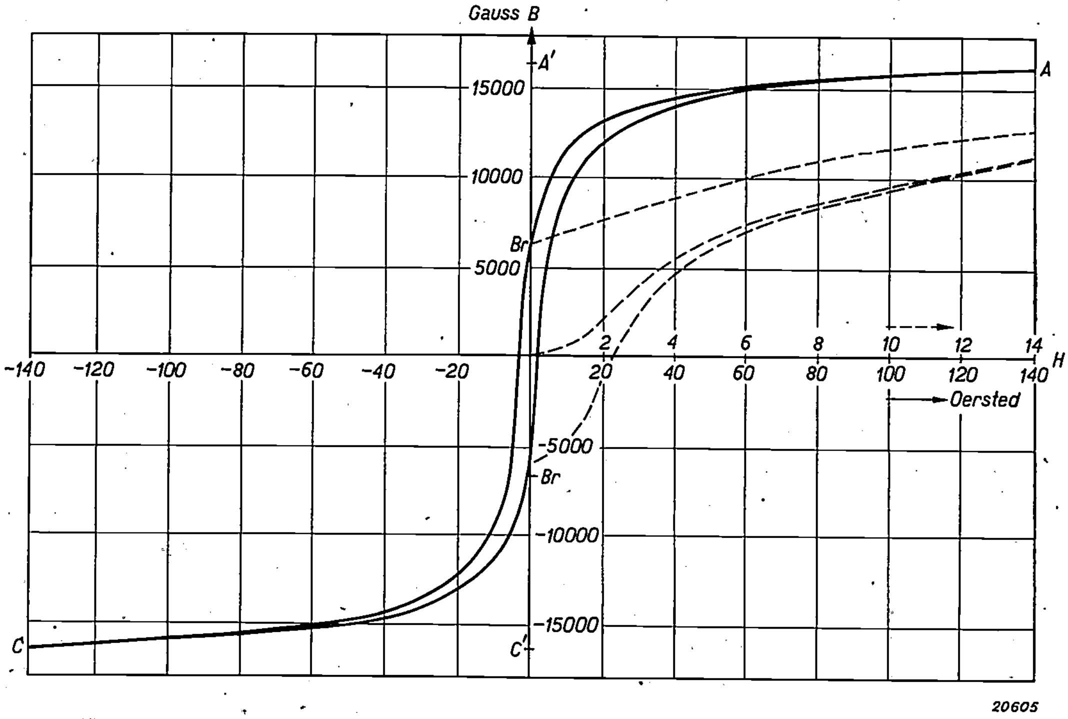


Fig. 3. Magnetisation curve for soft iron. The dash lines indicate the induction on a scale magnified 10 times.

GLASS FOR MODERN ELECTRIC LAMPS AND RADIO VALVES

By J. SMELT.

Summary. Modern views regarding the nature of the physical state of glass are discussed. The viscous state is accounted for if the glass in this state is regarded as a polymerisation product comprising large molecular groups. The glass used for bulbs and the internal components of electric lamps and gas discharge lamps must conform to very close specifications. How these requirements are met and which factors have to be taken into consideration are discussed.

What is Glass?

Technically glass is a transparent solid material highly resistant to chemical attack, which in the fluid or viscous state can be worked to any desirable shape. It is usually manufactured from very cheap raw materials, such as sand, soda, lime, etc.

Physically glass is a supercooled liquid, viz., a solution of silicates in silicic acid. Recent investigations have, however, shown that this simple description does not entirely agree with the facts, and the hypothesis in now advanced that glass is subject to reversible transformation from and into at least three states of aggregation with change in temperature.

Above at certain temperature which for an ordinary glass (Na_2O-CaO-$6SiO_2$) is approximately 1200 °C., glass must be regarded as a liquid. This temperature is termed the transition temperature, and above it the crystallisation process termed devitrification in glass technology can no longer take place. The viscosity varies only slightly with temperature in this region, and is approximately 400 c.g.s. units. For the sake of comparison it may be noted that this value is ten times the viscosity of glycerine.

Below the transition temperature, the viscous state is situated, in which the viscosity varies in close relationship to the temperature, since for every increase of approximately 9 °C in temperature the viscosity is doubled. This marked change in internal friction is a characteristic of this state of aggregation, which may in fact be termed the fourth state of aggregation of matter. In this state devitrification can take place, other conditions being favourable, viz., if the glass is kept for some time in the viscous state. Small crystals gradually form in the initially transparent glass which increase to such numbers that eventually the glass becomes opaque.

It is assumed that the viscous glass on further cooling becomes increasingly viscid and very gradually passes over into the solid state of aggregation.

Various properties of glass, such as the viscosity, the thermal expansion and the electrical resistance, however exhibit anomalies between 500 and 600 °C. Most types of glass must be reheated to these temperatures after shaping in order to avoid mechanical stresses, which frequently develop on rapid cooling and cause the glass to crack. This temperature, which must be fairly accurately determined for each kind of glass, is termed the normalising temperature of the particular glass. The various anomalies observed indicate that at this temperature discontinuous changes take place.

Figs. 1a, b and *c* show the changes in a number of properties at various temperatures [1]), and brings out clearly the discontinuous changes at the normalising temperature. At still lower temperatures the viscosity of the glass varies only slightly. The tendency to devitrification has disappeared, the glass has become brittle and is very fragile, its properties being those we generally associate with ordinary glass.

These characteristics are assumed to be due to a polymerisation of the silicates, i.e. in the formation of molecular complexes. Above the transition temperature, ions alone are present, such as in ordinary liquid solutions, viz., Na^+, Ca^{++}, SiO_3^{--}, as well as molecules of SiO_2. But at the transition temperature the molecules and ions commence to combine to form higher polymers, such as:

$$SiO_2\text{-}SiO_2\text{-}SiO_2 \quad \text{and} \quad Ca < \begin{matrix} SiO_2\text{-}SiO_3^{--} = SiO_2 \\ SiO_2\text{-}SiO_3^{--} = SiO_2 \end{matrix}.$$

At the same time Na^+ ions remain. In the viscous state polymerisation gradually spreads until at the normalising temperature the Na-ions are also linked up at the polymerised molecule and only polymerised silicates remain. Contrary to the hardening of artificial resins, the polymerisation process in glass is reversible, such that a solidified glass can be reconverted to the liquid phase by transition through the viscous state.

---

[1]) Cf. Berger, Glastechnische Ber. 5, 396, 1927: Beitrag zur Frage nach der Natur des Glaszustandes.

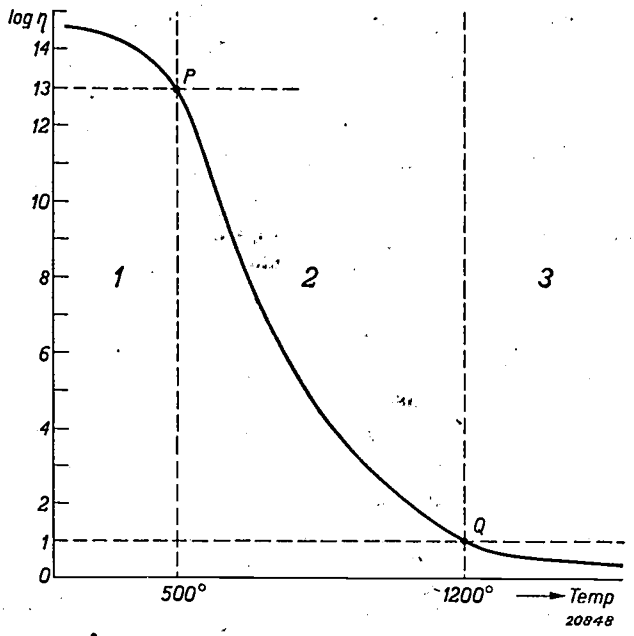


Fig. 1 *a*

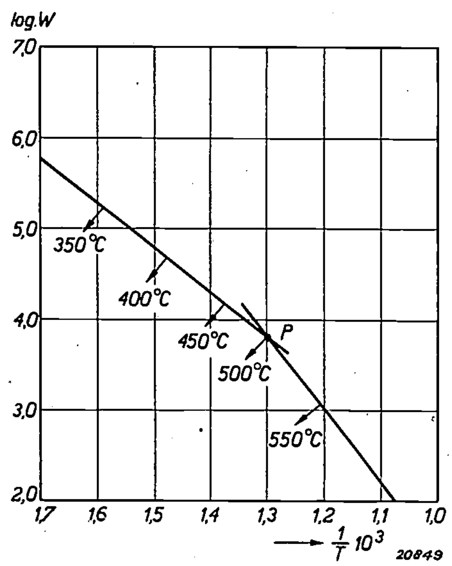


Fig. 1 *b*

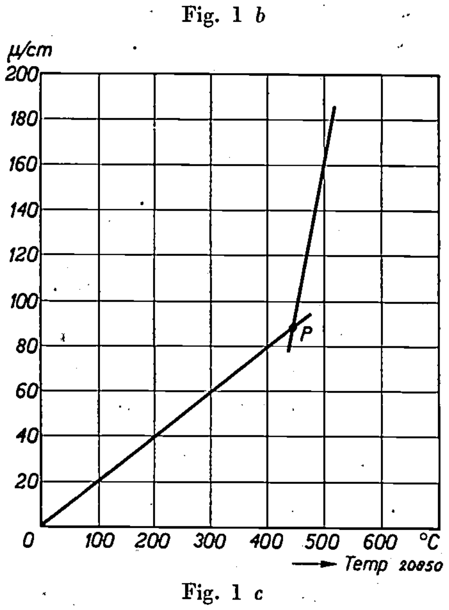


Fig. 1 *c*

Fig. 1. Curves of *a*) viscosity, *b*) specific electrical resistance W, and *c*) relative expansion in microns per cm, plotted as a function of the temperature (Temp. signifies °C, and T deg. abs.). In fig. 1*a*, *1* is the solid phase, *2* the viscous phase; *3* is above the transition temperature. The point P corresponds in all curves to the normalising temperature. In *b*) and *c*), which should be regarded as diagrammatic, the bends indicate that P is subject to abrupt changes. Actually the curves are continuous.

Before closer knowledge had been gained of the states of aggregation of glass, technical requirements made it necessary to produce different types of glass with specific optical properties, which led to the search for new oxides suitable for glass making. As a result of researches carried out to this end, the majority of the chemical elements are now employed in glass manufacture. Thus boric acid and phosphoric acid are frequently used in place of silicic acid, while the oxides PbO, K_2O, MgO, BaO, ZnO are also in common use. Fluorides which are difficult or impossible to dissolve and readily crystallise from solution are employed for making certain kinds of glass producing a dispersion of light. In addition, many metals, oxides, sulphides, sulphates and selenides are used for colouring glass or as auxiliary raw materials. The adoption of this wide range of new constituents of glass has enabled a large number of new glasses to be produced possessing specific optical, thermal, chemical and electrical properties.

Apart from the optical industry, no industry has gained more from the advances made in glass physics and technology than the electric lamp manufacturing industry.

In addition to a comprehensive knowledge of the properties of glass in the vitrified and viscous states (the rate of fusing the flanged stem tube to the lamp bulb is determined by the velocity with which the bulb can be heated without cracking, and by the time taken by the various components to fuse together), the electric lamp industry also requires a variety of glasses possessing different properties.

Electric lamps, radio receiving valves, X-ray tubes, mercury, sodium and neon lamps all require their own type of glass with well-defined properties. In electric lamp manufacture, the mechanical processes entailed require the glass to be easy and rapid to melt and shape. The use of these products in the Tropics necessitates a marked chemical resistance to moist heat. The action of sodium and mercury at high temperatures must not cause a deterioration of the glass used for these lamps. In view of the need for adequate insulation against high electric tensions, frequently at a distance of only a few millimetres, the electrical resistance has to meet strict specifications. Current leading-in wires of various types necessitate high or low coefficients of thermal expansion as the case may be. In addition different types of glass are needed which either transmit or absorb readily all waves in a specific wave band from the infra-red to X-rays.

Owing to the variety of technique employed in modern glass manufacture it has been possible to produce a suitable glass to meet almost every specific requirement.

The particular requirements from case to case are discussed below as well as the measures adopted to meet them.

Glass for Standard Electric Lamps and Receiving Valves

In glass for these products a distinction is generally drawn between that used for making the bulb and that for making the interior components of the lamp or valve.

The bulb glass must be cheap and hence made from inexpensive raw materials. Furthermore, it must have such a composition that it fuses readily at moderate temperatures without excessive wastage. It must also be adaptable to the modern mechanical methods of making electric lamp bulbs; it must fit the lamp base and give a close fused joint with ordinary flame heating. It must, moreover, be clear and in particular it must not devitrify or discolour when exposed for a comparatively prolonged period to the flame. It must also be satisfactorily resistant to moist heat.

Following this resumé of the essential requirements, it will be instructive if several points are discussed in further detail. As already indicated the raw materials from which glass is made are sand, soda and lime. The facility with which glass fuses can be measured by determining the softening point, i.e. the temperature at which the glass exhibits a specific viscosity. So that the viscosity of the glass is not raised too high, lime is in part replaced by barium oxide or magnesium oxide. Clear glass is made by using only pure raw materials free from iron. The tendency to devitrification is reduced by increasing the number of different components in the glass. The addition of Al_2O_3 also has a favourable effect in inhibiting devitrification. Although as a rule glass is highly weatherproof, its sensitivity to the action of water must not be underestimated, for during a long sea voyage in the Tropics glass which will remain unaffected in temperate climates will be found to have become completely weathered.

This action of water is due to the removal of the alkaline oxide from solution in the surface of the glass. This weathering action takes place in various stages, as follows:

1) Overall coating which can be readily removed with water.
2) Appearance of small crystals which as a rule cannot be removed by water.
3) The surface is more strongly attacked in certain localities; glass affected in this way can generally still be cleaned with hydrofluoric acid.
4) General and widespread attack of the surface, when the glass becomes quite useless; this occurs when the bulb has not been sealed during a sea voyage in the Tropics.

The first and second stages are due, as stated, to the action of water on the alkali contained in the surface of the glass, the carbon dioxide in the atmosphere forming Na_2CO_3 with the alkali. This carbonate in its turn dissolves the silicic acid in the glass (third stage), a process which spreads continuously (fourth stage).

The apparent contradiction that a finished lamp during a voyage through the Tropics undergoes no appreciable deterioration, while the inside of an open bulb is completely weathered, is due to the difference in the properties of the inner and outer surfaces resulting from the method of manufacture. Being polished by the influence of the flames the outside of the glass always exhibits a greater resistance than the inside surface.

Fortunately simple means have been devised to raise the resistance of the glass against atmospheric influences, viz., heat treatment of the bulb in an atmosphere of sulphur dioxide. By interaction of the alkaline oxide with sulphur dioxide and oxygen a thin coating of sodium suphate is formed on the surface of the glass, which can be readily removed from the bulb with water. The remaining surface, poor in alkali, has a greater resistance than the untreated surface.

If dilute hydrofluoric acid is allowed to act on the surface for a short time, the original composition of the glass is restored and the glass again exhibits its initial behaviour (see *fig. 2*). When bulbs which have been treated as described above are washed with hydrofluoric acid the greatest care must be exercised.

The requirements which have to be met by glass used in the interior components of lamps are as follows:

Since heat has to be applied externally for fusing the flanged stem tube to the bulb neck, the bulb becomes warm. Glass is therefore selected for the interior which has a low softening point in order that both the interior and exterior glasses soften at the same time. Usually the mounted foot is made of a

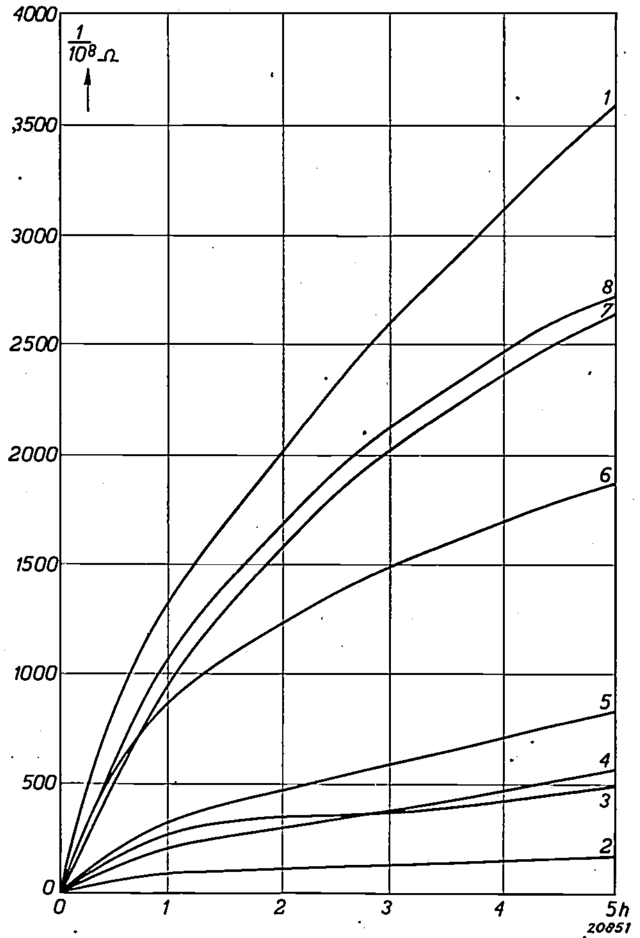


Fig. 2. Etching of glass bulbs after different pre-treatment. *1* - No pre-treatment, *2* - treated with SO_2 and rinsed out with water.
3 as *2* but treated with 3 per cent HF for $^1/_2$ min.
4 „ *2* „ „ „ 3 „ „ „ „ 1 „
5 „ *2* „ „ „ 3 „ „ „ „ 2 „
6 „ *2* „ „ „ 3 „ „ „ „ 5 „
7 „ *2* „ „ „ 3 „ „ „ „ 11 „
8 „ *2* „ „ „ 3 „ „ „ „ 18 „
The increase in the electrical conductivity of distilled water with which the bulb was filled for testing purposes over a period of 5 hours was taken as a measure of the degree of etching.

glass with a fairly high content of lead, whose coefficient of expansion must be the same as that of the metal wires which are passed through the glass. Furthermore, as regards electrical properties, the glass used for the base must also satisfy certain requirements respecting conductivity.

In general glass is one of the best electrical insulating materials. The electrical resistance W varies with the temperature according to the equation $\log W = B/T + A$, where A and B are constant for a specific composition. This equation is valid up to the normalising temperature. Above this temperature the same relationship applies with different values for A and B (cf. fig. 1*b*).

Conduction in glass is electrolytic, the current being carried by the alkalies and mainly by the Na-ions.

In radio valves and D.C. electric lamps socalled "lead trees" sometimes make their appearance at the negative pole after long use, together with fine cracks at the positive pole (*fig. 3*). Lead trees are

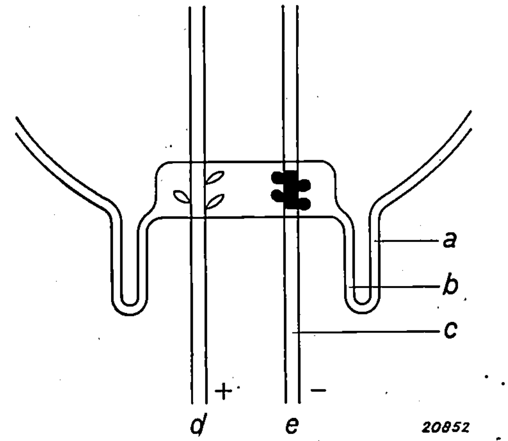


Fig. 3. Diagrammatic representation of the formation of lead trees in a D.C. electric lamp *a* bulb, *b*support, *c* lead-in, *d* cracks at positive lead-in, *e* lead tree at negative lead-in.

black discolorations in the glass branching in all directions which appear in the direct neighbourhood of the pole wire. It is apparent from illustration why these frequently very variegated markings are termed trees.

This phenomenon is obtained irrespective of the material of which the leading-in wire is made, whether of platinum, nickel-iron or chrome-iron wire, or dumet wire. Where sodium atoms are the current carriers, the positive pole must become reduced in sodium content. It is assumed that as a result a glass layer is produced which has a lower coefficient of expansion and is the cause of cracks appearing in the glass round the wires. At the negative pole, on the other hand, sodium is liberated and may lead to the reduction of the PbO in the glass to lead which builds up the lead tree. The amount of sodium liberated according to the electro-chemical equivalent, in a period of time of say 1000 hours, is however so small (with a current intensity of 0.1 μA approximately 0.07 mgr) that this action can hardly be held responsible for the cracking of the glass.

It has, however, been found that this phenomenon can be considerably reduced by using glasses with better electrolytic properties, i.e. glasses having a higher specific resistance at the temperatures in question.

The glasses with the best insulating properties are generally lead glasses which, in view of their low softening point, are also suitable for making lamp bases.

If in a particular lead glass Na_2O is replaced by K_2O or/and the PbO content is raised, a series of glasses is obtained with a considerably higher resistance and hence less liable to form lead trees or to crack. This alteration in the properties of the glass is readily explained by the fact that the mobility of the potassium ions is less than that of the sodiums ions.

Bulb Glass for Special Purposes

A. Opal Glasses. In these glasses the light is dispersed by devitrification or the precipitation of very small particles. The type of glass almost exclusively used at the present day is a cryolite glass in which the NaF particles crystallise out and disperse the light by diffraction. It is by nature more brittle than ordinary glass and is therefore used normally in the form of two layers the inner layer being ordinary glass.

B. Coloured Glass for photographic or illumination purposes. Coloured glass can be made by two processes:

1. By true (molecular) colouring with various coloured oxides, such as copper oxide, cobalt oxide, nickel oxide, chromium oxide or iron oxide.
2. By colloidal colouring.

While in the first process the glass solidifies in the coloured state from the viscous phase, glass coloured by the colloidal process must be submitted to subsequent heat treatment in the neighbourhood of the softening point, in order to render the submicroscopic particles visible. The media used for colouring include gold, selenium, cadmium sulphide (filter glass for motor-car and bicycle lamps), and copper (glass for darkroom lamps).

On the whole, glass technology has a very limited choice of colouring materials. In recent years the employment of the rare earths has enabled glasses to be produced which possess fairly sharp absorption bands at the same point at which the salts of these elements in solution exhibit sharp absorption lines. Thus e.g. neodymium oxide has an almost complete absorption in the yellow, although transmitting all other colours.

C. Hard Glass, which can be used for both bulbs and internal components.

If, for instance, the temperature of the filament in projector lamps is very high or the dimensions and construction are so restricted from practical considerations that the glass is raised to a very high temperature, the glass is liable to crack on switching on and off owing to the marked rapid temperature changes, or even become blown out following partial softening. The requirements which glass for these purposes has to meet are therefore:

a) The glass must have a higher softening point than ordinary glass; this as a rule can be obtained by reducing the alkali content.

b) The glass must be able to withstand considerable temperature changes; the thermal resistance is dependent on many factors of both a mechanical and thermal nature.

Winkelmann and Schott in 1894 by theoretical analysis deduced that the thermal resistance measured experimentally by suddenly raising the glass to a high temperature and then determining what temperature fluctuations it can withstand, depends on the following quantity F:

$$F = \frac{T}{\alpha E}\sqrt{\frac{\lambda}{SC}},$$

where: λ = the thermal conductivity, T = the tensile strength, α = the coefficient of expansion, S = the specific gravity, E = the modulus of elasticity, and C the specific heat.

This equation has been found to be not quite accurate; yet the thermal resistance of glass is indeed mainly determined by the coefficient of expansion. The smaller this coefficient the greater are the temperature fluctuations which the glass can withstand. The raw materials which reduce the coefficient of expansion are SiO_2 and B_2O_3, while the alkalies cause a marked increase in this coefficient.

For the various oxides used in glassmaking, data have been gathered for the standard ranges of mixtures employed by means of which the physical constants of a glass, such as the coefficient of expansion, the modulus of elasticity, the specific gravity and the specific heat, can be calculated when the composition is known. These properties vary in general linearly with the amount of oxide added, a behaviour which is in good agreement with the assumption that glass consists of a solution of silicates in silicic acid.

The behaviour of the borosilicates is not however in accord with theory in this respect, since the coefficient of expansion does not vary with the percentages of the constituents. For instance in the case of a glass with *a* per cent Na_2O, *x* per cent SiO_2 and (100—*a*—*x*) per cent B_2O_3 a minimum coefficient of expansion is found for a specific ratio of $SiO_2 : B_2O_3$. Similarly when the content of silica is kept constant, a pronounced minimum value is obtained for a specific ratio of $Na_2O : B_2O_3$.

This point has to be considered in determining the composition of a type of glass to be used for projector lamps which have to sustain heavy currents and high temperatures. The coefficeint of expansion has to be such that the glass always gives a tight joint with the metal wires fused through it, these wires being here of tungsten, molybdenum or iron nickel alloys.

Glass for Gas Discharge Tubes

The above considerations apply mainly to electric lamps with or without a gas filling and also to radio valves, in which stability at high temperatures (glass transmitting valves) is required in addition to good insulating properties of the glass.

Present day *gas discharge lamps* and *tubes* with their multitude of designs and types have introduced a series of new problems and requirements.

In the case of glass for *low-tension neon tubes* used for publicity and display purposes, the only condition requiring consideration is that the glass must not discolour in the blowpipe flame. This discoloration is due to the reduction of a number of oxides present in the glass, such as lead oxide, and the oxides of arsenic and antimony, which must therefore be avoided in glasses intended for these purposes. If *mercury* is added to the neon, the glass may become blackened after a time owing to the action of the mercury ions. Glasses with a relatively high electric resistance have been found by experience to be quite satisfactory for these applications. Lead glasses are also useful if they do not discolour in the flame, while barium glasses are exceptionally efficient.

Where it is a question of raising the efficiency of these light sources by transforming the ultraviolet radiation of the discharge into visible light, *fluorescent glass* may be used which usually contains a small quantity of certain metals, such as uranium.

An important problem is the transmission factor of glass for different wave lengths. Reference has already been made to the various methods employed in glass technology for improving or controlling transmission in the *visible spectrum*. The behaviour in the *ultra-violet region* is, however, also of importance. *Quartz* transmits the whole of the ultra-violet to 1850 Å. Different kinds of glass have also been produced which transmit a sufficient range of the U.V. radiation present in sunlight, while absorbing practically the whole of the shorter wave-lengths which have a deleterious action on the skin. One of the principal factors in making these types of glass is in fact a negative one, viz., the absence of iron and titanium; SiO_2, B_2O_3, P_2O_3 are the best oxides used in glassmaking which transmit U.V. rays. With low-purity glasses, i.e. those containing iron, the transmission factor becomes fairly rapidly diminished owing to the reduction of ferrous oxide to ferric oxide.

Nickel oxide glass is a glass which absorbs nearly all visible rays except the red. But it also transmits the longwave ultra-violet and again absorbs the short-wave ultra-violet rays. This glass is used in lamps for investigating fluorescence, where interference from visible light must be eliminated.

Ferro-glass, on the other hand, is bluish-green, has a marked absorption in the infra-red, and a low absorption in the visible spectrum. It is mainly used for absorbing heat rays.

In surveying the kinds of glass required for the transmission or absorption of *X-rays*, the well-known rule may be taken as a guide that the absorption increases with rising atomic weight of the constituents. The lighter the oxide used, the batter will the glass made from it transmit X-rays. Lindemann glass for instance is a lithium-beryllium borate. Glasses with a high lead content can, on the other hand, be used as a protection against X-rays, since they are impermeable to these rays.

The highly reactive element, sodium, which is used in sodium lamps, readily attacks ordinary silicate glasses with the result that in a short time the glass becomes brown owing to the reduction of the silica to silicon, and the light output is reduced in consequence. Boric acid offers a greater resistance to reduction than silica, and therefore borate glasses exhibit a greater resistance towards sodium than silicate glasses. Unfortunately these borate glasses are readily soluble in water and acids, so that they cannot be used for ordinary bulbs. They can however be used as enamels or in any other form as a thin coating on a bulb or a valve made of ordinary glass.

In the *high-pressure discharge tubes* with a mercury filling, the high pressures and temperatures obtaining require the glass to have a high softening point. The only kinds of glass of use for these purposes contain a high percentage of aluminium oxide and practically no alkali. A glass of this character can only be melted in the best furnaces, and the requirements to be met are at the limits of ordinary glass technology.

If the pressure of the mercury is still further raised, only one type of glass is available at the moment, viz., clear quartz glass.

Concluding Remarks

Glass must be free from its common defects, such as bubbles, streaks and grains, from considerations of appearance and to reduce the danger of cracking.

As indicated at the outset glass is shaped in the viscous state. The commonest method of making a glass bulb is by blowing into an iron mould, using either the mouth or machinery. This procedure makes it difficult to produce articles of a glass which has to conform to a specific composition. For by blowing, while the outer diameter is fixed, the inner diameter which depends on the amount and distribution of the glass used, is always somewhat uncertain. Every subsequent operation to which the glass must be submitted is difficult and tedious owing to the material's great hardness and brittleness, that it is advisable in making all glass objects to use only such designs in which the internal diameter can vary between wide limits. If the cavity is low and so shaped that it can be readily removed from a shaped die the shaping process can also be performed by pressing, when an exceptionally high standard of accuracy can be achieved.

PRACTICAL APPLICATIONS OF X-RAYS FOR THE EXAMINATION OF MATERIALS X

By W. G. BURGERS.

The example of the application of X-ray analysis discussed in this article shows how different physical states of one and the same material can be distinguished by X-rays; this is the same problem as was dealt with in the previous article in this series [1]).

22. Investigation of the Quality of Nickel-Iron Strip for Loading Coils

The material for making the cores of loading coils which has been evolved in this Laboratory consists of a thin strip of rolled nickel iron which is produced from the raw material (a strip several mm in thickness) by the following series of operations [2]):

Raw material

(Strip several mm thick)

↓ Rolled to

a strip of 0.1 mm in thickness.

↓ Recrystallised at 1000 °C to give a

Recrystallised strip, 0.1 mm thick

↓ Rolled to

a strip 60 μ thick:

↓ Reheated at a temperature of approx. 400 °C.

Final product.

---

1) No. IX in this series of articles, Philips techn. Rev. 2, 29, 1937.

2) Cf. the article by W. Six, Philips techn. Rev. 1, 357, 1936.

character can only be melted in the best furnaces, and the requirements to be met are at the limits of ordinary glass technology.

If the pressure of the mercury is still further raised, only one type of glass is available at the moment, viz., clear quartz glass.

Concluding Remarks

Glass must be free from its common defects, such as bubbles, streaks and grains, from considerations of appearance and to reduce the danger of cracking.

As indicated at the outset glass is shaped in the viscous state. The commonest method of making a glass bulb is by blowing into an iron mould, using either the mouth or machinery. This procedure makes it difficult to produce articles of a glass which has to conform to a specific composition. For by blowing, while the outer diameter is fixed, the inner diameter which depends on the amount and distribution of the glass used, is always somewhat uncertain. Every subsequent operation to which the glass must be submitted is difficult and tedious owing to the material's great hardness and brittleness, that it is advisable in making all glass objects to use only such designs in which the internal diameter can vary between wide limits. If the cavity is low and so shaped that it can be readily removed from a shaped die the shaping process can also be performed by pressing, when an exceptionally high standard of accuracy can be achieved.

PRACTICAL APPLICATIONS OF X-RAYS FOR THE EXAMINATION OF MATERIALS X

By W. G. BURGERS.

The example of the application of X-ray analysis discussed in this article shows how different physical states of one and the same material can be distinguished by X-rays; this is the same problem as was dealt with in the previous article in this series [1]).

22. Investigation of the Quality of Nickel-Iron Strip for Loading Coils

The material for making the cores of loading coils which has been evolved in this Laboratory consists of a thin strip of rolled nickel iron which is produced from the raw material (a strip several mm in thickness) by the following series of operations [2]):

Raw material
(Strip several mm thick)
↓ Rolled to
a strip of 0.1 mm in thickness.
↓ Recrystallised at 1000 °C to give a
Recrystallised strip, 0.1 mm thick
↓ Rolled to
a strip 60 μ thick:
↓ Reheated at a temperature of approx. 400 °C.
Final product.

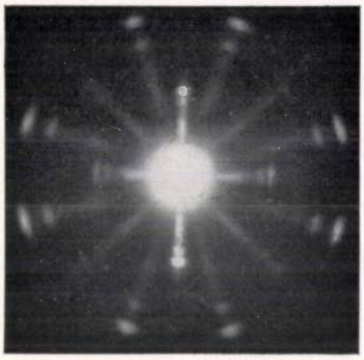
---

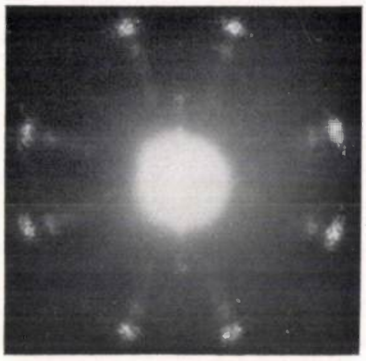
1) No. IX in this series of articles, Philips techn. Rev. 2, 29, 1937.

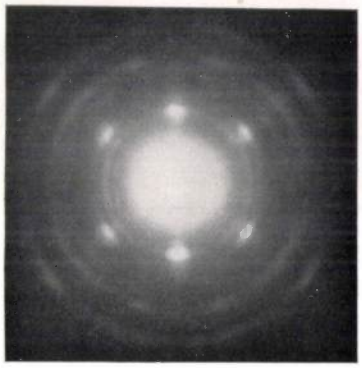
2) Cf. the article by W. Six, Philips techn. Rev. 1, 357, 1936.

After passing through this series of operations, a final product was obtained which exhibited very different qualities as regards the hysteresis losses. Since it was feasible that this was in some way related to differences in the orientation of the crystals in the material [3]), radiographs were produced of good and bad specimens of the material. regular and to a large extent symmetrical cubic configuration of points and lines indicates that the regular nickel-iron crystals are orientated in a very definite manner in the strip, such that a lateral surface of the elementary cube lies in the plane of the strip, while the edges of this lateral surface are parallel and perpendicular to the direction of rolling.

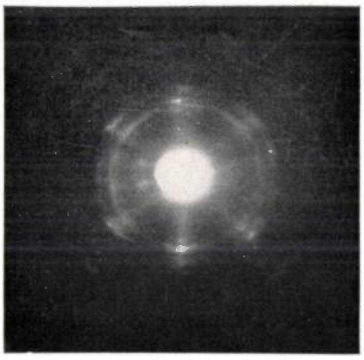
a)







b)



X-Ray Photographs of nickel-iron strip for making loading-coil cores during various stages of manufacture.

Fig. 1. Final Product *).
a) Strip with low hysteresis loss: Symmetrical cubical interference figure, due to a very marked preferential orientation of the crystals in the strip (socalled cubic configuration).
b) Strip with high hysteresis loss: Interference rings more or less complete; from this photograph it may be concluded that the crystals have not assumed a very pronounced preferential orientation.

Fig. 2. Initial material rolled to 0.1 mm and recrystallised subsequently:
a) On further rolling this material furnishes a strip with a low hysteresis loss (cf. fig. 1*a*.)
b) On further rolling this material furnishes a strip with a high hysteresis loss (cf. fig. 1*b*). At this stage also the X-ray photographs of the two materials differ in the same way as those for the final products.

Fig. 3. Initial material rolled to 0.1 mm and not yet recrystallised. At this stage the material gives a radiograph of the type illustrated, irrespective of whether on further working a loading-core material with a high or with a low hysteresis loss is obtained.

*) In all diagrams the direction of rolling is horizontal.

Fig. 1*a* reproduces a radiograph of a final strip with satisfactory properties. In making this picture the radiation impinged on the strip in a direction perpendicular to the surface so that the plane of the strip was parallel to the X-ray film. The direction of rolling is horizontal in the picture. The very Close examination of the picture shows that the crystals are closely packed about the cubical configuration.

That the cubic configuration is important as regards the magnetic properties of the loading-coil strip is shown in fig. 1*b* which reproduces a similar radiograph for a material with a high hysteresis loss: the interference rings are more or less uniformly

[3]) Cf the article by J. L. Snoek in this issue of the Review, p. 77.

blackened round the whole periphery, so that a predominating orientation of the crystals can only be assumed to obtain to a very slight degree: in any case the orientation differs entirely from that of a cubic configuration, which is indicated by the absence of any trace of a symmetrical cubic structure in the picture. It thus appears highly probable from the radiograph that the striking magnetic properties of the material are related in some way to a specific predominating orientation of the crystals.

It was important to determine at what stage in the finishing process this cubic configuration is obtained, and for this purpose radiographs were made of the material at intermediate stages. It was found, in the first place, that the same patterns were given by a 60 μ strip which had not been reheated, as by the final product, which result is not surprising since the temperature at which reheating was conducted was too low for any new crystals to be formed.

The X-ray photographs obtained with a strip rolled to 0.1 mm and then recrystallised at 1100 °C are reproduced in *figs. 2a* and *b*; fig. 2*a* relates to a material which on further working gave a strip with a low hysteresis loss comparable to the type shown in fig. 1*a*, while fig. 2*b* is comparable to fig. 1*b* and gave a material with a high hysteresis loss.

It was observed that already at this stage the difference between the two specimens of material was unmistakable: Fig. 2*a* reveals a well-defined symmetrical cubic configuration, with the interference spots concentrated in groups to a still greater degree than found in fig. 1*a* for the final strip rolled out to 60 μ. The photograph thus indicates the occurrence of a definite cubic orientation in the recrystallised strip, which on rolling down to 60 μ does not lose appreciably in the definition of orientation, the general configuration being retained.

On the other hand, the spots in fig. 2*b* are distributed over practically the whole of the film (on a careful examination a trace of the characteristic cubic pattern of figs. 1*a* and 2*a* can still be detected): The crystals in this material exhibit practically no definite preferred positions.

At this stage it is hence possible to distinguish by means of X-rays between good material and bad material, and with a little practice the quality can be judged from the definition of the symmetrical cubic configuration revealed.

This cannot however be done at an earlier stage in manufacture, as is indicated by *fig. 3* which reproduces an X-ray photograph of the 0·1 mm strip before recrystallisation at 1100 °C In this stage the good-quality strip in figs. 1*a* and 2*a* gives practically the same pattern as the low-quality strip shown in figs. 1*b* and 2*b*, and a diagram such as that shown is always obtained irrespective of whether on further working a strip with low or high hysteresis losses is produced.

It is evident therefore that for the X-ray control of the manufacturing process the 0.1 mm strip after recrystallisation is the most suitable. Closer investigation has demonstrated that the cubic configuration is the more pronounced in this stage, the greater the difference in thickness between the initial material and the strip when rolled down to 0.1 mm.

It is a very striking fact that two substances which on X-ray analysis reveal no difference (cf. fig. 3), yet recrystallise quite differently on being subjected to the same heat treatment. This must be due to recrystallisation being determined by a property which is not revealed by the radiograph of the worked material. Probably the cause for this is to be sought in the formation of new crystals during recrystallisation at those points which have undergone maximum deformation (cf. e.g. the comments on this question made in reference to Example **20** in Part IX of this series of articles), such that the course of recrystallisation is closely related to the nature of deformation. At highly-deformed points of the lattice there will occur no or only the very slightest interference between the diffracted rays, and as a result it is possible that just these places in the deformed material do not appear in the radiograph; hence differences in the nature of the abnormalities, which may cause an alteration in the course of recrystallisation, need not become apparent in the X-ray photograph of the rolled material given in fig. 3.

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

No. 1136*: R. Houwink: Synthetic Resins, their formation, their elastic and plastic properties and their possibilities (J. Soc. chem. Ind., **55**, 247 - 259, September, 1936).

This paper which was read before the annual meeting of the Society of Chemical Industry at Liverpool gives a survey of the methods of resin formation, with special reference to the following aspects:

1) The chemical nature of the reactions with a report of some results of recent investigations relating to the reaction kinetics;
2) The colloidal structure of the straight-chain and three-dimensional polymerides in their various stages of formation;
3) The plastic and elastic properties of resins in relation to their state of polymerisation.

An attempt is made to correlate the changes in elastic and plastic deformability with those in the colloidal structure of the resins, such correlation appearing possible in the light of current chemical theory regarding polymerisation reactions.

No. 1137*: J. R. J. van Dongen: Ermittlung der zulässigen Schnittgeschwindigkeit aus Plandrehversuchen (Stahl und Eisen, **56**, 1815 - 1187, September, 1936).

As full details have already been published in several previous issues of this Review of the results obtained in disk tests, reference should be made to these papers (Philips techn. Rev., **1**, 183 and 200, 1936).

No. 1138: R. Vermeulen: Modellen voor de studie van zaalacoustiek (Bouwkundige Bladen, **1**, No. 8, September, 1936).

Discussion of the application of optical models to the investigation of the acoustical properties of halls, of which details have already been published in Philips techn. Rev., **1**, 46, 1936, in reference to a hall in the Philips Theatre, is continued in this article with reference to the League of Nations building at Geneva.

No. 1139: J. A. M. van Liempt and J. A. de Vriend: Die Farbenwiedergabe beim Photographieren mit Natrium- und Quecksilberlicht (Photogr. Korresp., **72**, 142 - 144, October, 1936).

It is shown in this article how the already good reproduction of colours in photographs made with sodium light can be still further improved by suitable admixture of glowlamp or mercury light. (Cf. also Philips techn. Rev., **2**, 24, January, 1937).

No. 1140: P. J. Bouma: Verblinding (Natuur en Mens, **56**, 217 - 221, October, 1936).

For the subject matter of this article reference should be made to the previous papers by the same author published in this Review and relating to problems associated with roadway lighting (Philips techn. Rev., **1**, 102 and 225, 1936).

No. 1141: F. A. Heyn: Evidence for the expulsion of two neutrons from copper and zinc by one fast neutron (Nature, **138**, 723, October, 1936).

Bombardment of copper and zinc with neutrons gives rise to radio-active isotopes of these elements whose liberation cannot be accounted for by any of the known methods of nuclear transformation. From the behaviour of the particles emitted by these active bodies and from the fact that these reactions are initiated by fast neutrons, the author concludes that in the reaction investigated the expulsion of two neutrons takes place under bombardment with a fast neutron which is absorbed by the nucleus.

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

VOL. 2 No. 4 APRIL 1937

Philips Technical Review

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

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

NUCLEAR PHYSICS

by W. DE GROOT.

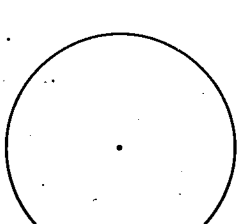
Summary. In this article a survey is given of the main principles of the physics of the atomic nucleus, a subject which is becoming increasingly important in modern technology.

Introduction

The growing importance of the physics of the atomic nucleus, or "nuclear physics", in the technical design of high-tension apparatus, discharge tubes, counters, relay valves, amplifiers, etc., calls for a brief survey of the development of this branch of knowledge during the last quarter of a century.

The "Atomic Nucleus" Conception

The concept of the atomic nucleus is due to Rutherford, who from observations made in 1912 on the dispersion in various substances of alpha-particles emitted by radium (alpha rays) deduced that at the mass centre of each individual atom there is a heavy nucleus at which the bulk of the atomic mass is concentrated.



20931

Fig. 1. Model of atom. The circle represents the extreme boundary of the atom and the dot the nucleus. Actually the diameter of the nucleus is about 1/30000 of the diameter of the atom.

This conception of atomic structure is shown schematically in *fig. 1*, where the large circle represents the extreme boundary of the atom and the dot at the centre the nucleus. The diagram is not drawn to scale, for this would require the circle to have a diameter of e.g. 3 m and the dot a diameter of only 0.1 mm.

The space between the nucleus and the boundary of an atom is occupied by the orbits of the electrons. The atomic theory propounded by Bohr in 1913 furnished a method for investigating these electronic orbits, and the general conclusion may be drawn from Rutherford's experimental work and Bohr's theory of the validity of the periodic classification of the elements that the atomic nucleus possesses a positive charge which according to van den Broek and Moseley is equal to

$$Z \cdot e$$

where e is the socalled elementary charge (approximately $4.8 \cdot 10^{-10}$ electrostatic units) and Z is an integer, the so-called atomic number. The atomic number and the laws governing the orbits of the electrons determine the chemical properties as expressed by the periodic classification of the elements, this classification being one of the corner-stones of modern chemistry.

Physical and Chemical Atomic Weight

In addition to the charge of the atomic nucleus its mass is also of fundamental importance. Since the mass of the nucleus and that of the associated atom bear a simple relationship to each other, and the atomic mass, contrary to the nuclear mass, can be determined directly by experimental means, it has become a common practice to introduce the atomic mass (M) in discussions of the nuclear mass (M_k). The relationship between these two masses is given by the expression:

$$M = M_k + Z\, m_e \quad . \; . \; . \; . \; . \quad (1)$$

where m_e is the mass of the electron. The chemist generally employs as the unit of atomic mass the sixteenth part of the mass of an oxygen atom. Expressed in this unit the electron has a mass of 0.00055, and since for all atoms $Z < M$ it follows that the nuclear mass accounts for more than 99.95 per cent of the atomic mass. In 1913, J. J. Thomson discovered that the majority of the chemical elements are mixtures of isotopes, i.e. compounded of various species of atoms with different masses and the same nuclear charge, hence with identical or nearly identical chemical properties. Since that time a distinction has been drawn between the chemical and physical atomic weight (isotopic weight). The latter is expressed in terms of a unit equal to $^1/_{16}$ the mass of the oxygen isotope O^{16}. The chemical element oxygen is thus itself a mixture of isotopes, viz., O^{16}, O^{17} and O^{18} in a ratio of 99.8 : 0.03 : 0.16, such that:

$$O_{\text{chem}} = \frac{16 \cdot 99.8 + 17 \cdot 0.03 + 18 \cdot 0.16}{16 \cdot 100} O^{16} = $$
$$= 1.00022\ O^{16}.$$

The relationship between the chemical atomic weight (M_{ch}) of an arbitrary element and the physical atomic weight of its isotopes, M_1, M_2 ... may be represented as follows:

$$M_{\text{ch}} = \frac{1}{1.00022} (p_1 M_1 + p_2 M_2 \ldots) : 100, \quad (2)$$

where p_1, p_2 ... are the percentages of the various isotopes present. The agreement between the atomic weights obtained by physical methods from equation (2) and the atomic weights determined by direct chemical methods is usually very close.

Composition of the Nucleus. Binding Energy

A characteristic of the isotopic weights discussed above is that to a very close approximation they are all whole numbers. To define the nucleus there has therefore been introduced, in addition to the integer Z, a second integer A, the mass number, which is the integer nearest to the atomic weight of the isotope. For the hydrogen nucleus $Z = 1$ and $M = 1.0081$, and hence $A = 1$. The hydrogen nucleus is the simplest nucleus known and is in fact regarded as a single elementary particle; as such it has been given the characteristic name of "proton".

In addition to the proton, physical research has during the last five years discovered still another elementary particle, the neutron (Chadwick, 1932). For the neutron $Z = 0$, $M = 1.0090$, hence again $A = 1$. Since the discovery of the neutron, the general assumption has progressively gained ground that other nuclei are built up from protons and neutrons; it is thus assumed that a nucleus with the mass number A and the atomic number Z is composed of

$$Z \text{ protons} + N \text{ neutrons},$$

where $N = A - Z$ [1]). It is evident that in this way we arrive at a nucleus with the correct charge and approximatively the correct mass. Yet on a closer examination it is found that the mass of the nucleus is a little smaller than those of its components. Take as an example the nucleus of an atom of the reference element O^{16}; for this atom $M = 16.000$ by definition, while the weight of its components is:

$$8 \cdot 1.0081 + 8 \cdot 1.009 = 16.1368 \text{ [2])}$$

The difference is due to the liberation of binding energy on the combination of protons and neutrons to form a single nucleus; the magnitude of this energy is given by a well-known equation in the theory of relativity as equal to the difference in mass multiplied by the square of the velocity of light. As the true mass of $^1/_{16}$ of the atom O^{16} is $(1/6.06) \cdot 10^{-23}$ gram and the velocity of light is $3 \cdot 10^{10}$ cm per sec, the binding energy of the oxygen nucleus is found to be:

$$\frac{0.1368}{6.06} \cdot 10^{-23} \cdot (3 \cdot 10^{10})^2 = 0.000203 \text{ erg}.$$

Another unit used for expressing the atomic energy is the electron-volt, which is the energy acquired by an elementary charge when it passes through a potential gradient of 1 volt. These units are related to one another as follows:

$$0.001 \text{ atomic-weight unit} = 1.49 \cdot 10^{-6} \text{ erg} =$$
$$= 0.93 \cdot 10^6 \text{ eV} = 0.93 \text{ MeV}.$$

1) The nucleus and hence the associated atom are completely determined by N and Z or also by A and Z. A nucleus could be adequately described by these two numerical values. Usually a specific atom or the corresponding nucleus is denoted by the chemical symbol, which is determined by Z alone, and the value of Z indicated by subscript figures on the left of the symbol with the value of A given by superscript figures on the right. The isotope of elementary oxygen (0) with a nuclear charge $Z = 8$ and an atomic weight of 16 is thus denoted by

$${}_8O^{16}.$$

2) In general it is required to compare the mass of Z protons plus N neutrons, i.e. $Z M_p + N M_n$ with the mass of the atomic nucleus M_k. If however to both is added the mass of Z electrons ($Z m_e$), we get since $M_p + m_e = M_H$ = the mass of the hydrogen atom, a comparison between $Z M_H + N M_n$ and the atomic mass M ($= M_k + Z m_e$).

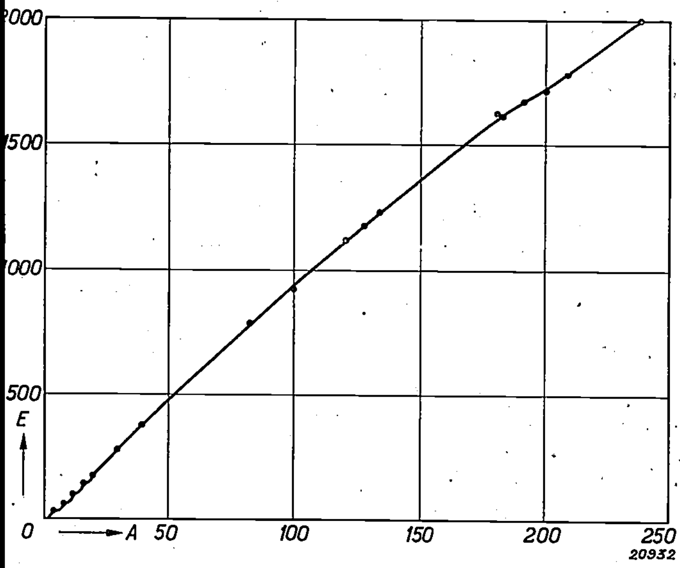


Fig. 2. Binding energy of the nucleus in 0.001 atomic-weight units (a smooth curve is drawn through the points most accurately known).

In *fig. 2* the binding energy of the atomic nuclei is plotted as a function of the mass number, and in *fig. 3* the same values are given for the lightest nuclei ($A < 50$). On the whole the proportionality is fairly good, although for the lightest nuclei there is a distinct oscillation in the curve, the binding energy being a maximum at $A = 4, 8, 12$, etc. Since $A = 4$ is the mass number of the helium atom $_2He^4$, the firm bond in nuclei whose weights are multiples of 4 is regarded as evidence of the presence of separate alpha particles in the nuclei.

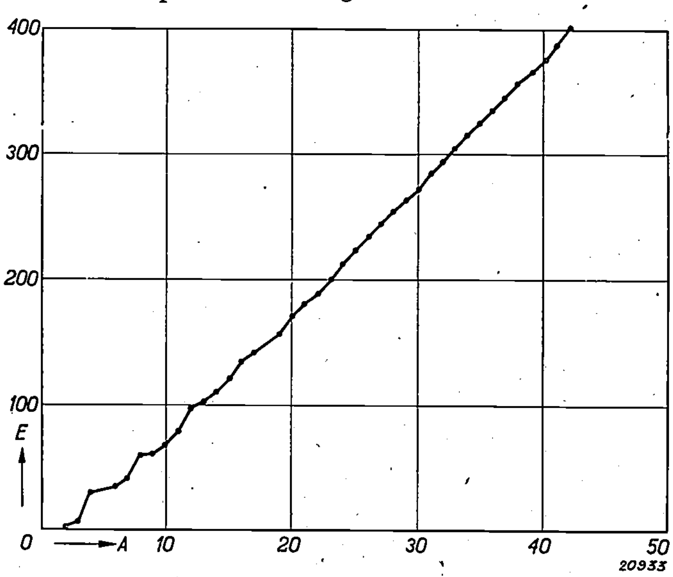


Fig. 3. Binding energy in 0.001 atomic-weight units of light nuclei (up to Ca^{42}). The oscillation may be followed up to $A = 24$.

It may also be asked to what maximum values the binding energy may assume for each elementary particle. This value may be found by dividing the energy given in fig. 3 by A. As shown in *fig. 4* this energy for $A > 20$ is approximately a constant, which already follows from the fact that the curve in fig. 2 is a straight line.

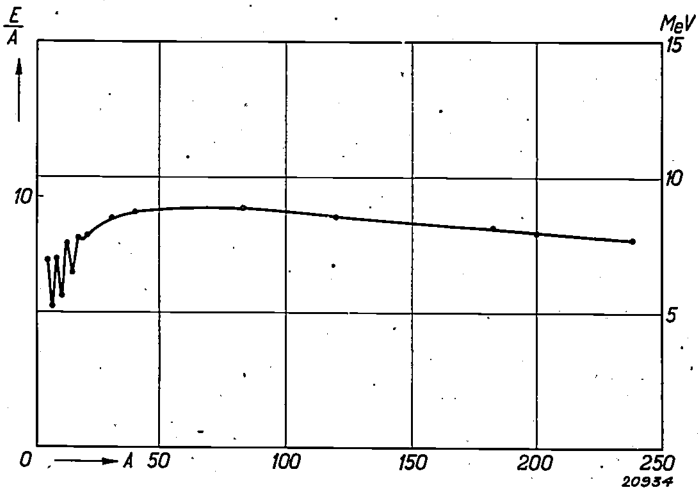


Fig. 4. Binding energy per particle for all nuclei (to left in 0.001 atomic-weight units; to right in MeV). Through the most accurately-known points a smooth curve has been drawn. The oscillation for $A < 20$ is indicated diagrammatically by a zigzag line.

Conception of the Nucleus as a "Liquid Drop"

The constant energy of approximately 0.008 atomic weight units ($= 12\ \mu$ erg $= 7.5$ MeV) which is required to expel a single elementary particle (proton or neutron) from the nucleus, has led to the atomic nucleus being compared to a "liquid drop" in view of the analogy between this energy value and a "heat of vaporisation". The application of this conception will be now discussed from various aspects. The comparison of the nucleus to a liquid drop can be developed in considerable detail and is even susceptible to quantitative expression. It postulates, for instance, that the volume of the nucleus is proportional to the "quantity of liquid" i.e. A. The radius of the nucleus, assumed to be spherical, must therefore be proportional to $\sqrt[3]{A}$ or $A^{1/3}$. Estimates made by other methods of the radius of the heavy nucleus actually give to a reasonable approximation a radius of:

$$r = 2.0 \cdot 10^{-13} \cdot A^{1/3} \text{ cm.}$$

In the case of a true liquid drop, the surface tension must be considered in addition to the energy of condensation; the former is an energy magnitude which apparently reduces the condensation energy in proportion to the area of surface. It is, indeed, found that the reduction in energy per elementary particle (*fig. 4*) observed at low values of A can be expressed numerically by a function which is proportional to $A^{2/3}$, i.e. to the square of the radius or to the surface (this energy per elementary particle is therefore proportional to $A^{-1/3}$, and hence increases with diminishing A). A plausible explanation is also forthcoming for the small downward

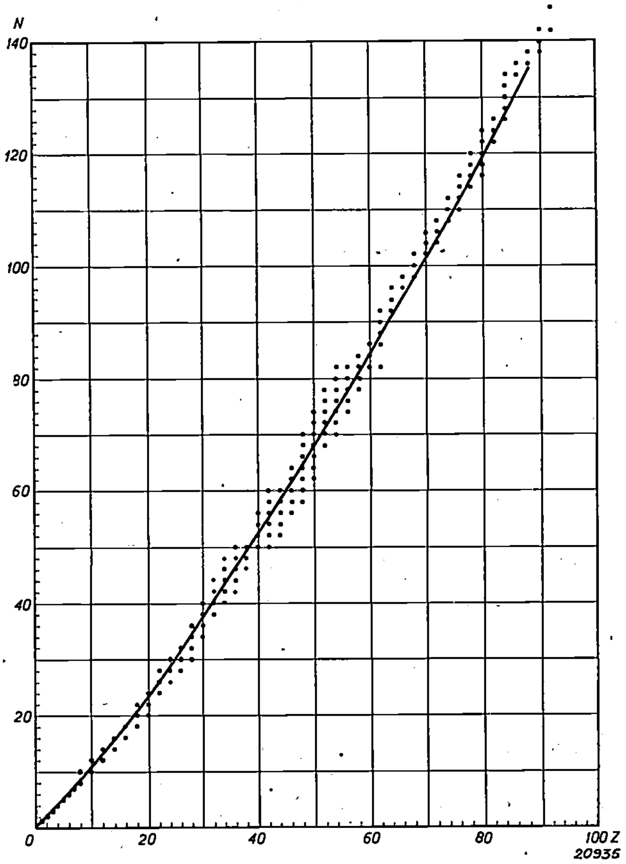


Fig. 5. N-Z diagram for nuclei with even values of A. Above $Z = 7$ only even values of N and Z are obtained. Isobars, i.e. nuclei with identical A values and different Z and N values (· .), are general features; isotopes, i.e. nuclei with identical Z and different A values, are frequent for the higher values of Z.

slope, shown in fig. 4 towards the highest values of A. This is accounted for by the mutual repulsion of the protons, the energy of repulsion being according to Coulomb's law proportional to Z^2/r, i.e. approximately to $A^{5/3}$. For each elementary particle this energy is proportional to $A^{2/3}$, and it increases with rising values of A.

Regarding the ratios of the numbers of protons and neutrons present in different atomic nuclei experience shows that these vary from $N/Z = 1$ for the lightest nuclei to $N/Z =$ approximately 1.5 for the heavy nuclei. In *figs. 5* and *6*, the known atomic nuclei are plotted in a system of co-ordinates according to their N and Z values. Fig. 5 relates to nuclei with even values of A, and fig. 6 to nuclei with odd values of A. The nuclei with even values of A possess as a rule even values of N and Z, as indicated in the figure, while nuclei with odd values of A have an equal number of even and odd values for Z and N. Space does not permit a detailed discussion of the important conclusions which can be drawn from these diagrams regarding the forces operating between the various elementary particles.

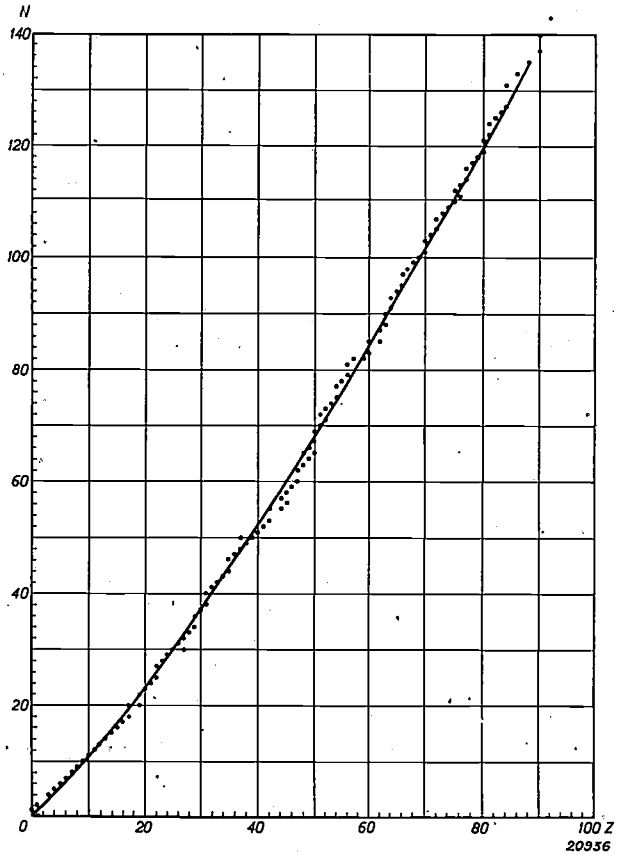


Fig. 6. N-Z diagram for nuclei with odd values of A. Even and odd values of Z are found with equal frequency. Isobars (· .) are rare. The number of isotopes is a maximum of 2 per element.

It must suffice to point out that by far the most powerful binding forces occur between protons and neutrons, and that a weaker bond exists between each pair of neutrons or protons (quite apart from the mutual repulsion between the protons as determined by Coulomb's law).

Nuclear Reactions

A brief survey has been made above of the stable nuclei and their binding energy. We will now discuss the reactions which are possible between atomic nuclei. In general these reactions consist of bombarding a static atomic nucleus (K_1) with a light nucleus which may be termed the projectile (P_1). The following may act as projectiles:

1. The hydrogen nucleus (proton) . . . . $_1H^1$
2. The nucleus of the heavy-hydrogen atom (deuteron) . . . . $_1H^2$
3. The helium nucleus (alpha-particle) . . $_2He^4$
4. The neutron . . . . $_0n^1$

The high velocity of the projectiles is derived from the intrinsic energy contained in radioactive substances or by means of canal or positive rays (positive ions accelerated in a suitable discharge tube).

The liquid drop concept is also of value here in depicting the mechanism of the processes taking place. The projectile P_1 penetrates the nucleus K_1 and forms an intermediate nucleus K_{int} with it. In this intermediate nucleus the particles will however not be in the same state as in a stable nucleus of the same weight. Both the kinetic energy of the projectile and the binding energy liberated are absorbed by this intermediate nucleus. It may be said that the drop has acquired a higher "temperature" and can therefore volatilise more easily, i.e. it tends to emit a projectile P_2. If this secondary projectile is identical with P_1, only a scattering process has taken place. But if P_2 differs from P_1, a nuclear reaction has resulted:

$$P_1 + K_1 \rightarrow K_{int} \rightarrow K_2 + P_2 ,$$

and a new nucleus K_2 takes the place of the initial K_1. This process is shown schematically in *fig. 7*.

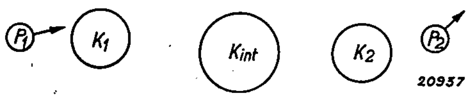


Fig. 7. Diagram of a nuclear reaction resulting on the bombardment of a nucleus K_1 with a projectile P_1, as a result of which an intermediate nucleus K_{int} is produced, which then disintegrates into a new nucleus K_2 and a projectile P_2.

We shall now discuss a few examples of nuclear reactions:

If nitrogen is submitted to bombardment with alpha-particles, high-speed protons will be emitted as a result thereof:

$$_2He^4 + {_7N^{14}} \rightarrow {_8O^{17}} + {_1H^1}$$

The importance of this reaction is that it was the first transmutation of this type discovered (Rutherford, 1919).

Another well-known reaction is that observed by Cockroft and Walton in 1932 which was initiated by bombarding lithium with high-speed protons (hydrogen positive rays):

$$_1H^1 + {_3Li^7} \rightarrow 2\ {_2He^4}.$$

In this case the secondary projectile P_2 was the same as the residual nucleus K_2.

Induced Radio-Activity

As our third example we shall refer to the action of alpha-particles on aluminium, resulting in two reactions taking place simultaneously, viz.:

$$_2He^4 + {_{13}Al^{27}} \begin{matrix} \nearrow & {_{14}Si^{30}} + {_1H^1}. \\ \searrow & {_{15}P^{30}} + {_0n^1}. \end{matrix}$$

Protons as well as neutrons may occur as secondary projectiles. It is interesting to note that in the second reaction the product $_{15}P^{30}$ represents an unstable nucleus which by the emission of an electron (viz., a positive electron or positron) is converted to a stable nucleus with a half-life period of 3.25 min. This was the first case of induced radio-activity and was observed by Irene Joliot-Curie and F. Joliot in 1934. The disintegration of the phosphorus atom takes place according to the following equation:

$$_{15}P^{30}\,{*} \rightarrow {_{14}Si^{30}} + e^+.$$

The asterisk * denotes that the product is unstable.

An important series of transmutations are those in which the primary projectile is the nucleus of the heavy-hydrogen atom $_1H^2$ (deuteron). These reactions were studied principally by Lawrence, and of which the following is an example:

$$_1H^2 + {_{11}Na^{23}} \rightarrow {_1H^1} + {_{11}Na^{24}}\,{*}.$$

Again here unstable, i.e. radio-active, products are frequently formed. In the case cited above the disintegration process is expressed by:

$$_{11}Na^{24}\,{*} \rightarrow {_{12}Mg^{24}} + e^-.$$

The sodium nucleus under consideration is therefore a beta-radiator.

Finally, there is an important group of transformations which are initiated by neutrons, the neutrons required for this purpose being generated by one of the two following methods:

1. From a radio-active preparation (alpha-radiator) and beryllium:

$$_2He^4 + {_4Be^9} \rightarrow {_6C^{12}} + {_0n^1}.$$

2. By the positive ray method in which lithium, beryllium or compounds of deuterium (D_3PO_4, ND_4Cl) [3]) are bombarded with deuterons.

The following transformations take place:

$$\begin{aligned} _1H^2 + {_1H^2} &\rightarrow {_2He^3} + {_0n^1}, \\ _1H^2 + {_3Li^7} &\rightarrow 2\,{_2He^4} + {_0n^1}, \\ _1H^2 + {_4Be^9} &\rightarrow {_5B^{10}} + {_0n^1}. \end{aligned}$$

Below some typical examples of neutron reactions are given; these have been mainly investigated by Fermi. We have intentionally selected three reactions which give the same end-product:

$$\begin{aligned} _0n^1 + {_{14}Si^{28}} &\rightarrow {_{13}Al^{28}}\,{*} + {_1H^1}, \\ _0n^1 + {_{15}P^{31}} &\rightarrow {_{13}Al^{28}}\,{*} + {_2He^4}, \\ _0n^1 + {_{13}Al^{27}} &\rightarrow {_{13}Al^{28}}\,{*}. \end{aligned}$$

[3]) D is the common abbreviation for $_1H^2$.

The $_{13}Al^{28}$ * disintegrates further according to the equation:

$$_{13}Al^{28}\,{}^{*} \rightarrow {}_{14}Si^{28} + e^{-}.$$

The last of these three reactions is of particular interest. This type of process in which no secondary projectile is produced, but the neutron remains bound in the nucleus, is initiated by artificially retarded neutrons, i.e. neutrons which have been passed through a medium rich in protons (e.g. a layer of water or paraffin wax a few cm thick). As a result the neutrons have lost so much energy by collision with hydrogen nuclei that they retain only a small energy of the order of 0.001 to 100 eV in place of an initial energy of several MeV.

Yields from Nuclear Reactions

Brief reference must still be made to the yields from the nuclear reactions discussed above. Since the nuclear diameter is of the order of 10^{-12} cm, it would be expected that the probability of striking the nucleus K_1 (fig. 7) with a projectile P_1 will depend on the effective sectional area per nucleus of 10^{-24} cm². If instead of investigating the yield per atom, we consider the total yield it is necessary to include also the range of the projectiles in the medium in question and the concentration of the nuclei K_1. If the medium contains n nuclei K_1 per cub cm and if the range is R cm, then the yield will be

$$\int_0^R n\ \sigma(x)\ dx.$$

where σ is the effective section at the velocity of the particles at the point in question. This effective section diminishes rapidly with the velocity. In the case of charged projectiles with a low energy the electrical field must also be taken into consideration, as it prevents the (positively) charged particle (proton or alpha-particle) from penetrating the nucleus. The reduction from this cause in the yield of the reaction will be the lower the smaller the charges of P_1 and K_1, which explains why in the reaction investigated by Cockroft and Walton (protons bombarding Li) measurable transmutation was still obtained at exceptionally low voltages (10 keV).

On the other hand, there are other factors which may increase the probability of reaction. In the case of neutrons, for instance, in which the repulsion by the nuclear field may be neglected, the probability of a projectile striking its target may be very high at specific low velocities. We are referring to a resonance phenomenon, where the effective section is in many cases 10000 times the normal target surface of 10^{-24} sq. cm.

Importance of Nuclear Physics in Technology

Nuclear physics has become of importance to technology not only in regard to the need for evolving technical apparatus suitable for investigating the atomic nucleus, such as positive ray tubes with their auxiliary high-tension units (as already indicated at the outset), but also in regard to the future possibility of preparing adequate quantities of artificial radio-active materials for practical use. The applications of these substances will be mainly for medical, biological, chemical and general technological purposes.

Bibliography:

F. Rasetti: Elements of nuclear physics (New York 1937).

Mad. P. Curie: Radioactivité (Paris 1935).

C. F. v. Weizsäcker: Die Atomkerne (Leipzig 1937).

Lise Meitner und M. Delbrück: Der Aufbau der Atomkerne (Berlin 1935).

P. Debye: Kernphysik (Leipzig 1935).

ALTERNATING-CURRENT CIRCUITS FOR DISCHARGE LAMPS

By E. G. DORGELO.

Summary. In contradistinction to the incandescent electric lamp, the gas discharge lamp possesses electrical characteristics which are entirely different for those of an ordinary resistance. Therefore, when calculating the data for a proposed installation of gas discharge lamps modified methods must be employed; these are discussed in some detail in this article. The characteristics of the simplest lamp installation, consisting of a gas discharge lamp with a choke coil connected in series with it, are analysed as an example.

Introduction

A fundamental difference between gas discharge lamps and incandescent electric lamps is that the former must always be connected to the mains supply through a current-limiting device.

In its simplest form this component consists of a series resistance or a choking coil which limits the current passing through the lamp. If the running or starting voltage of the lamp is higher than the mains voltage, a transformer which can be built in with the choke must be inserted in the circuit. A so-called leakage transformer with the same electrical characteristics can also be used in place of the transformer connected in series with a choke.

The current-limiting device not only limits the current intensity but also modifies the character of the current. It is useful to know also the variations in the intensity and character of the current with alteration in the characteristics of the lamp or with fluctuations in mains voltage; a detailed knowledge of the method of operation of the current-limiting device is also desirable to enable its dimensions to be kept small and its manufacture inexpensive.

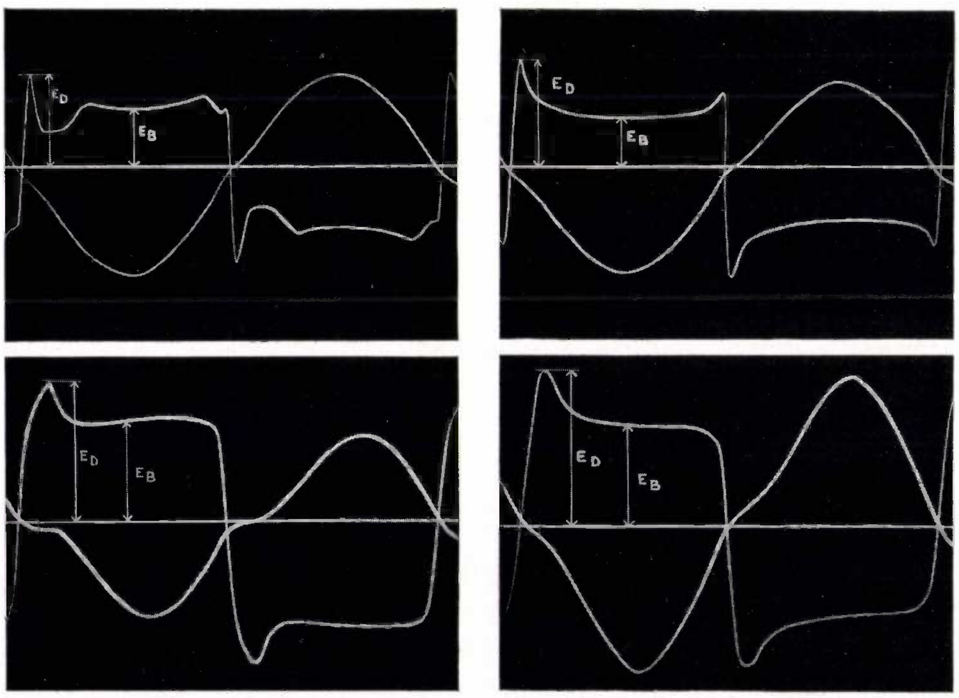
A brief survey is given below of the principal points which have to be considered in arriving at the electrical specification of discharge lamps and their current-limiting devices; it is also shown how — by adopting certain simplifications — the electrical characteristics of an installation can be calculated. It will be seen that each individual requirement of lamp voltage, power factor, the sensitivity of the lamp to voltage fluctuations, duration of the heating-up period, etc., introduces its own particular set of conditions, which cannot always be satisfied simultaneously, so that frequently a compromise must be made. These

Sodium lamp

Neon lamp

Mercury lamp SP

Mercury lamp HP



21109

Fig. 1. Voltage and current as a function of time for a sodium lamp, a neon lamp, a water-cooled SP mercury lamp and an air-cooled HP mercury lamp. Current commences to flow when the voltage has reached E_D; during the passage of current the voltage remains constant (to a first approximation) at E_B.

limitations are in part due to the simplicity of the circuit discussed here and which consists of a choke coil in series with the lamp. If, on the other hand, more complex circuits are used, a number of possibilities arise which cannot be discussed in this article.

Mathematical Simplifications

The following simplifications will be taken as the basis of our analysis:

1) The choke coil is assumed to combine a constant self-inductance L with an ohmic resistance R.
2) The mains voltage is assumed to be sinusoidal (with an instantaneous value of $E \sin \omega t$).
3) While the lamp remains burning, the voltage applied to the lamp is taken to be constant (E_B) with the current in opposition to it. To ensure that the lamp burns during each half-cycle, the voltage at its terminals must however first reach a value which is greater than E_B and which will be termed the striking voltage E_D.

The variation in the voltage as a function of the time is plotted in *fig. 1* for sodium, mercury and neon lamps. While the assumption that the running voltage is constant was found to be true in the case of mercury lamps, with sodium lamps on the other hand the oscillogram reveals a diminution of this voltage at the beginning and end of each half-cycle. But for most calculations it was found permissible to ascribe a constant mean value to the running voltage in these cases also.

Final Mains Current and Consumption Current

The inductive circuit described below is shown in diagrammatic form in *fig. 2*. It is assumed that the mains voltage ($E \sin \omega t$) is applied when its

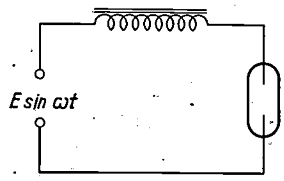


Fig. 2. The circuit under discussion. The lamp and choke coil are connected in series and the mains voltage $E \sin \omega t$ is applied. The choke coil has self-inductance L and resistance R.

instantaneous value is just zero ($\omega t = 0$); no current then passes through the choke coil and the lamp. Only when $E \sin \omega t = E_D$ (which is only possible when $E > E_D$), does ignition take place and a current is able to pass. The lamp thus behaves like a switch which is operated at the moment when $E \sin \omega t = E_D$.

We then have:

$$\omega t = \arc \sin \frac{E_D}{E}.$$

This value of ωt will be denoted by a. If we continue to compare the lamp with a switch, we must also assume that on closing the switch the circuit absorbs an e.m.f. with a value of E_D and a polarity in opposition to the applied mains voltage. *Fig. 3*

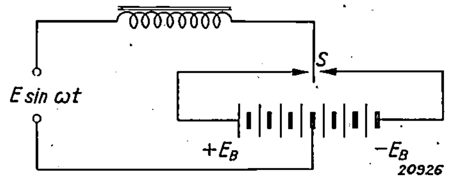


Fig. 3. Equivalent circuit for the lamp. An e.m.f. E_B which can be applied by closing the switch takes the place of the discharge lamp.

shows the equivalent circuit of the lamp described above. Switch S is closed as soon as the absolute value of the voltage exceeds E_D and remains closed until the current changes sign and becomes zero.

The voltage equilibrium obtaining in the equivalent circuit after closing the switch can be expressed by the equation:

$$E \sin \omega t = i R + L \frac{di}{dt} + E_B.$$

Solving this equation for i we get:

$$\left.\begin{aligned} i &= i_s + i_v, \text{ where} \\ i_s &= \frac{E}{Z} \sin (\omega t - \psi) - \frac{E_B}{R}, \\ \text{and } i_v &= \left[\frac{E_B}{R} - \frac{E}{Z} \sin (a - \psi)\right] e^{\frac{R}{\omega L}(a - \omega t)} . \end{aligned}\right\} \quad (1)$$

Here

$$Z = \sqrt{R^2 + (\omega L)^2} \text{ (total impedance)},$$

$$\psi = \arc \tan \frac{\omega L}{R}.$$

If in the equivalent circuit (fig. 3) the switch were kept closed for any arbitrarily long period, the current intensity i_v would become very small after a certain time ($\lim i_v = 0$). The current then flowing is given by:

$$i_{t=\infty} = i_s = \frac{E}{Z} \sin (\omega t - \psi) - \frac{E_B}{R}.$$

i_s is therefore termed the final current. As may be seen i_s is compounded of an A.C. component given by the mains voltage divided by the impedance of the circuit, and a D.C. component determined by the direct voltage E_B and the resistance R.

At the moment of closing the circuit ($\omega t = \alpha$), i_s will as a rule not become zero immediately. In this case the compensation current i_v ensures that no resultant current flows when the circuit is closed. Circuits are frequently encountered in electricity in which the circuit-closing phenomena occurring are of a transient nature. This is however not the case with gas discharges. Ignition is repeated twice in each cycle and the circuit-closing phenomena here have a continuous effect on the current curve, as will be further shown below.

Periodical Ignition and Extinction

As already indicated the switch S in the equivalent circuit (fig. 3) is opened the instant the current becomes zero ($\omega t = \beta$). It is shown in *fig. 4* that

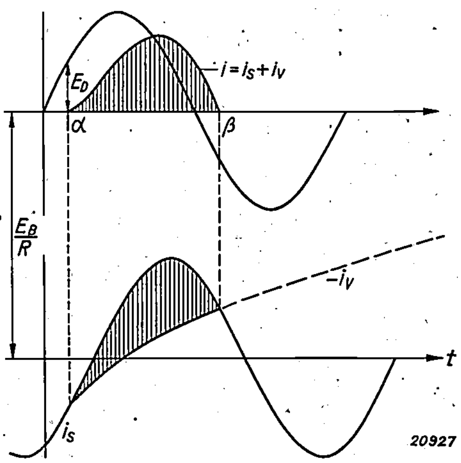


Fig. 4. Current intensity through the lamp. The final mains current derived from the mains voltage $E \sin \omega t$ and the impedance Z is represented by:

$$i_s = \frac{E}{Z} \sin (\omega t - \psi) - \frac{E_B}{R}$$

and the compensation current by:

$$i_v = \left[\frac{E_B}{R} - \frac{E}{Z} \sin (\alpha - \psi)\right] e^{\frac{R}{\omega L}(\alpha - \omega t)}$$

The hatched ordinates between the curves i_s and i_v represents the current intensity i. In the case illustrated the discharge period is exactly half a cycle ($\beta = \alpha + \pi$).

the current rapidly becomes zero, in fact so quickly that no marked diminution in i_v as yet takes place. When the current has once become zero, it remains at this value until the absolute value of the mains voltage has again attained the value E_D. During tbis interval, which may be termed the re-ignition interval δ, not only is $i = 0$ but also $di/dt = 0$. The only voltage remaining is then the mains voltage and it also determines the instant of re-ignition. In general the polarity of the mains voltage is then reversed; after re-ignition we can again make use of our equivalent circuit diagram, although now the switch must be closed in the opposite direction.

It is evident that the expression which we now obtain for the current again contains the two components i_s and i_v. i_v is again determined by the condition that at the instant of re-ignition $i = 0$. Hence contrary to circuits which do not contain a discharge lamp, the compensation current never entirely disappears here. It is this component which gives the current a permanent, non-sinusoidal character.

If the voltage E_D at which the lamp commences to burn is constant, re-ignition occurs when $\omega t = \alpha$, $\alpha + \pi$, $\alpha + 2\pi$, etc. The expressions for i_s and i_v are the same in all half-cycles, apart from the sign. To determine the character of the current it is therefore sufficient to consider a single half-cycle.

Discharge Interval and Dark Interval

We shall now return to equation (1). To simplify further analysis it will be assumed that R can be neglected with reference to ωL. The equation giving the current can then be derived from equation (1) by expanding the power function as a series, so that at the limit $R \rightarrow 0$ only the first two terms remain. We then get:

$$i = \frac{1}{\omega L} \left[E (\cos \alpha - \cos \omega t) + E_B(\alpha - \omega t)\right]. \quad (2)$$

By calculation or by means of the construction shown in *fig. 5*, the instant ($\omega t = \beta$) can then be determined at which i becomes zero for the first time. Fig. 5 shows that the value of β is determined by E_D and E_B and that the discharge interval ($\beta - \alpha$) can be both longer and shorter than a half-cycle.

If $\beta - \alpha < \pi$ the current will remain zero for a finite interval after extinction has taken place

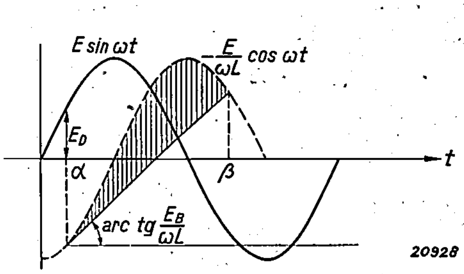


Fig. 5. The same diagram as fig. 4, but for $R = 0$, i.e. $Z = \omega L$. The e-function is resolved into a straight line with the slope $E_B/\omega L$.

and until re-ignition of the lamp takes place in the opposite direction. The dark interval is given by $\delta = \pi - (\beta - \alpha)$. Except for their algebraical signs the voltages and currents are exactly the same in the second half-cycle as in the first.

If $\beta - \alpha = \pi$, the dark interval disappears, as is generally desirable, since the lamp radiates very little light during this period and will be subject to much less flickering when the dark interval is eliminated.

Consider now the case when $\beta - \alpha > \pi$. It is seen that in this case also the lamp is immediately re-ignited in the opposite direction after extinction. The voltage at the instant re-ignition takes place is now greater than E_D in the negative phase, while in the positive phase the voltage is exactly equal to E_D. The characters of the two phases are therefore not identical. *Fig. 6* shows a sequence of half-cycles where it is apparent that the positive phase is longer and the negative phase shorter than a half-cycle. The sum of a positive and a negative phase is however slightly longer than a whole cycle, so that the variation of the current is not

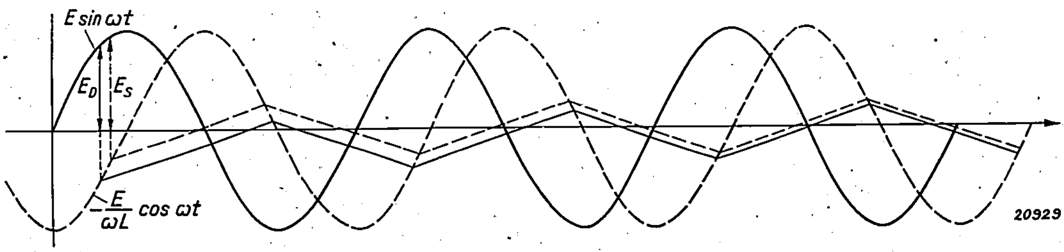


Fig. 6. The same diagram as fig. 5, but expanded to a number of periods. If the first discharge lasts longer than a half-cycle (full line) the third discharge will be shorter than the first, the fifth shorter than the third, and so on. The full line gradually approaches the dash line which represents the equilibrium condition in which the duration of the discharge is exactly half a cycle. At this equilibrium the voltage E_s available for re-ignition is greater than E_D.

periodic. As shown in the figure, the times of ignition are displaced progressively such that the equilibrium condition indicated by the dash line is attained. In this state the two half-phases are identical and are equal to a half-cycle. The voltage E at the instant of re-ignition is the same for both half-cycles in this equilibrium condition and is greater than the striking voltage E_D.

We thus arrive at the conclusions that the mains voltage may exceed the striking voltage at the instant the lamp is re-ignited. The excess value of the mains voltage is a measure of the reliability of re-ignition.

It is therefore important to know whether at given values of E_B, E_D and E the dark interval can be zero or not. The criterion for this is determined as follows.

Substituting $\omega t = \alpha$ in equation (2), we get as required $i = 0$. At the instant $\omega t = \beta$ where again $i = 0$ we get for the equilibrium state:

$$\omega t = \beta \leqslant \pi + \alpha,$$

where the sign of equality applies when the dark interval is zero. On inserting this value of ωt in equation (2) we get:

$$\cos\alpha \leqslant \frac{\pi}{2}\frac{E_B}{E}, \quad \ldots\ldots\ldots (3)$$

the sign of equality again applying to the absence of a dark interval.

The voltage E_s at the instant of re-ignition however satisfies the following equation which can be readily deduced from fig. 6:

$$\sin\alpha = \frac{E_S}{E} \geqslant \frac{E_D}{E}, \quad \ldots\ldots (4)$$

where the sign of equality now implies the existence of a dark interval, i.e. for just the converse case to that for which equation (3) is valid. We thus arrive at the following scheme:

With interval	Interval just zero	No interval
$\sin\alpha = \frac{E_D}{E}$	$\sin\alpha = \frac{E_D}{E}$	$\sin\alpha > \frac{E_D}{E}$
$\cos\alpha = \frac{\pi}{2}\frac{E_B}{E}$	$\cos\alpha = \frac{\pi}{2}\frac{E_B}{E}$	$\cos\alpha < \frac{\pi}{2}\frac{E_B}{E}$
$\left(\frac{\pi}{2}\frac{E_B}{E}\right)^2+\left(\frac{E_D}{E}\right)^2>1$	$\left(\frac{\pi}{2}\frac{E_B}{E}\right)^2+\left(\frac{E_D}{E}\right)^2=1$	$\left(\frac{\pi}{2}\frac{E_B}{E}\right)^2+\left(\frac{E_D}{E}\right)^2<1$

The third line in this table is obtained by squaring and adding the first two equations. It is found that the following expression may be taken as the criterion for immediate re-ignition:

$$\left(\frac{\pi}{2}\cdot\frac{E_B}{E}\right)^2 + \left(\frac{E_D}{E}\right)^2 \leqslant 1 \quad \ldots\ldots (5)$$

If the sign of equality applies, then α is determined both by the discharge interval given by equation (3) and by the dark interval given by equation (4); where the sign $<$ applies, α is determined by the discharge interval alone, when $E \sin \alpha > E_D$, and the difference $E \sin \alpha - E_D$ may be taken as a measure for the reliability of re-ignition.

Reliability of Re-ignition and Power Factor

It is interesting to examine what cases arise when E_B and E_D vary from 0 to E either together or independently of each other.

This variation can in fact be examined in the case of a mercury lamp, for on heating the lamp the pressure of the mercury increases and hence also the values of E_B and E_D.

At very low values of E_B/E and E_D/E, equation (5) is definitely satisfied. It follows from equation (3) that α is approximately 90 deg [1]). Ignition therefore takes place when the mains voltage has almost reached its peak value. The reliability is therefore very high.

If E_B and E_D increase, equation (3) indicates that α diminishes. Ignition is therefore obtained relatively earlier. The amount by which the mains voltage exceeds E_D is gradually reduced and eventually becomes zero (*fig.* 7). On further raising E_B and E_D, α must again increase, and the instant of ignition is then no longer given by equation (3) but by equation (4), at the same time a dark interval is obtained. We have thus passed through the minimum value of α. How small α can be made depends on the ratio E_D/E_B, i.e. merely on the lamp.

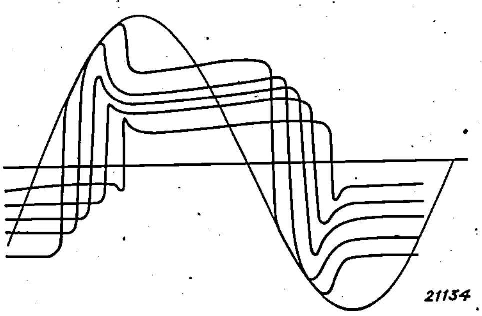


Fig. 7. This diagram is based on oscillograms recording the variation of the electrode voltage of a mercury lamp. The five exposures were made at short intervals during the heating-up of the lamp. The gradual increase in the striking and running voltages is indicated as well as the resulting displacement of the moment of ignition. In the case of an air-cooled mercury lamp this increase ceases before the striking voltage becomes greater than the mains voltage at the moment of extinction. In the case of a water-cooled mercury lamp with which the above diagram was obtained, heating-up continucs after the mains voltage is reached. A dark interval occurs, during which the voltage impressed on the lamp is exactly equal to the mains voltage.

[1]) An extreme case occurs when the lamp is short circuited; the current then passes through zero at the instant the mains voltage just passes through its peak value.

If, for instance $E_D/E_B = 1$, then it follows from equation (5):

$$\left[\left(\frac{\pi}{2}\right)^2 + 1\right]\left(\frac{E_B}{E}\right)^2 = 1.$$

Hence $E_B/E = 0.536$ and $(\cos \alpha)_{max} = 0.536 \cdot \pi/2 = 0.84$. With a higher ratio of E_D/E_B the maximum value which $\cos \alpha$ can assume is naturally lower.

The greater the value of $\cos \alpha$ the smaller will be the phase displacement between the current and the mains voltage, in so far as one can speak of a phase displacement in the case of a non-sinusoidal current. It may therefore be expected that there is a close relationship between the instant of ignition and the power factor. For a sinusoidal current the power factor is usually given as $\cos \varphi$, which we define as the ratio of the true power to the wattless or apparent power.

Limiting consideration to the case where the dark interval is zero, the power W is equal to the running voltage multiplied by the mean current intensity during a half-cycle:

$$W = E_B \cdot \frac{1}{\pi} \int_{\alpha}^{\alpha+\pi} i \, d(\omega t).$$

By means of equation (2) we can solve the integral and thus get:

$$W = E_B \frac{\frac{2}{\pi}\sqrt{E^2 - \left(\frac{\pi}{2} E_B\right)^2}}{\omega L}.$$

Inserting the effective value (E_{eff}) of the mains voltage, we have:

$$W = E_B \cdot \frac{0.9 \sqrt{E_{eff}^2 - (1.11\, E_B)^2}}{\omega L} \quad . \; . \; (6)$$

The apparent power VA can be calculated in the same way, thus:

$$VA = E_{eff} \cdot i_{eff} = E_{eff} \cdot \sqrt{\frac{1}{\pi} \int_{\alpha}^{\alpha+\pi} i^2 \, d(\omega t)}.$$

and we get the expression:

$$VA = E_{eff} \cdot \frac{\sqrt{E_{eff} - (1.09\, E_B)^2}}{\omega L} \quad . \; . \; (7)$$

As the terms under the root signs in (6) and (7) are almost the same, we have for the power factor η to a first approximation:

$$\eta = \frac{W}{VA} = 0.9 \cdot \frac{E_B}{E_{eff}} = 0.81 \cos \alpha.$$

In this circuit therefore the ratio of the running voltage of the lamp to the effective mains voltage is a measure of the power factor.

In the case discussed above where the maximum value of *cos* α is realised, the maximum power factor is also obtained, but is only 0.68. It has indeed not been found possible to obtain a higher value with the inductive circuit. Generally E_D/E_B is greater than 1, which results in a still lower power factor. The power factor can be improved by connecting a condenser, which compensates the wattless current component, in parallel with the lamp and choke coil.

Choice of Running Voltage

When designing the combination consisting of the discharge lamp and the current limiting unit, the first point to be determined is the voltage which has to be applied to the lamp and choke coil. A preference is usually shown for direct connection to the mains (e.g. 220 volts for HO-mercury lamps). Other lamps can be run with better efficiency on higher voltages, in which case the mains voltage must be stepped up by transformation (e.g. 440 volts for SO-sodium lamps, 600 volts for water-cooled SP-mercury lamps). To obtain a maximum power factor the running voltage is taken as high as possible for a specific mains voltage (when also the dimensions of the choke coil are reduced to a minimum).

Hitherto we have regarded the striking voltage E_D as having a fixed value, being determined by the characteristics of the lamp. Actually this is not the case, for with all discharge lamps the voltage required for re-ignition increases during the dark interval, as indicated in *fig. 8* by the three thin

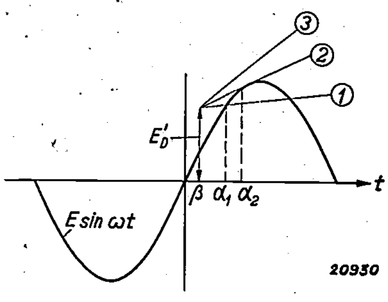


Fig. 8. In the example shown here the mains voltage at the instant β is smaller than the voltage E_D required for re-ignition. It is therefore necessary to wait until the mains voltage has reached this value. In general the voltage required for ignition increases during this interval. *1*, *2* and *3* represent points at which the mains voltage eventually predominates and starts the discharge, while in *3* this does not occur and no re-ignition results.

lines. It may then occur that the instantaneous value of the mains voltage, which at the beginning of the dark interval is smaller than E_D, does not increase sufficiently quickly to attain the voltage required for re-ignition, so that no re-ignition at all takes place (fig. 8, case 3). To prevent this happening the running and striking voltages of the lamp must not be made too high, so that condition (5) specifying the absence of a dark interval down to zero is satisfied.

Stability with respect to Fluctuations in Mains Voltage

The question of determining the lamp voltage has been discussed in principle above; there are however a number of other factors which necessitate a modification of the results arrived at.

The chief of these is a fluctuation in the mains voltage. With slow fluctuations the temperature of the lamp also may change, which in mercury lamps result in an alteration in the striking voltage. The variation of the latter is in the same sense as the fluctuations in mains voltage (the static characteristic is positive); with sodium lamps these variations are in opposite directions. If, for instance, the mains voltage drops, the striking voltage will rise, whereby re-ignition becomes less reliable. In this case the running voltage must therefore be taken lower than required by equation (5) in order to preserve full reliability in service.

On sudden variations in the mains voltage the running voltage may vary also in the opposite direction with mercury lamps (negative dynamic characteristic), so that again here it must be put slightly lower. Fluctuations in mains voltage affect also the consumption of energy and hence the luminous flux. By differentiating equation (6) (at constant E_B) with respect to E_{eff} we get for the percentage variation of the power input:

$$\frac{dW}{W} = \frac{1}{1-\left(1.11\,\frac{E_B}{E}\right)^2} \cdot \frac{d\,E_{eff}}{E_{eff}}\,.$$

The percentage variation in power is therefore always greater than the percentage variation in the mains voltage, the difference being the greater the greater E_B/E. If a considerable fluctuation in W is to be avoided, then according to the last equation, E_B/E must be made too high. This again leads to the condition that the running voltage of the lamp should not be taken too high.

Special importance attaches to the value of the power absorbed with variation in the running voltage. With both mercury and sodium lamps, there is a tendency for the running voltage to increase with the time the lamp has been in use.

This may result in a reduction in power consumption, in other words an increased reduction in the luminous flux; there is, however, also the possibility that the power consumption may also increase and the reduction in light referred to is then either partially or totally compensated.

It follows from equation (6) that with an increase of E_B from 0 to E_{eff}, ω at first increases, and later begins to diminish again. The maximum value is obtained at $E_B = 0.64\ E_{eff}$.

If, for instance, $E_{eff} = 220$ V, the power consumption will increase with the age of the lamp for E_B below 140 volts, while it will diminsh for E_B above 140 volts [2]).

Heating-up of High-Pressure Mercury Lamps

In conclusion reference must be made to the heating-up of mercury lamps. After switching on a certain time must elapse before the voltage across the mercury lamp has reached its full value (10 to 15 mins with an HP lamp, $^1/_2$ to 2 secs with a watercooled SP lamp).

[2]) If E_D is greater than 1.57 E_B, then for $E_B = 0.64\ E_{eff}$ the condition that no dark interval shall occur is no longer satisfied. Extending our analysis to include this case, we should get in place of equation (6) a generalised expression for the power containing an independent variable in addition to E_B and E_D. This aspect of the subject cannot be further pursued here.

As already pointed out in a previous article [3]), a minimum starting current is required for heating-up the lamp, which is two to three times the normal lamp current. The starting current is roughly equal to the short-circuit current i_k of the current-limiting component:

$$i_k = \frac{E_{eff}}{L\omega},$$

while the normal running current is:

$$i = \frac{\sqrt{E_{eff}^2 - (1.09\ E_B)^2}}{\omega\ L}.$$

For i_k greater than 2 i, we have $E_B > 0.79\ E_{eff}$, thus giving in this case a minimum running voltage, which is so high that it conflicts with the requirements set out above. This difficulty may be overcome by using a choke coil during the heating-up period with a lower self-inductance than required for normal running. A method in general use for this purpose provides a fairly high saturation of the iron in the series connected choking coil. The self-inductance then diminishes with increase in the current intensity, such that the short-circuit current i_k does actually reduce the impedance of the choke.

[3]) Philips techn. Rev., 1, 129, 1936.

PUBLIC LIGHTING. PRINCIPLES OF ROAD LIGHTING

By G. B. VAN DE WERFHORST.

Introduction

Under the term public lighting we shall include in the present article every class of permanent lighting installation on public highways, squares, roads, bridges, viaducts and subways, provided essentially for the convenience of road traffic as distinct from water-borne and railway traffic.

There is by no means a consensus of opinion in regard to the principles of road lighting and their practical applications. This divergence of opinion cannot occasion surprise when it is remembered that in practically all countries the provision of public lighting is left in the hands of local authorities, each of which has adopted an individual policy not in harmony with those of other bodies. The latitude in this direction is due to the absence of uniform official regulations and the lack of universally accepted standards to act as a guide in developing a public lighting scheme. Furthermore, at one place the responsible official is the power-station engineer, and at others either the municipal architect, the traffic inspector or the local surveyor.

Nevertheless, a marked similarity may yet be observed in certain directions. These points of agreement in public lighting schemes may be summarised as follows:

Of the luminous flux provided by the public lighting in all countries at least 30 per cent is directed upwards into the sky.

The appearance of the lighting unit during the day is considered to be more important than the illumination furnished by night; this is well exemplified by the acanthus leaves on gas lamp standards, by the crosier-like ornamentation of lamp standards on esplanades, and the concrete and cast-iron pediments of modern lamp standards and masts.

In the majority of public lighting schemes the lamp standards are only 13 to 16 ft. high, this mounting height having been adopted without alteration from the early oil lamps for which from the considerations of convenience and weight the length of the lamplighter's pole was limited to about 16 ft.

In the present article the endeavour is made to arrive at consistent and rational principles of public lighting shorn of all subordinate considerations and antiquated conceptions.

In the first place, it must be accepted as a general postulate that public lighting constitutes a part of the totality of exterior illumination and that its efficiency and utility to the road user is in many cases closely determined by the other sources of exterior illumination available.

Exterior lighting irrespective of its actual nature may be divided into two groups on the basis of its general type:

a) The lighting is designed to illuminate the surroundings so that the user can discriminate these surroundings; to the road user the surroundings are of principal interest, and the lighting unit itself of secondary importance;
b) The illuminant is designed to attract the attention of the road user and provide him with a signal; in this case the surroundings which may happen to be illuminated at the same time are of secondary interest, and principal importance attaches to the lamp itself, i.e. to the radiation which it directs on the eye of the road user.

An entirely different classification is obtained if lighting systems are classified according to their functions:

c) Lamps which are provided generally for the convenience, benefit, needs, safety, and pleasure of road users; the interests of the general public are here of primary consideration.
d) Lamps installed for the private convenience, safety and advantage of those installing them; private or personal considerations here take first place.

Comparing the classification into a and b with that into c and d, the practice of road lighting gives the following subdivision:

The lighting of streets *per se* with all crossings and viaducts, etc., i.e. illumination in the public interest (function c), where principal importance pertains to the illuminated surroundings and not to the lamps (type a);

Illuminated signals, including those in use on railways, in navigation and on public highways, and on highways embracing vehicle lighting, traffic signals and indications, again always in the public interest (function c), where the lamps and not the surrounding commands first consideration (type b);

Publicity lighting and illuminated signs, irrespective of type, clearly installed for private interests (function d) and usually so installed that the light source must be viewed directly (type b).

Private lighting, i.e. the lighting of factory premises and areas, vehicle headlights for the personal convenience of drivers, the illumination of drives and entrances to buildings lying back from the highway; these private lighting arrangements are installed for private use (function d), but at times may also operate in the public interest; the design of these lighting installations is extremely diverse (both types a and b, also frequently combined).

These divergent types of external illumination with all their contradictory requirements are found side by side, a juxtaposition which compels an order of priority to be determined. Without such specified priority, it is impossible to realise the desired results or to evaluate the efficiency of a given scheme of exterior illumination and particularly of a scheme of public lighting, in view of interaction between the various intrinsically different lighting schemes.

We shall assume that the public interest has precedence over purely private considerations, and that also where the public interest has alone to be served, safety and benefit shall be given priority over convenience and artistic effects, and that finally other things being equal, a disadvantage to one class of road user shall receive prior consideration over the advantages of another class.

The demands which are made, whether rightly or wrongly, on different public lighting systems differ so markedly that they do not fall exclusively within one or other of the subdivisions in the above classification. Thus the requirements in built-up municipal and urban areas differ so much from those ruling in country districts, that a discussion of public lighting must give adequate consideration to this difference.

In the present article we shall therefore confine ourselves to the public lighting of country roads (class I, II and III roads [1])) outside of built-up town areas, the illumination of which may be generally described as "road lighting". We will first analyse the conditions on class I roads.

Class I Roads

The lighting of class I roads is provided in the public interest. If the question is asked what in these cases is the public interest it will be found that:

Fast automobile traffic is the principal factor to be considered; that it should be possible to travel at speeds of up to 50 m.p.h. is not an undue requirement:

Bicycle traffic is an important factor in some countries where no separate cycling lanes are provided alongside the main highway;

Slow-moving vehicles (handcars and horse-drawn vehicles) if not prohibited from using this class of road must also be given special consideration.

Pedestrian traffic can be neglected.

The public interest requires that traffic should be able to proceed with reasonable safety under the conditions ruling on the highway. To arrive at this factor of safety a specific visibility is necessary which must be provided by the road lighting. How the lighting scheme must be devised and installed to afford satisfactory visibility on a class I road is determined by this need for adequate visibility and in the case under discussion is the sole determinant, since:

Aesthetic aspects bearing on the surrounding landscape are not being considered (on the other hand, in built-up areas the contours of surrounding buildings are important factors);

There is less likelihood of interference from light sources extraneous to the road, so that their existence can be neglected when planning the lighting scheme (in built-up areas entirely different conditions again obtain).;

No consideration need be given to the pleasing appearance of objects on the road, since the question of safety alone has absolute priority (as opposed, for instance, to the requirements of the lighting on esplanades and promenades).

There can be no question of depending on the lighting equipment of vehicles to provide adequate illumination of the highway, for the reason, among others, that however efficient this illumination may be it will inconvenience road users travelling in the opposite direction to such an extent that safe driving is impossible; this policy would therefore be against the condition of prior consideration stated above as well as contravene the need for maximum safety. Roads on which it is sought to avoid the danger inherent in dazzle produced by the headlights of approaching vehicles by providing two traffic lanes have not been sufficiently long in use to enable any definitive conclusions to be drawn [2]).

It is evident therefore that the road must be illuminated by means of permanent lighting units.

[1]) Which country roads outside built-up areas rank as class I, II and III roads depends, *inter alia*, on the regulations in force in each country.

[2]) Cf. van de Werfhorst, De Auto, Sept. 26, 1935: Enkele en dubbele wegen en hunne verlichting.

Since fast motor traffic is the determining factor, we must determine under what conditions the visibility of surrounding objects is sufficient to meet the needs of the average motorist. Economic considerations compel us to adopt the technical solution of the problem which will satisfy the minimum requirements.

Perception of an object on a road takes place either through the centre of the eye: foveal vision, or through a point outside of the centre: peripheral vision. An object is first consciously observed when we look in its direction, i.e. when the eyes are directed towards it and foveal vision is obtained. Whether we actually recognise the object in doing so depends on the definition of the object and this again on the conditions of illumination.

In this connection the two following points must be kept in mind: The eyes of a motorist when travelling at a high speed are directed far to the front (Baart de la Faille, van Lennep [3])), and are not rigidly fixed, but move up and down and to the left and right through a small visual angle. The field of vision of the driver therefore contains only a small cone in which foveal vision is maintained continuously [4]), yet at times his eyes are now and again directed to the nearer field, although continuing to look straight ahead, so that all objects outside the foveal cone of vision are viewed peripherally. Nextly, the question arises in how far we can stipulate that recognition must accompany perception of an object on the road irrespective of the nature of the object.

At a speed of about 50 m.p.h. the average driver focusses his eyes on a point at least 300 to 400 yards in front of his vehicle, provided the conditions of illumination of the field of vision permit. If the eye is focussed on a nearer point the road surface and curb appear with too high a relative velocity. The road surface close to the vehicle becomes converted into a swiftly moving band passing along and under the car; concentration of the eyes in this direction of vision then becomes very fatiguing. It is evident that such concentration is opposed to safety to driving, and the eye must therefore be focussed to a further point if safety is to be assured [5]). The field of vision thus covers only that portion of the road situated at a distance of 300 to 400 yards and even further in front of the vehicle. All objects within a distance of 300 yards are viewed peripherally unless they are accidentally located in the small foveal cone. In the case of peripheral vision alone, there can be no question of the conscious recognition of an object; moreover when the object is within the cone of vision and less than 300 yards away recognition will usually not take place immediately.

At moderate speeds a driver must carry out four to six operations almost simultaneously; the average person can perform only two operations consciously at the same time, so that with a good driver the majority of his reactions must take place automatically.

If an object appears in the field of vision at a distance of 300 to 400 yards or more, it is sufficient for the driver to become aware that some obstacle is located on the road at the point in question. Recognition of the actual nature of the object is not essential in the first instance, for 14 to 18 seconds must still elapse before he reaches the object, i.e. more than sufficient time for performing his subconscious manipulations, whilst his conscious vision is removed from the distant field of vision and focussed entirely on the approaching object in order to be able to recognise it within a distance of 300 yards.

Entirely different conditions arise when peripheral vision operates within a distance of 300 yards and an obstruction suddenly appears within the field of vision and the driver has his eye focussed on a point more than 300 yards away.

It is fortunate that a very weak stimulus in peripheral vision is already sufficient to cause the eye by reflex action to be directed immediately on the point from which the stimulus emanates. Yet just because a weak stimulus is already sufficient to do this, the lack of a peripheral stimulus leads to the unconscious conclusion that no obstruction is present. This alone is sufficient to indicate the extreme danger of lighting a class I road in such a way that only the road surface is illuminated. By installing lamps of special design, frequently with suitable reflectors, it is attemped to distribute the luminous flux of the lamps almost wholly over the road surface in a longitudinal direction. This method results in the more efficient utilisation of the luminous flux generated, practically none being "lost" in the lateral direction, and enables lamp standards to be placed at greater intervals, etc., apparently a decided economic argument although the cost of the lighting units or lamps themselves is high. But it should be remembered that it is not the function of the lighting to give the road

---

3) Cf., *inter alia*, Mededeelingen van de Nederl. Stichting voor Psychotechniek, No. 1.

4) Regarding the mechanism of ocular vision, see Bouma, Philips Techn. Rev., 1, 102, 1936.

5) Attempts to overtake a vehicle in front is due to the obstructive sensation referred to here. Cf. also van Lennep, footnote 3.

surface a pleasing appearance, but to provide a degree of visibility which will contribute to safety. From road edges which are left dark in an installation of this type the road user does not receive a single stimulus, and — we repeat — this leads to the unconscious and extremely dangerous conclusion that no obstacle may arise from these boundaries. Illumination of the curbs and pathways skirting the highway is not a waste of light, but on the other hand an indispensable necessity.

In addition to the rapid response to weak peripheral stimuli, the complex optical system of the eye also possesses the remarkable ability of not registering any sensation due to an apparently useless stimulus. The same also applies to many stimuli in peripheral vision. This is a fortunate circumstance, for otherwise we would have to determine the sequence with which the eye moving continuously and rapidly sweeps over the whole field of vision. All stimuli which we recognise and nevertheless have not to register are simply neglected. These considerations lead to the rejection of another, unfortunately frequently-used, method of road lighting. Many highways are still illuminated with unscreened lamps at distances of 60 and 70 yards at a mounting height of 16 to 20 feet. As a result of glare from the lamps, a driver travelling along a road illuminated in this way observes on passing each lamp dark patches appearing and disappearing alongside the traffic lane [6]), not in the form of true shadow produced by a material object but as ill-defined dark spots continually changing in shape. A driver knowing this type of road knows from experience that these evanescent peripheral stimuli following in quick succession are only the play of shadows and are not produced by actual obstructions. Unconsciously he suppresses all reaction to these stimuli; he removes them from his visual faculties. But if such a shadow is suddenly produced by an actual obstruction, the resulting stimulus makes just as little impression on the experienced driver; his necessarily automatic operations are performed just as if the obstruction were absent. The collision which then usually occurs is, however, not due to irresponsible and extremely careless driving", in the words of the courts, but to the deceptive illumination provided, which is the more deceptive the greater the experience possessed by the driver and the better he knows the road.

The peripheral area up to a distance of 300 yards in front of the vehicle must therefore be so illuminated that each obstacle produces a perceptible stimulus; in addition the illumination of the whole field of vision must be so devised that no disturbing elements are present within it which are liable to result in a false interpretation of the shadows or react to the detriment of visibility. Moreover the lighting must provide the driver with adequate means of foveal vision as soon as a peripheral stimulus has attracted his attention. In contradistinction to the more distant field just referred to, the rate of perception and recognision of the observed object are here important factors.

How these requirements can be met for both the near and distant fields of vision is beyond the scope of the present article. Nevertheless we may already conclude that the lighting of class I roads entirely falls within the first subhead of the above classification: The lighting of the road itself with its immediate boundaries is the main consideration, the type of lamp a secondary matter, even to the extent that it is desirable to make it as inconspicuous as possible. This method of lighting possesses none of the characteristics of light signals or indicators. The public interest preponderates to such an extent that all illuminated signs or private lighting units must take second place to the road lighting scheme. If the maximum efficiency attainable with this lighting is reduced by vehicle lamps (cycle lamps, or too bright head-lamps of approaching vehicles), these latter must be either entirely prohibited or restricted to an extent that they no longer constitute a nuisance. Every troublesome illuminated sign or private lighting installation should be absolutely prohibited. As long as these restrictions cannot be enforced, nothing definite can be achieved and it is impossible to install lighting on class I roads which will completely satisfy the specified requirements.

But when all these conditions have been met, it must also be remembered when judging the efficiency and suitability of a lighting system that when a driver experiences a definite stimulus by peripheral vision and he directs his eye on the point from which the stimulus emanates in order to recognise it, a short interval of 0.3 to 0.5 second must elapse before full recognition takes place, even when no movement of the head is necessary and the object is such that binocular vision is still possible; otherwise a much longer interval will elapse. In this interval of time the driver will have

[6]) On approaching, passing and after passing each lamp, simultaneous indirect glare and successive glare are experienced in rapid succession, while on a long straight road the lamps further away produce continuously a simultaneous indirect glare and those lamps in the remote distance even a simultaneous direct glare. Cf. Bouma: The problem of glare in highway lighting, Philips techn. Rev., 1, 225, 1936.

traversed a distance of 22 to 37 feet, assuming he is travelling at 50 m.p.h. Whatever lies within this distance is thus quite unable to produce a conscious reaction. It could still be stipulated that the peripheral stimulus due to an objcct within this distance of 22 to 37 feet should be so pronounced that a reflex action is immediately initiated to avoid the obstruction. But just this reflex action has led to such serious accidents that, on the other hand, its absence is desirable in these circumstances. This signifies that nothing within the 22 to 37 feet interval can be utilised to judge the efficiency of a lighting scheme or has any bearing on the lighting.

If we assume that under perfectly normal traffic conditions, i.e. no possibility of skidding, or fog, an obstacle appears at a distance of over 33 feet from the vehicle, such that a conscious manipulation by the driver must result, which in the extreme case would mean stopping the vehicle, then the vehicle to conform with the official regulations must pull up within a distance of 64 yards when travelling at a speed of 50 m.p.h. [7]). The authorities regard this distance for pulling up as safe at the speed in question. From the moment the obstacle produces a peripheral sensation to the instant the vehicle is brought to a stop the car would travel 74 yards. Let us examine what in this case could and could not be demanded of the lighting. If the obstacle is located within this distance of 74 yards, the illumination must be required to give an adequate peripheral stimulus and then permit the driver to see the obstacle satisfactorily by foveal vision. But in no case can the lighting be required to prevent a collision within this 74 yard interval, since this distance depends on factors extraneous to the illumination, so that such an accident could also take place in daylight. The opinion often expressed that then the driver should have been driving a little slower is untenable in ordinary circumstances. Means are not provided for safe driving at a speed of 50 m.p.h. and a maximum pull-up distance of 64 yards is not laid down for this speed, in order later to place the blame for an accident on either the driver or the given conditions of visibility, for which neither can be held responsible.

[7]) According to official regulations in the Netherlands, the braking power of vehicles must be such as to pull up the vehicle in a distance expressed in metres equal to the square of the speed expressed in tens of kilometers per hour, thus at 30 km.p.h. the vehicle must be pulled up in 9 m, and at 40 km.p.h. in 16 m, taking as a basis an average retardation of 3.86 m per sec. per sec. The Code de la Route in force in France adopts the same pull-up distance as well as the same retardation, rounded off to 4 m per sec. per sec.

Class II Roads

The type of illumination required on class II roads outside built-up areas depends on the importance attaching to these roads. If the road is considered sufficiently important for the same requirements to be laid down as on class I roads, as well as consideration to be given to pedestrian traffic [8]), then the same remarks as given above for class I roads also apply to class II roads: Lighting provided by permanent equipment, no dependence on the illumination from the vehicles themselves, no disturbing illumination from signalling lamps, traffic indications, illuminated signs or private lighting installations (petrol stations).

If a Class II road is not considered to merit this equivalence with a class I road (the volume of traffic is too small, for instance) and if the road must nevertheless be lighted, it becomes first necessary to require vehicles (and bicyles) to carry lamps; the permanent lighting can then act as a type of beacon lighting. As a result the nature of the lighting scheme is radically altered: In the lighting now needed the lamp itself must be visible, while the actual illumination of the surroundings is a secondary matter. The brightness of the light source may be made low, since it is even then sufficient to render the lamp visible; it also must be low, so low in fact that it never produces a pronounced glare and that the effects due to the evanescent shadows referred to above are entirely avoided. The lighting must follow all bends in the road and indicate their position and direction, but not illuminate them; it is sufficient merely to demarcate them. Where special points require to be made unmistakably conspicuous as road junctions, canal bridges, etc., a more powerful beacon should not be installed at these points as is frequently the case, as it will concentrate the notice of the road user on the lamp itself instead of on the dangerous point. At these points the surroundings must be extra-brightly illuminated, and the lamps used for this purpose should themselves be as inconspicuous as possible. It should be remembered that a beam directed straight ahead from a vchicle approaching from the left or right is now invisible. If this effect is to be retained then suitable reflecting beacons must be set up.

Class III Roads

On class III roads where an unbroken line of

[8]) The question arises whether pedestrians also should be required to carry a signal light in the same way as other road users, since it is pedestrians who are overtaken by vehicles far more frequently than other traffic elements.

beacons will prove too costly, no permanent lighting is provided. Nothing is more dangerous than isolated unscreened lamps at considerable distances apart, which necessarily appear as beacons to the road user likely to deceive him and moreover detract from the utility of his own lamps.

For points which must be specially demarcated the same considerations apply as for class II roads: More powerful illumination of these points, with adequate screening of the lamps.

DEVELOPMENT AND MANUFACTURE OF MODERN TRANSMITTING VALVES

By H. G. BOUMEESTER.

Summary. The improvements which have been made in the various components of transmitting valves, and the developments resulting in modern transmitting valves, are discussed; special reference is made to pentodes, transmitting valves for ultra-short waves and high-power transmitting valves.

Introduction

Since Philips manufactured their first transmitting valves in 1919 considerable advances have been made both in the general construction and applications of these valves. These early transmitting valves were provided with a tungsten filament as a source of electrons and had an anode dissipation measured only in tens of watts. A few years later attemps were made to cool the anodes of these valves with water in order to obtain a greater dissipation. For this purpose the anode was made of a chrome-iron alloy developed in Philips Laboratory and which has the same coefficient of expansion as ordinary glass (100.10^{-7}). In a short time these water-cooled high-output transmitting valves were in widespread use; thus in 1926 thirty valves of type were being employed in parallel at the Postthis Office Radio Station at Rugby. In the space of a decade a whole series of modern transmitting valves was evolved from this laboratory product, such valves now being made by mass production methods and conforming to very strict specifications. The larger valves of this type may be likened to machines which transform direct-current energy into alternating-current energy of practically any required frequency.

These advances were rendered possible, *inter alia*, by the discovery of substances for making the cathodes which were capable of readily emitting electrons. At the same time the grids and anodes were improved, new kinds of glass were evolved for the valves and considerable progress made in perfecting methods of evacuating the valves. Various physical and chemical phenomena which play an important part in the operation of transmitting valves were more closely investigated, such as primary and secondary electron emission, the radiation of energy, and the action of materials capable of absorbing gases (socalled getters), etc. Some general problems relating to the manufacture and the use of transmitting valves are briefly reviewed below.

Electronic Emission from the Cathode

The cathode as the primary source of electrons must satisfy two essential requirements:

1) The requisite rate of electron emission must be obtained with the minimum cathode heating power.
2) The cathode must have a satisfactory life.

beacons will prove too costly, no permanent lighting is provided. Nothing is more dangerous than isolated unscreened lamps at considerable distances apart, which necessarily appear as beacons to the road user likely to deceive him and moreover detract from the utility of his own lamps.

For points which must be specially demarcated the same considerations apply as for class II roads: More powerful illumination of these points, with adequate screening of the lamps.

DEVELOPMENT AND MANUFACTURE OF MODERN TRANSMITTING VALVES

By H. G. BOUMEESTER.

Summary. The improvements which have been made in the various components of transmitting valves, and the developments resulting in modern transmitting valves, are discussed; special reference is made to pentodes, transmitting valves for ultra-short waves and high-power transmitting valves.

Introduction

Since Philips manufactured their first transmitting valves in 1919 considerable advances have been made both in the general construction and applications of these valves. These early transmitting valves were provided with a tungsten filament as a source of electrons and had an anode dissipation measured only in tens of watts. A few years later attemps were made to cool the anodes of these valves with water in order to obtain a greater dissipation. For this purpose the anode was made of a chrome-iron alloy developed in Philips Laboratory and which has the same coefficient of expansion as ordinary glass (100.10^{-7}). In a short time these water-cooled high-output transmitting valves were in widespread use; thus in 1926 thirty valves of type were being employed in parallel at the Postthis Office Radio Station at Rugby. In the space of a decade a whole series of modern transmitting valves was evolved from this laboratory product, such valves now being made by mass production methods and conforming to very strict specifications. The larger valves of this type may be likened to machines which transform direct-current energy into alternating-current energy of practically any required frequency.

These advances were rendered possible, *inter alia*, by the discovery of substances for making the cathodes which were capable of readily emitting electrons. At the same time the grids and anodes were improved, new kinds of glass were evolved for the valves and considerable progress made in perfecting methods of evacuating the valves. Various physical and chemical phenomena which play an important part in the operation of transmitting valves were more closely investigated, such as primary and secondary electron emission, the radiation of energy, and the action of materials capable of absorbing gases (socalled getters), etc. Some general problems relating to the manufacture and the use of transmitting valves are briefly reviewed below.

Electronic Emission from the Cathode

The cathode as the primary source of electrons must satisfy two essential requirements:

1) The requisite rate of electron emission must be obtained with the minimum cathode heating power.
2) The cathode must have a satisfactory life.

Since these two requirements are directly in opposition to one another, a compromise has had to be effected, as is often necessary in technical problems of this kind.

The cathode must satisfy also a number of other requirements. Its behaviour must not depend too closely on the temperature of the surrounding electrodes, and it must also be suitable for use in valves with a comparatively high anode voltage. These conditions can be almost fully met by appropriately forming the electrodes and disposing them suitably within the valve.

Originally, a heated tungsten filament was used as a source of electrons, and even at the present day tungsten is still used in transmitting valves, particularly in valves with water-cooled anodes. The yield of thermal electrons which can be obtained with this form of tungsten cathode is greater the higher the temperature of the filament, but at relatively high temperatures the evaporation rate of the tungsten becomes large and the life of the filament is reduced as a result. An electron output of 5 to 8 mA per watt of cathode heating power can be achieved with a reasonable filament life. There has naturally been a search for materials giving a higher proportionate output of electrons; tungsten filaments with additions of 1.5 to 2 per cent thorium oxide have been made to furnish about 80 mA/W. A still higher yield may be obtained from the oxide-coated cathode consisting of barium-strontium oxide deposited on a metal base of either platinum, nickel, tungsten, copper or a variety of alloys. In the preparation of this cathode a thin layer of barium atoms is produced at the surface, as a result of which the work required for the emission of an electron is so far reduced that a yield of 200 to 300 mA/W is attained.

Still higher values can be obtained by using filament-type cathodes which are comparatively long, thus making the dissipation of heat at the ends of the filament practically negligible.

While oxide cathodes were successfully employed in transmitting valves up to moderately high outputs, their use in high-power and high-tension valves was initially limited by the above-mentioned secondary requirements. Nevertheless, it has been found possible to use oxide cathodes in the largest radiation-cooled transmitting valves, as in the PC 3/1 000 pentode. This valve operates with an anode voltage of 3 000 V, which can however be made higher without adversely affecting the cathode, even up to 4 000 or 5 000 V. But the use of oxide cathodes makes these high voltages quite unnecessary, since an adequate source of electrons is already available and comparatively high anode current and low anode voltages can be employed. This is frequently an advantage when using these valves in small transmitters.

So-called indirectly-heated cathodes are extensively used in receiving valves, but this type of cathode does not offer the same advantages in transmitting valves, although a number of valves, e.g. the PE 1/50 pentode, are fitted with them for other reasons.

Materials for Grids and Anodes

The function of the grid in a transmitting valve is to control the current intensity through the valve. If, in performing this function, the grid itself commences to emit electrons, for instance as a result of heating or of the impact of electrons upon it, the operation of the valve is adversely affected. By suitably cooling the grid, the emission of primary electrons from the grid can always be reduced to such a small value that no interference with the operation of the valve occurs. The dissipation of heat can be promoted by means of cooling ribs, cooling fins, heavy supports, etc., so that the temperature of the grid remains sufficiently low.

In the case of low-frequency amplifying valves, in which the grid has a negative average potential,

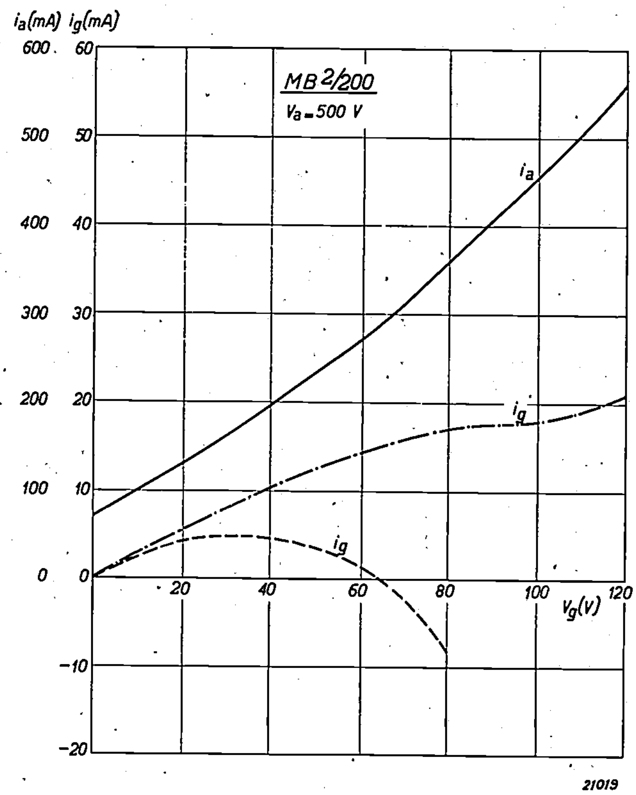


Fig. 1. Grid current i_g and anode current i_a plotted as a function of the grid voltage V_g for the MB 2/200 transmitting valve, for a zirconium —·—·— and for a molybdenum grid ------.

grid current is largely prevented and practically no heat is generated in the grid itself; temperature rise of the grid is almost entirely due to thermal radiation from the anode and cathode. If heat

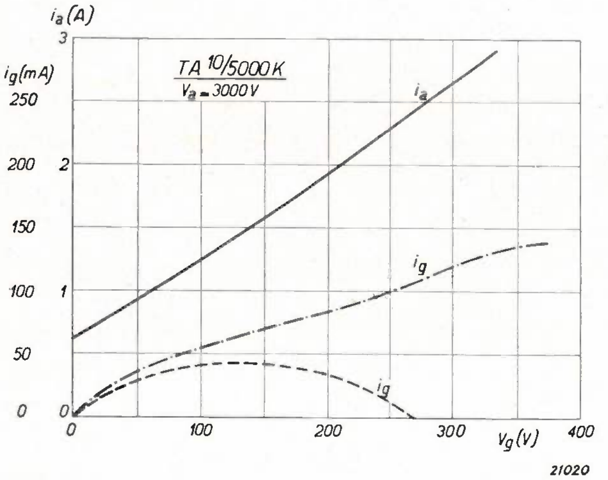


Fig. 2. Grid current i_g and anode current i_a plotted as a function of the grid voltage V_g for the TA 10/5 000 K transmitting valve, for a tungsten grid ------, and for a tungsten grid coated with zirconium oxide —·—·—·

were evolved at the grid itself, if should be made from a material which would easily radiate heat. On the other hand, in l.f. amplifying valves a substance must be used for the grid which readily reflects heat rays. Silver and copper are suitable for this purpose, and marked success has been obtained with grids wound from coppered wire.

The emission of secondary electrons from the grid can be effectively inhibited by suitably positioning the electrodes with reference to each other and by making the grids of certain materials. The anode current i_a and the grid current i_g in milliamps are plotted as a function of the grid voltage V_g in volts for a 200-watt amplifying valve (MB 2/200) in *fig. 1*. The dash curve was obtained with a molybdenum grid which at high positive grid voltages emits such a large amount of secondary electrons that the grid current does not increase continuously with the grid voltage and even becomes negative at grid voltages exceeding 62 V. On the other hand, the secondary emission is so small with a grid wound from zirconium wire that the grid current i_g rises continuously with the grid voltage, as indicated by the dot-dash curve.

A similar result can also be obtained with a tungsten grid coated with zirconium oxide. In *fig. 2* the anode current i_a in A and the grid current i_g in mA are plotted as a function of the grid voltage V_g in volts for a valve with a water-cooled anode. The dash curve relates to a tungsten grid whose secondary emission at grid voltages above 270 V is so great that the grid current i_g became negative. If the tungsten grid be coated with zirconium oxide, secondary emission is reduced to a reasonably small value, and grid

Fig. 3. The TB 2/250 triode.

current then increases uniformly with the grid voltage, as indicated by the dot-dash curve.

To keep the dimensions of the transmitting valve as small as possible the thermal loading of the anode must be as high as practicable, in other words the working temperature of a radiation-cooled anode must be raised to a level consistent with the need for permanently maintaining a satisfactory vacuum. If the anode is not water-cooled, it is desirable to make the anode of a material which at the specified temperature readily radiates heat. Nickel foil coated with carbon is a good thermal radiator and has been used for many years in valves with oxide-coated cathodes. Graphite anodes have also been used, but these introduce difficulties in valves run at high anode voltages, since occluded gases cannot be readily expelled from the graphite and there is a tendency for small particles to become detached from the anode and cause sparking in the valve.

If the anode is coated with a layer of very finely-divided metal in place of graphite it acquires a black surface also and will then readily radiate heat; hence it is very suitable for use in transmitting valves. Tungsten powder can now be prepared with a mean diameter of 10^{-6} cm per particle.

A triode TB 2/250 is shown in *fig. 3* which has a molybdenum anode coated with a firmly adherent layer of this pulverised tungsten.

Mechanical Construction

Special attention must be devoted to the mechanical construction of valves in view of the need for withstanding transport over long distances and for their use in portable transmitters, and ship and aircraft transmitters. The electrodes are frequently supported by being bonded together and to the glass bulb by means of ceramic insulating distance pieces having suitable electrical characteristics, or with mica supports. To enable a high vacuum to be created and maintained in the valve, these components must be submitted to special treatment before assembly.

The TC 2/300 valve designed for ultra-short waves is shown in *fig. 4* as an example of the mechanical construction of a transmitting valve. The grid and filament are supported by ceramic insulating pieces and insulated from the anode in like manner.

Before leaving the works, transmitting valves are submitted to a shock test which is applied to every large type of valve individually and to random samples in the case of the smaller types. In the arrangement shown in *fig. 5*, 240 shocks per minute are applied to 100 and 250 kW valves for a period of two hours. After this test a radiograph is taken to determine whether the grid and filament have sustained any mechanical damage or have failed to retain their relative positions; finally each valve has to undergo a complete electrical test.

Fig. 4. The TC 2/300 transmitting valve. The anode is welded at the top to a chrome-iron plate which is fused to the bulb; at the lower end it is supported against the bulb by a mica disc. The ceramic insulating distance pieces are visible above and below the anode.

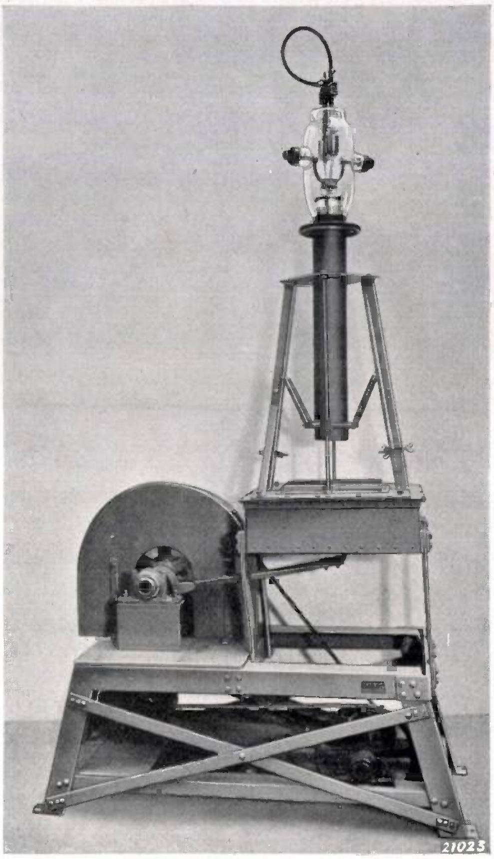


Fig. 5. Arrangement for shock tests on the large TA 20/250 transmitting valve.

Without entering into a detailed description of the various transmitting valves manufactured by Philips, a few examples of the principal types of transmitting valves made by us will be discussed in the light of the above considerations.

Tetrodes and Pentodes

Just as in the course of time, triode receiving valves have given place to screen-grid valves, for the principal reason that a smaller capacity was obtained between the grid and anode, so a similar development has also been followed in the design of transmitting valves. For the last eight years Philips have manufactured a series of tetrodes which dispense with neutrodyning in transmitters.

Another important advance made in recent years in the design of multi-grid valves has been the appearance of a complete series of pentodes *(fig. 6)*, possessing the advantage of simple and economical modulation at the third grid (suppressor grid). At the same time these valves give a higher efficiency than tetrodes, and are used, for example, in television transmitters. The smallest valve in this series (fig. 1) has an output of 15 watts, the largest radiation-cooled pentode is the PC 3/1000 rated for 1 kW, while the PA 12/15 is water-cooled and has an output of 15 kW.

High-Frequency Transmitting Valves

During recent years there has been a growing demand for transmitting valves operating with a satisfactory efficiency at frequencies exceeding $30 \cdot 10^6$ c/sec, i.e. on waves below 10 m. This class of ultra-short waves is being increasingly used for television purposes, in diathermy and in special signalling systems for military purposes or traffic control.

At these extremely high frequencies, all internal components of the transmitting valves must be made as small as possible, since the times of transit of the electrons cannot be neglected relative to the cycle of the electrical oscillation. In addition the capacities must be small and the leads very short, in order that their inductance remains within reasonable limits. Since the output of the valve must nevertheless be high, they cannot on the other hand be of too small dimensions, so that here again a compromise must be made.

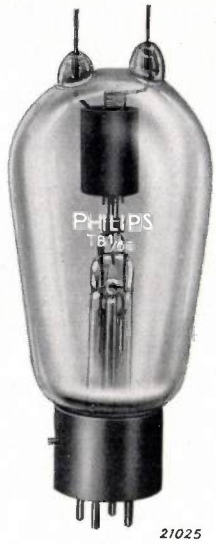


Fig. 7. The TB 1/60 transmitting triode.

A triode TB 1/60 for ultra-short waves is shown in *fig. 7* where it is seen that the leads are extremely short. The anode is made of carbon to obtain a high output (60 W with a wavelength of 4 to 5 m). This valve can also be used for wavelengths down to 1 m., although in these circumstances the output is much reduced.

Another point of considerable importance in ultra-short wave transmitting valves is the high dielectric losses in the insulators, and particularly in the glass forming the bulb. If no precautions are taken to eliminate these losses, a marked heating and the occurrence of electrolytic phenomena in

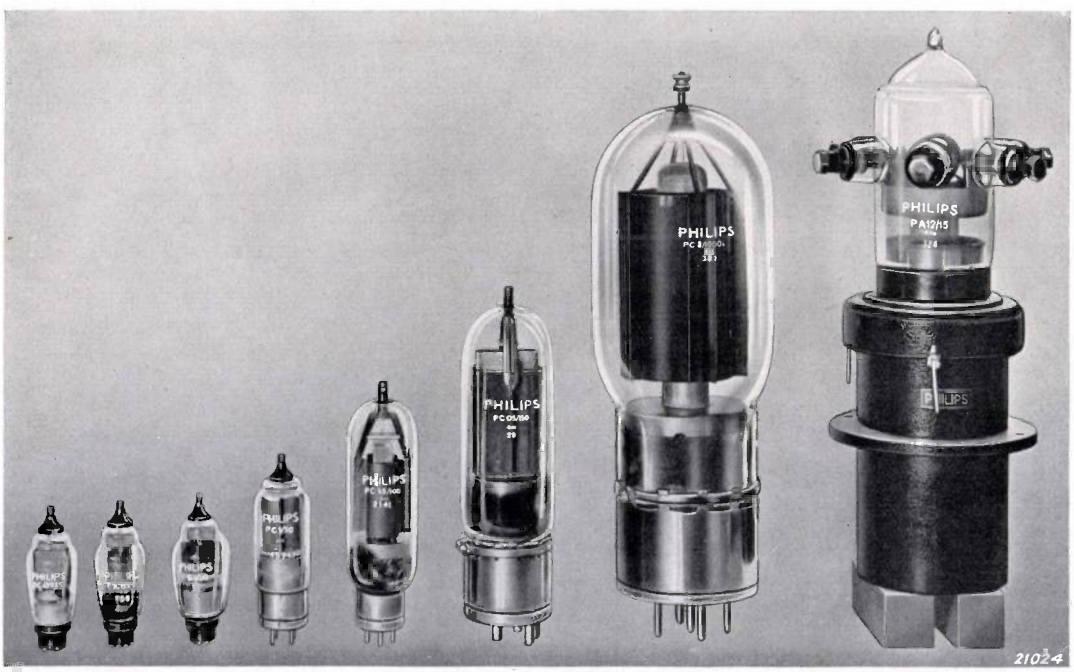


Fig. 6. Series of pentodes of which the smallest has an output of 15 W, the largest radiation-cooled valve PC 3/1 000 with an output of 1 kW and the water-cooled pentode PA 12/15 with an output of 15 kW.

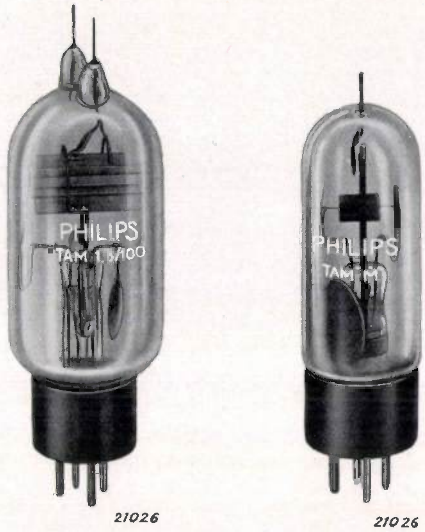


Fig. 8. The TAMM and TAM 1.5/100 magnetron valves.

the glass will result in the destruction of the bulb. In the TC 2/300 transmitting valve shown in fig. 4 the leading-in wires to the various electrodes have therefore been so positioned in the bulb with respect to each other that they are separated by glass walls over a distance which considerably reduces the electrical stress on the glass. This transmitting valve in which the permissible anode dissipation is 300 watts has been designed for use in diathermy apparatus.

For special signalling systems, the frequency band above $300 \cdot 10^6$ c/s is important, i.e. the wave band below 1 m. Single-grid and multi-grid valves are hardly suitable for these purposes since they can never furnish more than a few watts useful output in this frequency band irrespective of the circuit employed. On the other hand, a satisfactory efficiency can be obtained at these low wave-lengths by using valves in which the electrons are deflected during their passage through the axial field of a magnet which may be either an electromagnet or a permanent magnet. These valves are called m a g n e t r o n valves.

Two magnetron valves of this type are shown in *fig. 8*, viz., the TAMM with two anodes and TAM 1.5/100 with four anodes which are connected crosswise in pairs. At a frequency of $600 \cdot 10^6$ c/s the latter valve has, for instance, an output of 50 to 60 W, with an efficiency of approximately 40 per cent. The first-named valve can still be run effectively at a frequency of $1200 \cdot 10^6$ c/s, having then an output of 7 W and a 10 per cent efficiency. The magnetic field intensity in this case is approximately 1000 oersted. These special transmitting valves are being steadily developed towards higher outputs and higher frequencies.

Water-Cooled Transmitting Valves for Outputs of Over 100 Kilowatts

Philips have manufactured high-power water-cooled transmitting valves for a number of radio transmitters. The largest type, TA 20/250, shown in *fig. 9*, has a useful output of 250 kW on long waves; including the cooler this valve is 1.40 m. high. The filament is made in 12 parts each about $^1/_2$ m long. Owing to the use of a central rod which can slide through two insulators at the top of the valve, a perfectly rigid and self-supporting assembly is obtained. The filament consumes 15 kW (425 A at 35 V). Special measures have to be adopted to lead this powerful current through vacuum-tight seals in the glass, and include water-cooling of the filament supports. The 100-kW valve, the TA 18/100 000 type, operates with a

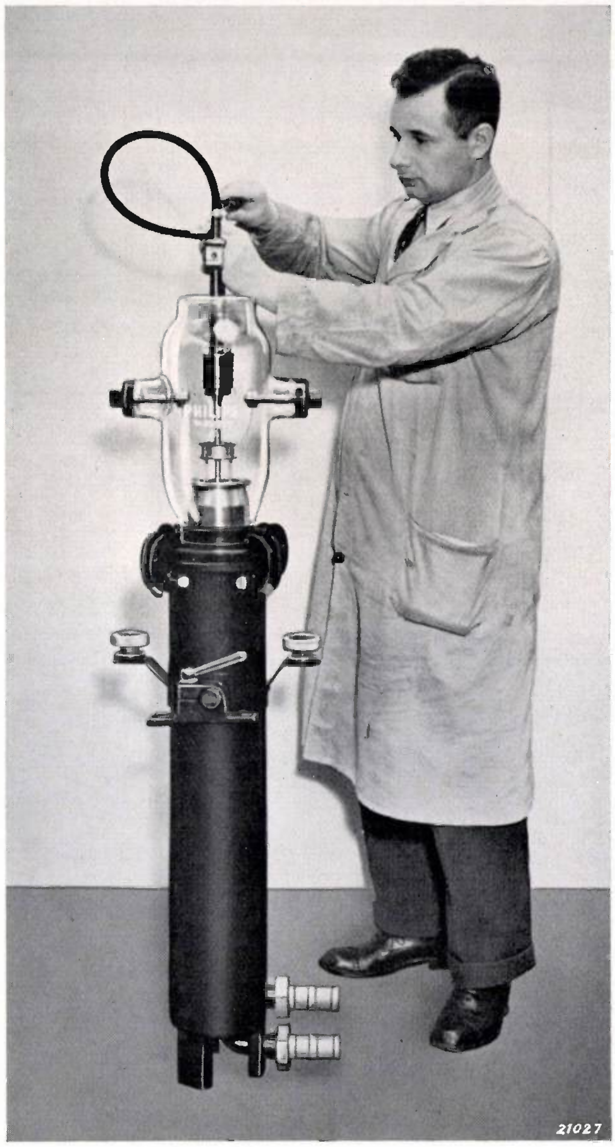


Fig. 9. The largest water-cooled transmitting valve TA 20/250 for an output of 250 kW.

filament current of 207 A. This large current can be passed into the valve without artificial cooling, either by water or air, owing to the provision of a massive chrome-iron cylinder fused to the glass.

To obtain maximum economy in running these valves the anode voltage has been increased to the greatest value consistent with proper functioning of the valve. Thus for the TA 20/250 valve shown in *fig. 9* a maximum anode voltage of 20 000 V is used. To obtain the same output at 10 000 V, the anode current, and hence the filament input, would have to be doubled, and 30 kW filament power would be required in place of the 15 kW actually used. With 5 000 working hours per *annum*, which is normal for a radio transmitter, this would amount to an extra annual consumption of 75 000 kWh per valve.

A special water-circulating system serves for cooling the anode, as shown in *fig. 10*, and ensures that the water flows over the whole of the anode at a uniform rate of 2 m per sec (i.e. a flow of 120 litres per min.). By using this method of cooling, the anode can dissipate 100 W per sq.cm.

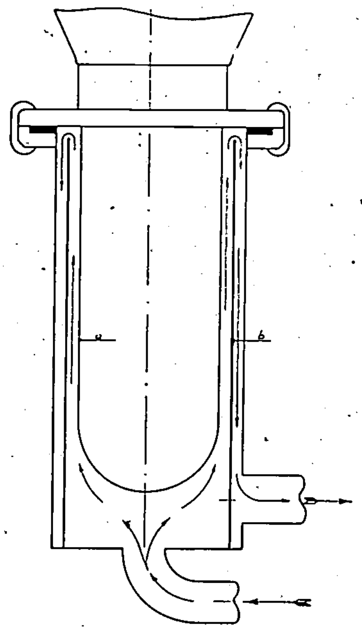


Fig. 10. Flow of cooling water for large transmitting valves. No regions of stagnant water at the anode surface are permissible.

PHYSICAL PRINCIPLES OF GASFILLED HOT-CATHODE RECTIFIERS

by M. J. DRUYVESTEYN and J. G. W. MULDER.

Summary. Some of the principal phenomena occurring in the operation of gas-filled rectifying valves are discussed on the basis of a general analysis of the voltage-current characteristic of a gas discharge. The construction of a number of rectifying valves for different current and voltage ratings is described.

Introduction

While solids and liquids can be classified according to their electrical behaviour into conductors and non-conductors (insulators), gases can behave as either conductors or insulators according to their physical condition. As a rule a gas is an excellent insulator, but if it contains sufficient ions it will become a good electrical conductor. In gas-filled rectifying valves the gas is rendered conducting when the p.d. applied is one particular direction while it acts as an insulator when the p.d. is reversed. This action is obtained in gas-filled rectifying valves by making the electrode acting as cathode during the passage of current in the form of a hot-cathode or by using a mercury cathode. In the present article we shall confine ourselves to hot-cathode rectifiers, and to the physical processes occurring in these valves, omitting for the moment any consideration of the circuits themselves.

Characteristic of a Gas Discharge

To gain an insight into the action of a gas-filled rectifying valve, let us examine in the first place what takes place in a simple gas-discharge tube fitted with two metal plates at a distance of, say,

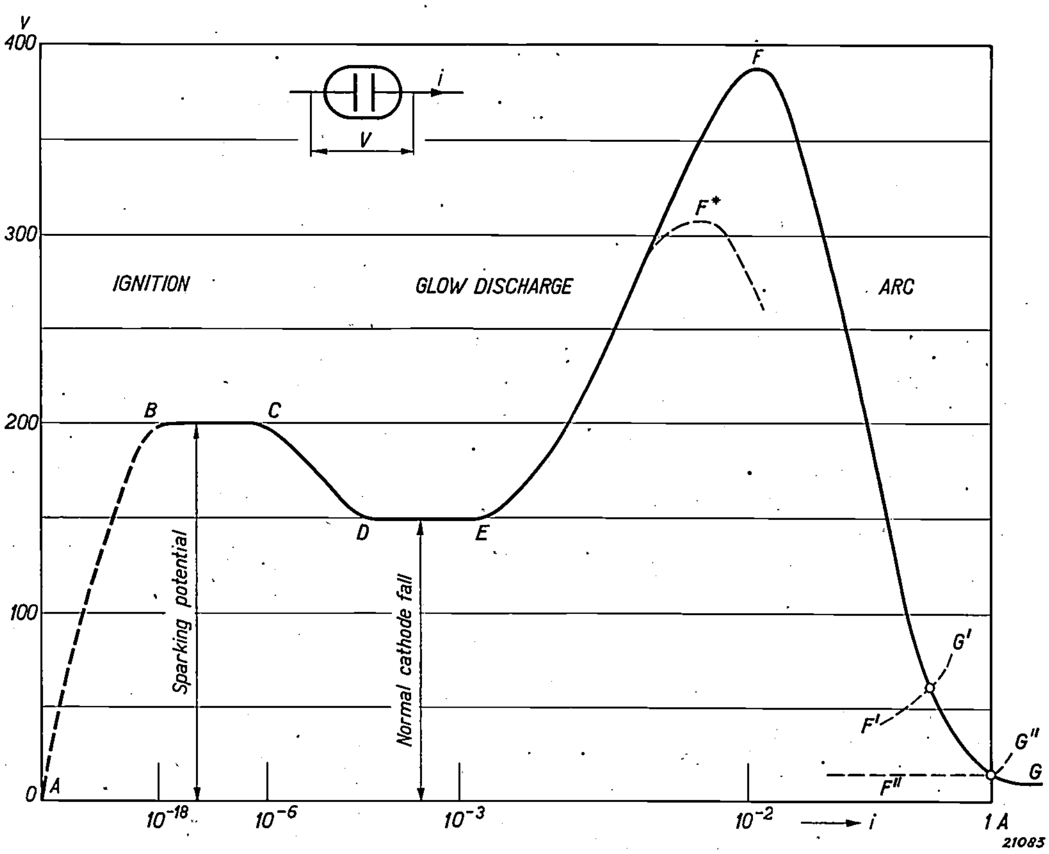


Fig. 1. Diagrammatic voltage-current characteristic of a discharge between parallel plates. Individual current values are indicated on an arbitrary scale to show the order of magnitude of the currents at which different discharge phenomena occur.

1 cm from each other and containing a rare gas atmosphere at a few mm pressure. The characteristic of a valve of this type is shown diagrammatically in *fig. 1*, the curve representing the connection between the impressed voltage and the current flowing through the valve. It is seen from this curve that a gas can act both as a conductor and as an insulator; thus at 20 V a current can flow of approximately 10^{-20} A (along the broken line *AB*) or one of several amperes (at *G*). At a potential difference of 180 V there are in fact four current values. This multiplicity of current values corresponding to a single voltage is a general feature of all gas discharges.

The particular curve illustrated applies not merely to a discharge between parallel plates but is also representative of electrodes having different shapes and disposed in various ways. Nevertheless the characteristic is subject to marked alterations when the electrodes are not contained in the same bulb but are placed, for instance, in two tubes connected by a long tube, as well as when the gas pressure is made abnormally high or low.

Analysing the nature of the different portions making up the characteristic it is found that the very weak current passing in the insulating portion *AB* has its origin in the field attracting towards the electrodes the ions which are present in every gas. At *B* the current increases considerably and so-called ignition takes place. At the striking voltage the field is powerful enough to impart to sufficient electrons a velocity adequate to ionise the gas atoms; the positively-charged ions pass to the cathode (negative plate) and liberate electrons from it. The current increases considerably (*BC*), and as a result space charges are generated which destroy the homogenity of the field; a glow discharge is then obtained (*DEF*). The field is now concentrated in a thin layer of gas at the cathode, where the cathode fall occurs. The potential gradient between the boundary of the cathode fall and the anode (positive plate) is comparatively small. The normal cathode fall [1]) is lower than the striking voltage, this being possible because with the field distribution in question here the electrons abstract more energy from the field for ionisation purposes than during ignition. In the portion *EF* of the curve the cathode fall again increases with increasing current; nevertheless the voltage *V* impressed on the valve and the cathode fall remain

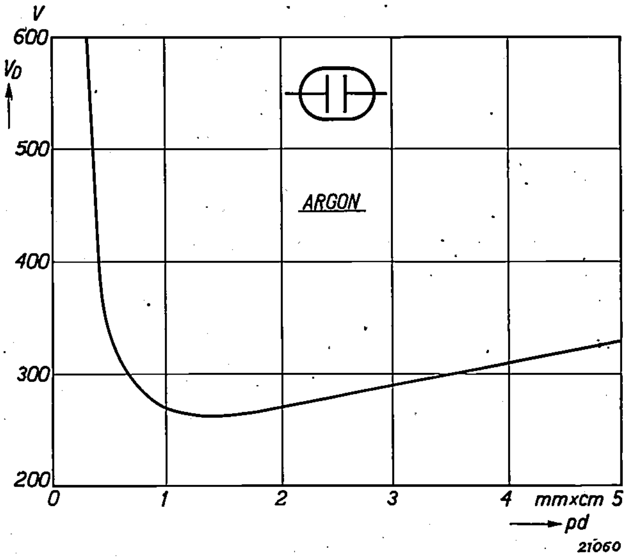


Fig. 2. Ignition voltage V_D for argon across parallel plates as a function of the product of the gas pressure and the distance between the electrodes (Paschen's curve).

roughly equal to one another (this applies from *D* to *G*).

Up to this point the electrons are liberated from the cathode by the impact of positive ions, which strike the cathode with a high velocity and heat the cathode more with increasing current. At F a new effect is observed: the temperature of the cathode has been raised so high (with tungsten to approximately 1750 deg. C.) that thermal emission of electrons now begins. Since the electrons are now liberated not merely by positive ions striking against the cathode, the cathode fall can diminish and an arc discharge is obtained at the hot cathode. The arc voltage may become very small, smaller even than the ionisation potential of the gas, and large currents can now flow through the gas. In the portion *FG* (sharp drop in potential with increase in current intensity) the temperature of the cathode rises. The maximum value of the current at a constant cathode temperature is determined principally by the thermal emission of electrons.

If, on the other hand, the current is not altered slowly as assumed above, so as to give time for an equilibrium to be reached, but the variation in *V* is determined for a rapid change in *i*, the cathode temperature being constant, since a certain time is required before this temperature becomes stable, then characteristics will be obtained similar to the dash lines *F′G′* and *F″G″*, the temperature for *F′G′* being lower than that for *F″G″*.

The shape of the characteristic in fig. 1 depends on a number of factors, such as the nature of the gas, the material of the cathodes and the gas pressure. The striking voltage is determined by

[1]) With a normal cathode fall (*DE*) current flows to only part of the cathode while with an abnormal cathode fall (*EF*) the current flows to the whole of the cathode.

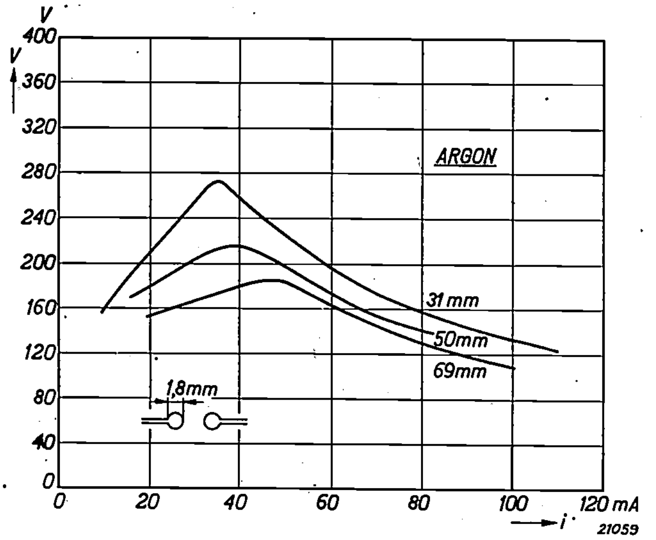


Fig. 3. Transition from a glow-discharge to an arc discharge between tungsten spheres.

the product of the gas pressure p and the distance between the plates d. The relationship for an iron cathode in argon is shown in *fig. 2*.

In *fig. 3* the valve voltage is plotted against the current in the portion EFG for different gas pressures, using as electrodes two small tungsten spheres, 1.8 mm in diameter. This shape of electrode was used because of the difficulty of reproducing measurements during the transition from the glow discharge to the arc discharge when using flat electrodes. It is seen from the curves that the slope of the curve portion EF is the greater and the maximum voltage higher the lower the gas pressure.

Hot-Cathode Rectifying Valves

In gasfilled rectifying valves, an arc discharge has to be obtained when the p.d. is applied in one direction, while the gas must act as an insulator or at most lead to a glow discharge with the p.d. in the opposite direction. When current flows in the latter direction a discharge is therefore always obtained on one of the two positive-slope sections (AB or EF in fig. 1). To facilitate the formation of an arc discharge in the direction of current-flow, a hot cathode, usually an oxide-coated cathode, is used in the rectifying valves under discussion here. By means of a heating current, applied either directly or indirectly, this cathode is raised to a temperature at which it can emit electrons. The arc will thus have a characteristic roughly of the type of $F''G''$ in fig. 1.

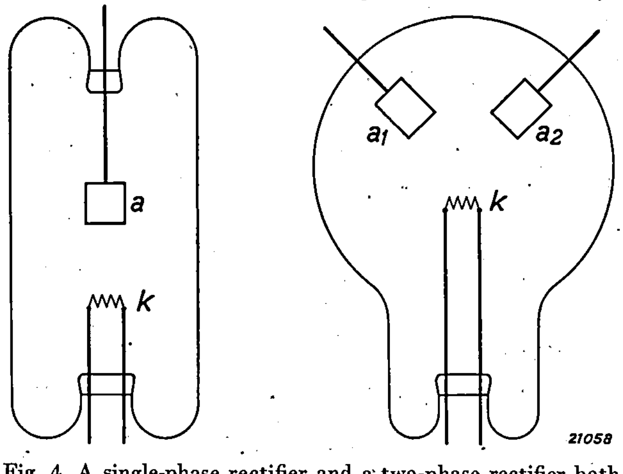


Fig. 4. A single-phase rectifier and a two-phase rectifier both with hot-cathodes.

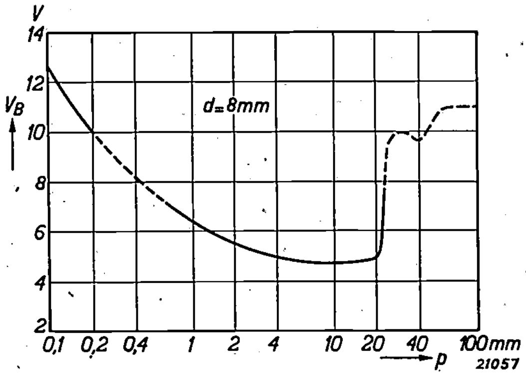
A single-phase rectifying valve with the hot cathode k and the anode a is shown on the left in *fig. 4*. We shall continue to refer to the "anode" a even when it is negative with respect to k and is hence able to act as the cathode for a glow discharge. On the right of fig. 4 is shown a two-phase rectifier with two anodes, a_1 and a_2. In the normal circuits adopted for these valves, at the moment an arc is obtained between k and a_1, a_2 becomes negative with respect to k, and *vice versa*.

The life of a valve is usually limited by the disintegration of the cathode, for during the arc discharge ($k - a$) the positive gas ions strike the hot cathode with such a velocity that they frequently cause the detachment of atoms. To minimise the disintegration of the oxide cathode the velocity of the ions on impact against the cathode must not be too great. Since at a pressure of less than a few millimetres the positive ions on passing through the cathode fall hardly ever collide with gas atoms and therefore store the total acceleration acquired, the cathode fall must be low to ensure a long life for the valve.

On the basis of fig. 1 we shall discuss the principal phenomena occurring during the arc discharge with a hot cathode (k—, a+) as well as the measures taken to prevent an arc discharge in the opposite direction (i.e. k+, a—); for should such an arc be obtained the valve will no longer function as a rectifier.

Arc Discharge

The variation in potential in the arc discharge of a hot-cathode valve of the type shown in fig. 4 is usually of the following type: Close to the cathode the potential increases sharply over a distance of less than 0.1 mm, so that with an arc discharge there is also a cathode fall (of approximately 15 V). Further away the electrical field intensity is much smaller and the drop in potential is only a few volts; in this region a potential maximum is usually obtained. To reach the anode the electrons must overcome an electrical retarding field, which is possible by



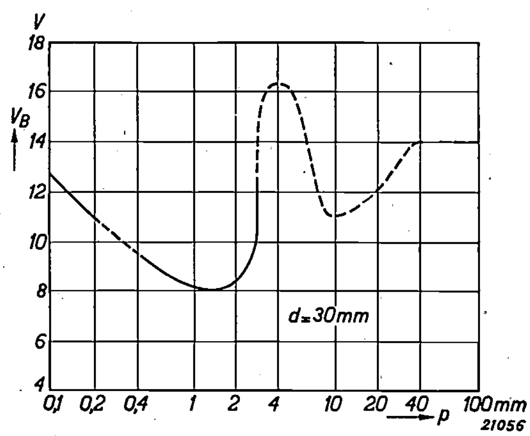


Fig. 5. Arc voltage V_B as a function of the argon pressure at 0.5 A and an electrode gap of a) 8 mm, b) 30 mm.

diffusion. The highly ionised part of the discharge where the field intensity is low is frequently referred to as the plasma. Sometimes the electrical field at the anode again increases in intensity and accelerates the electrons, when an anode fall is produced.

This anode fall, whose value often oscillates, is only obtained when the number of collisions between electrons and gas atoms exceeds a certain critical value. The frequency of these collisions is determined principally by the product of the gas pressure and the distance between the electrodes.

The arc voltage V_B is equal to the sum of the cathode fall, the potential gradient in the plasma and the anode fall. *Fig. 5a* and *b* shows two examples of the arc voltage in rectifying valves with oxide cathodes plotted as a function of the gas pressure (argon). Over the broken portions of the curves low-frequency oscillations in the arc voltage are obtained [2]). It is seen that a minimum value of the arc voltage obtains between pressures of 3 and 20 mm (and between 1 and 2.5 mm). The steep rise (in fig. 5a at 25 mm, and in fig. 5b at 3 mm) is probably due to an anode fall, which occurs at pressures above this value (see *Table I*).

Table I

Distance between cathode and anode	Pressure of argon above which anode fall occurs
8 mm	25 mm
20 mm	8 mm
30 mm	3 mm

The increase in the voltage with diminishing pressure below 1 mm is due to the cathode fall, which increases as the gas pressure is reduced and remains nearly constant at pressures above a few mm. The lowest value of the cathode fall is generally lower than the ionisation potential, which for the rare gases used here lies between 10 and 25 volts.

The increase in the cathode fall with diminishing gas pressure at constant current intensity is the smaller, the greater the dimensions of the cathode used.

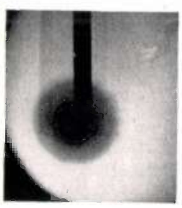
We shall endeavour to explain the variation in the cathode fall as a function of the gas pressure. It is assumed here that the cathode fall is determined by the fact that n electrons emitted from the cathode must ionise so many gas atoms that the formed positive ions which arrive at the cathode compensate the space charge of the electrons to such a degree that n electrons are again liberated from the cathode. Up to approximately 100 V the probability of ionisation per collision increases with the velocity of the electrons. It thus follows that at a lower gas pressure, in other words for less collisions, the requisite number of ionisations can only take place at a higher electronic velocity, i.e. with a higher cathode fall. The measured increase in the cathode fall with diminishing pressure is however greater than would follow from the above considerations. A second factor must therefore also operate. The electrons liberated from the cathode are accelerated in the cathode fall, such that a stream of electrons with a high velocity passes into the plasma. This electron stream loses energy both by collision with the electrons and the gas atoms; during the latter ionisation may result, and by the transfer of energy to the slower electrons. While the first means of losing energy is closely dependent on the gas pressure, the second is determined mainly by the concentration of the slower electrons, and will hence depend on the current intensity. Since at low pressures the electrons must traverse a considerable distance before striking

[2]) While the oscillations of V_B at high pressure shown in figs. 5 and 6 are probably due to the anode fall, this does not apply to the variations occurring at a pressure of 0.3 mm.

a gas atom, it is probable that at high current intensities they will already have parted with their energy to the slower electrons.

It is conceivable that when using a larger cathode the cathode fall will be smaller at a lower pressure, since in this case the current intensity, and hence also the concentration of the electrons in the neighbourhood of the cathode, will be lower, resulting in a reduction in energy transfer from the fast to the slow electrons. Since the cathode fall drops with rising gas pressure the life of the oxide cathode, which is determined by cathodic disintegration, will be increased. Yet if the gas pressure exceeds a few centimeters, the cathode life will again decrease for various reasons.

The interaction between the fast electrons in the electronic stream and the slower electrons in the plasma is clearly shown by a dark layer, the so called dispersion layer[3]), which was recently discovered and which occurs in an arc at a hot cathode. A directional stream of fast electrons issues from the cathode fall; this stream not only imparts energy to the slow electrons but is also dispersed by the latter, such that fast electrons are deflected in all directions and convert a large number of gas atoms into a radiating state. As a result it may occur that close to the cathode less light is radiated than further away. A photograph of this layer obtained at a pressure of 0.05 mm of argon and 2 A current is reproduced in *fig. 6*. The thickness of the layer diminishes with increasing current intensity, since at higher current intensities the concentration of the electrons is greater and dispersion can then more readily occur.



21077

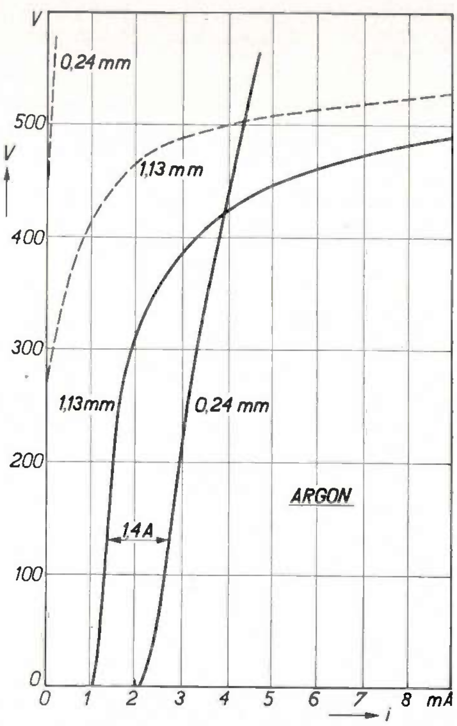
Fig. 6. The dark layer of the cathode fall (0.1 mm thick) and adjoining it the dark "dispersion layer" (2 mm thick) can be seen surrounding the indirectly-heated cathode, which is shown in the photograph as a black spot with bright edge (diameter 4 mm); outside these is the light of the plasma. The edge of the bulb may be seen in the left-hand bottom corner.

Back-Firing

Sudden back-firing may take place in the direction in which the valve should function as an insulator and thus destroy the rectifying action. The glow discharge produced will then become suddenly converted to an arc. Immediately back-firing occurs rectification ceases, so that precautions have to be taken to prevent backfiring or that it occurs at most only very rarely during the normal life of the valve. From fig. 1 it may be concluded that backfiring will not occur as long as the opposing voltage remains below a specified critical value (F, fig. 1).

This conclusion is however not quite valid, since fig. 1 relates to normal conditions of operation and it is specifically stipulated that the cathode is at the same temperature throughout and that this temperature must exceed a certain limiting value to permit an arc discharge to be obtained. Although this corresponds to normal conditions, the temperature distribution on the anode may become non-uniform owing to fortuitous circumstances, when an arc discharge (i.e. back-firing) will occur at the hottest part of the negative anode. Thus already at lower voltages an arc discharge may be obtained (F^* in place of F in fig. 1) than when the temperature distribution is uniform.

In practice back-firing may be produced by a variety of causes[4]). We shall discuss here only such backfiring as is caused by the partial heating



21055

Fig. 7. Connection between current i and voltage V of an anode of a two-phase rectifier, at pressures of 0.24 and 1.13 mm, the current flowing to the other anode being either 1.4 A (continuous line) or nil (dash line).

of the anode just referred to. The danger of back-firing may be minimised by making the anode of a material with a low thermal emission, such as graphite: Although the heating effect produced is greater in the direction of current flow, the current

[3]) M. J. Druyvesteyn and N. Warmoltz, Physica, 4, 51, 1937.
N. Warmoltz, Nature **138**, 26, 1936.

[4]) Another cause for back-firing should also be mentioned: As soon as a particle, preferably conducting, strikes against the cathode of a glow discharge, there is the possibility that the glow discharge will be transformed to an arc, the point of impact of the metal particle on the cathode acting as the cathode spot.

and voltage in the cut-off direction have also an important bearing on back-firing, since at a given anode temperature the liability of back-firing increases very rapidly as the current and voltage are raised. At lower pressures the danger of back firing is less, for according to fig. 3 the glow discharge current falls, while the maximum voltage increases as the pressure drops. In polyphase rectifying valves the current flowing to a negative anode is greater than the glow-discharge current, because the positive ions of the arc discharge between the cathode and one of the other anodes diffuse to the negative anode. *Fig. 7* shows these conditions for a two-phase valve of the type shown in fig. 4. The current of the negative anode is here plotted as a function of the p.d. at this anode, a current of 1.4 A flowing to the other anode.

Construction of Rectifying Valves [5])

In service, rectifying valves have to meet a variety of requirements, such as a long life and high efficiency (in other words a low arc voltage), as well as freedom from back-firing at the normal currents and voltages used. The various points discussed above have resulted in a variety of designs each conforming to specific practical requirements.

Rectifiers for different current ratings and the same voltage differ mainly in the dimensions of the hot cathode and of the anode and in the capacity of the bulb, particularly the latter since its surface area determines the temperature. The surface of

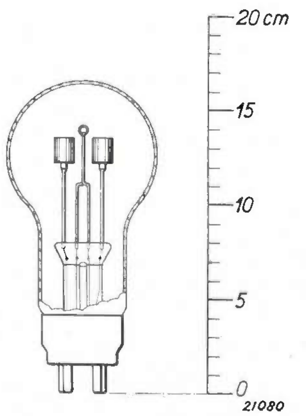


Fig. 8. Two-phase rectifier (type 367 for 6 A and 24 V).

the bulb must be the greater the greater the amount of energy liberated in the valve, i.e. the higher the current intensity.

Rectifiers for different voltage ratings and the same current differ mainly in the nature of the gas filling (gas mixture and pressure), as well as in the means adopted for the mutual screening of the electrodes against the action of discharges at undesirable points or at undesirable moments. In view of this the discharge gap is usually made much longer for high voltages, although at exceptionally low pressures (when as in the case shown in fig. 2 the valve is operated to the left of the minimum) the gap must be made very short.

One of the simplest types of two-phase valves, rated for 6 A direct current and 24 V direct voltage, is shown in *fig. 8*. The cathode is a tungsten coil wound with a second wire of thinner gauge which serves to carry the barium oxide. The wedge-shaped slots provide good adherence for the oxide. In rectifiers for higher currents the second wire is also coiled in order to increase the effective surface (e.g. the "Triarlita" cathode shown in *fig. 9*).

The two anodes consist of graphite blocks across which a maximum potential difference of 85 V is obtained, the gas filling being of argon. At the gas pressure chosen, there is no likelihood of a glow discharge between the anodes at this particular

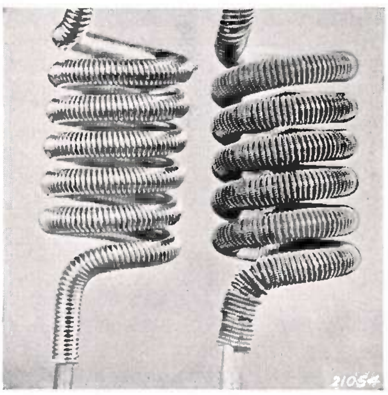


Fig. 9. Single coiled cathode (left) and "Triarlita" coiled-coil cathode (right). The windings have been removed at the bottom to show the cores.

voltage rating. The discharge gap is short and wide so that the above-mentioned theoretical considerations regarding the arc discharge are applicable here. Rectifiers of this type are made for various current ratings up to 60 amps.

For higher voltage ratings certain modifications must be made in the general design of the rectifiers described above. The gas pressure is usually reduced, and for very high voltages is made very low, down to about 10^{-3} mm of mercury, while the leads, especially those to the anode, have to be insulated to prevent undesirable discharges (see *fig. 10*).

[5]) For detailed data, cf. J. G. W. Mulder, Diss. Delft, 1934.

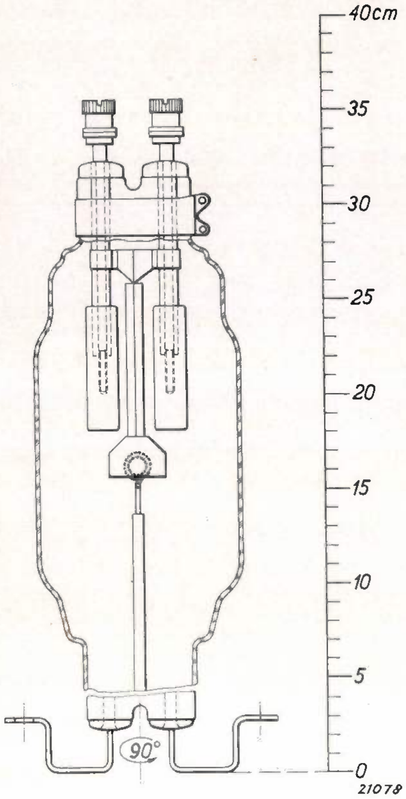


Fig. 10. Two-phase rectifier (type 1069 k for 60 A and 60 V).

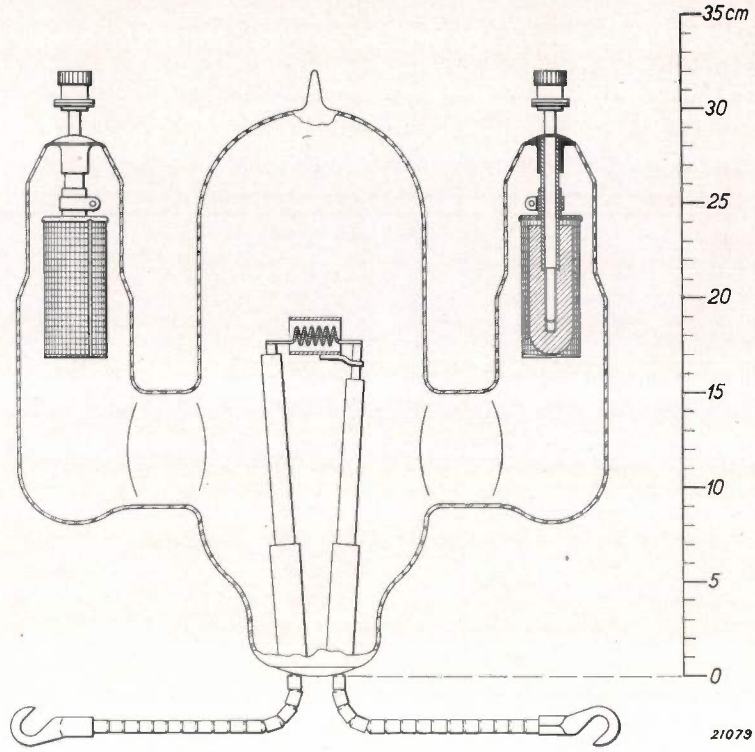


Fig. 11. Two-phase rectifier (type 1554 for 40 A and 250 V).

In all rectifiers for voltages above 40 volts the cathode is located in a space which is partly separated from the remainder of the discharge bulb. This may be done, for instance, by fixing a short tube or a small metal cap round the cathode (fig. 10). In this type of rectifier for current ratings of 60 amps and voltages up to 60 volts (Philips welding rectifier) the cap is open at the bottom, while for higher voltages the cathode is totally enclosed by a coaxial cylinder (up to ratings of about 110 volts). In both cases the effect produced is twofold: In the first place, particles which become detached from the cathode surface are prevented from reaching one of the anodes, these being liable to initiate back-firing, and secondly the current flow diverted from one anode to another by diffusion from the arc discharge is reduced (see fig. 7).

At still higher voltages the design is taken a step further and both anodes and cathodes are accommodated in separate chambers. *Figs. 11* and *12* illustrate examples of these rectifiers. The cathode in these rectifiers is located in the central chamber, and the anodes in fig. 11 are accomodated in two side tubes connected to the central chamber by narrow arms [6]).

In the rectifier shown in fig. 12 the cathode is situated in the top bulb which is connected to the main bulb by a short metal tube. This rectifier has already been illustrated and described in this Review (volume I, p. 163). The greater the divergence in the design of the bulbs from the simple type shown in fig. 4, the more will the discharge phenomena occurring differ from the conditions analysed for this simple type.

Fig. 12. Photograph of the single-phase rectifier DCG 5/30 II for 6 A and 6 kV.

[6]) In the rectifier shown in fig. 11 a part of the anode surface is screened against a glow discharge in the negative phase. Since this discharge (in the portion *EF*, fig. 1), contrary to the arc discharge in the direction of current flow, tends to spread over the whole surface, the glow-discharge current can be reduced by means of an adequately-insulated grid which partially surrounds the anode at a short distance; this arrangement does not affect the arc discharge and barely reduces the radiation of heat from the anode through the grid.

VOL. 2 No. 5 MAY 1937

Philips Technical Review

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

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

WELDING AND WELDING RODS

by J. SACK.

Summary. After a short description of the various methods of electric arc welding, the significance of the coating of the welding rod is discussed in relation to the quality of the weld.

Introduction

In the course of time various electrical processes have been developed to join metals, such as arc welding, resistance welding, spot and seam welding. In this article we shall discuss only arc welding, a method whereby the heat of an electric arc is used to join metal parts by fusion. Although in the following special attention has been paid to the welding of iron, most of the discussion applies equally well to the welding of other metals.

Typical examples of welded joints are the butt weld and the fillet weld, as represented in crosssection in *figs. 1, 2* and *3*.

Towards the end of the last century and in the present one various methods of electric arc welding were developed. The most important and most interesting of these are shown schematically in *figs. 4, 5, 6, 7* and *8*.

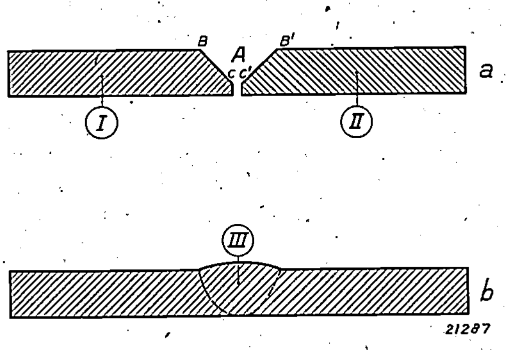


Fig. 1. Butt weld (so-called V-weld).
The sheets I and II (*a*), which are to be welded together, lie in the same plane. The edges of the sheets have been bevelled (*BC* and *B'C'*) so that a V-shaped groove is formed which is filled with molten iron. In the process the bevelled edges melt also, and the result is shown in *b*. In *b* the region III within the dotted line represents the metal which has been in a liquid state during the welding.

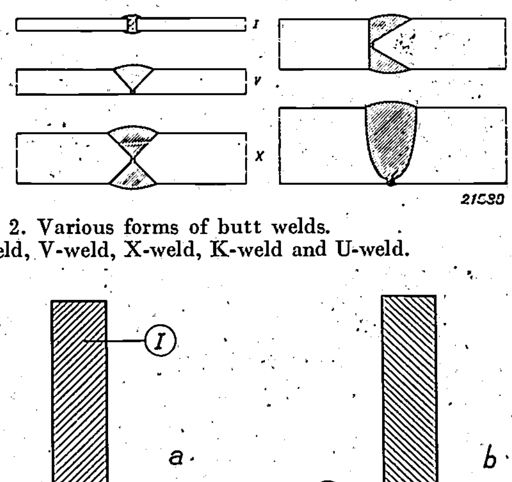


Fig. 2. Various forms of butt welds.
I-weld, V-weld, X-weld, K-weld and U-weld.

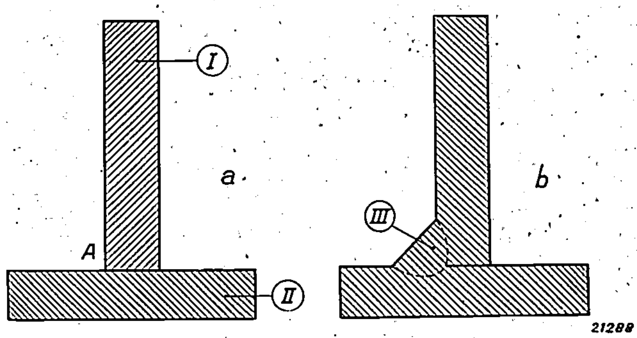


Fig. 3. Fillet weld.
In the fillet weld the parts to be joined, I and II, are perpendicular to each other. In the angle *A* (*a*) iron is melted down so that the joint *b* is made. Here again III represents the portion which was fluid during the welding.

The methods of Benardos (1885), Zerener (1889) and Langmuir (1926) have a certain resemblance to oxy-acetylene welding, particularly the last two; the electric arc has been substituted for the oxy-acetylene flame. Just as in gas welding, the filler rod may sometimes be omitted in electric welding. As an example, in *fig. 9* the welding of the bottom of a vessel by means of a carbon arc (B e n a r d o s method without filler rod) is represented. The edges of the bottom and sides are fused together by means of the heat of the arc.

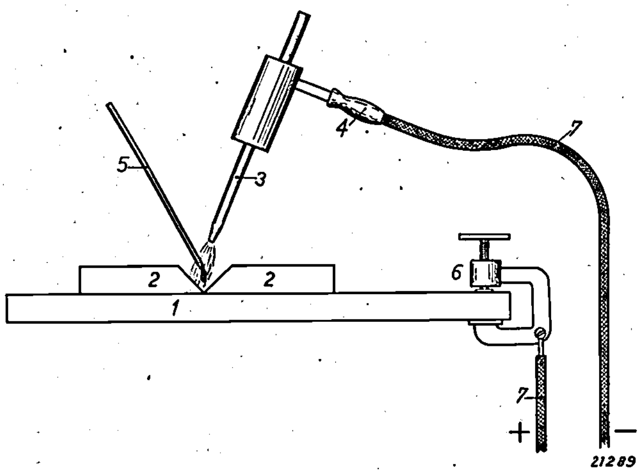


Fig. 4. Welding according to the method of B e n a r d o s (1885).

1: welding bench. *5*: filler rod.
2: piece of work. *6*: flexible clamp.
3: carbon electrode. *7*: flexible cables.
4: electrode holder.

Between the work (*2*) and the carbon flexible (*3*) an electric arc is maintained. In the arc a filler rod (*5*) can be brought to the melting point. The molten material fills the V-shaped groove, whose edges are also melted by the arc.

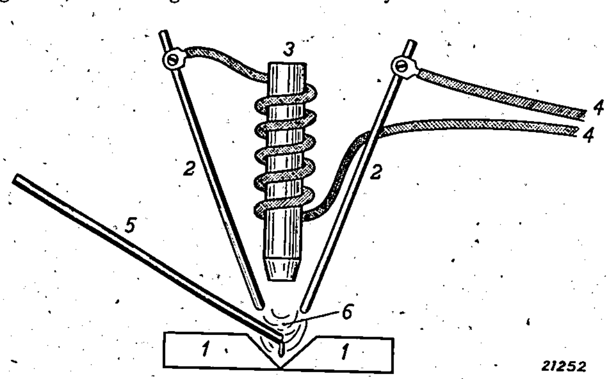


Fig. 5. Welding according to the method of Z e r e n e r (1889).

1: piece of work. *4*: flexible cables.
2: carbon electrodes. *5*: filler rod.
3: blowing magnet. *6*: welding arc.

Z e r e n e r allowed an arc to burn between two carbon electrodes (*2*). With the aid of an electromagnet (*3*) the arc was blown toward the work. An iron filler rod (*4*) introduced into the arc melts away and fills the groove between the parts of the piece of work (*1*).

In the methods of S l a v i a n o f f (1892) and A l e x a n d e r (1926) the separate electrode rod is omitted, and its place is taken by the filler rod itself, the welding rod.

[1]) P. A l e x a n d e r, Arc Welding in Hydrogen and other Gases, Gen. Electr. Rev. **29**, 169, 1926.

[2]) R. A. W e i n m a n n and I. L a n g m u i r, Atomic Hydrogen Arc Welding, Gen. Electr. Rev. **29**, 160, 1926.

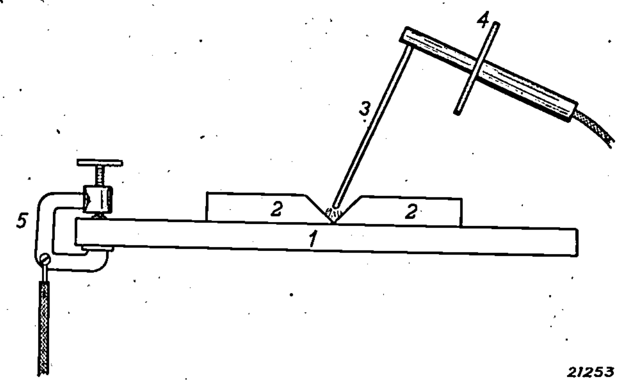


Fig. 6. Welding according to the method of S l a v i a n o f f (1892).

1: welding bench. *4*: welding rod holder.
2: piece of work. *5*: clamp.
3: welding rod.

The arc burns between the piece of work (*2*) and an iron electrode (*3*), the so-called welding rod. The arc causes the welding rod to melt as well as the edges of the groove. The welding rod thus performs a double function, it is electrode as well as filler rod.

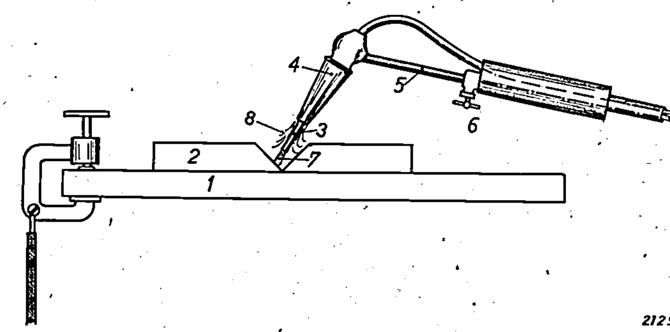


Fig. 7. Welding according to the method of A l e x a n d e r (1926) (semi-automatic model)

1: welding bench. *6*: tap for regulating the supply of gas.
2: piece of work. *7*: electric arc.
3: welding wire, automatically fed. *8*: atmosphere of protecting gas.
4: nozzle.
5: tube for the supply of protecting gas.

About the wire (*3*), which is preferably automatically fed, a protecting gas (for example, hydrogen) is blown by means of the nozzle (*4*) in order to prevent the bare welding rod from coming into contact with the oxygen and nitrogen of the air during melting. Otherwise this method is the same as that of S l a v i a n o f f.

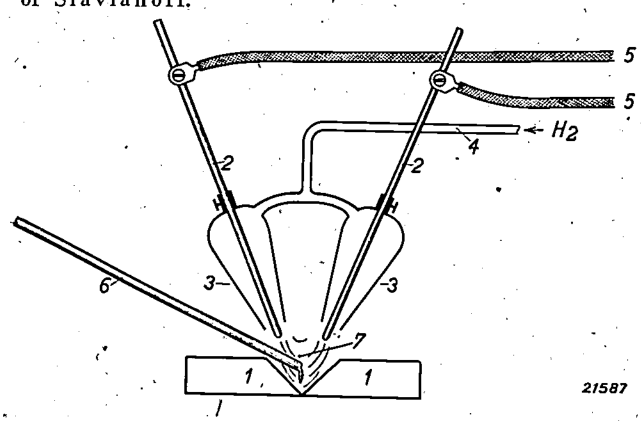


Fig. 8. Welding according to the method of L a n g m u i r [2]) (1926).

1: piece of work. *4*: tube for the supply of hydrogen gas.
2: tungsten electrodes. *5*: flexible cables.
3: nozzles for the supply of hydrogen gas. *6*: filler rod.
7: atomic hydrogen arc.

We shall not enter into the various advantages and disadvantages and the possibilities of application of the methods mentioned. It may however be noted that the method of Slavianoff is clearly the simplest, and is the one practically always used at present for manual welding.

Electric welding was developed extremely slowly at first, a fact which must be ascribed to the unsuitable electrical apparatus available, the slightness of metallurgic knowledge and probably also to the competition of oxy-acetylene welding. Only a few factories used electric welding, and even then exclusively for repair work. So-called bare wire, i.e. ordinary iron wire, was used for welding rods. This wire could only be welded with direct current. The generators available before 1914 were expensive and were not constructed especially for welding. A considerable advance was made by the use of coated welding rods which were introduced by the Swede Kjellberg.

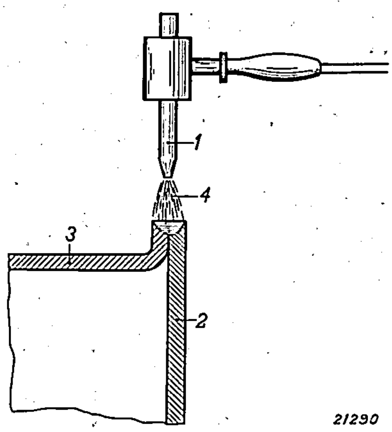


Fig. 9. Carbon arc welding.
1: carbon electrode. *3*: side of vessel.
2: bottom of vessel. *4*: welding arc.

After 1918 a rapid development not only of welding machinery [3]) but also of welding rods began. Electric welding developed from a repair method into a method of construction. Much which was previously done by oxy-acetylene welding, riveting or casting, is now done with success and advantage by electric welding. This development is still going on, and, more than with any other construction method, the application of electric welding is accompanied by intensive technical and scientific research. The following will give some information about the significance of the coating of the welding rod in relation to the quality of the weld. We shall begin by discussing the welding arc.

[3]) H. A. W. Klinkhamer, Philips techn. Rev. 1, 338, 1936.

The welding arc

Accurate determinations of the temperature of an arc have recently been carried out. The most important and most accurate method of measurung arc temperatures was worked out by Ornstein and his co-workers at the University of Utrecht [4]). The spectrum of the arc was observed and it appeared that not only the rotation spectrum but also the vibration spectrum of the molecules in the column of the arc provide the observer with sufficient data for the calculation of the temperature.

Suits [5]) follows another method. The velocity of sound in a gas is proportional to the square root of the absolute temperature of the gas. Thus it is possible to calculate the temperature of the arc from the velocity of propagation of a sound wave in the arc. The sound is manifested in the arc by an increased radiation at the place where the pressure wave is passing through (see *fig. 10*). This can be recorded on a rapidly moving film.

The temperatures found by investigators for the electric arc are surprisingly high. Suits found in the case of a welding arc between a coated welding rod with a diameter of 3/16″ as cathode and an

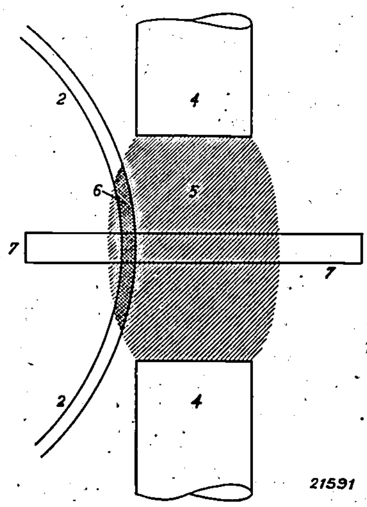


Fig. 10. Diagram showing Suits' method of measuring the temperature of an arc.
A spark discharge at (*1*) makes a report, i.e. a pressure wave (*2*) about one mm. thick. The movement of the pressure wave through the arc (*5*) between the electrodes (*4*) is visible because of an increased radiation (*6*). By means of a diaphragm only the portion (*7*) of the phenomenon is recorded on a rapidly moving film. From the measured velocity of propagation of the report through the arc the temperature of the arc may be calculated.

[4]) L. S. Ornstein and W. R. van Wijk, Temperaturbestimmung im elektrischen Bogen aus dem Bandenspektrum, Proc. kon. Akad. Wet. A'dam, 33, 44, 1930. L. S. Ornstein and H. Brinkman, Temperature Determination from Band Spectra, Proc. kon. Akad. Wet. A'dam, 34, 33, 498, 1931.

[5]) C. G. Suits, A Study of Arc Temperatures by an Optical Method, Physics, 6, 315, 1935.

iron plate as anode, with a current of 125 A and an arc-length of 1 cm, an arc temperature fluctuating between 4400 °K and 7000 °K with an average value of 6020 °K. For a welding arc between a tungsten cathode with a diameter of 3,2 mm and an iron plate, current of 80 A, arc length 0.5 cm, Suits found an average temperature of 6150 °K.

The following conclusions may be added. Oxygen and nitrogen molecules may break up into atoms (dissociation), and the dissociation probability increases with increase of temperature. At room temperature the equilibria:

$$O_2 \rightleftarrows 2\,O \quad \text{and} \quad N_2 \rightleftarrows 2\,N$$

are toward the left.

At high temperatures definite fractions of the oxygen and nitrogen molecules are dissociated. The gas mixture then consists of O_2 and N_2 molecules and O and N atoms. The partial pressure of the atomic oxygen and nitrogen as a function of the temperature is known.

At a temperature of 6000 °K 80 per cent of the atmosphere is atomic oxygen and nitrogen.

Bare welding rods

From the above it follows that with bare welding rods one actually welds in an atmosphere of "atomic air", which contains much atomic nitrogen. The presence of atomic nitrogen in the arc has been confirmed spectrographically by Séférian [6]). He showed further that nitrogen in the atomic state is readily taken up by the iron with the formation of iron nitride Fe_4N. It is therefore not surprising that the material melted down in welding with bare welding rods is very rich in nitrogen, even increasing from 0.15 to 0.16 per cent of N with a short arc (2 mm) to 0.19 to 0.20 per cent, of N with a long arc (6 to 8 mm; a longer arc presents more opportunity for the taking up of nitrogen [7]). Portevin and Séférian even make use of this material melted down from uncoated welding rods for the setting up of an improved phase diagram for the system iron-nitrogen. According to this diagram the iron can contain in solution its maximum amount of nitrogen, namely 0.14 per cent., at a temperature of 590 °C. At room temperature the solubility is, however, appreciably less: about 0.01 per cent.

If the iron contains more nitrogen than the amount soluble in it, this excess nitrogen may be precipitated as iron nitride Fe_4N, a hard component with a very bad influence on the ductility (deformability) and toughness of the material. This is seen clearly in *figs. 11* and *12*, in which the impact value (as a measure of the toughness) and the elongation (as a measure of the ductility) are plotted as functions of the nitrogen content. On the other hand the tensile strength is increased by nitrogen [8]).

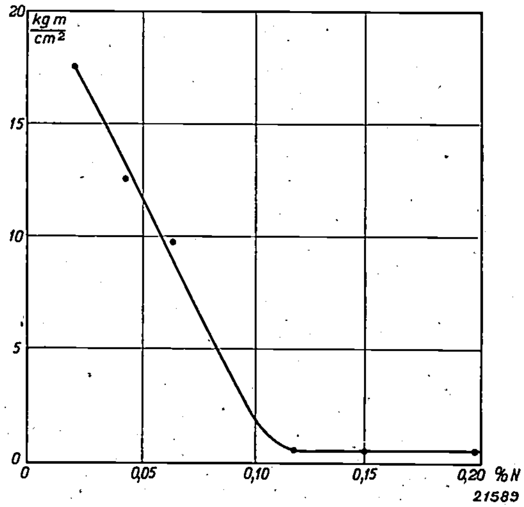


Fig. 11. The notched bar toughness of remelted iron in relation to its nitrogen content. The material was annealed at 900 °C. The notched bar toughness value is determined on notched specimens of the Mesnager model. The values are borrowed from Séférian's dissertation.

The material obtained in welding with bare wire can therefore best be described as hard and brittle, while what is desired is a material which is more or less hard and tough. In addition to the unfavourable influence of the nitrogen in welding with bare wire there is also the bad influence of the oxygen if this is present in a large amount in the weld. It has, however, been found that the mechanical properties are much more strongly influenced by nitrogen than by oxygen. While 0.25 per cent of O and more may be permissible in the weld, according to fig. 11 the impact value of a weld with about 0.06 per cent of N is about one half of that with 0.02 par cent of N.

The gases — and in this case particularly oxygen — also have an important influence on the density of the weld. Experience has shown that the material melted down from the bare welding rod

[6]) D. Séférian, Étude de la formation des nitrures de fer par fusion et du système fer-azote. Dissertation: Paris 1935.

[7]) On the other hand J. C. Hodge (The Chemistry of Low-carbon Arc Weld Metals, J. Amer. Welding Soc. 11, 36 Oct. 1932) gives lower values, for example 0.08, 0.11 and 0.12 per cent of N.

[8]) It has been found very difficult to examine separately the influence of nitrogen and of oxygen. For that reason some investigators prefer to plot the properties as functions of the total content of $O_2 + N_2$. Thus E. Söhnchen (Gase in Lichtbogenschweißung von Stahl, Arcos 13, 1440, (1936), finds that the tensile strength decreases with increasing content of oxygen plus nitrogen.

is porous due to slag inclusions and blow holes. The slag inclusions consist chiefly of iron and manganese oxides, the blow holes contain carbon monoxide which is formed from the reduction of iron oxide by carbon.

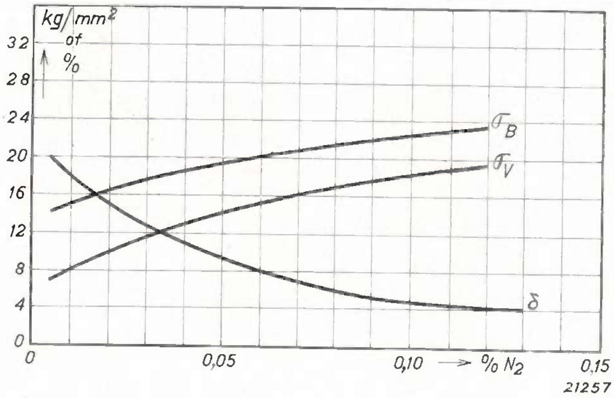


Fig. 12. Tensile strength (σ_B) and yield value (σ_V) in kg/mm², and elongation (δ) in per cent, in relation to the nitrogen content of nitride' steel wire (0.14 % C, 0.18 % Si and 0.45 % Mn) according to Tschischewski (The Occurrence and Influence of Nitrogen in Steel, J. Iron and Steel Inst. **92**, 47 1915).

In order to obtain good mechanical properties of the weld, care must be taken to prevent the material being attacked by the air during the welding process. This is done by coating the welding rods.

Coated welding rods

The coating is carried out by dipping the rods in a thin paste of ground minerals, sometimes mixed with organic substances, drying and then repeating the operation until the correct thickness of coating is reached. It is also possible to extrude the coating around the wire in one operation.

It has been shown by means of cinematographic X-ray recording [9]) that the material of the welding rod is transferred to the work to be welded in the form of drops. This holds for the bare as well as for the coated welding rod. In the latter case it also appeared that the formation of the drops takes place entirely or for the most part within the coating, and that the moving drops are covered with the flux (molten coating) (*fig. 13*). The drop form of the transfer of material is shown very clearly when the welding rod is drawn rapidly over a plate. The drops are then seen to lie separately on the plate (*fig. 14*). Each drop is found to be provided with a layer of slag. Metal and slag are thus transferred simultaneously, and the flux entirely surrounds the drop of metal, making difficult in this way the entrance of gases from the atmosphere.

c)
b)
a)

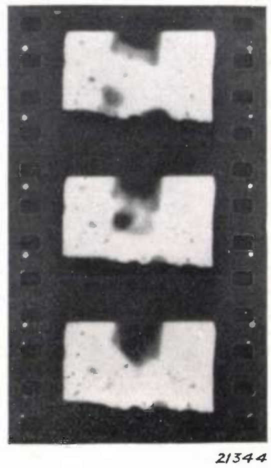


Fig. 13. Cinematographic X-ray record with 50 pictures per second of the transfer of a drop in welding:
a) drop upon formation entirely covered with flux,
b) drop passing over entirely surrounded by flux,
c) the coating forms a cup.

It is further desirable during the formation of the drop at the extremity of the rod that it be entirely covered with flux and thereby shut off from the air. The fact that this is actually the case with the Philips welding rods is proved by *fig. 15* which shows the form and condition of the coating after the interruption of the current during welding. A layer of flux shuts off the drop which is in the process of being formed from external gases.

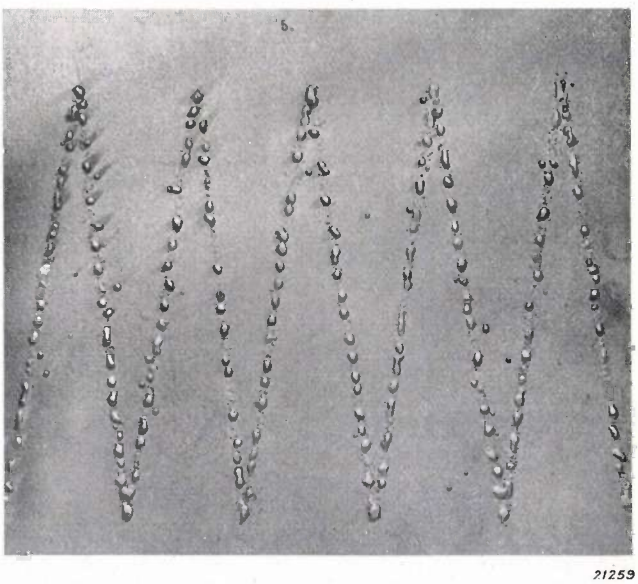


Fig. 14. Drops on an iron plate (reduced 4 times) obtained by a rapid relative movement of welding rod and plate. Automatically welded with PH-50. After welding the slag is removed and the plate is cleaned with a sand blast.

The molten metal is not only protected mechan-

[9]) J. Sack. The Iron and Steel Institute, Symposium on the Welding of Iron and Steel, **2**, 553, London 1935.
J. Sack, How Does a Welding Electrode Fuse? Philips techn. Rev. **1**, 26, 1936.

ically by the flux, but it is also screened by a mantle of gas. This protecting gas which displaces the harmful gases is formed by the evaporation of the flux and by the decomposition of some of the components of the coating. In *fig. 16* the welding process is represented schematically. The drops, entirely surrounded by flux, pass over from the welding rod to the pool in the piece of work. The density of the flux, or "slag", is about half that of iron, so that in the pool the slag easily floats to the surface and shuts off the liquid metal.

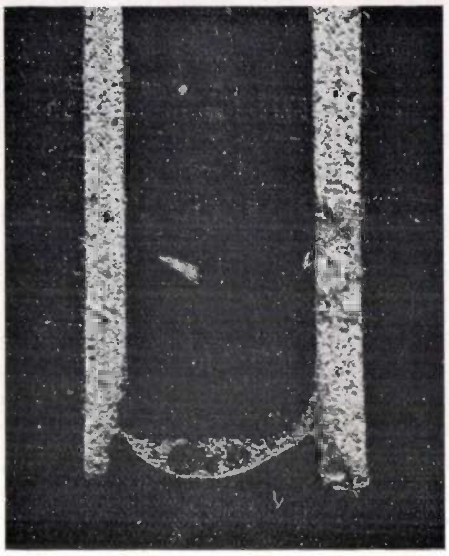


Fig. 15. Cross section of the coating after the iron core has been removed (enlarged 5 times). Condition after interruption of current during welding. A thick layer of flux covers the half-formed drop and prevents the liquid metal from being attacked by nitrogen or oxygen of the air.

It may be seen from fig. 16 that there is a fundamental analogy between what takes place in welding and what takes place in the manufacture of steel.

T h e m e l t i n g c o a t e d w e l d i n g r o d a n d t h e p o o l a r e i n f a c t m i n i a t u r e e l e c t r i c a r c f u r n a c e s.

The welding process as a method of "steel manufacture"

In the steel furnace, for example a basic open hearth furnace or a Thomas converter the molten metal is covered with a layer of slag, which serves not only to protect it against oxidation by the air, but at the same time fulfils all sorts of other functions such as purifying and reacting with the metal.

In welding, the flux must also serve similar "metallurgical purposes", in order that the weld shall satisfy the various requirements which will be made of it. While, however, in an electric furnace the time during which the slag is in contact and reacting with the liquid metal, is more than an hour, during welding all the metallurgical reactions must be complete in a few seconds.

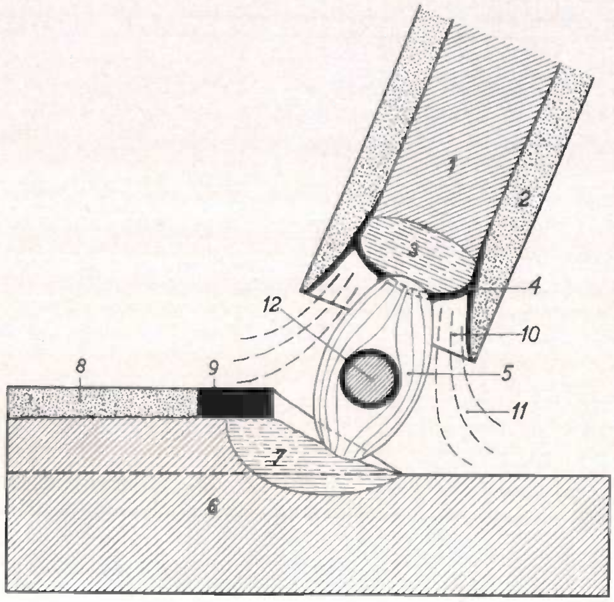


Fig. 16. Diagram of the welding process (cross section).
1: core of the welding rod.
2: coating of the welding rod.
3: drop of liquid metal (to be compared with the bath of molten metal in the manufacture of steel).
4: liquid flux (covers the molten bath 3).
5: electric arc.
6: piece of work to be welded.
7: bath of molten metal under tha arc, the so-called pool.
8: solidified flux which covers the metal already melted.
9: liquid flux.
10: cup (exerts a very favourable influence on the stability of the arc.
11: protecting gas.
12: drop in transit surrounded by flux.

This is made possible by two things:

1) During welding the temperatures are much higher than during smelting. The velocity of the reaction between slag and metal depends largely upon the temperature, and it is appreciably higher during welding than during smelting.
2) The relation of the size of the surface of contact to the amount of molten metal is, perhaps, 100 times greater in electric welding, which is more favourable to reaction than in smelting.

Other functions of the coating

The coating may have still other functions besides the above-mentioned ones:

a) It decreases the loss by oxidation and evaporation of the elements existing in the welding wire, such as C, Mn and Si. With bare wire the oxidation [10]) of

carbon amounts to 40 to 60 per cent,
manganese „ „ 50 to 85 per cent,
silicon „ „ 100 per cent.

With coated welding rods the oxidation is dependent on the nature and thickness of the coating. If desired the loss by oxidation of an element can be compensated for by the addition of this element to the coating. Thus in the case of the Philips 50 (a thickly coated welding rod for welding mild steel) the oxidation of:

carbon is 0 per cent
manganese „ 0 per cent,

while the content of silicon in the deposited material is higher than that of the wire.

b) The substances in the coating, especially those with a low work function, make possible the use of alternating current for welding, which is impossible with bare rods. A very thin coating is sufficient to make this possible.

c) The coating forms a cup at the extremity of the welding rod, which directs the arc (fig. 13*c*), so that the blowing of the arc due to the so-called magnetic wind is decreased. In welding with bare and thinly coated welding rods this phenomenon often gives much difficulty. Moreover the arc blow also depends upon various factors such as the intensity and nature of the current.

d) The closed cover of slag on the weld provides for a more gradual cooling of the material deposited.

Philips welding rods

The welding rods are provided in various kinds and thicknesses. For welding mild steel and wrought iron the welding rods PH-38, -40, -46 and -50 may be used. Each type has its specific range of application:

PH-38 For simple constructional work of mild steel of medium strength, and for reinforcing. Also suitable for the welding of cast iron, with and without preheating, but the junction cannot be machined.

PH-40 For constructional work of mild steel of medium strength. For welding sheet steel.

PH-46 For constructional work in which a tight, regular and smooth weld is required. For thin sheet welding, steel window frames, etc.

PH-50 For constructional work with high static and dynamic loads, such as bridges, locomotives, railway carriages and ships. Also for welding and reinforcing high-tensile steel and for maintenance welding on cast iron and malleable iron. The weld can be forged and has the following physical properties:

tensile strength	29 - 33 tons/sq.in.
yield	23 - 25 tons/sq.in.
elongation ($l = 5d$)	27 - 31 per cent
notched bar toughness on DVM specimens	9 - 12 kgm/sq.cm
resistance to fatigue	14 - 15 tons/sq.in.
angle of bending	180 degrees

For reinforcing with high resistance to wear, the welding rods PH-200, -250 and -600 can be used. The numbers are approximately the same as the Brinell hardness numerals of the layer built up.

For the welding of aluminium, bronze and cast iron (without preheating, the weld can be machined) the welding rods PH-AL, -BR and -GM may be used.

For the welding of stainless steel the following are suitable: PH-RS (especially adapted for fillet welds) and PH-RSB (especially adapted for butt welds).

[10]) G. M. Tichodeev, Physical-chemical Phenomena of the Transfer of Metal during Welding, J. Amer. Welding Soc. 15, 26 Mar. 1936.

[11]) K. T. H. Dag Du Rietz, Blaswirkungen bei der Lichtbogenschweißung, Elektroschweißung, 7, 101, 1936.

THE CAUSES OF NOISE IN AMPLIFIERS

by M. ZIEGLER.

Summary. Two causes of noise in amplifiers are considered in detail: the corpuscular nature of electricity and the thermal motion of the electric charges in conductors. Both of these phenomena cause in an electric circuit fluctuations of potential and current which are evenly distributed over the whole range of frequencies. By comparing the observed fluctuations with the result of theoretical considerations it is possible to determine the elementary charge e and the Boltzmann constant k.

Introduction

When the development of radio technique, particularly that of amplifier tubes, had made possible the amplification of electric potentials to almost any desired extent, the difficulty was soon experienced that the possibility of making very weak signals observable was not exclusively limited by the presence of disturbances of "atmospheric" or industrial origin. While the level of the disturbances mentioned can, under favourable circumstances be made so low as to be neglected, a continuous noise is, however, still observed, from which very weak signals can no longer be separated.

Even by the most careful choice of the parts to be used and the most careful mounting of them, by the elimination of all fluctuations from mechanical sources, among which may be included the microphonic effect and current fluctuations due to poor contact, it is impossible to lower the level of the disturbances below a definite limit. The phenomenon is indeed of a fundamental nature. It is a manifestation of the spontaneous irregular motion and of the corpuscular structure of electricity.

Spontaneous movement

Wherever an electric current can flow under the influence of an electromotive force, the electricity is always spontaneously in motion, even in the absence of such an EMF, because the electrons, which make conduction possible, take part in the heat motion of the atoms; in other words, they possess their own thermal velocity (this velocity which allows them to leave a heated filament, for example). The thermal movements of electrons are, like those of the molecules of a gas, entirely at random, with the result that on the average over short periods of time there will occur an excess of electron motion in the one or the other direction. One then speaks of the "thermal" fluctuations of current or voltage.

Whenever, for instance under the influence of a sinusoidal EMF, a regularly alternating swarming movement of particles of electricity is superposed upon the above-mentioned unavoidable irregular motion, the regular motion will be distinguishable only if that motion is sufficiently great compared with the irregular motion, no matter how great the amplification.

We shall see that the intensity of the thermal fluctuations is connected with the thermal energy kT, in which $k = 1.37 \times 10^{-23}$ joule/degree, the so-called Boltzmann constant, and T is the absolute temperature, and is independent of the size of the elementary charge, that is, it would be as great if this charge were infinitesimally small.

Corpuscular structure

It is known that the particles of electricity have a finite charge e, which is equal to 1.6×10^{-19} coulombs. In a vacuum tube when either by photoemission or by thermionic emission electrons leave an electrode independently of each other, and, also independently, pass over to another electrode, then even with a strong current through the tube the number of electrons which passes over in a given interval of time is finite. This number is however not always exactly the same, but exhibits accidental fluctuations about the mean value. The fluctuations of current depending on this are determined by the size of the elementary charge e. The phenomenon is called the "shot effect" because of its analogy with the pattering of shot.

[1]) In order to make clear the significance of k it may be remarked that the average kinetic energy of a free particle taking part in the thermal motion is $3/2kT$.

Fluctuations in the number of independent events

The theoretical treatment of the shot effect is nothing but a special case of the well known problem of the calculation of the probability of fluctuations in a number of independent events. In the calculation of the thermal fluctuations of electricity one may also, as we shall see, make use of the solution of this problem. The problem in probability is the following. Assume that one is considering a series of events which occur independently of each other, for instance the passing of vehicles in a continually evenly crowded street, the births over the whole world, falling hailstones, etc. One counts the number of events during equally long, successive intervals of time. This number n will in general be different for all the intervals. We limit ourselves to those cases in which the mean of n taken over a large number of observations, which mean we call $\overline{n}$, no longer changes when the number of observations is taken still larger. The series of the differences $(n - \overline{n})$ tells us something about the fluctuations in the time of the number of events. The numbers counted may not only be larger but also smaller than $\overline{n}$. On the average, according to the definition of $\overline{n}$, the deviation $(n - \overline{n})$ from the mean will of course be 0. One may however ask what is the mean square of $(n - \overline{n})$, which is always positive. The theory of probability shows that if the n events are independent of each other, the mean square of the deviation is just equal to the mean number of events. One obtains therefore the important formula

$$\overline{(n - \overline{n})^2} = \overline{n} \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

As the time intervals are chosen longer, n and $\overline{n}$ and therefore also the mean square of the fluctuations, become greater. Relatively, the fluctuations become steadily smaller, since equation (1) can also be written

$$\overline{\left(\frac{n - \overline{n}}{\overline{n}}\right)^2} = \frac{1}{\overline{n}} \, .$$

If by n we understand the number of electrons which in a vacuum tube pass over in a given interval of time τ from one electrode to the other, formula (1) gives us a picture of the fluctuations in the current strength.

Let us consider as an example the current in the photocell in the case of a television installation described previously in this periodical [2]). If we are interested in the accidental variations in the brightness of an image point, then τ is the time of illumination for one point, and is in this arrangement 2×10^{-7} sec. The current strength in the dark parts of the picture is about 10^{-8} A $= 6 \times 10^{10}$ electrons per second, so that 10 000 electrons are emitted per element of the image. The mean square of the relative fluctuations in the intensity is thus 10^{-4}. The average variation in the brightness of an image point resulting from this is given by the square root of the square of the fluctuations, and will therefore be about 1 per cent.

[2]) H. Rinia and C. Dorsman, Philips techn. Rev. 2, 72, 1937.

Application to the shot effect

As the above example shows, equation (1) is immediately applicable to the fluctuations which occur in a vacuum tube where the chance of the transition of an electron from one electrode to the other is independent of the behaviour at that moment of the other electrons. This is for example the case in a saturated diode (see *fig. 3*).

Just as in the above described example n is the number of electrons which pass over in a given interval of time τ from the cathode to the anode. The current strength in that time interval is on an average

$$I_\tau = \frac{n\,e}{\tau} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Now equation (1) can be written in the form

$$\overline{\left(\frac{n\,e}{\tau} - \frac{\overline{n}\,e}{\tau}\right)^2} = \overline{n}\left(\frac{e}{\tau}\right)^2,$$

from which, with the aid of (2) it follows that

$$\overline{(I_\tau - \overline{I})^2} = \frac{e}{\tau} \cdot \overline{I} \quad . \quad . \quad . \quad . \quad . \quad (3)$$

Here $\overline{I} = \overline{n}\,e/\tau$ is the mean value of the current over a very long time.

We see from this formula that the mean square of the deviation of the current so defined is determined by the charge e of the electron, and is furthermore proportional to the average current and inversely proportional to the time τ of the observation.

If we should let the time τ become infinitesimally small, then according to (3) $\overline{(I_\tau - \overline{I})^2}$ would become infinitely great. The fact that this is actually not the case is dependent on the fact that the current impulses at each transition are not infinitely short, as we have up to now tacitly assumed.

Spectral consideration of the shot effect

In equation (3) the theory of the shot effect is given, but the form of this equation is not yet such that one can immediately see from it how the shot effect will be manifested in practice. It seems however natural to consider the irregular variations of the current, which are superposed upon the mean current because of the shot effect, as composed of components with all frequencies. Actually one is of course not concerned with all frequencies at once, but with those which lie within a finite range of frequencies.

It is easy to understand that the proportionality of the fluctuations to $\overline{I}$, expressed in equation (3), also holds for the fluctuations included in any arbitrary range of frequencies. There is no difference in kind between the current I of one saturated diode and an equally large current coming from n saturated diodes connected in parallel, of which each one allows the current $I_l = I/n$ (where the subscript l indicates an arbitrary example of the n diodes) to pass, because both currents are built up in exactly the same way of electron transitions. It is also impossible to distinguish among the n parallel connected diodes themselves. If we represent the fluctuations included in a given frequency range by the letter i, then the following holds:

$$i = \sum_{k=1}^{n} i_k, \text{ so that } \overline{i^2} = \overline{(\Sigma\, i_k)^2}.$$

Since there is no correlation among the various currents

$$\overline{i_1\, i_2} = \overline{i_1\, i_3} = \ldots = 0,$$

so that

$$\overline{i^2} = \sum_{k=1}^{n} \overline{i_k^{\,2}} = n\, \overline{i_l^{\,2}}.$$

From this it follows that the desired proportionality is

$$\frac{\overline{i^2}}{\overline{i_l^{\,2}}} = n = \frac{\overline{I}}{\overline{I_l}}.$$

What finally is the spectral distribution of the components into which the fluctuations can be resolved? Since the fluctuations are governed entirely by change, the chance of rapid variations is just as great as of slow ones, so that it may be expected that all frequencies will be equally represented. It may, as a matter of fact, be calculated that the contribution to the mean square of the fluctuations due to the components included in the frequency range between ν and $\nu + \Delta\nu$, which component we shall indicate by $\Delta\ \overline{(I - \overline{I})^2}$, is independent of ν and proportional to the breadth of the frequency range and to the mean current:

$$\Delta\ \overline{(I - \overline{I})^2} = 2\, e\, \overline{I}\, \Delta\, \nu \quad . \quad . \quad . \quad . \quad (4)$$

In the derivation of this result it is assumed that the length of time necessary for the transition of an electron is short compared with the time of vibration at the frequency ν. If this condition is no longer fulfilled, the fluctuations are smaller. Instead of (4) we then obtain

$$d\ \overline{(I - \overline{I})^2} = 2\, e\, \overline{I}\, f(\nu)\, d\, \nu,$$

where $f(\nu)$ is a function which, in a diode of the usual dimensions, is a constant equal to 1 up to about 10^8 per c/s, and then gradually approaches 0, so that $\overline{(I - \overline{I})^2}$ always remains finite. In the case of most of the frequencies usual in the radio one is still concerned with formula (4).

The discussion up to now has referred to the case of the saturated diode. If the current of electrons in a vacuum tube is not saturated, that is if the current strength is determined by the space charge (thus by the mutual repulsion of the electrons), then the shot effect becomes appreciably smaller. In this latter case the transitions of the electrons are no longer independent of each other. A chance excess of emitted electrons is reflected back to the cathode by the space charge, so that the number of electrons passing over exhibits much less variation. Because of this fortunate circumstance the fluctuations of the current in radio tubes, where there is always an appreciable space charge around the cathode, is considerably less than would follow from formula (4). It is however impossible to make them arbitrariety smaller without limit; in electron tubes there will always remain fluctuations which are found to be connected with the distribution of velocities among the emitted electrons.

Theory of the thermal fluctuations

We have seen in the introduction that the thermal fluctuations of electricity have their source in the chaotic movement of electrons among the atoms. It may therefore be expected that those elements of an electric system will serve as origin of these fluctuations, in which electrons and atoms can exchange energy among themselves, that is, in electric resistances.

The external effect of these fluctuations may be considered to be caused by an EMF V in series with each resistance. It is easy to understand *that*: *this electromotive force is independent of the nature of the resistance.* Consider a resistance R connected with any arbitrary system Z (*fig. 1*), the whole in temperature equilibrium at the temperature T. The EMF of the thermal fluctuation, which we imagine to be in series with R, causes a certain current to flow through the circuit, whereby a certain energy is dissipated in Z. According to the principle of detailed equilibrium, due to fluctuations in Z, an equally large amount of energy must be supplied to R. This latter energy will not change if we choose resistances of different kinds one after another instead of R; conversely therefore all these resistances must give equally large energy fluctuations.

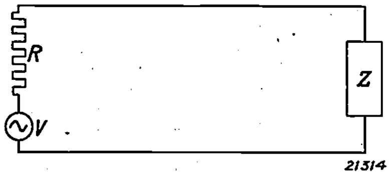


Fig. 1. The potential and current fluctuations which occur in an electric network due to the thermal motion of the electricity in a resistance R, may be considered to be caused by an alternating voltage V in series with the resistance. V is independent of the impedance Z which joins the ends of R, and is proportional to R and to the temperature.

It may further easily be deduced that the mean square of the alternating potential which is generated at the extremities of a resistance is proportional to R. If one considers a resistance R in a given circuit divided into n equal parts R_l connected in series, each resistance R_l is subject to the same conditions. The fluctuation EMF's V_l which may be considered to be in series with each R_l must therefore be equal. Together they must be equal to the alternating potential of the total resistance V. Since the potentials V_l are entirely independent of each other, and therefore $\overline{V_1 V_2} = 0$, $\overline{V_1 V_3} = 0$ etc. the mean square $\overline{V^2}$ of the sum is equal to the sum of the mean squares $\overline{V_1^2} + \overline{V_2^2} + \ldots$ so that

$$n\,\overline{V_l^2} = \overline{V^2}$$

and therefore

$$\frac{\overline{V_l^2}}{\overline{V^2}} = \frac{1}{n} = \frac{R_l}{R}.$$

Finally we desire to know how the alternating potentials are distributed over the various frequencies and how their magnitude depends upon the temperature. It is obvious that the spectral distribution will be the same as that of the shot effect, and that the strength of the fluctuations will be given by the thermal energy kT. The result of more detailed calculations is that, independent of the impedance connected with the resistance,

$$\Delta\,\overline{V^2} = 4\,kT\,R\,\Delta\,\nu \quad \ldots\ldots \quad (5)$$

Making use of the fact that potentials generated in a resistance by thermal causes are independant of the nature of the resistance, we propose to derive the above formula with the aid of an especially simple model. We consider as resistance a vacuum tube, to be described in more detail later, and are able by means of this imaginary experiment to show a relation between the thermal fluctuations and the shot effect. Considering the fact that we have already discussed the theory of the shot effect above, we may be permitted to use the results there obtained in the consideration of the thermal fluctuations.

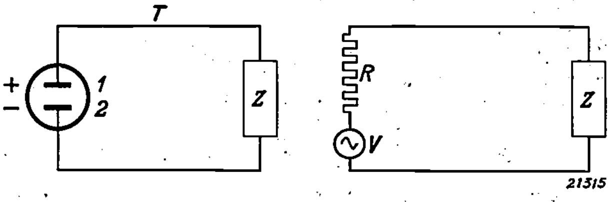


Fig. 2. A vacuum tube both of whose plates have the same temperature T (left) represents a very simple model of an ohmic resistance (right). From this model it is found that the thermal fluctuations can appear in a diode in the form of the shot effect.

Consider for instance a vacuum tube (*fig. 2*) with two similar electrodes *1* and *2* in the form of plates of the same metal which are kept at a temperature T sufficiently high to cause an appreciable emission of electrons in the electrodes. This tube behaves as a resistance: if the electrodes are joined with the poles of an EMF V, a very definite current I will flow through the tube. By definition the resistance of the tube R equals dV/dI. How large is dV/dI? Electrons leave each electrode with various velocities; as is known, they have the Maxwellian distribution of velocities corresponding to the temperature T. The (temporary) presence of electrons between the electrodes produces there a space charge which causes a minimum of potential. For the sake of simplicity we shall assume that the current density and separation of the electrodes are so small that the depth of the potential minimum which occurs is small enough in comparison with the average emission velocity of the electrons so that we may neglect its influence. In that case the average currents $\overline{I_1}$ and $\overline{I_2}$ in each direction depend only upon the emission from each electrode $\overline{I_{01}}$ an $\overline{I_{02}}$, respectively, and upon the potential difference $\overline{V_a}$ between the two plates. If the electrodes are shortcircuited externally the potential difference is zero, all the electrons emitted reach the opposite plate, and since the emissions are the same, the average current through the circuit is $\overline{I} = \overline{I_2} - \overline{I_1} = 0$. If by connection in series with a resistanceless battery V_a one plate, for instance plate *1*, is made positive with respect to the other, then the average number of electrons which pass from *2* to *1* per unit of time will remain the same, namely equal to the emission from *2* $(\overline{I_2} = \overline{I_{02}})$. The number which passes from *1* to *2* is smaller on the average since part of the electrons do not have sufficient speed to overcome the potential difference: $\overline{I_1} < \overline{I_{01}}$. The way in which this current depends upon the counter potential is given by the known starting function

$$\overline{I_1} = \overline{I_{01}}\,\varepsilon^{-eV/kT}.$$

For the differential resistance R of the tube we therefore find:

$$R = \frac{dV}{d\overline{I}} = \frac{dV}{d(\overline{I_2} - \overline{I_1})} = -\frac{dV}{d\overline{I_1}} = -\frac{1}{d\overline{I_1}/dV} = \frac{kT}{e\overline{I_1}}. \quad (6)$$

The electrons are emitted independently from each electrode; for the case when the diode is short-circuited externally all the electrons always pass over, and the fluctuations of the current in each direction may immediately be calculated with the help of formula (4) for the shot effect.

Thus we find

$$\Delta\,\overline{(I_1 - \overline{I_1})^2} = 2\,e\,\overline{I_1}\,\Delta\,\nu.$$

The fluctuations of I_1 and I_2 are not correlated; the mean square of the fluctuations of the total current is therefore exactly twice as much

$$\Delta\,\overline{I^2} = 4\,e\,\overline{I_1}\,\Delta\,\nu \quad . \quad . \quad . \quad (7)$$

From (6) it now follows that we may substitute $e\overline{I_1} = kT/R$ so that (7) takes the form

$$\Delta\,\overline{I^2} = \frac{4\,kT}{R}\,\Delta\,\nu \quad . \quad . \quad . \quad (8)$$

In order to cause the same current fluctuations an EMF V in series with the diode must satisfy the equation

$$\Delta\,\overline{V^2} = \frac{4\,kT}{R}\,\Delta\nu\cdot R^2 = 4\,kT\,R\,\Delta\,\nu \quad . \quad . \quad . \quad (5)$$

It can be proved that equation (5) also holds when the resistance under consideration is not short-circuited, but connected with a system having any arbitrary impedance.

If one has a circuit arrangement in which a number of resistances occur, one may thus calculate the fluctuations of the potential occurring between two given points A and B of the circuit simply according to the laws of Kirchhoff, if one imagines a fluctuation EMF introduced in series with each resistance, the value of the EMF being given by equation (5).

Since the various EMF's exhibit no correlation, the mean square of the total potential is equal to the sum of the mean squares of the contribution of each resistance.

Thermal fluctuations in potential are unavoidable; their proportionality to the temperature suggests however a method of making them as small as possible for a given circuit: namely by cooling the elements in which the fluctuations have their source.

Finally it must be noted that just as the shot effect cannot lead to infinitely large fluctuations, an ohmic resistance cannot have an infinitely large fluctuation potential, as would follow from (5) for an infinitely wide frequency range. Just as in a saturated diode, in an ohmic resistance finite times of displacement for the electrons will play a part, so that at very high frequencies (10^{14} c/s) the variations in electricity generated will decrease with the frequency.

Order of magnitude of the shot effect and thermal fluctuation

We have already noted that the formula for the thermal fluctuations exhibits much analogy with that for the shot effect. The agreement becomes still closer when in the shot effect one does not consider the fluctuating current, but the potential V which causes the curent in a resistance R: then for the shot effect one obtains

$$\Delta\,\overline{V^2} = (2\,e\overline{V})\,R\,\Delta\,\nu,$$

and for thermal fluctuations

$$\Delta\,\overline{V^2} = (4\,kT)\,R\,\Delta\,\nu.$$

In the first case the potential energy $e\,\overline{V}$ per electron appears, in the second the thermal energy per degree of freedom. The numerical factors have the same magnitude when for example $T = 300$ °K (room temperature) and $\overline{V} = 0.05$ volt. For $R = 10^5$ ohms this latter amounts to a current of only 0.5 μA. The mean square of the potential for a frequency band $\Delta\nu$ of 10^4 s/c is then more than 16 $(\mu V)^2$.

Experimental testing of the results

The testing of equations (4) and (5) comes down to this: the fluctuations are amplified and the amplified current is conducted through an instrument whose indication U is a measure of the mean square of the current, for example a combination of thermocouple and galvanometer. By comparison of the result with the theoretically calculated value the physical constants e and k can be determined. In *figs. 3* and *4* are represented schematically the arrangements for the measurement of thermal fluctuations

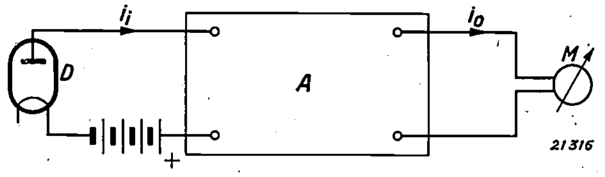


Fig. 3. Arrangement for the determination of the charge on an electron. D saturated diode (the anode current is determined by the cathode temperature). A amplifier. M instrument for measuring the mean square of the amplified current. A sinusoidal input alternating current i_i of frequency ν leads to a sinusoidal output alternating current i_o, the amplification γ_ν is i_o/i. γ_ν must be 0 when $\nu = 0$ since the direct current from the diode may not be amplified. The reading U_0 of the ammeter M caused by other sources of fluctuation besides D observed by cooling the cathode of D.

and shot effect respectively. The amplification γ_ν, further defined under the figures, may be any arbitrary function of the frequency, but must for each frequency be independent of the amplitude. It is permissible to apply the syperheterodyne principle any number of times so that a signal of frequency e can lead to a current of frequency $(\nu - p)$ at the output terminal of the amplifier.

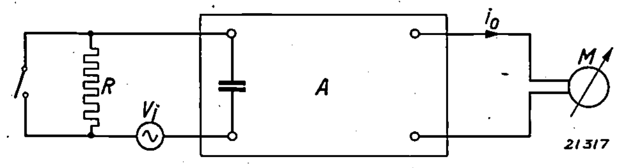


Fig. 4. Arrangement for the determination of the Boltzmann constant k. The resistance R which is the source of the thermal fluctuations to be measured is here shunted through the input capacity of the amplifier A. M is a measuring instrument for measuring the output current. A sinusoidal input alternating potential V_i of frequency ν leads to a sinusoidal output alternating current i_o: the amplification γ_ν is i_o/V_i. The reading U_0 of the measuring instrument M caused by other sources of fluctuation than R is observed by short-circuiting R, whereby care must be taken that γ remains unaltered.

In the setup for the measurement of thermal fluctuations (fig. 4) a potential V_i with a frequency ν gives a current i_o such that

$$i_o^2 = V_i^2 \gamma_\nu^2.$$

If we take for V_i the alternating potential which is generated by the thermal motion in a frequency interval $\Delta\nu$ in the neighbourhood of ν, then

$$\overline{V_i^2} = 4\ kT\ R\ \Delta\ \nu,$$

so that the contribution to the mean square of the output current is

$$\Delta\ \overline{I_{out}^2} = 4\ kT\ R\ \gamma_\nu^2\ \Delta\ \nu.$$

The total contribution of the thermal fluctuations we obtain by the summation of the above result over all the frequency intervals $\Delta\ \nu$, thus

$$\overline{I_{out}^2} = 4\ kT R \int_0^\infty \gamma_\nu^2\ d\ \nu.$$

If we indicate the contributions of all possible other (entirely independent) sources of fluctuation (for instance other resistances or tubes than the one under consideration) by U_0, then

$$U - U_0 = 4\ kT\ R \int_0^\infty \gamma_\nu^2\ d\ \nu.$$

In exactly the same way one finds from the setup for the shot effect

$$U - U_0 = 2\ e\ \overline{I} \int_0^\infty \gamma_\nu^2\ d\ \nu.$$

For a given arrangement the amplification γ_ν for each frequency is easily determined experimentally; the integration can readily be carried out graphically.

The constants e and k have been determined in a similar way by various investigators. In this laboratory also, accurate experiments were carried out several years ago. The result is uniformly satisfactory, since the deviations remain within the per cent maximum permissible error of measurement. It may be seen from this that it is possible in either a diode or a resistances to avoid all sources of fluctuations except those which are necessarily present for fundamental reasons; namely, the thermal effect and shot effect. One must, however, always take into consideration the phenomena discussed, and the formulae derived are of application to all branches of radio technique.

PUBLIC LIGHTING. PRINCIPLES OF STREET ILLUMINATION

by G. B. VAN DE WERFHORST.

In a previous article [1]) we noted how the influence which the total outdoor illumination exerts on the public lighting differs so much inside the built-up areas from that outside the built-up areas that the subject must be divided into two parts. While in the article mentioned the lighting of the highways outside the built-up area was discussed, we should now like to discuss the public illumination installed within the built-up area, and combine the two under the term "street lighting", under which, therefore, not only the lighting of streets in the narrower sense but also the lighting of squares, boulevards, main traffic thoroughfares and the like will be understood.

In treating this subject we shall take the case of a large city. We may then leave it to the reader to apply the general principles which we discover to the case of a smaller city and to that of a large or a small town.

The large city presents all forms of outdoor illumination: lighting of the roadway itself and of its surroundings, signals, advertising illumination and private illumination. The first two in the public interest, the last two for private interests. The public lighting of a large city is certainly installed for the benefit of the public. But in this case it is not so simple as for the highways through the country to determine what the interest of the public is. The large city exhibits within its limits much variation. One is forced to make a distinction among different streets. In *fig. 1* we have drawn a schematic representation of the streetplan of a large city. While the outermost circle indicates the limits of the urban district (the built-up area), the actual centre of the city lies within tbe inner circle. The main thoroughfares A_1B_2, A_2B_2 etc. lead to this centre as continuations of the intercommunal highways *1, 2, 3* ... These main thoroughfares are often connected with each other by ring highways R_1, R_2. Between these latter are the residential districts and the factory districts, the better residential districts (W) are usually built out toward the west, southwest and south, the factories and working men's houses (J) toward the east, northeast and north.

The public lighting, wherever it is introduced in this total complex of streets and squares, has as one of its subjects the promotion of safety by the creation of a certain degree of visibility. We are involuntarily inclined tot hink here exclusively of traffic safety, influenced as we are by the development of modern traffic. Such onesidedness

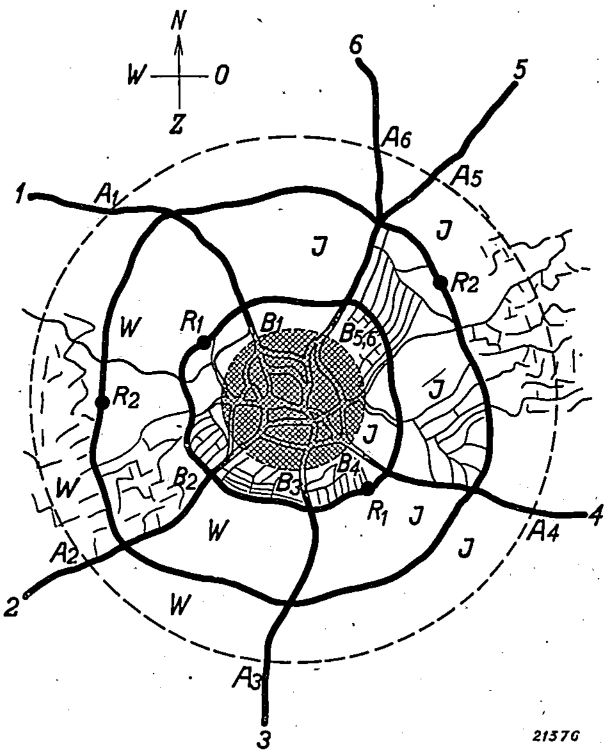


Fig. 1.

however, would lead us to a false standard. The original object of public lighting was, namely, to provide safety against hold-up and burglary by the criminal elements; we shall call this the public safety. In most of the streets of the residental sections, and in the smaller streets of the centre of the city, the street lighting has more this object of public safety than that of traffic safety.

Next to safety, the appearance of the city is of importance. Especially in the centre one wants to see the fronts of the buildings in the evening also. In this part of the city, which is usually the oldest and most characteristic, these buildings give a special character to each street in the daytime. One would like to preserve this street scene in the evening, be it in an altered form. The majority of the streets in the centre of the city are furthermore shopping streets. The modern brightly lighted show-windows are an important con-

[1]) Philips techn. Rev. 2, 110, 1937.

tributing factor in the street lighting, however only until ten or eleven o'clock in the evening, when the lights in most show-windows are extinguished. Because of the traffic intensity, which just then increases again due to the emptying of theatres and cinemas, the street lighting left to itself must have sufficient capacity. Such conditions will be found on most of the main thoroughfares and they will increase in degree from outside the city to its centre.

At their beginning (A_1, A_2 etc.) most of the main thoroughfares have the character of the rapid traffic roads through the country. The illumination of the house fronts is superfluous there. But toward the city the character is gradually changed, unless such a traffic highway is lined chiefly by factory complexes (A_4B_4), in which case the character of highway dominates into the centre of the city. This is even more the case with the ring highways, especially the outermost; there one may speak of a fast traffic like that on the highways through the country. These ring highways can be considered by the passing traffic as the links between intercommunal roads. For the passage of this through traffic it is favourable to have the lighting of the outermost ring highway of the same character as that of the connecting highways. On this ring highway, just as on the highways outside the city, the illumination of the surroundings is of no practical importance.

In the residential sections, especially in those with open building (garden village building, among others), the lighting of the surroundings is howe ververy important once more. From the street one must be able to distinguish sufficiently the houses behind their gardens, in the first place to orient oneself, and also for the sake of public safety. There is no question here of intense fast traffic.

While for the large intercommunal traffic highway we may fail to consider the pedestrian, he is very important in the case of all city streets, although not everywhere in the same way. In the centre of the city the shopping streets are to a great extent promenades, particularly in the evening hours. The pedestrian here is not a sort of traffic obstacle, he himself represents a very important traffic. The pedestrian wants to see and to be seen. not from considerations of safety, but because he considers it a pleasure. Here the general function of street lighting may be very well described as "for the pleasure of the road user", to use the words of our previous article in which we classified these general functions. In the residential sections however the lighting must provide the walker chiefly with public safety and the possibility of orientation.

On the main traffic thoroughfares and the ring highways the requirement of traffic safety dominates so far as the pedestrian is concerned.

While it may be seen from the above how diversified are the demands which are made on street lighting, in different streets, the problem shows still many more sides, when one considers the influence on the street lighting in a city exerted by the other outdoor illumination, and how varied this outdoor illumination is. In order to take this into account in the proper way, it must first be stated that from economic considerations a much higher level of illumination may be applied in the important streets of the centre of the city than in the unimportant streets of the residential districts, and further that in the centre of the city, for most of the streets, there is present an illuminated vertical limiting surface (the lighted fronts and show-windows), while in the residential districts, besides the so much less intensely illumined road surface, there is practically no question of a vertical lighted background. The road user is therefore in these two extreme cases (the other streets usually show transition states between these two) presented with entirely different degrees of brightness and therefore to very different sensitivity for glare.

Of the total illumination there are first the signal lights, which contribute to the public interest. Of these the traffic signal proper gives the least difficulty. It is however another question with the signals of the traffic island. Lying as they do in the average line of vision of the road user, they may be of considerable intensity in the centre of the city without being blinding, since the total brightness of the surroundings permits a fairly great intensity of these signals. The same traffic island signal in the suburban districts, however, has often a blinding effect in the much darker surroundings and thereby cancels to a large degree the effect of an otherwise adequate street lighting.

This is even more the case with the signal lights on vehicles. Since driving with "blinding lights" is usually forbidden by law within the built-up area, and this means in other words that the vehicle itself may not provide for the illumination of the way, the other lights carried by the vehicle may only be signal lights (serving to be seen and not to allow the driver to see). This signal lighting may also not be glaring. Even the "parking lamps" of many autos do not satisfy this condition and have

	1	2	3	4
Factors to be considered / Type of street	Colour of the light	Level of illumination	Background and fronts of buildings	Visual brightness light-source (la
Promenades and boulevards in the centre of the city	1) White or corrected to appear white for pedestrians	2) Very high	3) Continuous background of illuminated building fronts	4) May be high due and (3)
Smaller sidestreets in the centre of the city	11) White or corrected to appear white in order to correspond with (1)	12) Medium	13) Continuous background of illuminated building fronts	14) Must be rather low
Main thoroughfares from the limit of the built-up area to the innermost ring highway	21) Correspond to the colour used on the highways outside the built-up areas because of (25) and (26), no aesthetic considerations	22) Medium	23) Little background	24) Light-sour must be in ble
Main thoroughfares from the ring highway to the centre of the city	31) Gradual transition from (21) to (1)	32) High	33) Transition from (23) to (3)	34) Transition (24) to (4
Ring highway, outermost	41) No aesthetic considerations, for (45) and (46) most suitable kind of light to be chosen	42) Medium	43) No background	44) Light-sour must be i ble (43) (46)
Ring highway, innermost	51) Partial correction toward white	52) High	53) In some cases fronts which must be illuminated. Sometimes no background, gardens, park	54) When bor by buildin rather low When bor by garden source inv
Main thoroughfares through factory complexes	61) No aesthetic considerations, (65) and (66) determine the kind of light	62) High	63) No background: fronts of factory buildings are too dark	64) Light-sour invisible be of (65) and
Residential streets with continuous blocks of houses	71) White or corrected toward white because of pedestrians	72) Low	73) House fronts which may be dark	74) Light-sour invisible be of (72) an
Residential streets with separate houses	81) White or corrected toward white because of pedestrians	82) Low	83) No background	84) Light-sour invisible be of (82) an
Main thoroughfares through the residential districts	91) Rapid traffic the chief concern, therefore only slight corrections toward white for the benefit of pedestrians	92) Medium	93) No background	94) Light-sour invisible be of (92) (93 (96)
	By white light is meant the light of the ordinary sources of light having a continuous spectrum: glowlamp, gas light, arc light. The "correction toward white" of light which is not white from gas discharge lamps means either the application of light mixtures or transformation by means of fluorescence.	Under "medium" level of illumination is meant the level of illum. of the well-lighted highways through the country, for which one must calculate the light-sources to give about 200 lumens per metre length of road, from which follows meaning of low, high and very high.	By "little background" or "no background" is meant that the background is dim with no specification of whether it is the dark sky, vague groups of trees etc.	Visible means here within the angle α which for practic poses is considered 15°. As visual bri only the brightne within the angl meant

5	6	7	8	9	10
Type of safety desired	Type of traffic	Visibility	Speed of observation	Contrast	Glare due to street lighting
		also dependent on the various kinds of light			
Public traffic	6) Pedestrian, slow	7) Only white light to be considered, see (*1*)	8) May be neglected because of (*2*) and (*6*)	9) Need not be considered because of (*2*)	10) Due to (*2*) and (*3*) a fairly high degree is permissible
Public	16) Pedestr. slow	17) Of no importance	18) Of no importance	19) Need not be considered	20) Must be avoided as much as possible
Traffic	26) Fast traffic	27) Very important, therefore kind of light which gives greatest visibility. This is possible because of (*21*)	28) Very important, therefore best kind of light because of (*21*)	29) Very important, therefore best kind of light because (*21*)	30) May not occur
Traffic	36) Slow and fast traffic	37) Transition between (*27*) and (*7*)	38) Transition between (*28*) and (*8*)	39) Of less importance due to (*32*)	40) To be avoided as much as possible
Traffic	46) Fast traffic	47) Very important, choose best sort of light, permissible because of (*41*)	48) Very important, choose best sort of light, permissible because of (*41*)	49) Very important, choose most favourable level of illumination	50) May not occur
Traffic	56) Pedestrian, slow and fast traffic	57) Important but kind of light is already determined	58) Important but kind of light is already determined (*51*)	59) Need not be considered	60) To be avoided as much as possible
Traffic	66) Slow and fast traffic	67) Important, best kind of light possible because of (*61*)	68) Important, best kind of light possible because of (*61*)	69) Level of illumination too high to take advantage of this	70) May not occur
Public	76) Pedestr., slow	77) Of little importance	78) Of little importance	79) Need not be considered (*71*)	80) To be avoided as much as possible
Public	86) Pedestr. slow	87) Of little importance	88) Of minor importance	89) Need not be considered (*81*)	90) May not occur
Public & traffic	96) Pedestr. slow and fast traffic	97) Important but kind of light is already determined (*91*)	98) Important but kind of light is already determined (*91*)	99) Need not be considered	100) May not occur

We have divided traffic into 3 groups pedestrians, slow traffic with a speed up to about 12 m.p.h. under which bicycle traffic falls and fast traffic with a speed higher than about 12 m.p.h.

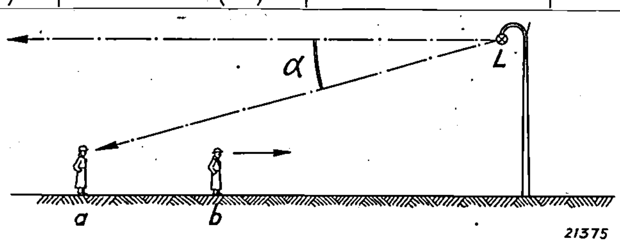


Fig. 2. Visible here means whether or not the light-source L is visible to the observer a and to all observers farther than a from L, thus within the angle α. Although observer b can see the light-source, the visibility for this observer (outside angle α) need not be considered.

much too great an intensity. It is, however, much worse with the signal lights of motorbusses, trams and bicycles [2]). Even in the lighted surroundings of the centre of a city, they usually have a very disturbing effect. But especially on the main thoroughfares and the ring highways every good intention of the street lighting is often made illusory by the blinding, effect of the signal of this kind of traffic.

Of still greater influence, however, is the advertising illumination.

Signal lighting (particularly the traffic signal) and street lighting — installed as they are in the interest of the public — are both in numerous cases rendered completely useless by advertising signs, which are for private benefit. The density of advertising signs can become so great, as in the heart of the greatest cities, that they begin to form a contributing factor in the lighting of the street, and even increase the level of total brightness so much that the blinding effect of all kinds of other sources of light becomes relatively less. This effect, however, remains practically confined to a very small district. Immediately outside that district the unbridled advertising signs make it senseless to ponder over what distribution of light one should apply to the street lighting, or what level of brightness is desirable for the road surface. At the present time the traffic arteries of large cities give a chaotic picture of motionless, moving and changing advertising lights, signal lights and traffic lights, with often an attempt by the street lighting to dominate all this with rows of extraordinarily strong street lamps which shine out and blind the road user [3]). The application of the law which prescribes the signal lights of the vehicles should be rigorously extended to include a prohibition against blinding by these lights also. The unbridled advertising illumination need not necessarily be limited in quantity or in extent, but it truly needs a limitation in permitted brightness depending upon its location.

Without such measures the street lighting, if it remains within economically acceptable limits, will never be able to satisfy the requirement of furnishing visibility on the road, even though the street lighting itself is carried out in the most perfect way.

The private lighting on a vehicle is, as was already mentioned, usually forbidden within the built-up areas. The private lighting, however, of factory property and emplacements along the street, of entrances, of petrol stations, of loading and unloading wharves, of playgrounds, often with festoons of coloured lights, is bound by no law. It is very often so blinding that the effect of the street lighting is very much reduced for the road user. These disturbing elements are, however, seldom to be found in the centre of the city and in the residential districts, but more along the main thoroughfares and the ring highways.

The factors which must be taken into account in the judging or projecting of the street lighting are therefore very different for the various streets.

Two other factors for which this is also the case must also be mentioned, namely the intensity and the speed of the traffic. The intensity of the traffic needs no explanation. The traffic speed, even when it is not bound by a prescribed maximum speed, is relatively low in the centre of the city, in any case so low that for the determination of the value of the street lighting proper the speed of observation, already influenced by so may other factors, may be left out of consideration. On the outermost portions of the main thoroughfares, however, on the ring highways and on the main roads through the extensive residential districts, the lighting must take into account the importance of high speed of observation with high traffic speeds.

The way in which the street lighting actually must be adapted to the various circumstances and conditions, what must be the standard for choosing among the various light sources available, in connection with colour, visibility, speed of observation, contrast [4]), how the glare [5]) must be taken into account, all these factors may be read from the summarizing table accompanying this article (page 144).

In order to make clear the use of this table let us consider a single type of street included in the table, for example the inner ring highway.

On such a street there are usually important, often impressive buildings. It is desirable to use a colour of light which does not lead to too great colour distortion of the whole (51), because the pedestrian traffic and the slow vehicle traffic (56) as road users do not relinquish all aestetic demands. Because of the importance of this street the level of lighting employed is high (52), so that the special contrast which can be obtained with certain sources of monochromatic light would in any case

[2]) This last only of course for those cities where there is an intensive bicycle traffic.

[3]) London, Paris and Berlin present striking examples.

[4]) See P. J. Bouma, Philips techn. Rev. 1, 102, 166, 1936.

[5]) See P. J. Bouma, Philips techn. Rev. 1, 225, 1936.

not be evident here (59). This consideration thus no longer holds as a loss factor in the choice of colour for the lighting (51). Since in this case account must be taken of fast traffic (56), visibility (57) and speed of observation (58) are important, but when the colour of the light has once been chosen, it is of no further importance for such a street to consider how a different kind of light might possibly increase the sharpness of vision and the speed of observation. Glare (60) is dependent not only on the source of light itself, but also on the total brightness of the surroundings, and this latter in its turn on the background. If this latter consists of the fronts of buildings (53), it is then desirable to keep them illuminated for the sake of the aspect of the street. In order to do this lamps must be used which send out light to the side and to a limited degree even vertically upwards, which again involves that the lamps shall show a certain visual brightness (54) for the road user.s This visual brightness may not be too blinding (60) but will be the less so because of the presence of a lighted background (53). If there is no background (53) then the necessity for the lighting of the fronts of the buildings is removed, but at the same time the requirement concerning glare becomes much more rigorous (54) (60).

The table (which might be extended to include still more types of streets) shows such a variety of problems set before those who must plan the street lighting, that it is manifestly impossible to speak of the public illumination of a large city as simply "street lighting".

Without a division of the subject — the one given here or some other — every dicussion or treatment of the subject leads to endless misunderstanding.

THE RECORDING OF RAPIDLY OCCURRING ELECTRIC PHENOMENA WITH THE AID OF THE CATHODE RAY TUBE AND THE CAMERA

by J. F. H. CUSTERS.

Summary. At different current strengths of the electron beam the maximum scanning speed for a green and a blue-fluorescing cathode ray tube was determined. The experiments carried out with various photographic material give as result that the greatest speed can be recorded by the use of a green-fluorescing screen and a suitable green-sensitive emulsion. With a tube for 2 kilovolts and a beam-current strength of 200 microamperes a scanning speed of about 1 kilometre per second was found with the use of Agfa Isochrom material. The camera lens was used at a stop of 1/4.5 and the object-image size reduction was 6 to 1.

Introduction

The cathode ray tube is a modern measuring instrument which is finding everwidening application for the most diverse uses. The phenomena which are being investigated with the help of this instrument may be divided into two categories:

1. Periodic phenomena.
2. Non-periodic phenomena.

In the case of periodic phenomena a motionless oscillogram can be obtained on the fluorescent screen of the tube with the aid of a synchronized "sawtooth" potential, by applying the potential to be investigated to one pair of deflecting plates, and the saw tooth potential to the other pair. These potentials deflect the electron beam in two mutually perpendicular directions, whereby the form of the unknown potential appears on the screen as a function of time. The establishment of a motionless oscillogram presents no difficulties, if only the potentials vary in the same way with time during every period, so that the oscillogram really is motionless. When a camera is used one can simply expose so long that a sufficient blackening of the plate is produced. Often it will be unnecessary to photograph, and the desired values can immediately be read from the image on the screen.

In the case of non-periodic phenomena which take place only once, on the other hand, photographic recording becomes essential. Here, also, it seems obvious to let the potential appear on the screen as a function of time, by applying to the other pair of deflecting plates a potential increasing linearly with the time. This deflecting voltage does not, however, need to be repeated periodically.

While with a motionless oscillogram it is not essential that the brightness of the spot should reach high values, it is of primary importance in the case of rapidly occurring non-periodic phenomena that it should do so. The brightness in this case must be so great that a sufficient blackening is obtained on the sensitive emulsion, although the image of the light spot acts on a surface element of the emulsion once only. The amount of light which falls upon such an element will therefore depend upon the speed of the spot (so-called scanning speed), while in addition still other factors such as the emulsion sensitivity and the lens aperture of the optical system, among others, play a part which we shall consider in some detail.

With these considerations in mind investigations were carried out with two definite types of cathode ray tubes having green- and blue-fluorescing screens respectively, but otherwise similarly constructed. From the experiments we could deduce the maximum value of the scanning speed. Moreover the results were given in such a form that the maximum scanning speed to be expected from other types of tubes could be fairly approximately calculated.

Derivation of the expression for the maximum scanning speed

When the light sensitive material and the camera type are given, the maximum scanning speed v_{max} for a given cathode ray tube can be calculated. By v_{max} we understand the speed of the moving spot at which is obtained a blackening of the photographic plate, giving a density $S = 0.1$ above "fog". The formula reproduced below expresses the determining factors in outline form. In *fig. 1* the following symbols are used:

Sch the fluorescent screen of the tube,

D the diameter of the light spot having the area $F = \pi/4 \cdot D^2$,

- O the lens of the camera with surface O, lens diameter d and focal distance f, so that the stop is $1/L$, where $L = d/f$,
- P the photographic plate, which receives the image of the spot.
- D' the diameter of the image of the spot with area $F' = \pi/4 \cdot D'^2$,
- b the image distance.

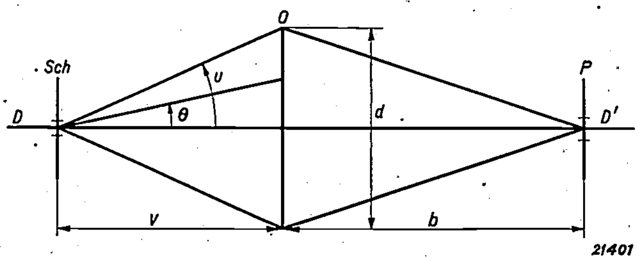


Fig. 1. Image of the fluorescent screen *Sch* formed by the lens O of the camera on the light sensitive plate P. The light spot of diameter D is reproduced in the image with diameter D'.

Let the brightness of the light spot be B. Under the condition which is well fulfilled by ordinary fluorescent screens, that the spot radiates according to Lambert's cosine law, B is independent of the direction.

The blackening of an element of the photographic plate is determined by the amount of light which strikes the element, which in its turn is determined by the product of the intensity of the illumination E and the exposure time t. The intensity of the illumination E is connected with the brightness B in the following way:

$$E = a\,B \cdot \frac{O}{b^2} \quad \ldots \ldots \quad (1)$$

Due to reflection and absorption not all the light falling upon the lens will pass through it, but only a part which is accounted for by the factor a [1]). If we introduce the enlargement N by writing $b = (N+1)\,f$, the expression becomes

$$E = a \cdot \frac{\pi}{4} \cdot B \cdot \frac{L^2}{(N+1)^2} \quad \ldots \quad (2)$$

The exposure time is found as follows. In *fig. 2* let p be the emulsion element over which the centre of the moving spot moves along the line $L\,L'$, so that when we indicate the speed of the image by v_b and that of the spot by v, the following holds:

$$t = \frac{D'}{v_b} = \frac{D}{v} \quad \ldots \ldots \quad (3)$$

For the amount of light we thus find:

$$Et = a \cdot \frac{\pi}{4} \cdot \frac{L^2}{(N+1)^2} \cdot \frac{BD}{v} \quad \ldots \quad (4)$$

The maximum scanning speed is determined by the smallest amount of light $(Et)_{min}$ which is

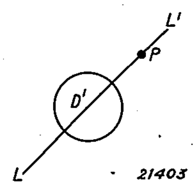


Fig. 2. The image of the spot with the diameter D' moves along the line $L\,L'$ on the light-sensitive emulsion and falls upon an element p of the emulsion during the time t.

just sufficient to cause the already mentioned minimum blackening $S = 0.1$. From this condition it follows that

$$v_{max} = a \cdot \frac{\pi}{4} \cdot \frac{L^2}{(N+1)^2} \cdot \frac{BD}{(Et)_{min}} \quad \ldots \quad (5)$$

Let η be the light yield of the fluorescent screen in candles per watt. If we introduce in watts the energy of the electron beam and assume that η is independent of this energy, the following holds:

$$BF = V_a \cdot I_a \cdot \eta\,, \quad \ldots \ldots \quad (6)$$

where V_a is the final anode potential and I_a the current strength of the beam.

If we substitute this in the expression for v_{max}, we find:

$$v_{max} = a \cdot \frac{L^2}{(N+1)^2} \cdot \frac{V_a I_a}{D} \cdot \frac{\eta}{(Et)_{min}} \text{ km per sec} \quad (7)$$

when D is expressed in mm.

Under given circumstances almost all the factors in the above expression are known, among them the diameter of the spot and the candle power per watt. $(Et)_{min}$ is not known. None of the various methods of indicating the sensitivity of a photographic emulsion, such as the Hurter and Driffield number, °Scheiner or °DIN, makes it possible for us to calculate $(Et)_{min}$. All the methods mentioned are based upon sources of more or less white light, while cathode ray tubes emit blue or green light limited to a fairly narrow region of the spectrum, as may be seen from *fig. 3*.

An emulsion sensitive only in the blue will have a certain sensitivity expressed in any of the ways

---

[1]) In the above expression for the intensity E of the illumination the substitution of $b = f$ (for infinite distance) transforms it into the well-known formula

$$E = aB \cdot \frac{\pi}{4} \cdot \frac{d^2}{f^2} = a \cdot \frac{\pi}{4}\,B\,L^2\,.$$

mentioned, while it will be extremely small in green fluorescence light. Therefore for $(Et)_{min}$ we write simply $1/\beta$, a constant factor which depends

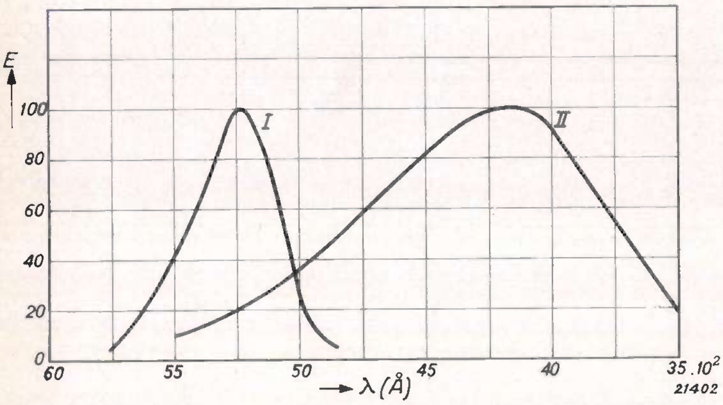


Fig. 3. Curves representing the relative energy in the spectrum of the light spot as a function of the wavelength.
Curve I is for the green-fluorescing screen.
Curve II is for the blue-fluorescing screen.

upon the spectral distribution of the energy of the fluorescence light and upon the spectral sensitivity of the emulsion. For each emulsion in combination with a given fluorescent screen β must be determined experimentally. For v_{max} we finally obtain:

$$v_{max} = a \cdot \frac{L^2}{(N+1)^2} \cdot \frac{V_a I_a}{D} \cdot \eta \cdot \beta \quad . \; . \; . \; . \; (8)$$

We will discuss first the results of the measurements of the maximum scanning, and then later and in more detail study the different variables which appear in the above expression.

Experimental determination of the maximum scanning speed

For a tube with a green-fluorescing screen (Philips DG 16 - 2) and one with a blue- fluorescing screen (Philips DB 16 - 1) the blackening of the recorded oscillogram was studied in its relation to the writing speed. The diameter of the screens of these tubes is 16 cm, the maximum potential of the final anode 2 kV. A camera was used with $L = 1/4.5$ and $f = 10.5$ cm. A reduction in size of six times was used in every case. The light sensitive material was the Agfa Isochrom Plate.

The phenomenon which took place on the screen was sinusoidal vibration passing once across the screen. With the aid of a circuit arrangement especially designed for the purpose it was possible to apply the required beam-current and to determine the velocity of the linear motion of the spot, and finally not only the frequency but also the amplitude of the sinusoidal alternating potential could be regulated.

In *fig. 4* one of the photographs is reproduced. The given scanning speed applies to those parts of the curve where the spot passes through the equilibrium position of the sinusoidal vibration. It is expressed as follows in terms of the wavelength λ of the sinusoidal curve, the frequency ν of the vibration and the amplitude A:

$$v = \nu \sqrt{\lambda^2 + 4\,\pi^2 A^2} \quad . \; . \; . \; . \; . \; (9)$$

Care was taken in every case that λ^2 was sufficiently small with respect to $4\pi^2 A^2$ to be neglected, so that

$$v = 2\,\pi\,\nu\,A \quad . \; . \; . \; . \; . \; (10)$$

The amplitude $A/6$ was measured on the photographs, while ν was read off on a calibrated tone generator.

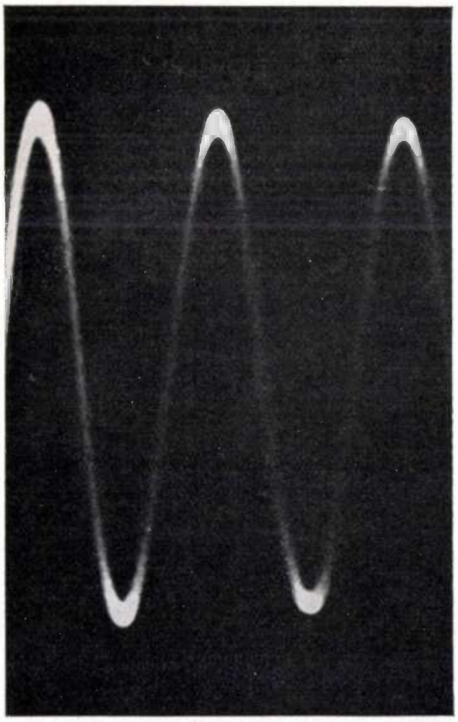


Fig. 4. Photograph of a sinusoidal vibration passing once over the green-fluorescing screen. The frequency was 4000 c/s, the beam-current 193 microamperes, while the speed at the points where the vibrating spot passed through the equilibrium position was 850 m/sec. The reduction in size of the photograph was 6 times. The small figure is a direct copy of the original photograph, while the other figure is a sixfold enlargement from the negative.

The blackening at the above-mentioned equilibrium positions of the sine curve was measured with the help of a M o l l microphotometer. In *fig. 5a*, for the blue-fluorescing screen, the blackening S

is plotted for three different current values. At low speeds the blackening increases rapidly with decreasing speed, while on the other hand it decreases very slowly with increasing speed in the interesting region of high speeds.

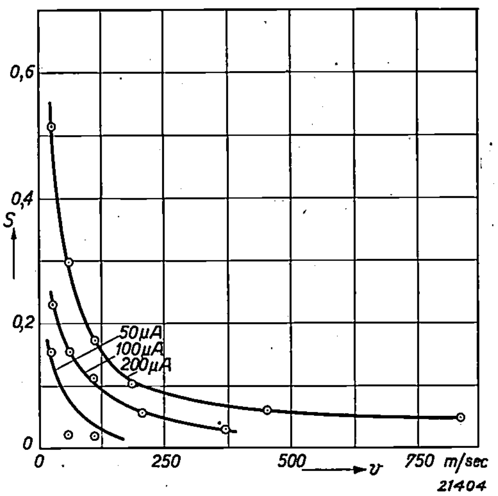


Fig. 5a. The blackening in those points of the recorded sine curve where the vibrating spot passes the equilibrium position is plotted against the scanning verlocity v of the spot with the beam-current strength as parameter. Blue-fluorescing screen.

The curves of this figure are actually only characteristic curves for the light sensitive material, but plotted in a special way. For the amount of light striking an element we found the expression:

$$Et = a \cdot \frac{\pi}{4} \cdot \frac{L^2}{(N+1)^2} \cdot \frac{BD}{v} \quad . \; . \; . \quad (4)$$

With the condition that D with a given I_a is, in the first approximation, independent of v (which condition is fairly well fulfilled, when v is not too small), and since with I_a constant the brightness B is also constant, we obtain

$$Et = \text{const} \cdot \frac{1}{v} \quad . \; . \; . \; . \; . \; . \quad (11)$$

If we plot the blackening as a function of $\log Et = \log \frac{\text{const.}}{v}$, we expect to abtain a blackening curve. In fig. 5*b* the blackening is plotted in relation to $\log 1/v$; the curves plainly show the well-known form of the characteristic of photographic emulsions. The slow decrease with increasing speed in the regions of high speeds results from the curved shape of the beginning of the blackening curves and it follows that even with large changes in the scanning speed in this region the blackening changes but little.

From the curves of fig. 5a, at $S = 0.1$, we may immediately read off the maximum scanning speed, which corresponds to the various beam-current values. These values are collected in *Table I.*

The corresponding measurements for the green-fluorescing screen are shown in *figs. 6a* and *6b*, while in *table II* the maximum scanning speeds corresponding to different current strengths are

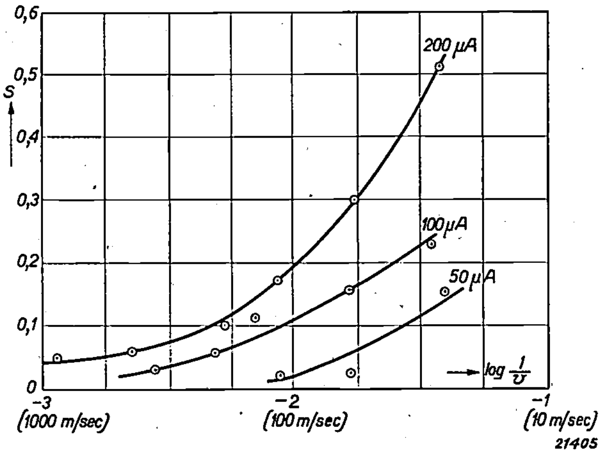


Fig. 5b. The blackening in the equilibrium points of the recorded sine curve is plotted as a function of $\log 1/v$ with I_a as parameter. The curves show the form of the well-known characteristic curve of a photographic emulsion. Blue-fluorescing screen.

given. The region investigated in this case was more extensive, with regard to v as well as to I_a. From fig. 6*b* it appears even more clearly than from fig. 5*b* that we were determining nothing else but the characteristic curves of the light sensitive emulsion upon illumination with fluorescence light; the curves belonging to various values of I_a can be made to coincide with each other by a displacement of the zero point of the $\log 1/v$ axis.

A glance at both tables is sufficient to show that with Agfa Isochrom as light-sensitive material the green-fluorescing screen is far better than the blue-fluorescing one, when it is a question of recording rapidly occurring phenomena.

Is Agfa Isochrom much more light-sensitive in the green than in the blue? On the contrary, at 535 mμ the sensitivity of this emulsion is only just over 40 % of that at 400 mμ [2]). The energy sent out in the form of green light by the green-fluorescing screen is thus considerably more than the energy emitted by the blue-fluorescing screen in the form of blue light, when the electron beam has the same energy. A discussion is given below of several other emulsions investigated.

Consideration of the variables in the expression for the maximum scanning speed

Now that v_{max} is determined experimentally we propose to study the different variables in expression (8) for v_{max}.

[2]) M. Biltz, Agfa-Veröff. Bd. IV, p. 32. Leipzig 1935.

Table I.

Blue-fluorescing screen (Tube DB 16-1)	
Beam-current strength in microamperes	Maximum scanning speed in metres per second
50	50
100	115
200	210

Stop 1/4.5
Reduction in size 6 times
Agfa Isochrom

The value at 50 μA is rather uncertain, since the number of points on the curve is so small.

Table II.

Green-fluorescing screen (Tube DB 16-2)	
Beam-current strength in microamperes	Maximum scanning velocity in meters per second
25	165
50	360
100	615
200	990

Stop 1/4.5
Reduction in size 6 times
Agfa Isochrom

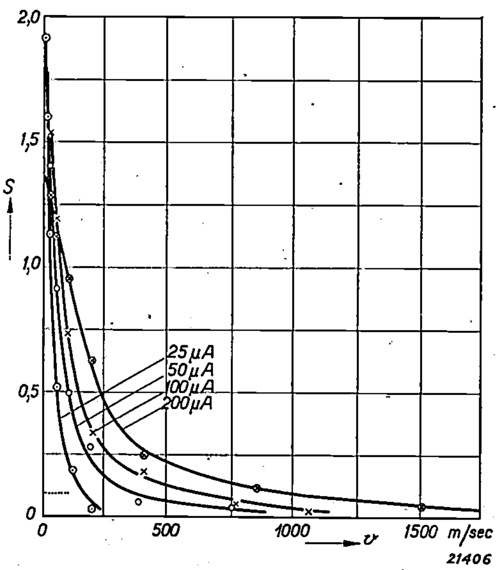


Fig. 6a. The blackening in those points of the recorded sine curve where the vibrating spot passes through the equilibrium position is plotted against the scanning speed v of the spot with the beam current I_a as parameter. Green-fluorescing screen.

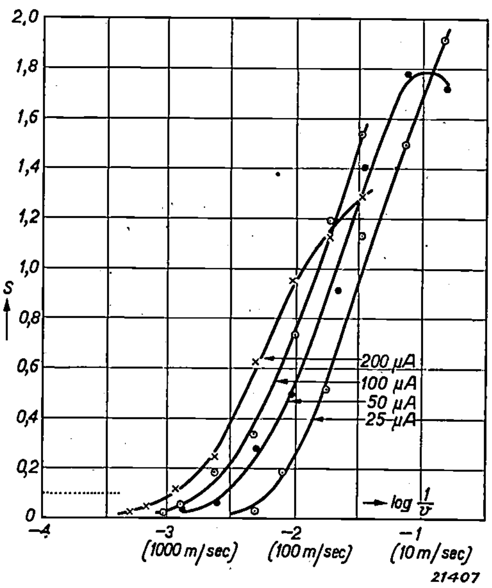


Fig. 6b. The blackening in the quilibrium points of the recorded sine curve is plotted as a function of log $1/v$ with I_a as parameter. The curves show the form of the well-known characteristic curve of a photographic emulsion.
Green-fluorescing screen.

Some of the curves begin to deviate very much in the region of low velocities (10—100 m/sec) from the straight part of the blackening curve. This is not the well-known bending of the curve at greater densities, but is caused by the local electric charging of the fluorescent screen at low scanning speeds, whereby the light spot does not become more intense, but much larger.

a, L and N are determined by the camera. The coefficient a will differ from one lens to another; moreover it depends, although to a slight degree, upon the wavelength. In general a will be somewhat smaller for blue light than for green light. With objectives of simple construction a may be set approximately equal to 3/4. We investigated a at $\lambda = 5461$ Å and 4358 Å by allowing a very narrow beam of light to pass through the lens in the direction of the axis and we found $a = 0.72$ and 0.705 respectively.

The stop of the widest aperture lenses is approximately 1/2.

We shall dwell at greater length upon the enlargement N. In *fig. 7* v_{max} is plotted as a function of N, while all the other factors in the expression are kept constant and $v_{max} = 1$, when $N = 1$. We see from this curve that v_{max} increases with decreasing N, that is to say, according as the reduction in size is greater. When $N = 1/8$ for instance, one may record phenomena which occur 3.16 times as rapidly as when $N = 1$, for the same plate blackening. Too great a reduction, however, serves no useful purpose, since the gain in speed is small from the point $N = 1/10$, and since the recorded oscillogram becomes more difficult to read.

Let us now consider $V_a I_a / D$. At the beginning we may keep I_a constant and see how $V_a D$ varies with V_a. It is known that with increasing potential V_a on the final anode the diameter D also increases, but less than V_a, so that in order to reach a large value of v_{max} one must take the maximum value for V_a which is permisslble with a given type of tube. Thus if we maintain V_a constant at this value v_{max} still depends upon I_a/D.

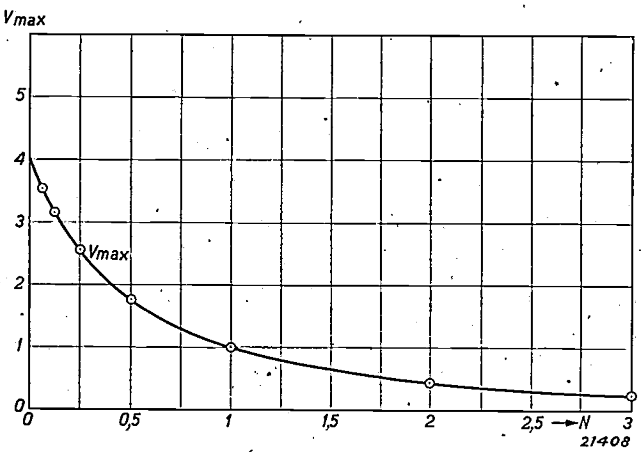


Fig. 7. Maximum scanning speed v_{max} as a function of the enlargement N. When $N = 1$, v_{max} is set equal to 1.

The scanning speed v_{max} therefore appears not to be determined solely by the brightness B of the spot but depends upon the product of spotdiameter and spotbrightness. Since in connection with the resolving power of the tube a small spot diameter is required, great brightness is important. As for the dependence of I_a/D on I_a, we may note that D increases slightly with increasing I_a. We have examined this relation separately.

On the photographs which were taken, with constant V_a but varying I_a and v, the diameter of the spot D was measured at the equilibrium points of the sine curve under a visual comparator. We thereby obtained the surprising result that, especially for the greenfluorescing screen, in the region of high speeds I_a/D was found approximately constant, independent of I_a. This cannot, however be correct, since then v_{max} would be in dependent of I_a.

With high current strengths and the consequent strong blackening of the negative the error in the visually determined D will be slight, but at low current strengths the measured value can be much smaller than the true. This occurs partly, because with small illumination only a part of the spot in the centre is sensitized above the threshold value and moreover because the apparent breadth of a line estimated non photometrically is found to small as a result of the contrast.

Let us assume on the other hand that D has a constant value independent of I_a, then we finp with the green-fluorescing screen the following calculated values for v_{max}, beginning with $v_{max} = 990$ m/sec, when $I_a = 200$ μA, so that the expression is now:

$$v_{max} = \frac{990}{200} \cdot I_a \ (I_a \text{ in } \mu\text{A}). \quad . \quad . \quad (13)$$

Table III.

I_a in micro-amp.	v_{max} in m/sec calculated	v_{max} in m/sec measured
25	124	165
50	248	360
100	495	615
200	990	990

From *table III* it is seen that we also may not consider the diameter of the spot as constant, although D changes only slightly with increasing I_a. The table shows us the relation between D and I_a. As we already noted the visually determined value of D is correct for high current values. It is found in the case of the green-fluorescing screen to be equal to 1.44 mm, when $I_a = 200$ μA and $v = 850$ m/sec. When $I_a \doteq 100$ μA, D will therefore be (see Table III) $\frac{495}{615} \cdot 1.44 = 1.16$ mm, etc.

In *fig. 8* the D calculated in this way is plotted as a function of I_a. With the blue-fluorescing screen we find a similar relation. We did not examine the latter in sufficient detail, however, to enable us to reproduce a curve with adequate degree of certainty.

Finally D depends upon the scanning speed v. Here again we encounter the disadvantage of visual measurement, so that in *fig. 9* we reproduce only the curve for the green-fluorescing screen at the highest current used, $I_a = 200$ μA, where the deviations from the true spot diameter are small. As might be expected D increases strongly with decreasing v. The cause of this lies in the local electric charging of the fluorescent screen. In the interesting region of high speeds D changes only very slightly with v and the calculated diameters in fig. 8 refer to this region.

As the last variable let us consider η, the light in candles per watt. It was measured with the aid of a flicker photometer. A television raster having a definite area was made on the fluorescent screen, while the speed of the spot was of the same order of magnitude as in the above-mentioned measurements of the blackening (200 m/sec). The potential of the final anode was kept constant, while the brightness of the screen was determined for different beam-currents. Within the limits of accuracy of the measurements, η was found approximately constant, namely equal to 1.2 candles per watt with the green-fluorescing screen, while with the blue-fluorescing screen η was found equal to 0.038 candles per watt. From these data we may calculate the factor β by means of the expression for

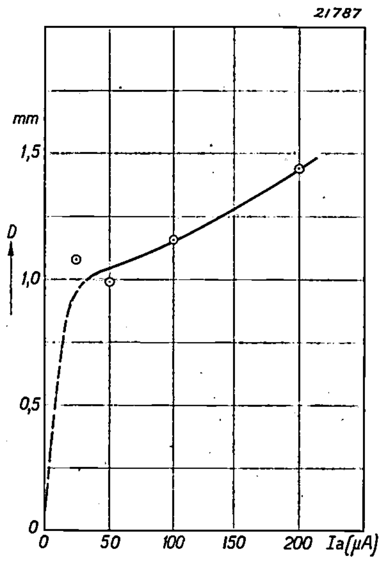


Fig. 8. Dependence of the diameter of the spot D on the beam-current I_a for the green-fluorescing screen in the region of high scanning speed.

v_{max}. With Agfa Isochrom as light sensitive material we find

green-fluorescing screen: $\beta = 110$, when $I_a = 200\ \mu$A, $V_a = 2095$ volts, $v = 990$ m/sec and $D = 1.44$ mm.

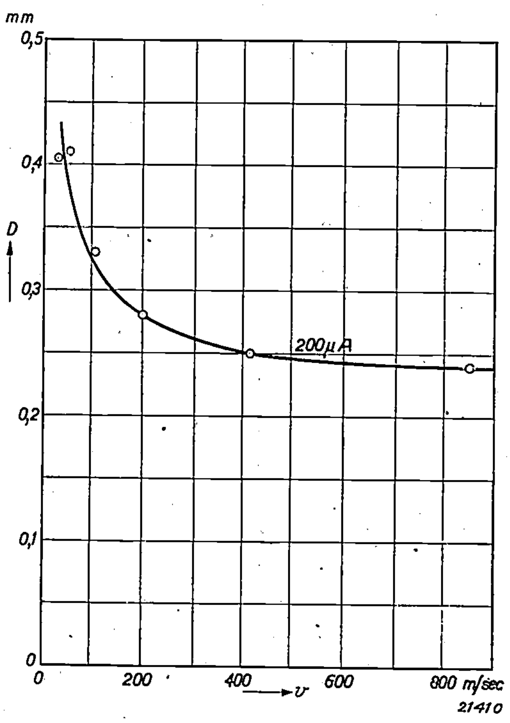


Fig. 9. Dependence of the spot diameter D on the speed of scanning v for a definite beam-current in the case of the green-fluorescing screen.

blue-fluorescing screen: $\beta = 735$, when $I_a = 200\ \mu$A, $V_a = 2095$ volts, $v = 210$ m/sec and $D = 1.44$ mm.

It is obvious that β is much greater with the blue-fluorescing screen than with the green-fluorescing screen, since the photographic emulsion is relatively much more sensitive in the blue region of the spectrum than the human eye.

Resolving power

It is clear that the knowledge of the maximum scanning speed does not suffice to answer the question as to which details occurring with a certain speed can still be observed in the recorded oscillogram. In this question the diameter of the spot plays a part. The resolving power of a cathode ray tube has been defined in various ways. None of these definitions is particularly lucid. The simplest way of indicating the resolving power of a tube seems to us to lie in the relation v_{max}/D, which multiplied by a constant is equal to the highest frequency which can still just be observed, as is discussed below in some detail.

On an oscillogram changes in the amplitude of the order of magnitude of the diameter of the spot can still be fairly well observed and read off. In order to discover what frequency can just be observed we refer back to the relation derived above between scanning speed, amplitude and frequency

$$v = 2\,\pi\,\nu\,A \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

If in this equation we write $v = v_{max}$ and $A = D$, then the maximum frequency ν_{max} follows:

$$\nu_{max} = \frac{v_{max}}{2\,\pi\,D} \quad . \quad . \quad . \quad . \quad . \quad (14)$$

For the green-fluorescing screen with $v_{max} =$ 990 m/sec and $D = 1.44$ mm, when $I_a = 200\ \mu$A, we find $\nu_{max} \approx 110\,000$ c/s. It is true that the above expression is not strictly valid, since it has been assumed that the wavelength is much less than $2\,\pi\,A$, nevertheless the error is restricted to 10 per cent.

Knoll [3]) attempts to indicate the resolving power in another way. He starts from a "normal oscillogram" for which the relation between amplitude and spot diameter is equal to 50, so that an accuracy of reading of 2 per cent is required, since variations of the order of the spot diameter may still be observed. Knoll now refers v_{max} as found from measurements to this normal oscillogram and introduces the "relative maximum scanning speed"

[3]) M. Knoll, Z. techn. Phys. 12, 54, 1931.

v_{max} (bezogene maximale Schreibgeschwindigkeit), which is defined in the following way:

$$v_{r_{max}} = v_{max} \cdot \frac{1}{50} \cdot \frac{A}{D}, \quad . \quad . \quad . \quad (15)$$

where A is the amplitude, D the diameter of the spot and v_{max} the experimentally determined maximum scanning speed. In the case of the green-fluorescing tube we find, when $I_a = 200$ μA, $A = 4.0$ cm, $D = 0.144$ cm and $v_{max} = 990$ m/sec (experimental data)

$$v_{r_{max}} = 550 \text{ m/sec}, \quad . \quad . \quad . \quad . \quad . \quad (16)$$

while the accuracy of reading is

$$100 \cdot \frac{D}{A} = 3.6\,\% \quad . \quad . \quad . \quad . \quad . \quad (17)$$

Light-sensitive material

After the relation between blackening and scanning speed had been carefully examined with Agfa Isochrom as the light sensitive emulsion, several other kinds of material were studied, in order to see whether or not they also are suitable. For this purpose photographs were made on plates from various sources, in every case under the same conditions, namely with $I_a = 100$ μA and $v = 400$ m/sec with the blue-fluorescing screen and with $I_a = 100$ μA and $v = 800$ m/sec with the green-fluorescing screen. It follows from these measurements that for the green-fluorescing screen the following materials are suitable:

Gevaert Ultra Panchro [4]), Agfa Isopan ISS, Agfa Isochrom, Kodak Supersensitive Pan Film, Perutz Braunsiegel, Ilford Commercial Ortho Film, and Lumière Super S.E., while for the blue-fluorescing screen the following proved serviceable:

Gevaert Superchrom, Agfa Ultra Spezial, Lumière Extra Rapide, Perutz Braunsiegel, Ilford Iso Zenith, Mimosa Oscillographpaper and Kodak Supersensitive Pan Film.

From a comparison of the measurements carried out with the various photographic materials it appears that the green-fluorescing screen in combination with a suitable green-sensitive emulsion is considerably better than the blue-fluorescing screen in combination with the most blue-sensitive material. The blue fluorescing screen certainly offers an advantage if one works with a moving film instead of on a time basis, since this screen has only an extremely short after-glow. In most cases however one may usually work to advantage with a stationary camera, and use one of the two parts of deflecting plates for a potential varying linearly with the time.

4) With Gevaert Ultra Panchro v_{max} (2 kV, 200 μA and DG 16-2) was found equal to 2 km/sec.

PRACTICAL APPLICATIONS OF X-RAY ANALYSIS TO THE TESTING OF MATERIALS XI.

by W. G. BURGERS.

In general the mechanical properties of a crystal are dependent upon the direction with respect to the axes of the crystal. A good example is the mica crystal which is extremely easily split along a definite crystal plane (the cleavage plane), while perpendicular to this plane it is only broken with great difficulty. Differences of this sort may, however, be found to a greater or a lesser degree with all crystals; for example, for an aluminium crystal the tensile strength in a direction along a body diagonal of the cubic crystal lattice is almost twice as great as in the direction along an edge of the cube.

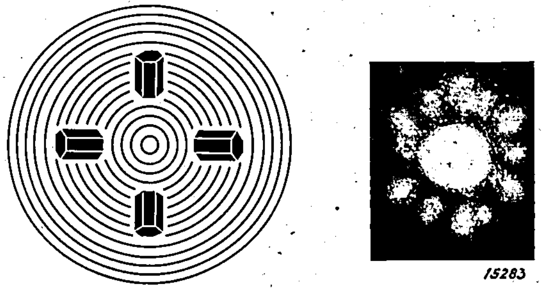
What has been stated above about the mechanical properties is also true for physical and chemical properties. Here too differences often appear when the properties are measured in different crystallographic directions. As an example we may point to the unequal resistance to attack of different crystal planes by chemical reagents, which, in the case of the etching of pieces of metal, results in the crystalline structure becoming visible.

In a crystalline material which is built up of very many small crystallites, this "anisotropy" of the properties of the crystal will be lost if the crystals lie at random as regards the position of their lattices. In that case of course all possible crystallographic directions occur in a given direction of the material, and therefore the mutual differences between these crystallographic directions are not manifested. Such a material is called "quasi-isotropic".

If however the crystallites assume definite preferential orientation as regards the position of their axes, then the anisotropy of the individual crystallites will be maintained, to a certain degree, also for the material as a whole depending upon the nature and the accuracy of the orientations. Materials which exhibit this phenomenon are said to have a "texture", which expresses the fact that the distribution of the directions of the crystal axes over the space is not uniform. Textures may occur naturally in a material or they may be induced in it by factors connected with the working or preparation.

The occurrence of a texture in a material may either be desired or even required because of a possible practical application, or it may be undesired. This depends entirely upon the nature of the practical application. In one case one may desire a material with properties as near as possible equal in all directions, in another case advantage is taken of the presence of a very pronounced anisotropy which makes the material suitable for the purpose in view. It is therefore clear that the determination of the presence or absence of a texture (and when present, of its nature) may be very important for judging the utility of a material. As we have already explained previously, this can often be deduced from the character of the interference lines on X-ray photographs. If the crystallites of a material have absolutely no preference for particular positions, then the lines are equally blackened along their entire course (either regularly or, in the presence of larger crystallites [1]), in the form of separate points). A texture on the other hand manifests itself in general in the following way. Along each interference line parts with a heavy blackening alternate with parts where the blackening is much slighter, due to the fact that in a given direction there are many more crystal planes which reflect the X-rays than in other directions. From the way in which light and dark alternate the X-ray crystallographer can moreover deduce how the crystal planes are distributed in space as regards their position, that is to say, the character of the texture present.

Examples of natural textures have already been encountered in the course of this series of articles

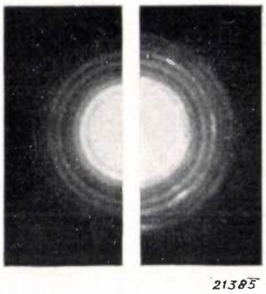


a) b)

Fig. 1. *a)* Diagrammatic representation of the texture of a real pearl. It consists of concentric layers of calcium carbonate. Each layer is built up of sixsided symmetrical crystals, whose axes are always perpendicular to the layers.
b) X-ray diagram of a real pearl; six-fold symmetrical pattern.

[1]) Cf. example 21, Philips techn. Rev. 2, 29, 1937.

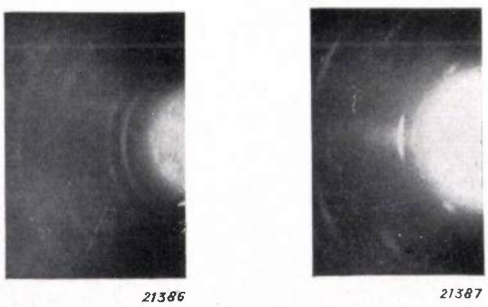
in the investigation of the genuineness of pearls [2]) (*fig. 1*), where it was shown that pearls are built up of crystallites of calcium carbonate arranged in a very definite way. In soapstone [3]) (*fig. 2*) portions were shown to occur in which the crystallites exhibit a greater or smaller preference for certain positions, a fact which has resulted in the occurrence of cracks upon heating the material. Such natural textures are the result of processes which have been active during the growth or formation of the substance considered. They are very often present, especially with substances which occur in the plant and animal kingdoms, and it may indeed be considered rather an exception when such substances are entirely anisotropic.



a) *b)*

Fig. 2. X-ray photographs of soapstone.
a) a portion which does not crack upon heating: regular distribution of the intensity along the rings indicates a quasi-isotropic material.
b) a portion where cracks appeared upon subsequent heating: interference rings with maxima and minima of intensity indicate texture.

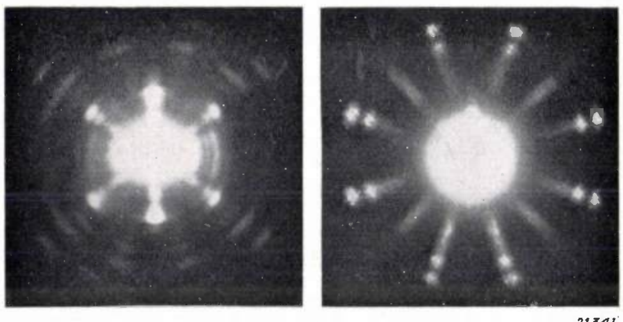
An "artificial" texture has already also been discussed, namely that of the nickel films which were deposited electrolytically on a copper rod [4]) (*fig. 3*). The occurrence of the texture was in this case probably the result of the fact that the nickel crystals grow most rapidly in a definite direction which may vary with the electrolysis conditions (composition of the bath, density of the current etc), and therefore only "favourably" oriented crystals appear on the surface of the deposited nickel film. Properties of the films such as gloss and hardness may depend upon the texture produced.



a) *b)*

Fig. 3. X-ray photographs of electrolytically deposited nickel films.
a) shiny film, shows no texture.
b) dull film, shows pronounced texture.

A second example was formed by the nickel-iron strip which is used for the construction of the cores of **Pupin** loading coils [5]). In this case not only the rolled, unannealed material, but also the subsequently recrystallized material exhibits a pronounced texture, as may be seen from the X-ray photographs reproduced in *fig. 4*. The texture caused by rolling is a direct result of the fact that the individual metal crystals are distorted by means of a sliding movement along definite crystallographic planes (the so-called slip planes), and these latter have a tendency to take up positions parallel to that in which the material flows most easily, that is the direction of rolling. The fact that upon annealing a material with such a rolling texture, a very definite recrystallization texture occurs, is connected with the fact that the places where the crystal lattice is most distorted during slipping, and thus where during annealing, the nuclei of the new crystallites are formed [6]), lie along the slipping planes. Consequently the nuclei also have a definite orientation which is connected with the rolling texture and is manifested in the X-ray picture of the heated material as recrystallization texture. As was explained in the previous article the occurrence of this recrystallization texture is essential for the quality of the Pupin loading strip obtained by further rolling.



a) *b)*

Fig. 4. X-ray photographs of nickel-iron for Pupin cores (direction of rolling horizontal).
a) rolling texture.
b) recrystallization texture.

Still another case is given below in which the

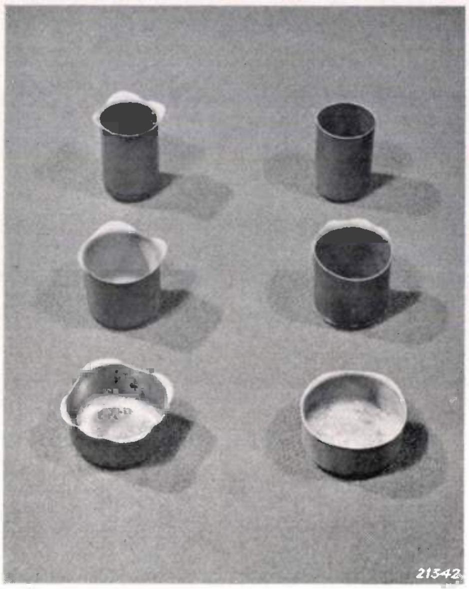
---

[2]) Example **4**, Philips techn. Rev. **1**, 60, 1936.
[3]) Example **3**, Philips techn. Rev. **1**, 60, 1936.
[4]) Example **5**, Philips techn. Rev. **1**, 95, 1936.
[5]) Example **22**, see Philips techn. Rev. **2**, 93, 1937.
[6]) Cf. example **20**, Philips techn. Rev. **2**, 29, 1937.

X-ray texture examination makes understandable the suitability or non-suitability of a material for a certain purpose.

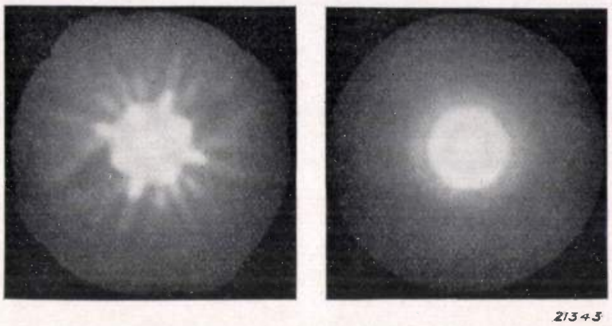
Drawing of chrome-iron cups

In the drawing of cups from round plates of chrome-iron sheet so-called lips usually appear on the rims (*fig.* 5), because the material stretches much more in one place than in another. If this lip formation is too pronounced it leads to the appearance of cracks. In a certain instance this was the case to only a slight degree with old stock material, yet upon the use of a fresh supply much larger lips appeared. Considering that such a phenomenon, according to the literature of the subject, was possibly to be ascribed to anisotropy of the material, X-ray photographs were made of both materials. These are reproduced in *fig. 6*. From the figure it may clearly be seen that in this case also the formation of the lips must be ascribed to the cause mentioned. While photograph 6*b* of the old material is regular in intensity, photograph 6*a* of the new material exhibits alternating dark and light parts. The distribution of the intensity is symmetrical with respect to two mutually perpendicular bisecting lines of the photograph. From this it may be deduced that the properties of the sheet in these two directions will be different from those in the directions lying between (the "45° directions"). From drawing tests on material with such a texture it actually appears that the yield value is unequal in different directions. It is understandable that, as a consequence, the material will flow more in certain directions than in the others during drawing, which may lead to rupture. By a change in the rolling process of this sheet, and particularly by not carrying cold-rolling too far, but alternating it with an intermediate heating, the appearance of a texture was prevented.



a) *b)*

Fig. 5. Cups drawn from round plates of chrome-iron sheet.
a) with } clear lip formation.
b) without }



a) *b)*

Fig. 6. X-ray photographs of chrome-iron sheet.
a) cups with pronounced lips. The material has pronounced texture.
b) cups with only slight lips. The material is quasi-isotropic.

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

No. 1142: J. H. de Boer and H. H. Kraak: Die elektrische Leitfähigkeit von dünnen Metall-, insbesondere von Molybdänschichten (Rec. trav. chim. Pays-Bas **55**, 941 - 953, Oct. 1936).

During and after the deposition of thin films of molybdenum the conductivity of the films in relation to their thickness was determined at various temperatures of the support. At temperatures lower than the highest temperature previously reached by the film the changes with temperature are practically reversible. At still higher temperatures, however, the conductivity rises irreversibly. If we consider thin films of metal as intermediate between vapour and solid metal, we may deduce a relative position of the energy levels of the electrons such that the anomalous temperature coefficient of the conductivity becomes logical. This scheme of the energy levels is similar to the scheme for semi-conductors according to Wilson.

No. 1143: F. M. Penning: Die Glimmentladung bei niedrigem Druck zwischen koaxialen Zylindern in einem axialen Magnetfeld (Physica **3**, 873 - 894, Nov. 1936).

Following an introduction dealing with the motion of the electrons in a vacuum between coaxial cylinders under the influence of an axial magnetic field, the breakdown voltage and the characteristic of the glow discharge at very low gas pressures are examined. For this purpose, among other factors, the influence of the inclination of the cathode with respect to the direction of the magnetic field and the influence of inhomogeneities of the magnetic field were studied.

No. 1144: W. Elenbaas: Die Elektrodenverluste in der Quecksilber-Hochdruckentladung (Physica **3**, 947 - 953, Nov. 1936).

The sum of the anode and cathode drop is found to increase with increasing vapour pressure and decreasing current, while it depends only slightly upon the phase in the case of alternating current.

No. 1145: F. Coeterier and M. C. Teves: An apparatus for the transformation of light of long wavelength into light of short wavelength. II: The influence of magnetic fields (Physica **3**, 968 - 976, Nov. 1936).

The definition of the picture obtained with the light transformer may be considerably improved by the use of a simple accelerating electrostatic field combined with the magnetic field of a long coil. The rotation of the picture which would otherwise occur is avoided by the use of a permanent ring magnet instead of an electromagnet.

No. 1146: J. F. H. Custers and J. H. de Boer: Elektrostatische und van der Waals'sche Adsorption von Jod an Fuoridschichten (Physica **3**, 1021 - 1034, Nov. 1936).

With a partial covering of a surface of calcium fluoride by iodine molecules of $^1/_2$ % at most, the absorption of light is found to be more than 100 times as great as for free iodine molecules, while the absorption spectrum is shifted toward shorter wavelengths. If more iodine molecules are adsorbed they are no longer held by electrostatic binding forces as were the first molecules adsorbed, but by van der Waals' forces. The absorption coefficient then decreases sharply and passes over into that for free iodine, although the absorption spectrum remains shifted to shorter wavelengtbs. Upon continually increasing occupation of the surface the maxima of the absorption curves are shifted toward longer wavelengths. This is especially the case with the maximum at the long-wave end. The absorption spectra for iodine adsorbed on strontium and barium fluorides have also been determined, and they were found to be very similar to those for calcium fluoride. In the case of the adsorption in calcium fluoride the spectrum of the iodine molecules is shifted farther toward shorter wavelengths than in the adsorption on strontium fluoride, and in the latter case more than barium fluoride. This supports the contention that the absorption spectrum is shifted more in the presence of a small cation than in the presence of a larger one.

No. 1147: H. Bruining: The depth at which secondary electrons are liberated (Physica **3**, 1046 - 1052, Nov. 1936).

The secondary emission of a surface exhibits a maximum depending upon the velocity of the primary electrons. The decrease in the secondary emission may be explained by assuming that the primary electrons penetrate deeper into the material with increasing speed, so that an increasing proportion of the electrons freed at those depths are unable to reach the surface. By bombarding surfaces of varying degrees of roughness and at various angles of incidence with speedy primary electrons it has actually been ascertained that the secondary emission of a sufficiently smooth surface is greater for shearing incidence than for normal incidence. With primary electrons of 500 volts the average depth to which secondary electrons are freed in a nickel plate is estimated at about 30 Å.

No. 1148: M. J. O. Strutt: Moderne Mehrgitterelektronenröhren (Schweiz. Arch. angew. Wiss. Techn. **2**, 183 - 199, Aug. 1936).

In this lecture given before the Physikalische Gesellschaft in Zürich a survey was given of the modern development of radio tubes with several electrodes. After an introduction dealing with the properties of triodes and pentodes, the tubes with variable amplification were discussed, namely the pentode with a control grid of "variable" pitch and the hexode. Such a hexode may also be used as a frequency changer. A simplified form of such a changer, which unites all functions in one tube, is the octode (six grids). The requirements of little noise and few other disturbances are discussed. Due to the vibrating space charges in such electron tubes a number of interesting effects occur, of which the so-called "induction effect" is discussed in detail.

No. 1149: W. G. Burgers (in collaboration with J. A. A. Ploos van Amstel): An apparatus for "optical" demonstration of some geometrical features of electron diffraction photographs (Z. Kirstallogr. A **95**, 54 - 73, Oct. 1936).

In this article a description is given of the construction and use of an apparatus by means of which different geometrical features of electron diffraction photographs may be shown visually with the help of the so-called reciprocal grating.

No. 1150: W. G. Burgers and F. M. Jacobs: Röntgenographische Spannungsbeobachtungen an Nickeleisenblech (Metallwirtschaft **15**, 1963 - 1066, Nov. 1936).

The reflection of Co-Kα radiations on 100-planes of a recrystallized nickel-iron plate with about 53 atom per cent of iron is very sensitive for small changes in the lattice constant. In this connection an attempt is recorded to obtain some insight into the state of tension which exists in the material of Pupin cores. Information is given about a recrystallization texture with (100) parallel to the axis of a drawn nickel wire.

No. 1151: P. J. Bouma: Vervorming van geluidstrillingen bij voortplanting door de lucht (Ned. T. Natuurk. **3**, 241 - 254, Oct. 1936).

In this article a systematic treatment is given of the various terms ordinarily neglected in the solution of the vibration equations of sound. A discussion is further given of the result obtained by the taking into account in a first approximation of these neglected effects, whereby the occurrence of non-linear distortions is particularly studied in some detail.

No. 1152 *) R. H. Houwink: Kolloidchemie und Kolloidphysik organischer plastischer Massen (Koll. Z. **77**, 183 - 201, Nov. 1936).

The relation between plastic and elastic deformation and the intra and intermolecular forces is discussed. The fundamental difference between pure elastic deformation and that of substances like rubber, in which plastic deformations occur internally, is given. Various types of diagrams for the velocity of flow as a function of the shearing stress are dealt with. Four postulates are proposed for the occurrence of such a highly elastic deformation, which postulates refer chiefly to the four possibilities of storing up potential and kinetec energy in the various bonds. The transition from liquid to solid by cooling and polymerization is discussed. The properties of plastic substances obtained by cooling or polymerization are treated separately for globular colloids and for linear colloids. Additional methods of altering plastic and elastic properties are the formation of links between linear molecules, chemical modification of the surfaces of molecules and the addition of fillers. Finally the significance of the orientation of the particles and of the formation of gels is briefly examined in relation to the properties of the plastic substances.

*) There is not a sufficient number of reprints available for distribution of the publications marked *. Reprints of the other publications will on request gladly be supplied by the Administratie van het Natuurkundig Laboratorium, Kastanjelaan, Eindhoven.

VOL. 2 No. 6 JUNE 1937

Philips Technical Review

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

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

A GENERATOR FOR VERY HIGH DIRECT CURRENT VOLTAGE

by A. KUNTKE.

Summary. Description of a high-voltage installation for a no-load voltage of 1.25 million volts and for a permissible loading of 4 mA, which was set up in the Cavendish laboratory. This installation corresponds in principle to the installation in the Philips laboratory previously described in this periodical [1]).

Some time ago, in a new section of the Cavendish Laboratory at Cambridge which is under the direction of Lord Rutherford, a generator was installed with which can be built up a continuous positive, or upon commutation, negative direct current voltage of a maximum of 1.25 million volts with respect to earth. This generator, which is to be used for investigations in the sphere of nuclear physics, was developed in the X-ray laboratory of the Philips factories. A detailed description will be given of the generator and the way in which it works.

The high voltage part of this generator is built according to the previously described[1, 4]) cascade arrangement a kind of multiplier circuit which was first given by Greinacher [2]), in which in addition to a high tension transformer, a number of condensers and rectifiers are so connected that all the condensers are charged to double the maximum voltage of the transformer, and the total voltage appears as the sum of the voltages of one half of the condensers. Greinacher himself worked only with relatively low voltages; for voltages of several hundred kV the arrangement was first used by Cockcroft and Walton [3]) and by Bouwers [4]), independently of each other and in different ways.

We may be permitted to repeat briefly the description of the principle on which the arrangement works.

The condenser C_6' (*fig. 1*) is charged through the valve V_6' to the maximum potential e of the transformer T. There occurs thereby a tension on the valve V_6' which oscillates between 0 and $2e$ with the frequency of the AC potential of the transformer.

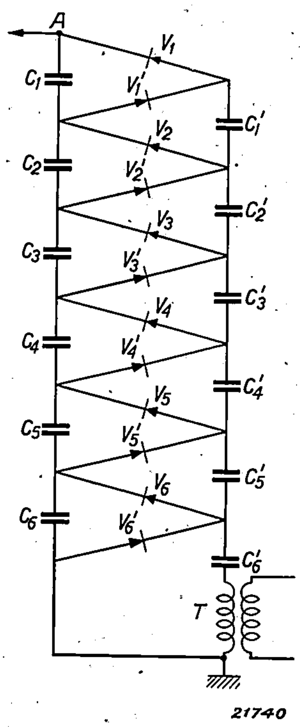


Fig. 1. Principle of the Greinacher arrangement. When the transformer T delivers an alternating current potential with the amplitude e, then in the absence of any current load, point A attains a constant direct current potential $12e$ with respect to earth.

With the help of the valve V_6 the condenser C_6 is charged to the DC potential $2e$. The tension at the valve V_6 oscillates in its turn between 0 and $2e$ and the condenser C_5' is likewise charged through valve V_5' to a DC potential $2e$. The process may be repeated and one finds — provided no

1) Cf. S. Gradstein, Philips techn. Rev. 1, 6, 1936.
2) H. Greinacher, Z. Phys. 4, 195, 1921.
3) J. D. Cockcroft and E. T. S. Walton, Proc. Roy. Soc. 136, 619, 1932.
4) A. Bouwers, Lecture American Congress of Radiology Chicago 1933 (Radiology 22).

current is taken — that all the condensers are charged to a potential of $2e$ except those which are in series with the transformer T. The valves also reach a maximum potential of $2e$. The total potential at point A is then the sum of the potentials of the condensers C_6, C_5, C_4, C_3, C_2 and C_1, therefore $12e$. This holds only when the apparatus is not loaded.

If now at A a current i is taken off, one finally reaches a state in which each condenser receives the same charge as it delivers in one period of the alternating current voltage. One then finds that, when equilibrium has been established, the condensers C_1 and C_1' are charged and discharged with iT in each period, the condensers C_2 and C_2' with $2\,iT$, the condensers C_3 and C_3' with $3\,iT$, etc.[5]).

All the valves transport the charge iT (T is the time for one period) in one period.

From this one may calculate the magnitude of the potential to which the various condensers are charged and how far the total voltage now reached lies below the no-load voltage. A detailed calculation would lie outside the scope of this article and we shall therefore only give the result of such a calculation [6]).

If one assumes that all the condensers have the same capacity C, except the first one which is in series with the transformer and has the capacity $2C$, one obtains for the voltage loss of the cascade circuit:

$$\Delta V = \frac{2}{3} n^3 \frac{i}{fC}.$$

In this expression n is the number of stages (a stage consists of 2 condensers and 2 valves), f is the frequency of the AC voltage, C is the capacity of the condensers and i the current taken off.

It may be seen from the above equation that the loss in voltage increases with the third power of the number of stages; it has been found better not to use more than 8 or 10 stages. It may also be seen that an increase in frequency as well as enlarging the capacities leads to a decrease in the loss of voltage.

When current is taken off, the DC voltage has a slight ripple proportional to the current taken off. One calculates for the amplitude δV of this ripple:

$$\delta V = \frac{i}{fC} \frac{n(n+1)}{2}.$$

in which it is again assumed that the condensers are equal in capacity C.

With the load permissible for the apparatus here described, this ripple is only 1 to 3 per cent of the maximum total potential, depending upon the load current.

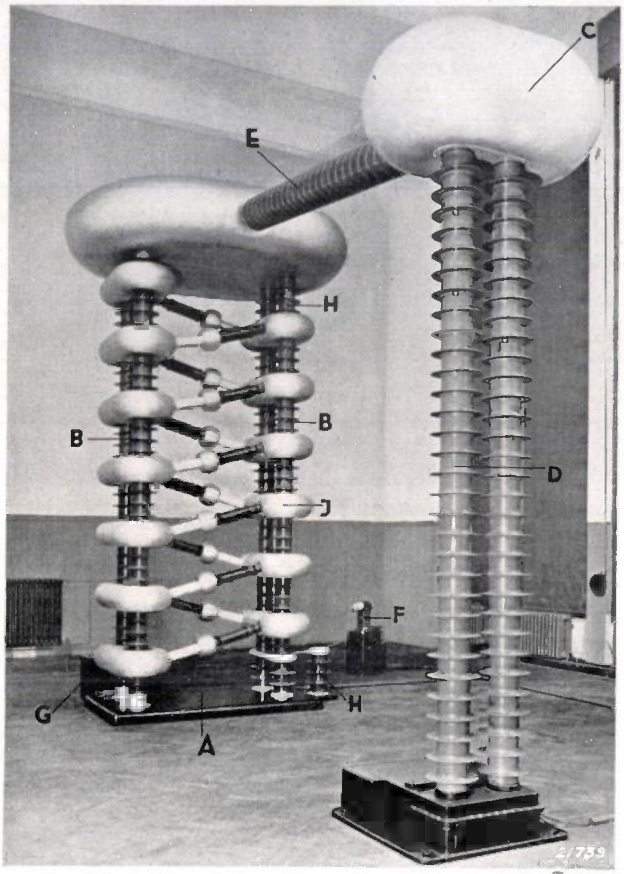


Fig. 2. Photograph of the high tension installation constructed for the Cavendish Laboratory at Cambridge. A pedestal, B columns built up of the condensers of the cascade circuit, C shielding electrode of the measuring column D, E damping resistance, G high tension transformer, F high frequency generator for the heating current.

The cascade generator here described has 6 steps, that is 12 valves with maximum peak inverse voltages of 225 kV, and 12 condensers. The values of the condensers are the following:

C_1, C_1', C_2, C_2', C_3, C_3' = 0.01 μF with 240 kV maximum DC voltage,

C_4, C_4', C_5, C_5', C_6 = 0.02 μF with 240 kV maximum DC voltage,

C_6' = 0.04 μF with 120 kV maximum DC voltage.

The initial voltage is delivered by a transformer of 120 kV peak voltage; the frequency is greater than the ordinary frequency of the mains, in order

[5]) The way in which the Greinacher circuit works when current is taken off is rather complicated, and will be dealt with in more detail in another article [6]). The fact that the charge of the condenser C_6' in each period must be larger than the average discharge of the condenser C_1 follows from the principle of conservation of energy; the condenser C_6' obtains its charge with the maximum transformer potential; the discharge follows with a much higher direct current voltage at the top of the cascade. The energy supplied (the product of charge times voltage) must, however, be equal to the energy removed.

[6]) A. Bouwers and A. Kuntke, to appear in Phys. Z. 1937.

to decrease the loss in voltage of the cascade generator. In the case here considered the frequency is 200 c/s.

The primary tension of the high tension transformer is supplied by a motor generator; by regulating the excitation of this the high DC tension can be continuously regulated.

The loss in voltage of the generator can be calculated from the given values of the capacities and is found to be 40 kV/mA, while the ripple of the total potential is found to be 7 kV/mA. Not loaded, the cascade generator gives a DC voltage of 1.25 million volts. The maximum current which may be taken is 4 mA, at which the voltage is 1.1 million volts.

In *fig. 2* the high tension part of the apparatus is shown together with the voltage measuring apparatus which is described below, and a damping resistance. *A* is the cascade generator proper; the columns *B* - *B* provided with ribs are built up of the condensers, which are thus used mechanically as structural elements. The valves may be seen mounted between the columns in a zigzag, with their damping resistances in series. As in the high tension installation of this laboratory previously described [1]) large shielding electrodes on the tops of the columns prevent electrical losses from the metal parts below; therefore these latter may have relatively small curvatures.

The consuming device, in the case under consideration a discharge tube, is connected to the electrode *C* of the measuring column *D*. Between the measuring column and the cascade generator is suspended the damping resistance *E*, which has a value of 5 megohms, and is of such dimensions that it can take up the whole voltage of the cascade generator for a short time upon breakdown of the electrode *C* to earth, or upon disturbances of the discharge tube connected to it. In the background may be seen the high tension transformer *F*.

The heating of the valve cathodes in this case, as in the apparatus already described (see footnote 1), is by means of high frequency currents. A high frequency generator (*G* in Fig. 2) of 250 watts power with a constant frequency of 500 000 c/s sends a current of about 0.7 A through the series of condensers *B* - *B*. The highest valve and the high tension transformer are short circuited for this high frequency by a bridge *G* consisting of a condenser of low capacity and an inductance in series, which are in current resonance. Inside the metal shields *J*, between which the valves are introduced, are small transformers with iron cores which are connected to the circuit and which transform the current from 0.7 A to the heating current of 3.6 A for the valves. All the valves are heated in series in this way. This arrangement has proved to be very effective for a greater number of valves.

The valves are a special type of the gasfilled rectifiers which are used with such good results in the Philips X-ray apparatus. The photograph on the right of *fig. 3* shows the external appearance of such a valve, and that on the left the valve proper which is mounted inside, and which has 8 metal cylinders between the anode *1* and the cathode *2*. These cylinders receive potentials from condensers mounted concentrically around the valve such that the tension is divided linearly over the valve during the non-conducting period.

The valve contains a drop of mercury; the vapour pressure is determined by the room temperature. The permissible temperature for the mercury lies between 15 and 45° C; this limitation is found in practice to present no difficulties. When the valves allow current to pass they have a small loss in voltage of about 50 volts; the striking potential is about 3 to 4 kV.

The measurement of such high tensions

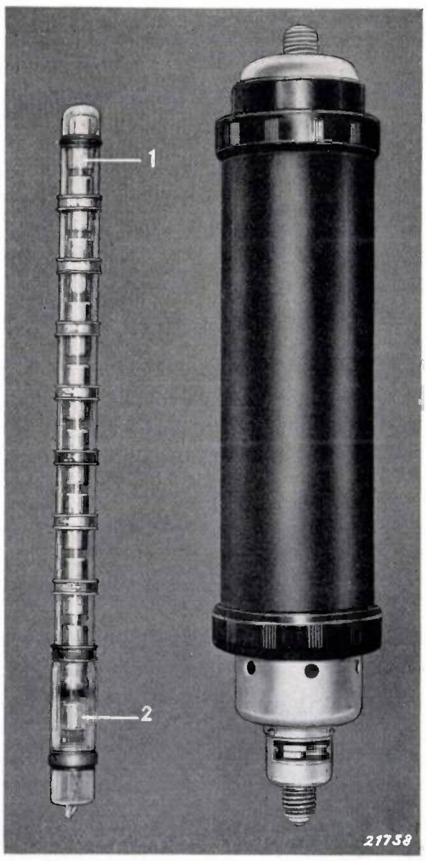


Fig. 3. High tension valves, *1* anode, *2* cathode. The path of the discharge between anode and cathode is subdivided by means of 8 metal cylinders in order to divide the inverse voltage uniformly.

as those with which we are here concerned is usually done with the aid of spark gaps between spheres. Apart from the fact that considerable deviations from the standard calibration values have recently been found [7]) — we ourselves found with the potential measuring apparatus described below that calibration values given in the VDE specifications for spheres with diameters of 100 cm are 8 per cent too high — a direct method of measurement is far preferable to the method of measuring by means of spark gaps. Moreover, for DC voltages of the order of a million volts no spark gap measurements are yet known, and the influence of the polarity is also unknown.

We therefore designed a potential measuring device which, in the case here considered, is built into the apparatus. In fig. 2 the measuring column *D* is in the foreground. A resistance of 1500 megohms immersed in oil is introduced into one of the two columns built of oiltight "Philite" moulded plastic. This resistance is composed of 2000 carbon resistances of 0.75 megohms each. Near the earthed end, by means of a tap on the resistance, a portion of the total potential is conducted to an electrostatic voltmeter. The change of resistance of the carbon resistances under the influence of the potential and the temperature is about 1.5 per cent. In electrostatic measurements, however, this change exerts no influence, since the ratio between the potentials remains the same.

Fig. 4. Control table of the high tension installation. The excitation of the alternator which acts as source of potential, and thereby also the DC voltage reached, is regulated by means of the hand wheel. The voltage reached can be read off immediately from the vertically placed electrostatic voltmeter above.

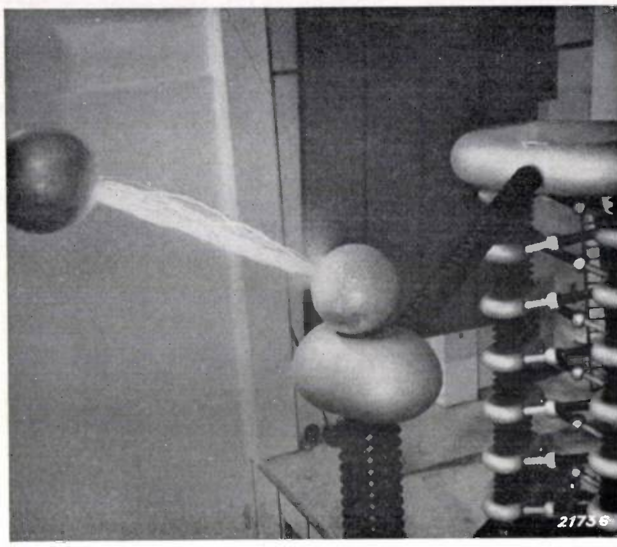


Fig. 5. A spark discharge at 1.1 million volts potential.

The ratio in our case is accurately adjusted at 1 : 1000; the 1500 V electrostatic voltmeter which is mounted on the control table (see *fig. 4*) is scaled to 1500 kV. A pump in the pedestal of the measuring column causes the oil to circulate and forces it through a cooler, thereby preventing an uneven distribution of temperature which might influence the measurement. The accuracy of this high voltage measuring apparatus is about 1 percent.

Fig. 4 shows the control table of the apparatus. When the button at the left is pressed the motor generator is started automatically and the high frequency generator for the heating of the valve cathodes is switched on. The excitation of the 200 cycle alternator, and thereby the DC voltage, is regulated with a hand wheel.

When the left-hand button is pressed the transformer is put in circuit. The cascade generator is then switched on and the electrostatic voltmeter indicates the DC tension.

In addition may be seen several instruments for measuring the power consumption, the voltage and current of the high tension transformer and a measuring instrument which indicates the current which is being taken from the cascade generator.

In *fig. 5* the cascade generator is seen in action. As a demonstration sparks are being made to pass between spheres of 75 cm diameter at a voltage of 1.1 million volts.

---

[3]) See for example W. Dattan, Elektrotechn. Z. 57, 377 and 412, 1936; E. Hueter, Elektrotechn. Z. 57, 621, 1936.

WATER-COOLED MERCURY LAMPS

by E. G. DORGELO.

Summary. The water-cooled mercury lamp is a very powerful light-source of high efficiency and small dimensions. In this article the properties of the new light-source are discussed, and the following are dealt with as examples of its application: film studio lighting, television, searchlights and air-port lighting.

Introduction

A year ago in this periodical [1]) a mercury lamp was described which excels because of its high efficiency and especially small dimensions. The investigations which led to its construction showed that the luminous efficiency (lumens/watt) of a high-pressure mercury vapour discharge increases steadily with the energy which is supplied per unit length of the column. By keeping the dimensions of the lamp sufficiently small, it has been possible to reach an energy dissipation of 40 W/cm with a total energy of 75 W. The vapour pressure of the mercury is about 20 atmospheres and the wall of the quartz tube assumes a temperature of the order of 1000° C.

Besides this mercury lamp, which is not artificially cooled, a water-cooled mercury lamp has been developed, which also works at very high mercury pressures and has very small dimensions, and which allows a considerably greater energy input (W/cm). Water-cooled mercury lamps are provided for various inputs from 200 W to 16 kW. We shall describe chiefly two types with inputs of 500 W and 800 W. These have been given the distinguishing symbols SP and SSP respectively. In *fig. 1* the mercury lamp SP 500 W with holder for water-cooling is reproduced.

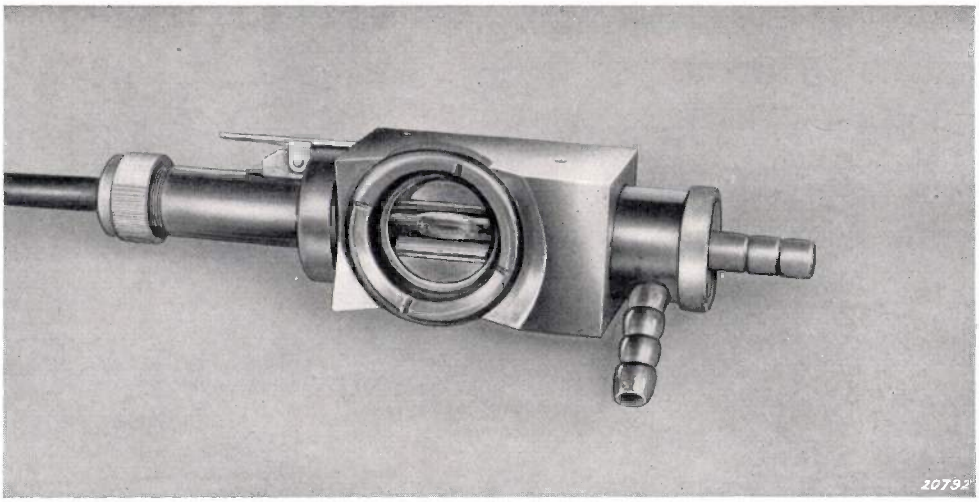


Fig. 1. The mercury lamp type SP 500 W with holder for water-cooling.

Construction of types SP and SSP

Lamps of the SP and SSP types may be used with alternating as well as with direct current. In both cases the data in the table below are valid.

Table I. Data of the super high pressure mercury lamps SP and SSP.

	SP 500 W	SSP 800 W
Length of the discharge	12.5 mm	10 mm
Internal diameter . . .	2 mm	1 mm
External diameter . .	6 mm	3 mm
Mercury pressure . . .	75 atm.	120 atm.
Input	500 W	800 W
Current (with A.C.) .	1.5 A	1.5 A
Current (with D.C.) .	1.3 A	1.3 A
Voltage	420 V	600 V
Light flux	30 000 lm	50 000 lm
Surface brightness (max)	33 000 c/cm²	91 000 c/cm²
Luminous efficiency . .	60 lm/W	62 lm/W

The SP lamp is normally mounted in a metal boat, in which also a mirror may be fastened (*fig. 2*). The whole is slid into a metal block through which cooling water is conducted. The block is

[1]) The mercury lamp HP 300, Philips techn. Rev. 1, 129, 1936.

provided with a circular window through which the radiation is emitted.

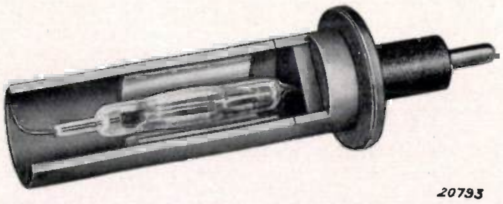


Fig. 2 Water-cooled mercury lamp (type SP 500 W) in its boat.

The SSP lamp may be used in the same holder. Usually, however, three SSP lamps are mounted close to one another on one block in order to obtain a broader light source (*fig. 3*). During the initial experiments with this method of construction it was found that the lamps quickly cracked.
The cause of this lay in the fact that each lamp absorbed a portion of the energy radiated by the other two lamps, so that the load became abnormally great. Since ultra-violet radiation in particular was absorbed, partitions were introduced between the lamps, made of a kind of glass which does not transmit ultra-violet light. In this way cracking was prevented.

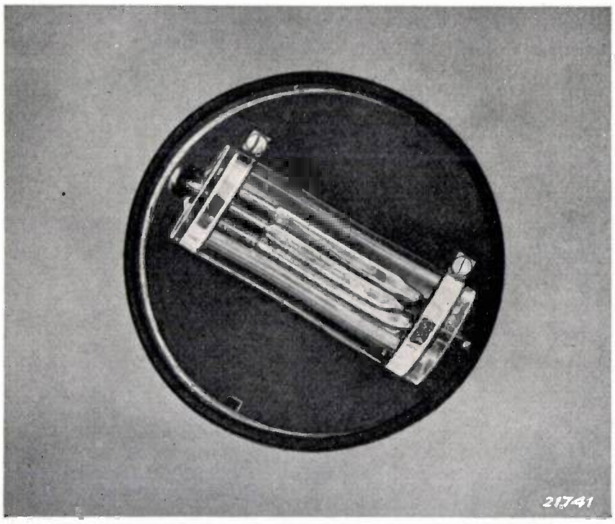


Fig. 3. Lamp holder designed for use in searchlights. Three SSP lamps are mounted on a block of "Philite". Between the lamps there is a glass partition which absorbs ultra-violet radiation. The whole is surrounded by a glass cylinder through which the cooling liquid flows.

Properties

The fact that with the lamps under discussion a large amount of energy is transformed into radiation within a small volume is apparent from the high value of the s u r f a c e b r i g h t n e s s. It may be seen from Table I that the maximum brightness (along the axis of the discharge) is 33 000 candles/cm^2 and 91 000 candles/cm^2 for the two types respectively. For the sake of comparison it may be noted that the brightness of an ordinary carbon arc is about 18 000 candles/cm^2. The brightness of the mercury lamps is not uniform over all parts of the luminescent surface. In *fig. 4* it may be seen that it becomes smaller toward the edges. The discharge thus does not fill the whole tube but is concentrated at the centre.

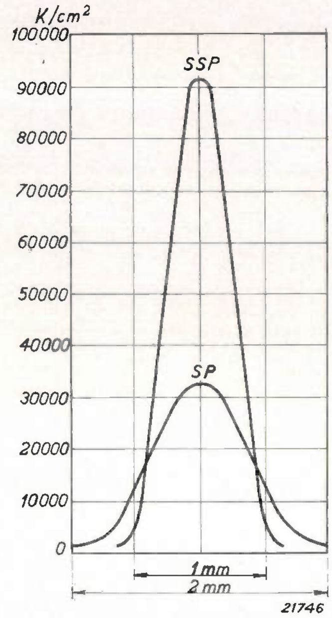


Fig. 4. Curves giving the surface brightness along a cross section of the lamp. The type SP has a maximum brightness of 33 000 candles/cm^2 with an internal diameter of 2 mm; with type SSP these values are 91 000 candles/cm^2 and 1 mm respectively.

The s p e c t r a l c o m p o s i t i o n of the light deviates from that of low pressure mercury tubes. The changes undergone in the spectrum with increasing pressure consist briefly of the following: an increase in intensity of the continuous background and a shift of the intensity maximum toward the long-wave end [2]). These facts are already clearly observable with the mercury lamp HP 300 [3]), with the water-cooled lamps they are observable to a greater degree (*fig. 5*). The modified spectral composition of the light results in considerably better colour reproduction than with ordinary mercury lamps. An idea of this may be obtained from *Table II*, in which a schematic distribution of the light intensity over a number of wavelength regions is given for several light sources.

[2]) Refer to Philips techn. Rev. 1, 2, 1936 for the causes of these phenomena.

[3]) See page 133 of the article cited in footnote 1.

The invisible part of the spectrum here consists mainly of ultra-violet radiation: 46 per cent of all that is radiated lies in the ultra-violet. Further 27 per cent of the light lies in the

Table II. Distribution of the light flux of different sources of light over four wavelength ranges of the visible spectrum in per cent of the total light flux. The regions are chosen so that, with an intensity evenly distributed over all wavelengths, they contribute the same amounts to the total light flux [1]).

Wavelength range in Å	Spectrum of constant intensity	Sun	Glow lamp	Mercury lamps		
				HP 300	SP 500 W	SSP 800 W
4000—5300	25	26	14	8	5	8
5300—5580	25	26	22	58	54	48
5580—5880	25	25	28	31	30	30
5880—7000	25	23	36	3	11	14

[1]) Cf. the table in Vol. 1, page 134 of this periodical.

visible region and 27 per cent of the energy radiated lies in the infra-red. With the high pressures occurring here, a very strong absorption band for ultra-violet radiation appears to the long-wave side of the line 2537 Å.

If one wishes to use the lamp for irradiation with ultra-violet light, the window must be made of quartz, or of a suitable kind of glass which transmits ultra-violet light. Very little ultra-violet light is absorbed by the water, while the opposite is true of the infra-red (heat) radiation. The light of water-cooled mercury lamps is much "colder" than, for instance, that of glow-lamps.

The luminous efficiency of the lamp is high. A new lamp gives about 60 lumens per watt. When the lamp is used the luminous efficiency falls because of the fact that the quartz does not remain completely clear. As a rule the light intensity decreases still more because of the fact that not only the luminous efficiency but also the energy taken up diminishes. This is particularly the case with alternating current supply [4]).

Cooling

The simplest way of cooling is connection with the water supply. The consumption is about 2.5 liters/min. The necessary pressure is 4 lbs per sq. inch. In cases where the water from the main is too hard, where one must economise on water or where the whole must be portable, a circulation system consisting of a pump and a radiator is suitable.

In *fig. 6* may be seen an apparatus for cooling mercury lamps up to a maximum size of 1 kW. The centrifugal pump has a "Philite" shaft bearing.

[4]) The current supplied by the source is smaller the greater the voltage of the lamp. Upon increase of voltage the product of current times voltage will initially, i.e. as long as the voltage is low, rise, and then finally fall again. With the lamp voltage used here we are on the falling branch. As the lamp becomes older, its voltage increases, and thus its input falls. When alternating current is used, the increase in the reignition voltage also plays a part when the lamp becomes older, so that in this case the energy falls more rapidly than when direct current is used. See in this connection: The Mercury lamp HP 300, Philips techn. Rev. 1, 129, 1936; Alternating current circuits for discharge lamps, Philips techn. Rev. 2, 103, 1937.

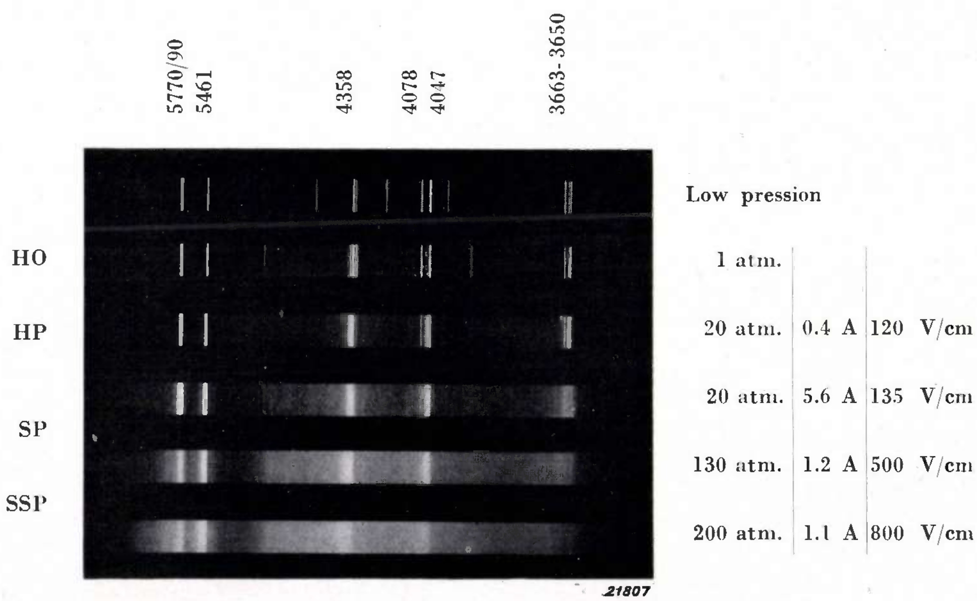


Fig. 5. Spectrum of mercury vapour at various pressures, photographed on a panchromatic plate. With increasing pressure the lines become broader and a continuous background appears. The lines of short wavelength become weaker, those of long wavelength stronger. The accompanying figures indicate, above: wavelength in Å, to the right: technical data of the mercury lamp, to the left: approximate position of the most important types in this series. From W. Elenbaas, Physica 3, 866, 1936.

This is lubricated with water so that contamination of the cooling liquid by oil or grease is out of the question. The capacity is 1.1 liter. If the pump must work in temperatures below the freezing

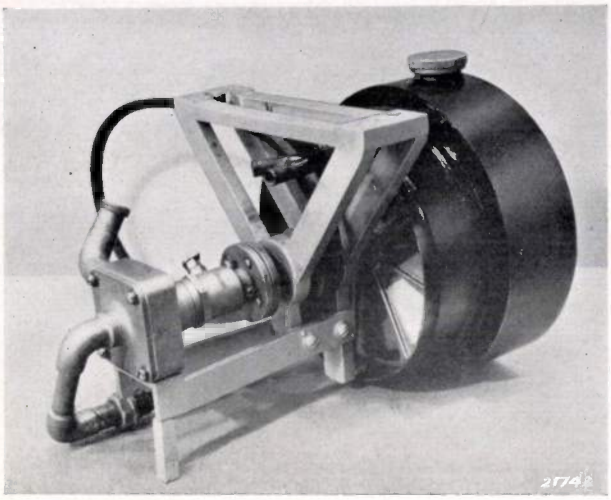


Fig. 6. Cooling unit consisting of centrifugal pump, radiator and fan. The input to the lamp(s) may be a maximum of 2 × 500 W.

point, one may use as cooling liquid a mixture of 2 parts water and one part alcohol. Since contamination of the liquid would cause a deposit on the lamp, it is desirable to use distilled water and to keep the whole system very clean. A good method of preventing the formation of a deposit is to add 0.1 % of sodium phosphate (Na_3PO_4) to the liquid.

In certain cases the noise caused by the pump and fan may be disturbing. In those cases one may use siphon cooling (*fig.* 7), where the heating due

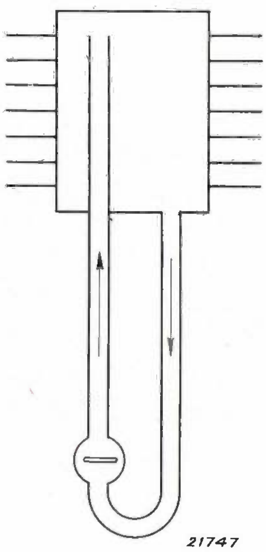


Fig. 7. Diagram of a siphon cooling apparatus. The flow of liquid is maintained by convection currents from the lamp. In the reservoir which is provided with cooling fins the liquid is cooled again.

to the lamp provides for the circulation of the liquid. The liquid cools off again in the reservoir. Because of the low circulation velocity it is impossible to prevent the formation of vapour bubbles in the cooling water. This is accompanied by a gradual decrease of clearness of the quartz tube.

Source of current

As we have already mentioned, the lamps discussed here can be used on alternating current as well as on direct current. The voltage of the supply must be higher than the working voltage of the lamp the difference being taken up by a resistance or, with alternating current supply, by a choke or a transformer with magnetic leakage. Immediately after the lamp has been switched on, while the mercury pressure has not yet reached its final value, the voltage across the lamp is very small, so that practically the whole voltage of the supply is borne by the series resistance or choke. The current then flowing (starting current) should not be too small or the lamp does not warm up at all [5]). A starting current which is too great is also harmful; a good value is 2 to 3 times the normal current.

The leakage flux transformer used with alternating current feed gives a terminal voltage for the SP and SSP lamps of 600 V and 900 V, respectively, with a short circuit current of 2.5 A. At this value of starting current, the lamps light themselves and reach their full intensity within a few seconds. For a detailed explanation of the connections we refer to the description of the mercury lamp HP 300 (footnote 1) and to the article about alternating current circuits for discharge lamps in an earlier number [6]).

Direct current can be obtained from a rectifier or a converter. With direct current one must use a series resistance and the losses are much greater than is the case with a transformer having leakage inductance. For the lamps SP and SSP one must have a source of current with no-load voltage of 500 V and 750 V respectively. As may be seen from Table I, when the SP lamp is fed with direct current, 80 V are taken up by the resistance. Since the working current is 1.3 A, the resistance must be $80/1.3 = 61.5$ ohms. If the supply were connected directly to the lamp and this resistance the starting current would be almost $500/61.5 = 8.1$ A. Such a high current would destroy the electrodes. In

[5]) For a detailed explanation see Philips techn. Rev. **1**, 129, 1936.

[6]) Philips techn. Rev. **2**, 103, 1937.

order to limit the starting current to a reasonable value, 4 A for example, the resistance upon closing the circuit must be 125 ohms. Only after ignition may it be brought to its final value.

In order to keep the losses in the series resistance low, it is desirable to keep the no-load voltage as small as possible. This may cause the ignition to be less reliable than is the case with alternating current. A circuit in which this disadvantage has been overcome may be seen in *fig. 8.* When the

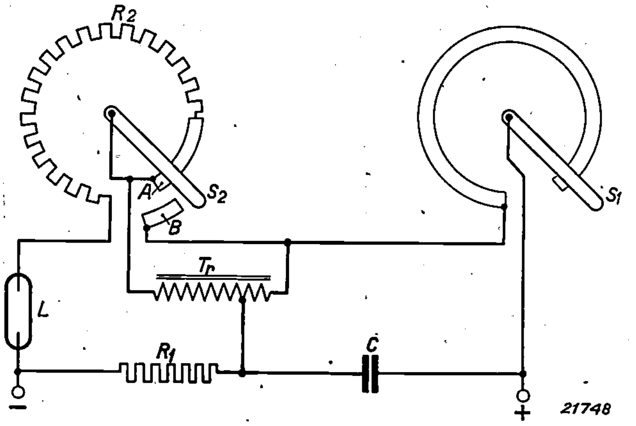


Fig. 8. Circuit for a mercury lamp with direct current. Both switches S_1 and S_2 are mounted on one spindle. S_1, however, has a certain play, so that upon moving it forward as well as backward, it remains behind S_2 by a certain angle. Upon turning the switches forward the contact between A and B is first broken and then the contact S_1 is closed, thereby connecting the lamp to the supply. At the same time the condenser C, which was previously charged, now discharges through a section of the transformer Tr. The voltage surge thus occurring causes the lamp to light. By turning farther, the resistance R_2 is decreased until the lamp takes up the correct amount of energy. Upon switching off, the resistance is first increased again. S_1, now remains so far behind S_2, that the contact between A and B is made before the circuit is broken with S_1. The inductive winding of Tr is thereby short-circuited so that upon switching off no high tension surges can occur. Upon switching on, however, A - B is again first opened.

switch S_2 is rotated the contact A - B is first broken, and then the direct current voltage is switched on by the switch S_1, which is coupled with S_2. At the same time the condenser C, which was charged previously, is discharged through a section of the transformer Tr, so that the voltage between the terminals of the lamp is temporarily raised and the lamp lights. When the knob is turned farther, the resistance R_2 is diminished gradually to the correct value. Upon switching off, R_2 is first again increased. Just before the contact S_1 is broken, the transformer is short circuited by means of the contacts A - B of the switch S_2. The occurrence of high inductive voltages upon breaking the circuit is thereby prevented.

Possibilities for application

In many cases where at present arc lamps or large glow-lamps are used, one may to advantage substitute SP or SSP lamps. It is true that the necessity of water-cooling may be considered a disadvantage, but to offset this there are great advantages, such as: small current consumption, slight heat radiation, small dimensions, absence of smoke or dirt, little upkeep, etc., while one may often be able to make use of their more special characteristics such as shape of the light-source, colour, smallness of the discharge inertia, etc.

It is still an open question which application will prove to be the most important in future. The following description of some applications, of which we already have experience, may be interesting.

Film studio lighting

Due to the high efficiency one can obtain an illumination by means of mercury lamps which is equivalent to that of existing arc lamps, but taking only about 1/3 of the power. The illumination is equivalent in the sense that it gives the same number of lux. In photography it is the actinic effect, and not the visual quantity, the lux, which is the standard. This actinic effect on panchromatic material is for mercury light about twice as high as for arc-lamp light of the same visual brightness, so that "equivalent" in the photographic sense means that one can work with 1/6 of the original energy. As is known a panchromatic emulsion has a rather low sensitivity to green light, as regards its colour reproduction in photography. The light of the water-cooled mercury lamp contains a large amount of green, so that in photography with mercury light the various colours are reproduced in a good natural relation. The reproduction of red is also satisfactory.

A very important advantage of the mercury lamp is the small amount of infra-red radiation. In order to give some impression of this, measurements were made of the increase in temperature undergone by the human skin upon irradiation, with mercury and incandescent lamps respectively, at different light intensities (*fig. 9*). The results depend of course very much on the characteristics of the skin of the person in question and on the size of the surface irradiated, and may only be considered as an illustration. In the case mentioned 4.5 times as much mercury light as incandescent lamp light could be sustained for the same increase of temperature.

Television

The SP lamp has proved suitable for fulfilling two functions in the Philips television transmitter.

Only after the installation of mercury lamps in the studio was it found possible to have sufficient illumination without the actors being too much inconvenienced by the heat. The weakness of the

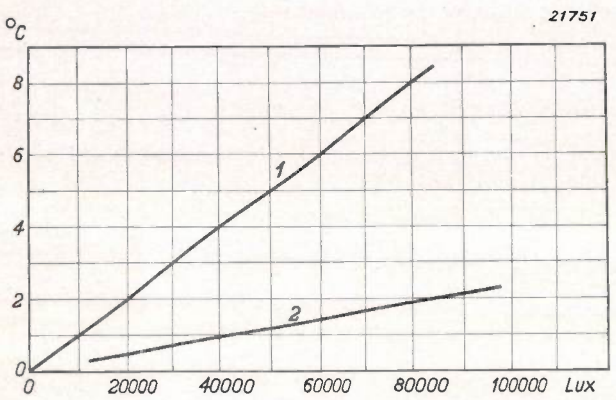


Fig. 9. Increase in temperature of a portion of the skin (lower arm) upon irradiation with light from incandescent lamps (*1*) and water-cooled mercury lamps (*2*) respectively. For equal increases in temperature $4^1/_2$ times as much mercury light can be employed as incandescent lamp light.

infra-red radiation was found to offer still another advantage. The iconoscope used for the broadcasting is very sensitive to infra-red. This is by itself no disadvantage, but the lens used was not chromatically corrected for infra-red, and therefore when mercury light was used sharper pictures were obtained. The installation contains 10 mercury lamps of 500 W and 3 of 3 kW.

Water-cooled mercury lamps were also used for scanning films in transmitters with a Nipkow disc. As explained in a previous article [7]) about this installation, the linear form and the great brightness are advantages which are lacking with other light-sources.

Searchlights

The first requirement of the light-source of searchlights is a high surface brightness. The SSP lamp, which has a surface brightness of 91000 candles/cm², is clearly suitable. By increasing the loading the surface brightness may be considerably raised. For example, with an energy of 1400 W/cm a surface brightness of 160 000 candles/cm² can be attained. The increase in energy is of course accompanied by a decrease in life. The SSP lamp offers a compromise in this respect and has a life of about 25 hours.

A difficulty in the application to searchlights is presented by the linear form of the mercury lamps. It is desirable to place the light-source in the reflector in such a way that all the light is radiated as close as possible to the focus. The extremities of the lamp give rise to strongly diverging rays, and contribute little to the intensity of the searchlight. It therefore serves no useful purpose to increase the energy by making the light-source longer. One may, however, attain greater intensity by placing several lamps next to each other, so that the light-source is made broader. A suitable holder is shown in Fig. 3. It contains three lamps with a total energy of 2.5 kW.

Air-port lighting

A very good adaptation of the reflector to the linear form of the light-source is obtained with cylindrical parabolic mirrors. These mirrors have a focal line, and by allowing the path of the discharge to coincide with the focal line a flat disc-shaped beam with an angle of divergence of 180° is obtained which is particularly suitable for the lighting of landing fields.

In this case, the length of the light source is not limited as with circularly symmetrical mirrors, so that longer tubes with higher energy may be used. The largest projector constructed in this way has a lamp 30 cm in length, with an input of 16 kW. It has, however been found that with much less energy very good results may also be obtained. With the 2.5 kW system shown in *fig. 10* completely satisfactory illumination was obtained at "Welschap", the Eindhoven air-port. By setting up such an apparatus on all four sides of the field,

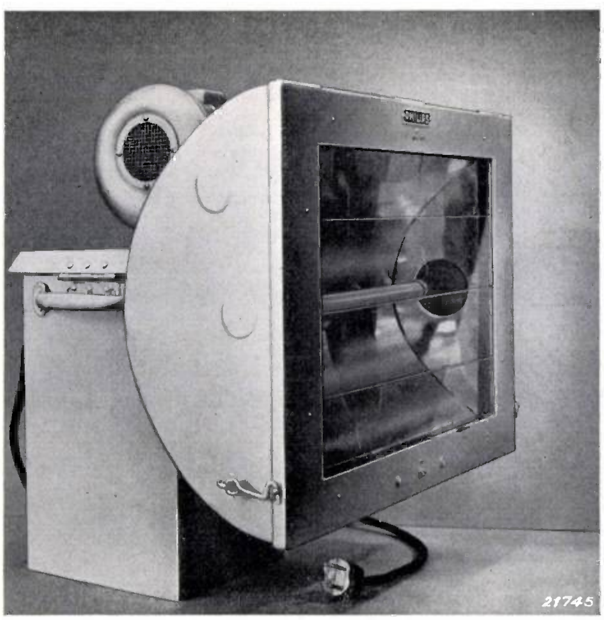


Fig. 10. Apparatus for the lighting of landing fields. The cylindrical parabolic mirror is built on to the transformer housing. Input 2.5 kW.

[7]) H. Rinia and C. Dorsman, Philips techn. Rev. 2, 72, 1937.

the direction of radiation can be adapted to the landing course for every wind direction.

Grass reflects mercury light particularly well, this is a special advantage of the unusual intensity of the green mercury line.

This concludes our account of some applications of water-cooled mercury lamps: there are undoubtedly many more. We hope, however, that we have been able to give an impression of the possibilities offered by this new light source.

AN ULTRA SHORT WAVE TELEPHONE LINK BETWEEN EINDHOVEN AND TILBURG

by C. G. A. VON LINDERN and G. DE VRIES.

Summary. A telephone connection between Eindhoven and Tilburg is described, in which use is made of waves of about one metre in length. Triode transmitters and "autodyne-superhet" receivers are employed. Directional aerials of the Y a g i type with an amplification of 3.5 are used. The field strength was recorded during the three months that the installation has been in use.

Introduction

The results of experiments carried out in the Philips laboratory on waves of about 1 metre in

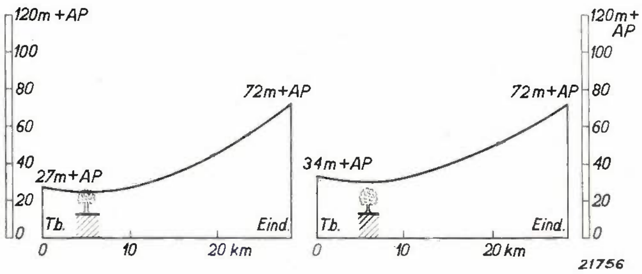


Fig. 1. The curved line represents the line of direct vision between the transmitter and the receiver in the connection between Eindhoven and Tilburg. The earth is here drawn flat, as x-axis, and represents sea-level. On the left an aerial of only 2 meters was erected on the roof of the factory building in Tilburg, while in the right-hand figure it was 9 metres high.

length led to the attempt at a practical application in the domain of telephone communication. Thanks to the collaboration of the Dutch National Post, Telephone and Telegraph Service permission was given us to carry out an experimental wireless connection between Eindhoven and Tilburg.

The first requirement of a reliable radio connection by means of ultra short waves is that the receiving aerial must be visible from the point where the sending aerial is placed. By "visible" in this case is meant that the line joining transmitting and receiving dipoles must be at least 10 metres above the tops of trees and buildings. In *fig. 1* the x-axis represents sea-level. The distance from Tilburg is set off horizontally, and the height above sea level vertically. In Eindhoven the aerial is erected at a height of 72 m on the roof tower of one of the Philips factories.

In the left-hand figure is represented the situation when the aerial in Tilburg was only 2 m above the roof of a factory building there. The curved line, which represents the line of direct vision, is found to pass through the tops of trees between the two towns. As a matter of fact very little was received on these ultra short waves with such a short aerial.

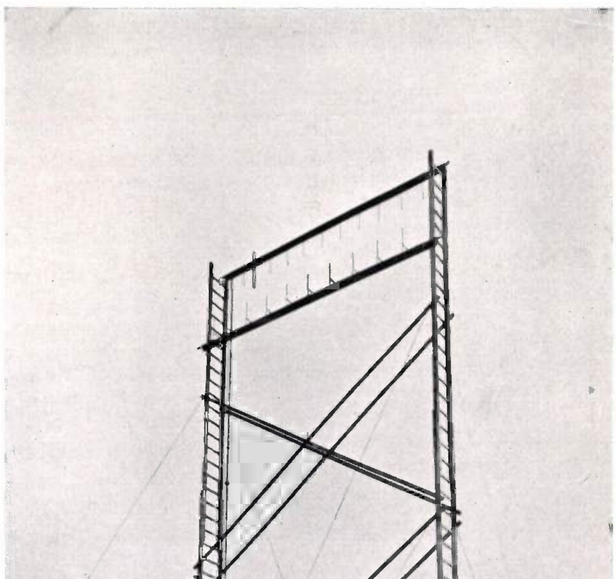


Fig. 2. The aerial arrangement in Tilburg. The aerials are mounted between two ladders; the receiving aerial is 13 m above the roof, and the transmitting aerial 15 m.

the direction of radiation can be adapted to the landing course for every wind direction.

Grass reflects mercury light particularly well, this is a special advantage of the unusual intensity of the green mercury line.

This concludes our account of some applications of water-cooled mercury lamps: there are undoubtedly many more. We hope, however, that we have been able to give an impression of the possibilities offered by this new light source.

AN ULTRA SHORT WAVE TELEPHONE LINK BETWEEN EINDHOVEN AND TILBURG

by C. G. A. VON LINDERN and G. DE VRIES.

Summary. A telephone connection between Eindhoven and Tilburg is described, in which use is made of waves of about one metre in length. Triode transmitters and "autodyne-superhet" receivers are employed. Directional aerials of the Y a g i type with an amplification of 3.5 are used. The field strength was recorded during the three months that the installation has been in use.

Introduction

The results of experiments carried out in the Philips laboratory on waves of about 1 metre in length led to the attempt at a practical application in the domain of telephone communication. Thanks to the collaboration of the Dutch National Post, Telephone and Telegraph Service permission was given us to carry out an experimental wireless connection between Eindhoven and Tilburg.

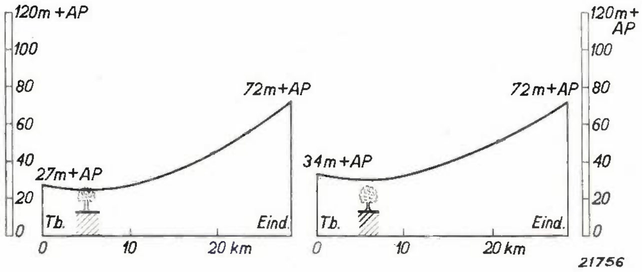


Fig. 1. The curved line represents the line of direct vision between the transmitter and the receiver in the connection between Eindhoven and Tilburg. The earth is here drawn flat, as x-axis, and represents sea-level. On the left an aerial of only 2 meters was erected on the roof of the factory building in Tilburg, while in the right-hand figure it was 9 metres high.

The first requirement of a reliable radio connection by means of ultra short waves is that the receiving aerial must be visible from the point where the sending aerial is placed. By "visible" in this case is meant that the line joining transmitting and receiving dipoles must be at least 10 metres above the tops of trees and buildings. In *fig. 1* the x-axis represents sea-level. The distance from Tilburg is set off horizontally, and the height above sea level vertically. In Eindhoven the aerial is erected at a height of 72 m on the roof tower of one of the Philips factories.

In the left-hand figure is represented the situation when the aerial in Tilburg was only 2 m above the roof of a factory building there. The curved line, which represents the line of direct vision, is found to pass through the tops of trees between the two towns. As a matter of fact very little was received on these ultra short waves with such a short aerial.

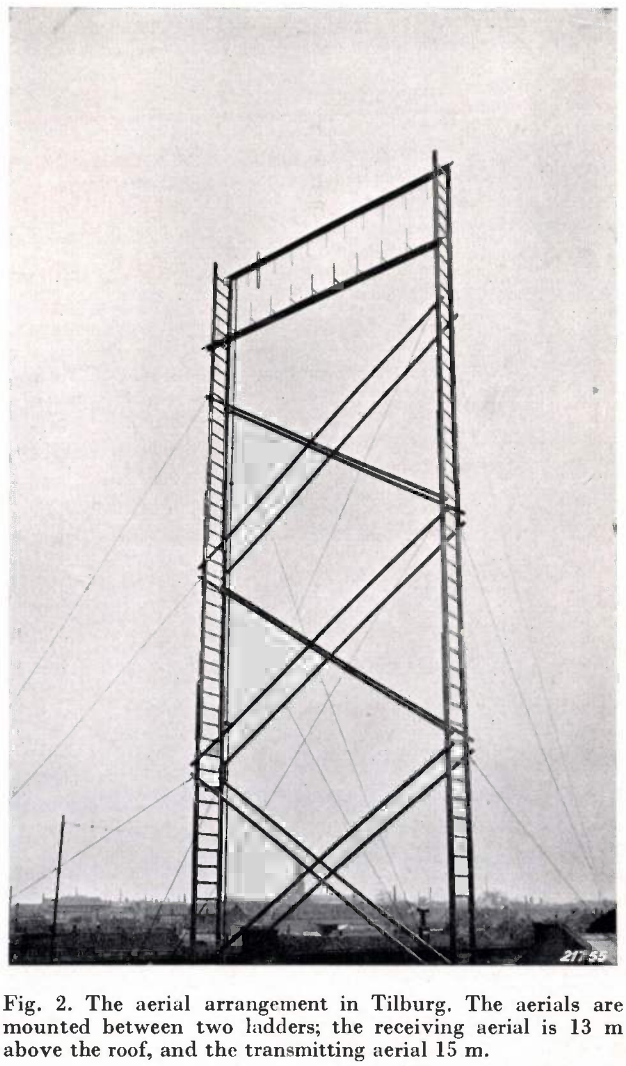


Fig. 2. The aerial arrangement in Tilburg. The aerials are mounted between two ladders; the receiving aerial is 13 m above the roof, and the transmitting aerial 15 m.

In the right-hand figure the situation is sketched when the dipole in Tilburg was erected on a pole 9 m in height. The line of direct vision was then everywhere at least several metres above the treetops and reception was satisfactory. With relatively little extra cost the aerial poles could be made still longer. The final aerial, shown in *fig. 2*, is 15 m high. The transmitting aerial is at the very top, while the receiving aerial is 2 m below.

In order to have two equivalent circuits the receiving aerial in Eindhoven was also mounted 2 m below the transmitting aerial. In order to avoid mutual disturbances the transmitters in Eindhoven and Tilburg work on wavelengths of 140 and 123 cm respectively.

The aerials

In Eindhoven as well as in Tilburg so-called Yagi aerials were used for transmitter and receiver. The Yagi aerial consists of a number of parallel rods set up in one plane, which are coupled by the

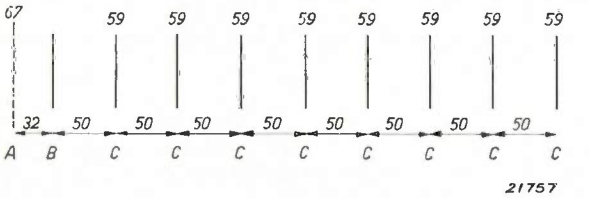


Fig. 3. Sketch with dimensions of a Yagi aerial for a wavelength of 136 cm; *B* is the rod connected with the receiver (or transmitter). The rods are supported in the middle. The numbers above the rods indicate their length in cm, while the other numbers represent their mutual separation.

electromagnetic field with each other and with the dipole which is connected to the transmitter (or receiver). The length of the rods and their mutual distances are determined experimentally so that the maximum amplification is obtained in the direction in which the rods are situated one behind the other. *Fig. 3* gives an example for a

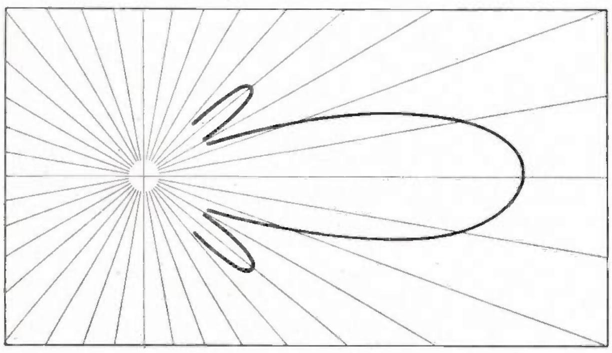


Fig. 4. Polar diagram for the Yagi aerial of Fig. 3.

wavelength of 136 cm. The action consists briefly of this, that the currents coming from the correct direction induce such tensions in the various rods that the components caused thereby in the rod connected to the receiver are added. If a reflector rod also is introduced, such as is indicated by the dotted line in fig. 3. one obtains as result that the signal which now reaches the receiver has an amplitude greater by a factor of 3.5 than if a single dipole were used for receiving. In the experiments for determining this amplification, for a Yagi aerial, use was made of a thermocouple as indicator in the dipole. By introducing and again removing the rods which make the aerial directional, the abovementioned factor was measured in a simple manner. In the same way the transmitting aerial was also investigated; by measurement of field strength at a great distance it was also ascertained that the amplification is 3.5.

In addition a study was made of the way in which the field strength decreased with the distance. For these measurements the sensitivity of the receiver had to be calibrated, which was done by means of a signal of known strength. For this a generator was used with an aerial whose radiating part is small with respect to the wavelength, while a horizontal plate as condenser at the top of the dipole provides for a uniform distribution of current over the vertical part. The current in the vertical part can be determined from the capacity, and the potential on the condenser measured with an acorn diode voltmeter, and thus the field

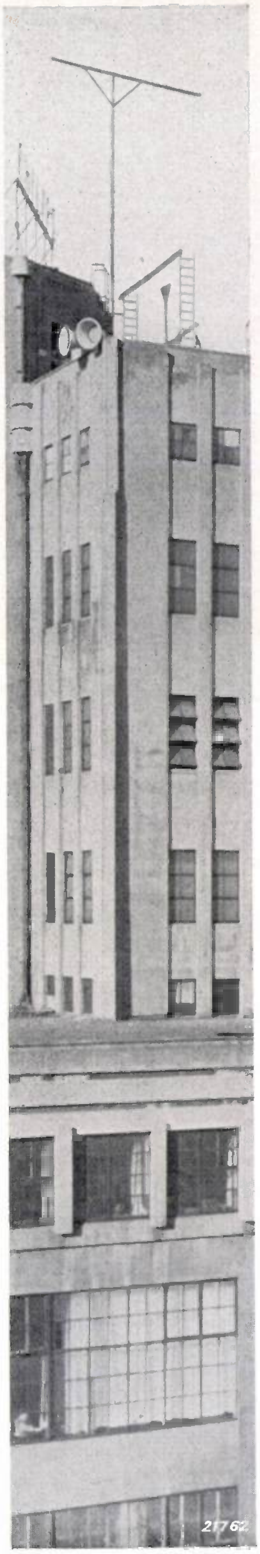


Fig. 5. The high aerial in this photograph shows the experimental Yagi aerial referred to under Figs. 3 and 4, which may be turned about a vertical axis. In addition, lower and to the right, may be seen the aerial arrangement for the connection with Tilburg.

radiated is known [1]). The result of the measurement with the receiver calibrated in this way was that the field given by the Yagi aerial at different distances corresponds satisfactorily with what one would roughly expect on a basis of the available energy.

In *fig. 4* the polar diagram is given for the experimental transmitting aerial which was used in Eindhoven. This aerial can be turned about a vertical axis and is shown in *fig. 5*, in which also the final aerial for the connection with Tilburg may be seen. This latter is shown separately in *fig. 6*.

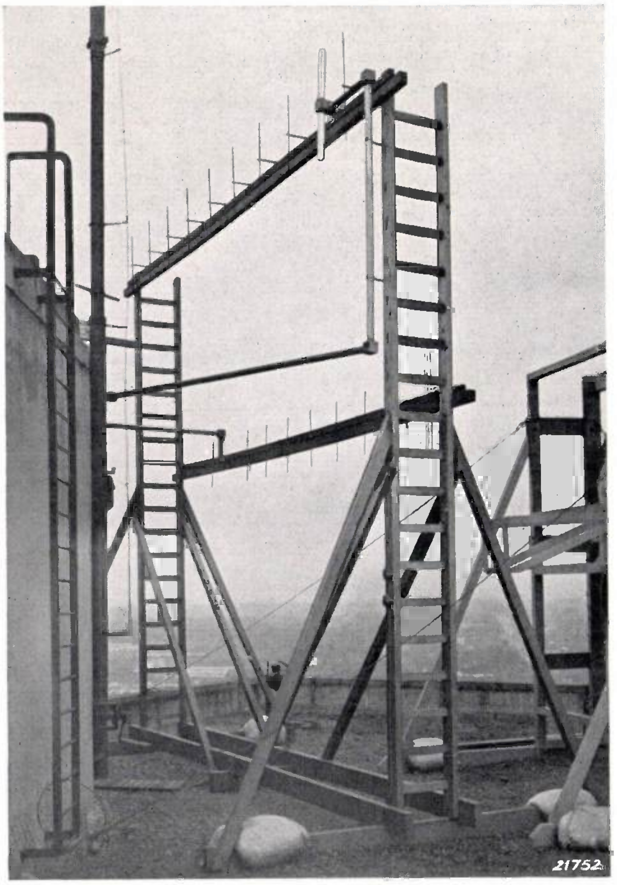


Fig. 6. Larger scale photograph of the aerial arrangement for the connection with Tilburg.

The feeders

Since in these experiments we are concerned with very short waves, the self-inductance and capacity of the connections between transmitter and aerial play a part. In the first experiments the feeders were two parallel wires which were held at the correct distance apart by "Micalex" distance-pieces every 1.5 m. In such feeders of about 20 m in length the loss is 50 per cent at the highest. The feeder transmission is completely disturbed by a glaze of ice or melting snow, while in damp weather considerable losses occur. We therefore shielded the feeders and surrounded the dipole with a glass tube, while arrangements were made to warm the whole system. With these precautions the difficulties were found to be overcome.

The transmitter

The transmitter uses triodes which are so constructed that they can generate short waves of about one metre in length. *Fig. 7* shows such a valve in which grid and anode terminals are at the

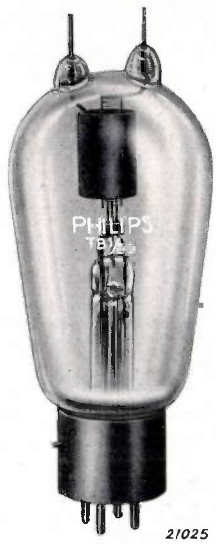


Fig. 7. The transmitting tube TB 1/60 used with this apparatus

top, the electrodes having only the glass wall as support and insulator. The filament terminals are in the usual position. The dimensions are such

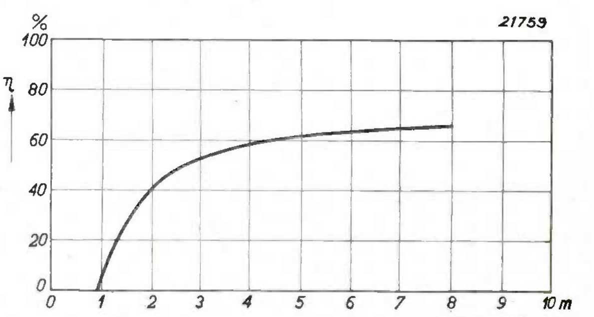


Fig. 8. Efficiency η of the transmitting valve TB 1/60 as a function of the wavelength λ.

that the damping due to the transit time effect [2]) is not yet serious at these frequencies.

In order to obtain a transmitter which can be easily adjusted, a push-pull circuit was used in

---

[1]) B. Trevor and R. W. George, Notes on propagation at a wavelength of seventy-three centimeters, Proc. Inst. Radio Eng. **23,** 461 - 469, 1935.

[2]) C. J. Bakker and G. de Vries, Vacuum tube electronics, Physica **2,** 683 - 697, 1935.
C. J. Bakker, Several characteristics of receiving valves in short-wave reception, Philips techn. Rev. **1,** 171 - 178, 1936.

which the input leads for the filament current are made so that the system is tuned. The transmitter can thereby be made to work at various wavelengths in a simple way, and always with the highest possible efficiency. The measurements, the results of which are reproduced in *fig. 8*, show how the efficiency depends upon the wavelength: with the wavelengths used by us the efficiency is about 20 per cent.

Modulation is achieved by varying the anode potential. The modulator is constructed as a push-pull amplifier and has six F 410 valves. In this method of modulation the frequency fluctuates. In order to keep this fluctuation sufficiently small, an *LC* circuit with low losses is used. Just as the frequency is kept constant in longer wave transmitters by a piezo-electric crystal, this is accomplished for these ultra short waves by the *LC* circuit, which is much cheaper and sturdier. This *LC* circuit is made of copper, and constructed as a solid of revolution; a cross section is shown in *fig. 9*. It is clear that the magnetic lines of force

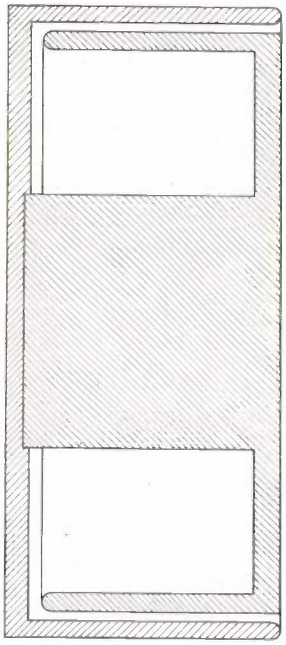


Fig. 9. Sketch of the loss-free *LC* circuit used for stabilization of the frequency.

pass through the space between the core and the surface of the outer tube; the self-inductance L_{cm} in cm will be about $2\, l \ln R_1/R_0$, where l is the length of the axis in cm, while $2\, R_0$ and $2\, R_1$ are the diameters of the core and the outer tube respectively. The resonance wavelength then follows from the formula

$$\lambda_{cm} = 2\pi \sqrt{L_{cm}\, C_{cm}}\,.$$

Since no insulation material is used in the construction the losses are very small. With such short waves the radiation losses of such an *LC* circuit play an important part, but they are kept sufficiently small in the construction shown in *fig. 9* since the radiation here takes place only through a narrow slit. In this respect therefore the circuit quality can be very good. The result is, in fact, that with the correct coupling of this circuit to the transmitter the changes in frequency consequent upon full modulation remain small.

The receiving system

For the reception of very short waves one has a choice of various receiving systems. Apparatus with direct high-frequency amplification, super-regenerative receivers as well as superheterodyne and autodyne superhet receivers may be used. The desire was that the receivers should be able to be fed with alternating current exclusively, that they should need no attention, should as far as possible accomodate themselves to the peculiarities of ultra short wave transmitters and should operate with ordinarily available valves. Furthermore it was desirable to be able to introduce such modern im-

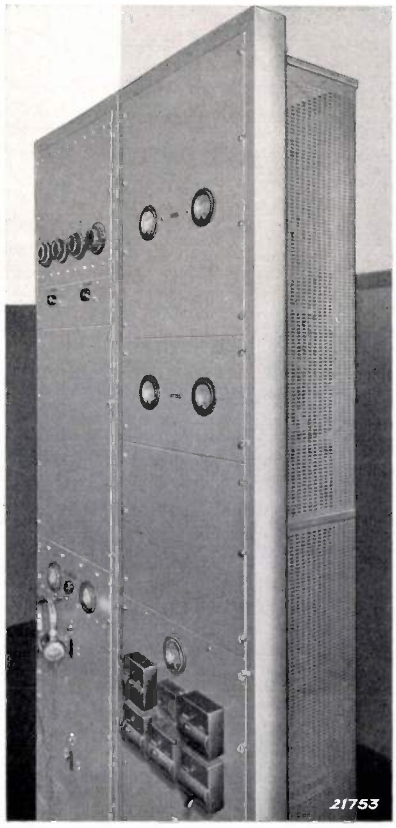


Fig. 10. Apparatus for the connection set up in Tilburg. The left-hand panel contains two receivers, one of which acts as reserve. The four tuning knobs, which may be seen near the top of this panel, are the control for the four oscillator-modulators used. The right-hand section is the transmitter.

provements already in use at longer wavelengths, as visible tuning, automatic volume control, etc.

In consideration of the above autodyne superhet receivers were chosen. In this case, as is at present usual in receivers for longer waves, there are no separate oscillator valves or separate modulator valves. There are not yet, however, combined oscillator-modulator valves available for such short wavelengths counterparts of the well known hexode and octode "mixing valves", for longer waves. We used "acorn" triodes as mixing valves. These are triodes with very small dimensions, so that the transit times of the electrons remain sufficiently small. These "acorn" valves also have much smaller capacities than the ordinary valves and it is consequently possible to generate considerably shorter waves with them than those used here. The left-hand panel in *fig. 10* shows the receiver which was used.

The receiving aerial is also of the Yagi type and is connected to the receiver with a double wire feeder. As shown in *fig. 11* this double wire feeder

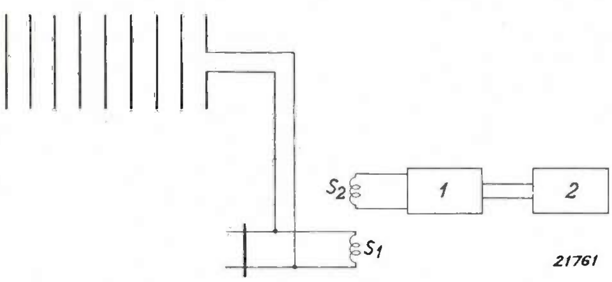


Fig. 11. Diagram of the oscillator-modulator with double wire feeders for coupling to the aerial. The oscillator-modulator is indicated by *1*, while *2* represents the intermediate frequency amplifier, detector and low frequency amplifier.

is connected with another double wire feeder (Lecher system) about a half wavelength long and ending in a coil S_1, which is coupled with S_2 to which the oscillator-modulator *1* is joined. *Fig. 12* is a photograph of the oscillator-modulator with the wires of the Lecher system wound in a helix. At the top may be seen the bridge with which the natural frequency of the Lecher system is tuned to the frequency received. Adjustment of the receiver may be made by:

1) changing the length of the Lecher system, which is about half a wavelength, by means of the bridge,
2) changing the place at which the double wire feeder, coming from the aerial, is connected,
3) changing the coupling between the coils S_1 and S_2.

The formation of a glaze of ice on the double wire feeders gives rise to difficulties with receivers, but to a smaller extent than with transmitters. The receiver feeders are therefore also entirely protected.

The intermediate frequency amplifier

The intermediate frequency amplifier differs from those in use for broadcasting waves mainly in the greater width of its frequency range. Neither the frequency of the transmitter nor that of the oscillator is absolutely constant. If the beat frequency is to remain within the same limits as in the reception of longer waves, the absolute variations may also not be greater than those for longer waves, and the relative variations must therefore be much smaller. Even with ordinary short waves

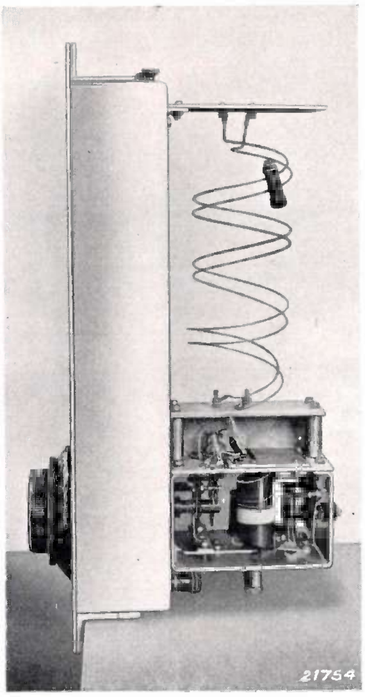


Fig. 12. Photograph of the oscillator-modulator.

(15 m) one must already take certain measures in order to keep the beat frequency sufficiently constant. Assuming that it is possible to obtain the same relative constancy of frequency on our shorter waves, then the beat frequency will fluctuate 10 to 15 times more than for the 15 m wave. This already requires an appreciably wider band than with an ordinary receiver. Moreover — and this is much more serious — the wavelength of the transmitter is much less constant than that of the broadcasting cristal - controlled transmitters. In this connection the intermediate-frequency amplifiers are made to have a bandwidth of 400 kc and are tuned to a wavelength of about 40 m. This

does not offer special difficulties: for television purposes intermediate frequency amplifiers with still greater bandwidth are made. If necessary, automatic volume control can be introduced in the usual way at the grid of the first intermediate frequency valve.

A record taken over a period of 24 hours is shown in *fig. 13*. It is sufficient to retune the receiver once per 24 hours.

Mutual disturbance

It was found possible from the beginning to set up the transmitter and the receiver next to each other. The mutual disturbance which might occur could only result in the production of faint "sidetone" and this leads to the speaker hearing his own speech, as on many telephones. One notices practically nothing of this phenomenon. If it were desired to connect a telephone system to the apparatus, push-pull circuits should be used, which by their own imperfection would cause greater disturbances than those we have just mentioned.

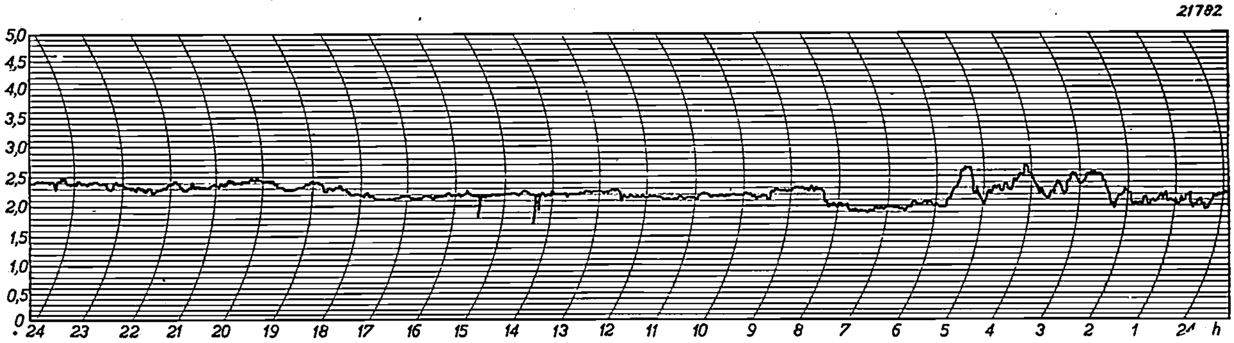


Fig. 13. Record made in Eindhoven of the intensity received there from the transmitter in Tilburg.

Recording of the carrier wave

The intensity of the carrier wave received in Eindhoven from Tilburg is recorded in the following way. The direct current supplied by the detector in the receiver flows not only through the coupling resistance for the last valve, but also through a second resistance. The voltage drop over this second resistance, after amplification by means of a direct current amplifier with two pentodes AL 4 in a bridge connection, supplies enough energy to feed a recording milliammeter.

THE MEASUREMENT OF THE HARDNESS OF METALS AND ALLOYS

by E. M. H. LIPS.

Summary. A survey is given of the commonest methods of measuring the hardness of metals and alloys. In particular a micro hardness meter is described, with which the hardness of structural components can be measured.

Introduction

The hardness of a solid substance was considered by the most ancient peoples to be an important property and was connected more or less with the idea of workability. A mineral, for instance, was called hard when it was impossible, or only with difficulty possible, to shape it by means of scraping, grinding or other similar processes.

It was therefore only natural in indicating the hardness of a substance to compare it with another substance of which the degree of difficulty of its "working" was known by experience.

Expressions like "as soft as butter" and "as hard as stone" show that this conception still has a certain survival at the present time.

It was a long time, however, before people began to measure the hardness of solid substances according to a systematically worked-out plan.

After experiments by Barba in 1640 and Réaumur in 1722, Mohs in 1822 first set up a scale of 10 known minerals which showed a relatively increasing hardness. The hardness of a solid substance expressed by one of the numbers from 10 to 1 was determined, beginning with diamond, by scratching the unknown substance with the successive standard minerals, until one of the standard minerals was itself scratched by the substance to be examined.

The desired hardness then lies between those of the last two minerals used.

The hardness scale of Mohs (*Table I*) has served and still serves its purpose very well in mineralogy.

Table I.

Mineral	Scale Number
Talc	1
Gypsum	2
Calc-spar	3
Fluor-spar	4
Apatite	5
Orthoclase	6
Quartz	7
Topaz	8
Corundum	9
Diamond	10

The hardness scale of Mohs was no longer adequate for the metal industry arising in the second half of the previous century. It was too inexact; the differentiation was too slight.

Brinell's method

Using as a basis the investigations of Hertz and Auerbach in particular, Brinell now developed the ball pressure method which was discussed before the "Congrès International des Métaux d'Essai" in 1900.

The great value of Brinell's investigations lay chiefly in the fact that the definition of hardness as "the capacity of being scratched"[1]) was relinquished, and instead was taken the degree of denting which is undergone by the material examined upon being pressed by another material.

Brinell's test, which even now is often carried out, consists substantially of the following:

A hardened steel ball is pressed with a given pressure against a ground surface of the material being examined. The load is then taken off and the ball removed. In most cases the ball leaves an impression in the form of a spherical segment, as is shown in *fig. 1*. In this figure *I* is the ball by means

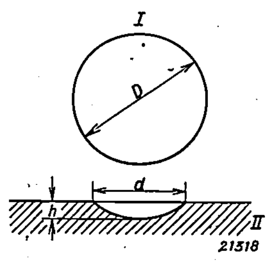


Fig. 1. Concavity in the form of a spherical segment obtained by Brinell's test.

of which the spherical segment is impressed in the material, *II* the hardness or the Brinell value is

[1]) This definition of hardness (which as will appear later has been given up) forms the basis of the hardness meter of Martens in which a cone-shaped diamond with a vertex angle of 90° is drawn over a polished surface of the object to be measured under a load of 20 grams. The width of the resulting scratch, which is measured under the microscope, is inversely proportional to the hardness.

obtained by dividing the pressure by the area of the surface of the spherical segment, or

$$H = \frac{P}{\pi D h}, \quad \ldots \ldots \quad (1)$$

where H is the Brinell hardness in kg/mm^2,
P is the pressure applied in kg,
D is the diameter of the ball in mm, and
h is the height of the spherical segment.

For practical reasons it is easier to measure d the diameter of the spherical segment instead of the height h. The formula then becomes:

$$H = \frac{2P}{\pi D(D-\sqrt{D^2-d^2})} \quad \ldots \quad (2)$$

In order to secure reproducible measurements it is necessary that the pressure P be adapted to the hardness of the material to be examined, while the time during which the ball is pressed against the material is also important. Various considerations of this kind have led to standardization of the diameter of the ball with the corresponding pressure and time of loading. The standardized values are given in *Table II*.

Table II

Thickness of testpiece in mm	Diameter of the ball in mm	Force in kg for: steel & cast iron	Brass, bronze, etc.	aluminium white metal, etc.
>6	10	3000	3000	250
6—3	5	750	250	62.5
<3	2.5	187.5	62.5	15.6

A Brinell hardness of 200 for instance, which is measured with a ball of 10 mm in diameter, under a pressure of 3000 kg, which pressure is kept constant for 30 sec. is expressed briefly as follows:

$$H.10/3000/30 = 200 \quad \ldots \ldots \quad (3)$$

Several advantages of Brinell's test are:

1) It is easy to carry out,
2) An impression is obtained which can serve as a permanent record; that is to say, the impression can always be re-measured,
3) The impressions made are so large that it is also possible to measure non-homogeneous materials such as cast iron,

The following are disadvantages:

1) The impression of the ball is not always of the same shape for all materials, so that the calculation formula (2) holds only within fairly narrow limits; one may therefore compare values found with the same material only when they were obtained under exactly the same conditions [2]).
2) Materials whose hardness approaches that of the hardened steel ball, cannot be tested by the Brinell method. In such a case the ball itself would be flattened, whereby a (no longer spherical) impression would occur which is too large, and gives too low a value for the hardness upon calculation.
3) The impression of the ball sometimes represents a non-permissible damaging of the surface.

Because of the flattening of the steel ball the measuring range of Brinell's test is confined between 0 and the hardness of hardened steel. The hardnesses of tool steels, for instance, which in the state in which they are used possess a hardness in the neighbourhood of hardened steel, cannot therfore ordinarily be measured by the Brinell method.

Only in recent years has this disadvantage been partially overcome by taking, instead of the hardened steel ball, a ball of a special alloy such as "Widia", which has a considerably greater hardness than hardened steel. A high degree of inaccuracy however is caused by the slightness of the depth of the impression.

Rockwell's hardness meter

The hardness meter of Rockwell which came on the market in 1920, and with which particularly hard steels can be measured, has found extensive application.

Rockwell's method of measurement is as follows (see *Fig. 2*):

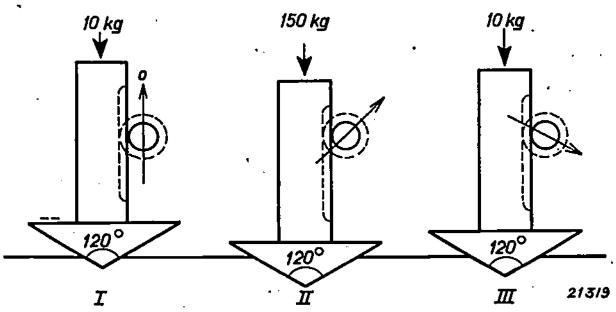


Fig. 2. Diagram of the course of a measurement by Rockwell's method.

A cone-shaped diamond with a vertex angle of 120° is placed on the material to be tested with its point against the material, and is then loaded with

[2]) If for instance the hardnesses H 10/3000/30 and H 5/750/30 of cast iron are determined they will be found to be appreciably different.

a pressure of 10 kg. The indicator, which is joined by a transmission to the diamond, now assumes a definite position. The movable dial, around which the indicator moves, is now set at 0, and the load is increased to 150 kg; after the indicator has become stationary 140 kg are removed. The indicator then assumes a definite final position, which is read off from the calibrated scale on the dial as the Rockwell value. The Rockwell value is thus the difference between the depths *I* and *III* of penetration.

For soft materials the diamond may be changed for a ball with a diameter of 1.59 mm (1/16″). Instead of 150 kg a pressure of 100 kg is then applied. The measurement proceeds otherwise in the same way.

In order to distinguish whether the Rockwell value has been measured with the dimond or with the ball, one speaks of R_C (Rockwell value measured with the diamond *cone*) and R_B (Rockwell value measured with the *ball*).

The chief advantages of the Rockwell method are the following:

1) The measurement takes only 10 to 20 seconds, compared with the Brinell method which takes 1 minute.
2) The hardness of very hard materials can also be measured, which is not the case with the Brinell method.

The pressure applied is 150 kg at a maximum compared with the 30 000 kg in the Brinell test. The impressions made with the Rockwell method are thus smaller and less deep, which is an advantage when one wants to test materials of slight thickness such as sheet material.

Disadvantages of the Rockwell method are the following:

1) The impressions made cannot be used as permanent records. The zero point (fig. 2 *I*) can not be adjusted again after the impression is once made.
2) The measuring range for R_C as well as for R_B is relatively small.
3) Since a difference in depth is measured, it is necessary to fasten the object to be measured in such a way that it cannot bend through under the pressure. Elastic deformations, which may occur chiefly between the restpiece and the underlayer, thus give rise to inaccuracies in the measured results.

Vickers' hardness meter

A hardness meter which combines the advantages of the Brinell and Rockwell method is the Vickers' meter. Vickers' hardness meter is fundamentally like a Brinell apparatus in which the ball is replaced by a diamond in the shape of a four-sided pyramid, whose side planes make an angle φ of 136°.

The pyramid is pressed into a polished surface of the material being examined with a pressure between 1 and 120 kg, depending mainly upon the thickness of the material to be tested. If the impression left, after the pressure has been removed, is examined with a microscope, a square is observed whose diagonal is measured. The hardness V, measured according to Vickers [3]) is obtained, as in Brinell's method, by dividing the load by the area of the surface of the impression, or

$$V = \frac{2L \sin \varphi/2}{d^2}, \quad \dots\dots \quad (4)$$

where L is the weight in kg,
φ is the vertex angle of the diamond (136°)
d is the diagonal of the base of the pyramidal impression.

Vickers meter possesses the great advantage of always giving similarly shaped impressions, so that it makes practically no difference whether one measures the hardness of a homogeneous material under a pressure of 5 or 50 kg. A great disadvantage is that a plane surface must be ground and polished on the material to be measured, while for the Brinell and Rockwell methods a ground surface is enough.

In *fig. 3* a comparison is given of the impressions

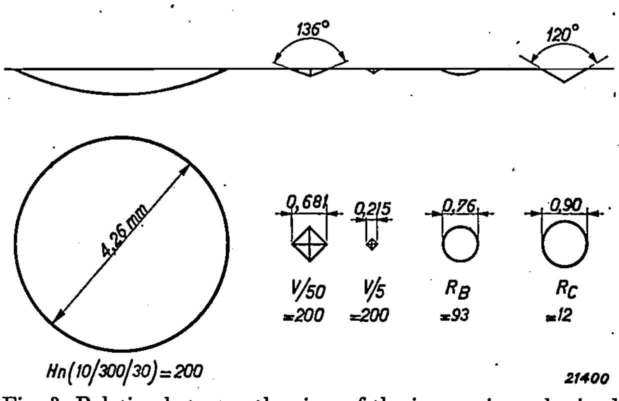


Fig. 3. Relation between the sizes of the impressions obtained in measuring the hardness of a given material by the Brinell, Vickers and Rockwell methods.

obtained when the hardness of a given steel is measured successively by the Brinell, Vickers, and Rockwell methods. It is obvious from the figure that the Vickers meter gives the least

3) The load employed is often added to the symbol V, for instance $V/10$ means Vickers hardness measured with a load of 10 kg.

depth of penetration, which is a great advantage especially when the hardness of sheet material or even of thin surface layers must be measured.

In *fig. 4* the different hardness scales are shown under each other in their correct relation. In addition the positions of several commonly known materials are indicated. If one measures the hardness of the metals with the minerals referred to in the Mohs scale, one obtains very varied values, whose averages are practically proportional to the logarithms of the Vickers hardness. It may be seen from the figure that the Vickers meter alone

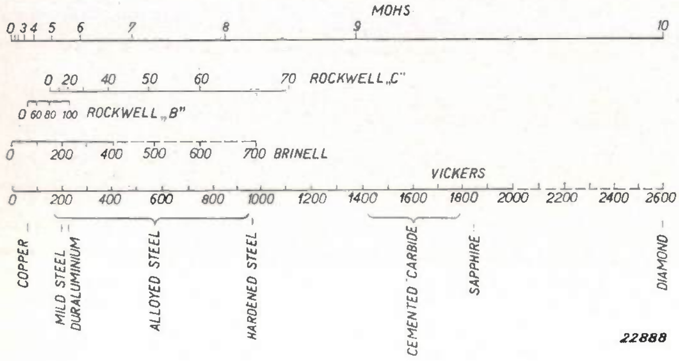


Fig. 4. Relation between the hardness scales of Mohs, Rockwell, Brinell and Vickers.

has a practically unlimited measuring range with a uniform scale division. The great advantage of this is that the hardness is expressed in the same quantities for soft materials like copper as for the hardest cutting metals such as "Widia".

Other methods of determining hardness

The three methods last described of measuring the hardness may to a certain degree be considered as standard methods. Apart from those already mentioned there are various other methods of determining the "hardness" 4) of metals and alloys.

These methods often have the advantage of causing no damage to the material examined, but, however, the disadvantage is that all of them have only a short measuring range. With the scleroscope for example, use is made of a hardened steel ball or hammer upon which a specially ground diamond is fastened. This hammer is allowed to fall from a height of 25 cm upon the material to be investigated and the height of rebound is measured, which height is greater the harder the material. The values so obtained are very often expressed in normal Brinell, Rockwell or Vickers units by means of comparison tests.

The micro hardness meter

The methods described of measuring the hardness of materials all give macroscopically visible impressions. For the practical testing of materials, this is no disadvantage. The scientific investigation of materials, however, sometimes makes greater demands, as will be shown in the following.

The pure metals which may be considered for practical use have such low values for their solid properties that they are practically useless for application as structural material.

If to a molten metal one adds another metal or a non-metal, an alloy is obtained upon cooling, which usually consists of hard and soft components which, if they are in the right proportion, give to the whole solid properties which may be a multiple of the components separately. If a surface of such an alloy is ground and polished and then etched with a suitable reagent, it is possible to distinguish these components from each other under the microscope.

Fig. 5 shows an etched surface of an aluminium

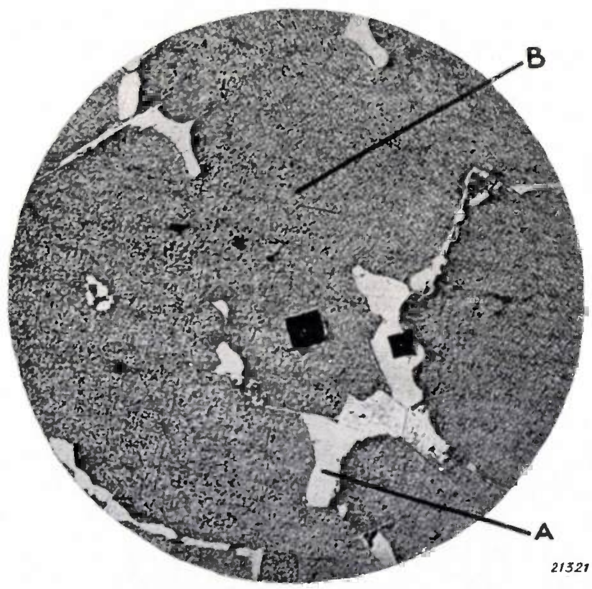


Fig. 5. Etched surface of an aluminium-copper alloy, magnification 150 times.

alloy magnified 150 times. A uniformly etched foundation mass *B* may be seen in which white inclusions *A* are scattered.

If the hardness of this alloy is measured by one of the methods mentioned, it is always measured over such a wide area that not only the soft foundation mass but also a greater or a smaller number of inclusions falls within the area chosen, in other words, an average hardness is measured.

3) Hardness is here placed in quotation marks because these methods do not define the hardness according to the depth of penetration, but are based on other properties, which, it is true, have a certain connection with the hardness, but whose values cannot in general be converted into Brinell and Vickers values.

If one takes a material like the one whose photomicrograph is given in Fig. 5, where the inclusions are harder than the foundation mass, one can make the foundation mass harder by means of the correct technique in alloying and at the same time decrease the number of inclusions. In this way a material is obtained whose average hardness is the same, but whose physical properties have been changed. It is therefore important to dissociate the average hardness obtained by the Brinell, Rockwell or Vickers method into the hardness of the foundation mass and that of the inclusions.

A hardness meter [5]) developed in the Philips laboratory makes this possible (see *fig. 6)*. The

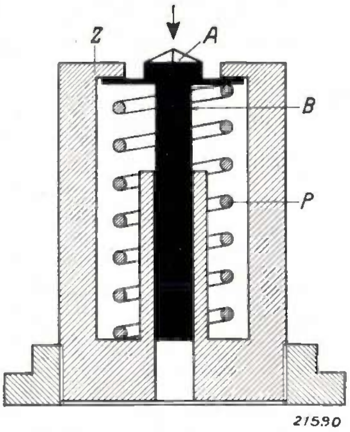


Fig. 6. Sketch of a micro hardness meter for measuring the hardness of structural elements.

diamond A, which has the same form as that for measurements by Vickers' method is fastened upon a shaft B which may be moved freely in the direction of its length. The spring P has dimensions such that when the diamond is loaded in the direction of the arrow a pressure of 0.020 kg is just sufficient to free the contact at Z. The whole is about as large as a normal microscope objective (*fig.* 7) and is mounted on a normal objective fitting.

If during the microscopic examination of an alloy components are observed whose hardness it is desired to measure, the part to be investigated is brought into position by means of the cross hair in the eyepiece. The objective is then replaced by the hardness meter, and the microscope stage and diamond are screwed toward each other until the specimen just presses the diamond sufficiently to free the contact, which is indicated by a built-in electrical signal arrangement. After for instance

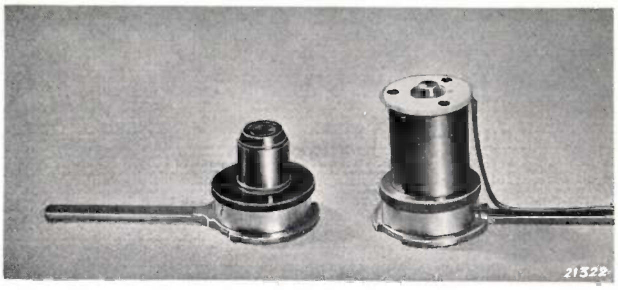


Fig. 7. Dimensions of the micro hardness meter compared with those of a normal microscope objective.

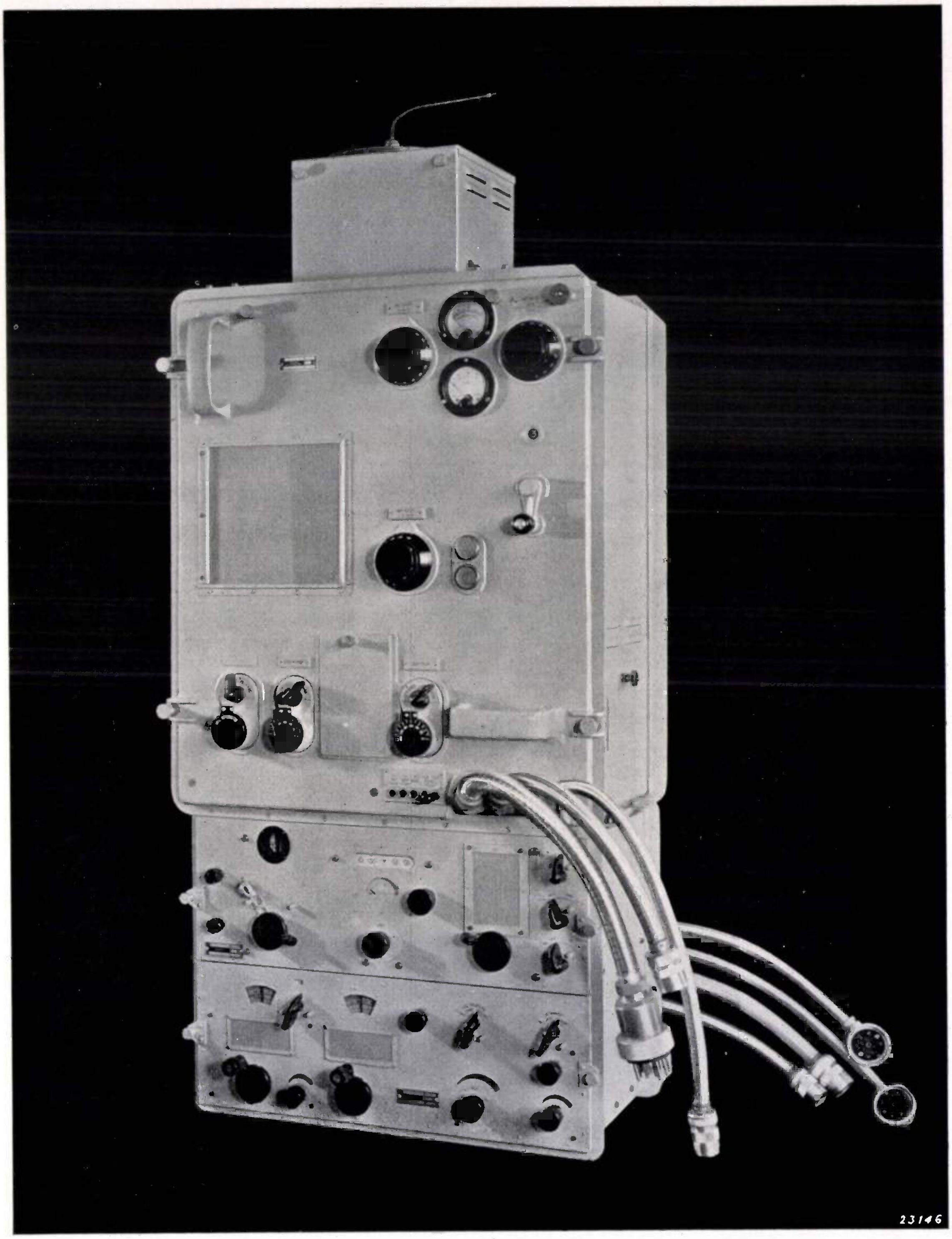
30 sec. the microscope stage is screwed free of the diamond, the hardness meter is replaced by the microscope objective and the impression is then measured with a measuring eyepiece. The value obtained is then recalculated into normal Vickers units.

In Fig. 5 two impressions may be seen which were made in the manner described. The impression in the matrix corresponds to a Vickers hardness of 75 units, while the inclusions have a hardness of 395 units. The "average" measured with the Vickers apparatus was 78 units.

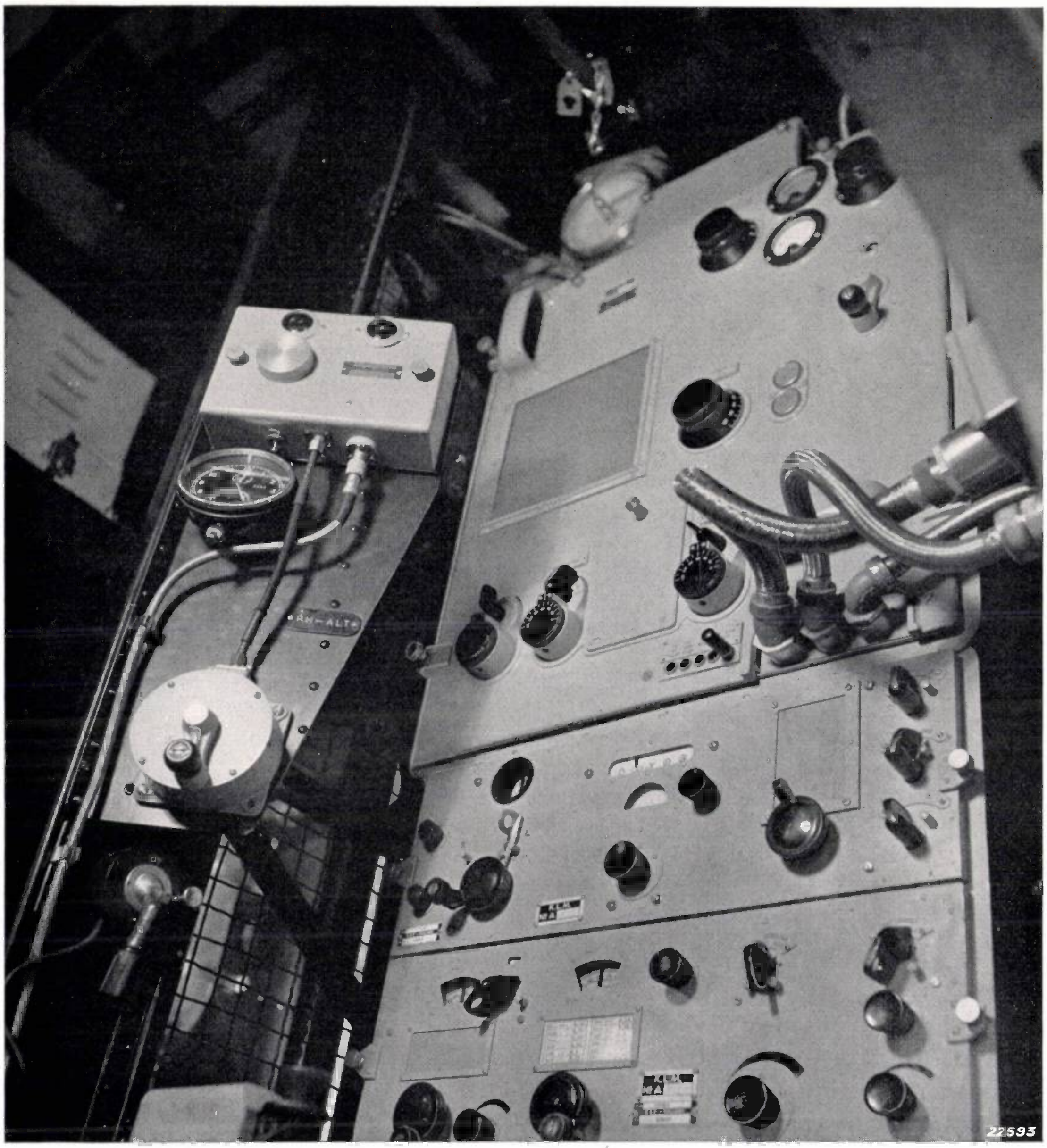
The application of the micro hardness meter is limited of course to the laboratory. It may however perform valuable services when it is necessary to examine the structure of alloys, for example in the transition of hardness in spot welding, or the identification of structural components. It may in addition serve for the measurement of the hardness of very thin surface layers such as those which occur by electrolytic deposition.

[5]) E. M. H. Lips and I. Sack: A Hardness Tester for Microscopical Objects, Nature, **138**, 328 (1936).

TWO PHOTOGRAPHS OF THE TRANSMITTING AND RECEIVING EQUIPMENT V.R. 35 COMBINED WITH THE AUTOMATIC PILOT V.P.K. 35.



The three section switchboard contains from the top downwards: the transmitter, the automatic pilot and the receiver.



The same transmitter and receiver installation installed in a Douglas aircraft, combined with the automatic pilot or course finder. On the left next to the receiver is the Morse key. The crank on the left next to the automatic pilot serves for controlling the frame aerial, whose position is indicated at the round windows in the small box above the crank. The equipment is installed behind the pilot. The wireless operator sits sideways so that the switchboard is close to his right hand side. A detailed description of the course finder is given in the article commencing on pag. 184 of this issue.

POSITION FINDING AND COURSE PLOTTING ON BOARD AN AEROPLANE BY MEANS OF RADIO

Summary. In this article several radio receiving arrangements are described, with which it is possible to plot the course and take radio bearings on board an aeroplane.

Introduction

The necessity of determining the position of aeroplanes by means of radio has become more urgent now that more and more flying is being done in the upper air. Originally a pilot had to depend for his navigation entirely upon the ground service, which at his request gave his position and the compass course he should follow. To do this the aeroplane asks one of the direction-finding stations on the ground to determine his position with at least one other D-F station. The aeroplane transmits a continuous signal for some time so that the D-F stations can determine the direction from which the signals are received. One of the ground stations then works out these cross-bearings on the map and signals the resulting position to the aeroplane. Even with long practice the taking of such cross-bearings of an aeroplane occupies at least two ground stations for several minutes. If in a given region there are other aeroplanes which are demanding their position or their course, they must await their turn. With the increase in speed of aeroplanes such a wait became more and more of a disadvantage, while because of the increasing density of air traffic it has become more and more unavoidable. An attempt is being made to limit as far as possible the demand for bearings from the ground service by equipping the machines with instruments which make it possible for the pilots to determine their own course and position.

If there is a radio transmitter in the line of flight of an aeroplane, then the correctness of the course followed can be checked with a relatively simple instrument, a so-called course finder, which can be installed in even small private aeroplanes. If, however, one's course is toward an aerodrome where there is no transmitter in the neighbourhood the course can be checked by finding the position at short intervals by means of a radiogoniometer. In order to take such radio bearings one must be provided with a more highly perfected measuring instrument, a so-called homing device which can be installed on board the larger aeroplanes. When bearings are being taken, however, the radio-telegraphist must neglect his ordinary radio traffic service duties, such as receiving weather reports and the like, and enquiring about the landing possibilities.

Course finder

With the aid of the Philips' course finder V.P. 4, shown in *fig. 1*, which can be used in series with an ordinary receiver, one is able to check the correctness of the course followed in a simple way. The action of this instrument is based upon the directional reception by means of a loop aerial, which is situated on or in the case of wooden construction inside the body of the aeroplane with its plane perpendicular to fore and aft line of the machine.

Fig. 1. Course finder, type V.P. 4.

Fig. 2 shows the horizontal polar diagram of a vertically placed loop receiving a constant signal. If the North-South direction is perpendicular to the plane of the loop, the polar diagram consists of two circles touching each other along the line of that direction as drawn in fig. 2 (figure-eight diagram). If we choose a course directly toward a radio station, it will be received with minimum strength on the loop which is directed with its plane perpendicular to the direction of flight. If one flies in any other direction, the signal is received

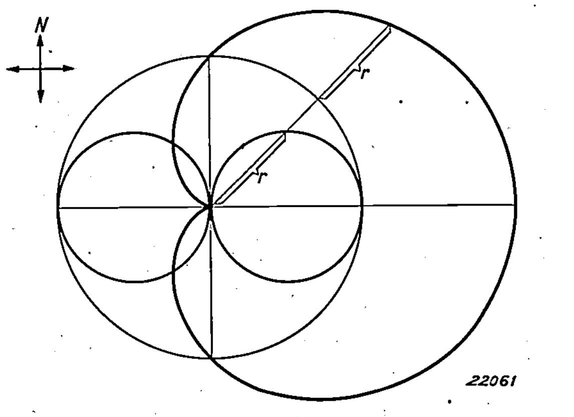


Fig. 2. Diagram of the receiving strength of a loop aerial (figure-eight shape) and non-directional aerial (circle). The heart-shaped diagram is obtained by combining these two in an appropriate way.

with greater intensity. The strength of the signal, however, gives no indication as to whether one must turn to the right or the left in order to regain the correct course.

In order to obtain an indication of the direction in which one must steer, the course finder is so arranged that the electromotive forces which are generated in the loop are combined with those induced in the ordinary receiving aerial of the airplane. The polar diagram of the received signal strength on an entirely non-directional aerial is a circle as shown in fig. 2. The ordinary aerial of an airplane, it is true, does not give a truly circular diagram, but an oval one, which is symmetrical with the axis of the aeroplane, so that in principle the following considerations may be applied to it. At a given position of the combination switch the aerial and the loop electromotive forces are in phase with each other for the right half of the polar diagram, and just in opposite phases for the left half of the diagram. We therefore obtain a heart-shaped diagram (cardioid) for the receiving strength of the course finder, as is shown in fig. 2. When the combination switch is reversed, the heart-shaped diagram is obtained on the other side of the north-south direction. If in fig. 2 the signal does not come exactly from the north, but a little to the east, then in the course finder the electromotive force in the ordinary aerial is increased by a small electromotive force in the loop. When the switch is reversed, the latter is subtracted from the former and the signal thus becomes fainter. A signal which in fig. 2 comes from west of north, however, becomes stronger on commutation.

In this simple course finder the commutation is done by hand, while reception is exclusively by ear. A second commutator may be introduced at some distance from the course finder, for instance on the control stick. This reversing of the switch by hand offers no great practical difficulty, since it is only necessary while the course is being found. When the course in once found, the course finder may be disconnected from the ordinary aerial so that the course is kept by means of the minimum loop reception.

Direction finder-homing device

The pilot can be provided with a visual indication of his course by providing for a periodical commutation. This can be done by means of an electrical, mechanical or electromagnetic circuit arrangement. The Philips direction finder-homing device V.P.K. 35 (*fig. 3* and photographs on page 182 and 183) is a combination of a direction finder and a course finder, which is furnished with an automatic commutation arrangement and a visual indicator. As the name indicates, the apparatus serves not only for "homing", but also for taking bearings. It is used in all the new Douglas DC 3 machines of the K.L.M.

The apparatus is suitable for:

1) Normal, non-directional reception when a special vertical aerial alone is connected to it. This aerial is introduced into a hollow aerial mast made of insulation material on top of the fuselage.

2) Taking bearings, whereby the angle between the direction of the sending station and the axis of the aeroplane is measured with the help of a loop aerial, which is erected on top of the aeroplane body, and rotatable about a vertical axis. In fig. 2 the figure-eight diagram of the receiving strength is that when the loop plane is assumed to be perpendicular to the north-south direction. At minimum reception the plane of the loop is perpendicular to the direction from which the signal comes. In order to obtain a sharp minimum the action of the loop as a non-directional aerial, the so-called aerial effect of the loop, is compensated. The signal may still, however, have come from one of two opposite directions, and one can decide which of these is the correct one by:

3) Sense determination. As with the course finder the signals received on the loop and on the

[1]) In a later number of this periodical, in an article about the measurement of field strengths, a more detailed explanation will be given of the way in which the aerial effect can be compensated.

vertical aerial must be combined to give the heart-shaped diagram of the receiving strength as in fig. 2. The loop is now, however, not fixed immovably with its plane perpendicular to the longitudinal direction of the aeroplane, but may be rotated about the vertical axis, so that the north-south direction in fig. 2 can be arbitrarily oriented with respect to the axis of the machine.

If, when the position of the plane of the loop is perpendicular to the north-south direction in fig. 2, we have a minimum reception on the loop alone, the signal may still come either from the north or from the south. If the loop is turned slightly to the right, the combined receiving strength on loop and vertical aerial will decrease for a signal coming from the north and increase for a signal from the south. If the phase of the loop signal is reversed, the heart-shaped diagram is mirrored with respect to the north-south direction of the diagram, and the changes in the received strength upon turning the loop in the neighbourhood of the position of minimum reception are exactly opposite.

4) Homing. The phase of the electromotive force generated in the loop is periodically according to a dot-dash rhythm having a time interval relation of 1 : 7. During the long dash interval the received signal strength is as shown in the heart-shaped diagram of fig. 2, while during the short dot interval this diagram is mirrored with respect to the north-south direction. If the signal in fig. 2 comes from the east, i.e., from the right when the plane of the loop is perpendicular to the axis of the aeroplane, we then hear strong dashes and weak dots, while a signal coming from the left half plane gives strong dots and weak dashes. When the course is directly toward the transmitter one hears a continuous dash. Not only is the pilot able to hold his course toward a radio transmitter, by ear, but also he has the course-line toward the transmitter concerned given on the visual course indicator which is installed on the dashboard.

5) Beacon reception of the Philips long-wave beacon, B.R.A. 101, which is installed at all the aerodromes in the Netherlands. This beacon has an aerial system which, like that of the bearings finder, consists of a loop and a vertical rod which are both fixed with respect to the earth. The radio beacon sends with loop and rod, and the phase of the electromotive force in the loop is reversed in a dot-dash rhythm. In a direction perpendicular to the plane of the loop a continuous dash occurs, while to the right of the beacon line, which may for the present be assumed to be along the north-south direction of the diagram of fig. 2, the dashes are heard more clearly, and to the left the dots. If the aeroplane receives the beacon signal only on the vertical aerial, then the same effect occurs in the detector as is caused by a non-directional transmitter with the help of the automatic commutator arrangement in a bearings finder. It may thus be seen by means of the visual course indicator on which side of the beacon line one is.

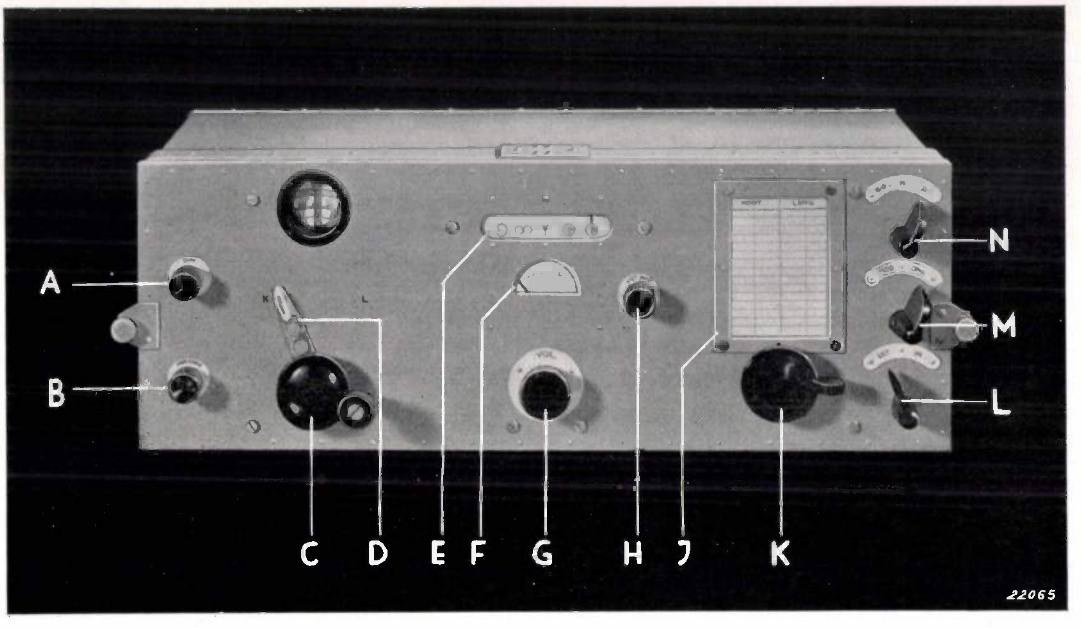


Fig. 3. Direction finder-homing device (Type V.P.K.P.. 35).

In the methods of reception described in 4) and 5) simultaneous indication by eye and ear is possible. Experience has shown that some pilots can keep a steadier course by ear than by visual indication; with others the reverse is true. All of them, however, appreciate the simultaneous information of sight and hearing, where one of the two acts as control and complement of the other. When receiving with atmospheric disturbances the visual course indicator is "restless", while in the telephone the desired signals may be distinguished adequately, since they may be adjusted to give a constant tone of any desired pitch, which can be plainly distinguished from the disturbances.

Flying in the direction of a broadcasting station, the carrier wave can be heard as a continuous dash with the aid of a beat oscillator. After the beat oscillator has been switched off, the modulation of the transmitter concerned may be listened to, without disturbance of the visual course indication. This may be of importance for the reception of weather reports, which are sent out periodically by many broadcasting stations. Furthermore one can check whether or not one's course is toward the right transmitting station, for instance by means of the announcements between the items of the programme.

Construction of the direction-finder homing-device

On the front plate of the apparatus various controls may be seen (fig. 3). The lighting of the various scales for night flying can be regulated with knob A. By compensation of the aerial effect of the loop with knob B, one can attain a sharp minimum. The wavelength region 250-670 m (1200-488 kilocycles) is chosen by setting the switch D in the left-hand position, while in the right-hand position wavelengths 790-2000 m (380-150 kilocycles) are received. The apparatus V.P.K. 35 is thus not only suitable for the shipping and air traffic bands, but also for the most important part of the medium-wave broadcasting band. Tuning is done with knob C, while the scale in the upper left-hand corner of the front plate is the wavelength scale.

On scale E is shown the kind of reception for which the receiver is adjusted by means of the switch K. On the measuring instrument F one may control the strength of the signal which is added to the visual course indicator situated on the pilot's dashboard. This signal strength is adjusted by means of knob H, while G operates the volume control for radio reception. On the plate J the position of the wavelength scale for reception from the most commonly used transmitting stations may be noted. The whole installation may be switched off with switch L, while with M the beat oscillator for carrierwave reception can be switched in. In order to make a distinction between the knobs frequently used and those seldom used, the former are larger.

The mixing switch N makes it possible for the wireless operater to listen to the direction-finder homing-device or the ordinary radio receiver (for traffic and weather reports) or to both at once, while at the same time the pilot, by means of a bridge circuit, receives only and without disturbance the signals of the direction-finder. Through the possibility of simultaneous listening to the receiver on board and the direction-finder, the operator may be called by a ground station via the receiver on board during the tuning of the direction-finder homing-device or during the taking of radio bearings.

A meter for checking the strength of the signal added to the visual course indicator is introduced not only on the front of the apparatus itself, but also on the instrument board of the aeroplane. In this way the pilot also can read the signal strength received.

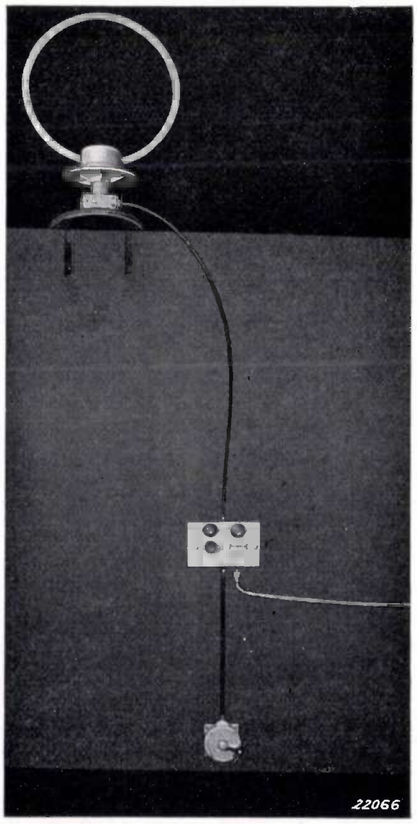


Fig. 4. Rotating loop aerial, angle indicator and arrangement for rotating the loop, for apparatus type V.P.K. 35.

Fig. 4 shows the loop and the arrangement for rotating it. In order to make it possible to read from the scale at eye-height and at the same time turn the loop easily with the hand, reading scale and handle are kept apart. The course direction with respect to the fore and aft line of the aeroplane can be read off from an angular scale which turns past an index mark. On a pelorous circle which is adjustable with respect to the course circle, and whose north-point can be set according to the compass, magnetic bearings can be found. The positions of the scales can be read clearly by means of a lens. With the help of a second lens one may read the deviation curve, which gives the amount by which the measured bearing direction must be corrected for every position of the loop with respect to the aeroplane as a result of the distortions of the electro-magnetic field by the body of the aeroplane. The two scales are lighted clearly by a small lamp, care having been taken that in the case of night flying no glaring light falls directly into the pilot's eyes.

Circuit of the direction-finder homing-device

The action of the apparatus is explained below with the aid of the block diagram (*fig. 5*).

The aerial signal is amplified by the high-frequency amplifier stage, the loop signal by the high-frequency amplifier stage, the loop signal by the high-frequency stage in which there are two valves with grids in parallel and anodes connected in opposite phase. By blocking one or the other of the valves in these high-frequency amplifier stages by means of the switch which determines the kind of reception, mixing takes place in the way required for the various kinds of reception, namely:

a) For non-directional reception only the aerial signal must be allowed to pass. For this purpose both the valves of the loop stage *2* are blocked by a high negative grid potential.

b) The same is done for beacon reception.

c) For blind flying the phase of the loop signal with respect to the aerial signal must be reversed periodically in a dot-dash rhythm. This is attained by blocking the valves of loop stage *2* alternately in this time relation. It is done by means of a commutator operated by the converter for the anode potential.

d) For the determination of the directon (heart-shaped diagram) one of the two valves in the loop stage *1* is blocked.

e) In taking bearings (figure-eight diagram) only the loop signal must be allowed to pass. Not only one of the valves of stage *2*, but also the valve of the aerial stage *1*, is blocked.

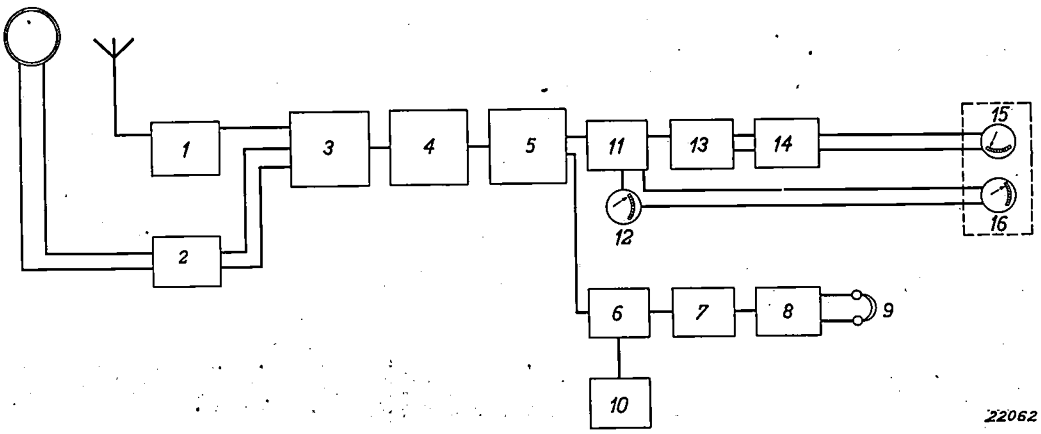


Fig. 5. Diagrammatic representation of the circuit of the apparatus VK. 35.

The signals thus allowed to pass and amplified are combined in stage *3*, where the high-frequency signals are also converted to a constant intermediate frequency by means of an auxiliary oscillator. The stages *1* and *2* and the input circuit of stage *3* are tuned to the frequency to be received, and a special circuit in the aerial stage *1* provides for the correct phase relation of acrial and loop signals.

The intermediate-frequency signals are now first amplified in the intermediate-frequency amplifier stages *4* and *5*, and then division takes place to the detectors *11* and *7* for visual and auditory reception respectively.

For auditory reception the intermediate-frequency signal is detected in stage *4*, and the low-frequency signal therby obtained is amplified by the low-frequency amplifier stages *7* and *8* and fed to the telephone *9*.

For the purpose of listening to unmodulated signals an oscillator (*10*) is added, which gives an audible beat with the intermediate frequency. Since the longwave beacon signals are always unmodulated in the position for "beacon reception" of the switch governing the kind of reception, this oscillator is automatically connected, while in the positions for other kinds of reception it may be connected when necessary by means of a separate switch.

For the visual indication in course and beacon flying the intermediate frequency signal is rectified in stage *22*. For the most satisfactory functioning of the indicator part it is necessary that the intermediate-frequency signal to be rectified be of a definite strength. This signal strength may be adjusted with the knob *H* on the front plate of the apparatus, and may be checked on a

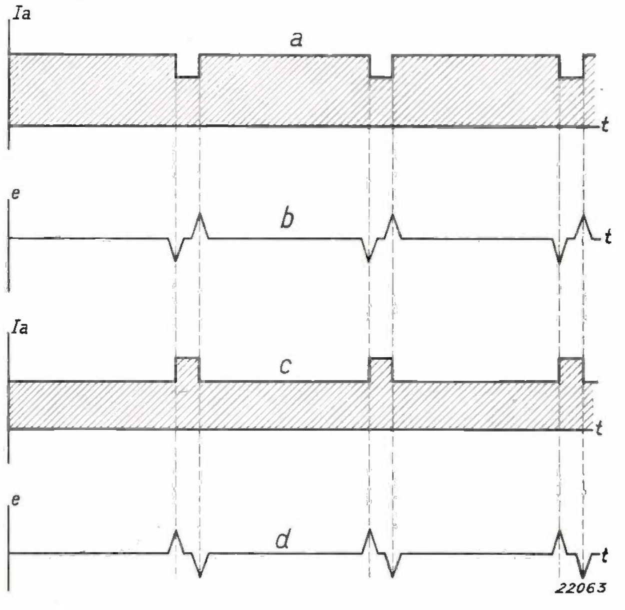


Fig. 6. Diagrammatic representation of the action of the visual indicator of the direction-finder homing-device, type V.P.K. 35.

current meter in the detector circuit (*F* in fig. 3, and *F* in fig. 5; in addition, a meter *16* in series with this latter is installed on the dashboard in front of the pilot.

Further it is desirable that in flying a course and in beacon reception the strength of reception remain the same upon approaching a transmitting station. An automatic volume control is introduced for this purpose. It controls the conversion stage *3* and the intermediate-frequency stages *4* and *5*, and is set in operation by the switch controlling the kind of reception. The automatic volume control has such a great time lag that the difference in intensity of reception between dots and dashes is not eliminated. The current pulses obtained after detection are conducted to a direct-current amplifier in stage *13* via a filter, which suppresses any audio-frequency modulation. In the anode circuit of this amplifier no current variations will occur when flying on the course line. When of the course, however, such variations will occur. In the region of dashes the plate current will have the character of *fig. 6a*, in the dot region of *fig. 6c*.

In the secondary windings of a transformer in the anode circuit, potentials are generated by these current variations, which have a character like that of fig. 6*a* for the dash region and of fig. 6*d* for the dot region. From the figs. 6*b* and *d* it may be seen that the first impulse of the impulse combination in the dash region is oppositely directed to that in the dot region. Therefore, if only the first impulse of every impulse combination is passed through, the indicating instrument will react oppositely for dots and dashes.

This is attained in stage *14*, where two valves in bridge connection are connected to the secondary winding of the above-mentioned transformer. In the grid circuit there is a special arrangement for blocking. The valves are so adjusted that their operating point lies in the region of the greatest curvature of the characteristic. As soon as the anode current on one valve increases due

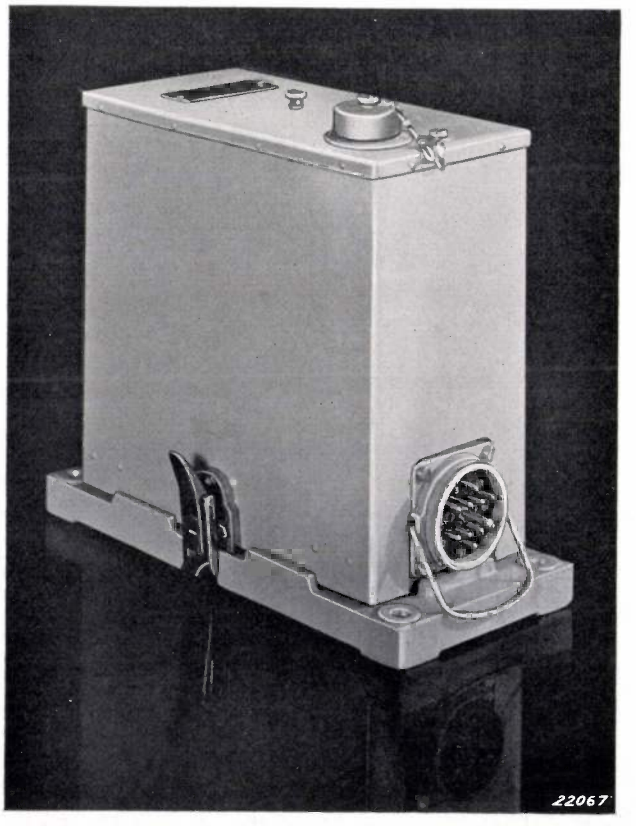


Fig. 7. Converter with interference suppressor for feeding the direction-finder homing-device, type V.P.K. 35.

to an impulse both valves are blocked during at least the time period of one dot, so that impulses occurring in this time are not passed through.

As a result only the one valve operates for dots and the other for dahes. The pointer of the indicator *15*, which is connected between the anodes of both valves, will take up the middle position when on the line of course and show a deviation for dots and dashes respectively in opposite directions.

The energy for the direction-finder is taken from the accumulator on board. The filaments of the indirectly heated valves are connected directly to the battery, while the anode potentials are obtained from a converter which runs from the battery on board. This converter is installed in a box with its interference suppressor and fuse (*fig. 7*). The suppression of interference is such that when the direction finder is adjusted to its greatest sensitivity, no disturbance, for example from commutator sparking, is observable even on the shortest waves.

The apparatus is very light in construction and weighs in total only 21.3 kg, to which the direction finder contributes 11.4 kg and the converter 4.2 kg. The rotating loop aerial which is fitted to the outside of the aeroplane is only 43 cm high.

Compiled by G. P. ITTMANN.

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

No. 1153: J. A. M. van Liempt: Die Anätzung von Molybdän durch alkalische Ferricyankalilösungen (Rec. Trav. chim. Pays Bas, **55**, 989 - 990, Nov. 1936).

The etching of molybdenum by ternary mixtures of water, sodium hydroxide and potassium ferricyanide was studied and the results were given in a diagram. From the experiments a formula was deduced for a solution which is very suitable for metallographic etching and for the investigation of the so-called intermediate substance.

No. 1154: H. C. Hamaker: A gencral theory of lyophobic colloids I. (Rec Trav. chim. Pay Bas **55**, 1015 - 1025, Nov. 1936).

The point of departure of this article is the common assumption that the reciprocal action between the particles of a colloid whose hydration may be neglected is caused by the superposition of an attraction according to v. d. Waals and London and an electrostatic repulsion. When this simple conception is worked out to its logical conclusion, it leads to essentially more complicated phenomena than were hitherto expected. With the conceptions given here a more extended range of phenomena can be treated than could be explaned theoretically up to the present.

No. 1155*: W. Elenbaas Zur Frage der Berechtigung des Minimumprinzips in der Theorie der Bogenentladiung (Elektrotechn. Z. **57**, 1497 - 1498, Dec. 1936).

In this article objection is raised to the minimum principle in the theory of the electric arc defended by Kesselring nad Koppelmann, according to which the arc will burn with a temperature and a diameter such that the current flows under the influence of the smallest possible drop in potential. Elenbaas shows that in some cases the arc is already wholly determined before one has made use of the minimum principle from which fact it follows than one no longer has the liberty of then introducing such a principle.

No. 1156: W. G. Burgers and J. J. A. Ploos van Amstel: "Oriented" oxidation of barium (Physica, **3**, 1057 - 1063, Dec. 1936).

Films of metallic barium were obtained by evaporating the metal in a vacuum and then condensing it on a plane polished copper plate. Depending among other factors upon the temperature

*) There is not a sufficient number of reprints of articles marked * for distribution. The administration of the Natuurkundige Laboratorium, Kastanjelaan, Eindhoven, will on request be glad to send reprints of the other articles.

to an impulse both valves are blocked during at least the time period of one dot, so that impulses occurring in this time are not passed through.

As a result only the one valve operates for dots and the other for dahes. The pointer of the indicator *15*, which is connected between the anodes of both valves, will take up the middle position when on the line of course and show a deviation for dots and dashes respectively in opposite directions.

The energy for the direction-finder is taken from the accumulator on board. The filaments of the indirectly heated valves are connected directly to the battery, while the anode potentials are obtained from a converter which runs from the battery on board. This converter is installed in a box with its interference suppressor and fuse (*fig. 7*). The suppression of interference is such that when the direction finder is adjusted to its greatest sensitivity, no disturbance, for example from commutator sparking, is observable even on the shortest waves.

The apparatus is very light in construction and weighs in total only 21.3 kg, to which the direction finder contributes 11.4 kg and the converter 4.2 kg. The rotating loop aerial which is fitted to the outside of the aeroplane is only 43 cm high.

Compiled by G. P. ITTMANN.

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

No. 1153: J. A. M. van Liempt: Die Anätzung von Molybdän durch alkalische Ferricyankalilösungen (Rec. Trav. chim. Pays Bas, **55**, 989 - 990, Nov. 1936).

The etching of molybdenum by ternary mixtures of water, sodium hydroxide and potassium ferricyanide was studied and the results were given in a diagram. From the experiments a formula was deduced for a solution which is very suitable for metallographic etching and for the investigation of the so-called intermediate substance.

No. 1154: H. C. Hamaker: A gencral theory of lyophobic colloids I. (Rec Trav. chim. Pay Bas **55**, 1015 - 1025, Nov. 1936).

The point of departure of this article is the common assumption that the reciprocal action between the particles of a colloid whose hydration may be neglected is caused by the superposition of an attraction according to v. d. Waals and London and an electrostatic repulsion. When this simple conception is worked out to its logical conclusion, it leads to essentially more complicated phenomena than were hitherto expected. With the conceptions given here a more extended range of phenomena can be treated than could be explaned theoretically up to the present.

No. 1155*: W. Elenbaas Zur Frage der Berechtigung des Minimumprinzips in der Theorie der Bogenentladiung (Elektrotechn. Z. **57**, 1497 - 1498, Dec. 1936).

In this article objection is raised to the minimum principle in the theory of the electric arc defended by Kesselring nad Koppelmann, according to which the arc will burn with a temperature and a diameter such that the current flows under the influence of the smallest possible drop in potential. Elenbaas shows that in some cases the arc is already wholly determined before one has made use of the minimum principle from which fact it follows than one no longer has the liberty of then introducing such a principle.

No. 1156: W. G. Burgers and J. J. A. Ploos van Amstel: "Oriented" oxidation of barium (Physica, **3**, 1057 - 1063, Dec. 1936).

Films of metallic barium were obtained by evaporating the metal in a vacuum and then condensing it on a plane polished copper plate. Depending among other factors upon the temperature

*) There is not a sufficient number of reprints of articles marked * for distribution. The administration of the Natuurkundige Laboratorium, Kastanjelaan, Eindhoven, will on request be glad to send reprints of the other articles.

at which the evaporation took place, the crystallites are randomly arranged or show a preference to having their (111) direction perpendicular to the surface. Upon oxidation of the oriented films of barium, oriented films of barium oxide are obtained, whose (100) direction fluctuates 10 to 15° about the normal to the surface. The direction of most compact arrangement in the metal lattice (111) is thus found to be practically parallel to that of the metal atoms in the oxide lattice (110).

It was found further that the (111) direction in the oriented barium films which were formed by diagonal incidence of the vapour beam deviates slightly from the normal to the underlayer, and deviates in the direction of the incident beam, as was previously observed with layers of calcium fluoride.

No. 1157: W. G. Burgers and J. J. A. Ploos van Amstel: Texture of thinly rolled tungsten foil (Physica **3**, 1064 - 1066, Dec. 1936).

From electron diffraction images of a very thin rolled tungsten plate it appears that it possesses a pronounced texture as a result of the rolling. The crystallites lie with a cube face parallel to the tungsten plate and with a rib at about 45° to the direction of rolling. Upon oxidation an oxide film is formed in which the crystallites are randomly arranged.

No. 1158: G. Holst and P. J. Bouma: Ein neues Messgerät zur Beurteilung der Güte einer Strassenbeleuchtung (Physica **3**, 1159 - 1163, Dec. 1936).

Since this visibility meter was described in the last volume of this periodical (cf. Philips techn. Rev. **1**, 349, 1936), it will not be discussed here.

No. 1159*: A. E. van Arkel and J. H. de Boer: La valence et l'électrostatique (435 pages; 1936, Librairie Felix Alcan; 108 Boulevard Saint Germain; Paris VIe).

This is the French edition of the well-known book by these authors which has already been published in Dutch (1930) and German (1931): Chemical combination as an electrostatic phenomenon. In this edition the book has been considerably enlarged. Among other new features are several applications of the Franck-Condon principle to chemical combinations. Not least important, according to the preface of Prof. Victor Henri, is the working out of the relation between the volatility and the electric dipole moments of the molecules of the homologous series of organic substances.

No. 1160: W. F. Brandsma and E. M. H. Lips: Über die Erkennung von Umwandlungen bei Metallen im festen Zustande (Z. Metallk. **28**, 381 - 382, Dec. 1936).

The manner in which the velocity of torsion at constant moment of torsion is influenced by changes in the solid state in the case of metals is investigated by means of a torsion meter previously described. With steel containing 0.16 per cent of carbon a sudden increase in the velocity of torsion is observed when the change from α- to γ-iron takes place, while two characteristic points were found with cold-worked aluminium. The first point may be explained as a transition from the crystalline plasticity to amorphous plasticity, while the second point may be ascribed to the conclusion of the recrystallization.

No. 1161: W. Elenbaas: Lampe à mercure à surpression (Rev. Opt. **15**, 343 - 350, Sept. 1936).

The construction and operation of the discharge tube with super high mercury pressure are discussed. The high surface brightness and the spectrum of the discharge are treated. Finally the various methods are given by which the pressure in these discharges can be determined.

No. 1164: W. de Groot: Het natuurlijke systeem der elementen van het standpunt der kernphysica (Chem. Wbl. **34**, 3 - 7, Jan. 1936). (The natural system of elements from the point of view of nuclear physics).

Considering the fact that a somewhat analogous article by the author has appeared in this periodical (Philips techn. Rev. **2**, 97, 1937), we shall not review this article from the Chem. Wbl.

No. 1165*: M. J. O. Strutt: Moderne electronenbuizen met meerdere roosters (Ingenieur, **52**, E 1 - 7, Jan. 1937) (Modern electron tubes with several grids).

This article is an excerpt of No. 1148.

No. 1166: M. J. O. Strutt: Verzerrungseffekte bei Mischröhren (Hochfrequenztechn. und Elektroakust. **49**, 20 - 23, Jan. 1937).

In this article a brief survey is given of the common mixing valves. The characteristics of the valves may be written as the sum of exponential functions by means of which the distortion effects may easily be calculated in general. Several beat tones are calculated and then the relation is studied between the distortion effects and these tones.

No. 1167: W. G. Burgers and J. J. A. Ploos van Amstel: Electron-optical observation of metal surfaces. I. Iron: Formation of the "Crystal Pattern" on activation (Physica 4, 5 - 14, Jan. 1937).

The construction of the electron microscope is briefly described. On the basis of photographs the phenomena are discussed which take place in the activation of iron with strontium obtained from strontium carbonate. A clear emission pattern is obtained of the crystalline structure. The method of activation used has a certain similarity to the activation of thoriated tungsten (cf. Philips techn. Rev. **1**, 321, 1936).

No. 1168: W. G. Burgers and J. J. A. Ploos van Amstel: Electron-optical observation of metallic surfaces. II. Phenomena observed on transition of α-into γ-iron (Physica 4, 15 - 22, Jan. 1937).

On the basis of numerous photographs the changes are discussed which occur in the emission patterns of iron at the transition from the α into the γ modification and the reverse. The subject is treated not only from the metallographic standpoint but also from the point of view of the emission of electrons. Ahead of the growing crystallite appears a zone which has temporarily an abnormal emission. This is an indication that the transition process is accompanied by a displacement of activated atoms in the region of growth (cf. Philips techn. Rev. **1**, 317, 1936).

No. 1169: T. Jurriaanse: Eine Glühkathodengleichrichterröhre mit Magnetfeld und niedrigem Gasdruck (Physica **3**, 23 - 27, Jan. 1937).

For rectifier valves with heated cathodes the gas pressure used must be as low as possible in order to avoid back-arcing. At such low pressures the working voltage would be so high that one would have trouble with cathode sputtering and too great losses. This troublesome increase of the working voltage is avoided by introducing a magnetic field between cathode and anode.

No. 1170: F. Coeterier and M. C. Teves: An apparatus for the transformation of light of long wavelength into light of short wavelength. III. Amplification by secondary emission (Physica **4**, 33 - 40, Jan. 1937).

An amplification whereby secondary emission is applied one or more times is possible by making use of the deflection of cathode rays through large angles (up to 180°) by magnetic fields while the sharpness of the image is retained.

No. 1171: M. J. Druyvesteyn and N. Warmoltz: Die Elektronenemission einer Oxydkathode in einer Bogenentladung (Physica 4, 41 - 50, Jan. 1937).

By measuring the voltage of the arc as a function of the current the electron emission of an oxide cathode in an arc discharge was determined. Since the current flowed through the arc only for 10^{-4} sec. the temperature and other properties of the cathode were not changed by the measurement. The phenomena observed may be explained by assuming that an accelerating field for the electrons does not occur at the same current at all spots on the oxide surface because the surface is very rough.

VOL. 2 No. 7 JULY 1937

Philips Technical Review

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

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

NON-LINEAR DISTORTION PHENOMENA OF MAGNETIC ORIGIN

by J. W. L. KÖHLER.

Summary. The nature of distortion, and the components of an amplifier in which it can be produced, are discussed. The distortion due to choke coils with iron cores is analysed in detail, and using the results obtained the distortion in amplifier transformers is considered.

Introduction

The subject of amplifiers has been already discussed from various aspects in this Review. In the present article certain characteristics of these units will be analysed in greater detail.

The general purpose of an amplifier is to amplify or magnify an electrical alternating voltage. An amplifier would have an ideal performance if every signal were amplified without any distortion whatsoever, a point which will be considered more closely below. Let us assume that the incoming signal is a certain function of the time, viz., $V_1(t)$; the amplified voltage can generally be represented by another function, $V_2(t)$. Amplification is free from distortion when

$$V_2(t) = c\, V_1(t)$$

for all values of t and any arbitrary form of V; c is here the amplifier gain. In other words, the ratio of the output voltage to the input voltage must remain constant. Deviations from this ideal case can be most easily investigated by assuming that the signal is a pure sinusoidal alternating voltage. The possible deficiencies of an amplifier can be classified under the following heads:

a) The amplifier gain is dependent on the frequency;
b) Phase displacements occur in the amplifier resulting in the input and output signals being out of phase;
c) The amplifier gain varies with the intensity of the incoming signal, with the result that the output voltage is not proportional to the input voltage; this is termed non-linear distortion.

a) This form of distortion is found in every case provided a sufficiently wide frequency range is explored. But it can be always reduced to a satisfactory minimum for any particular frequency range in use.

b) Phase displacement is not an important factor in low-frequency amplifiers, since the ear is unable to distinguish phase displacements in this range. On the other hand, it cannot be neglected in television amplifiers and high-frequency amplifiers in radio receiving sets.

c) Non-linear distortion, which is generally referred to as plain distortion, is obtained in all cases, since amplifiers always contain amplifying valves and transformers which as a rule are made up of non-linear circuits and components. In the present article this particular form of distortion will be discussed.

General Considerations on Distortion

In what respect do distortion phenomena detract from the quality of reproduction? Consider the relationship between the input and output voltages, which are connected by the expression

$$V_2 = f(V_1).$$

If $\qquad V_1 = V_0 \sin \omega t$

then $\qquad V_2 = f\,(V_0 \sin \omega t).$

While V_2 is thus still periodic, it is in general no longer sinusoidal, although compounded of sine and cosine functions. It may be recalled that Fourier's theorem states that a periodic function can be written as the sum of sine and cosine functions. If, therefore, V_2 is a function of the time, then:

$$V_2(t) = b_0/2 + \sum_1^{\infty} (a_n \sin n\omega t + b_n \cos n\omega t) \quad (1)$$

where a_n and b_n are coefficients given by the equations:

$$a_n = \frac{1}{\pi}\int_{-\pi}^{+\pi} V_2(t)\ \sin\ n\omega t\ \mathrm{d}\ \omega t,$$

$$b_n = \frac{1}{\pi}\int_{-\pi}^{+\pi} V_2(t)\ \cos\ n\omega t\ \mathrm{d}\ \omega t.$$

The output voltage is thus made up of the fundamental frequency and its derived harmonics, so that *non-linear distortion is found to consist in the creation of new frequencies which are absent in the input signal.* The following example will demonstrate this. Assuming that

$$V_2 = AV_1 + BV_1^2 + CV_1^3.$$

then for the case where

$$V_1 = V_0 \cos \omega t:$$

we have:

$$V_2 = AV_0 \cos \omega t + BV_0^2 \cos^2 \omega t + CV_0^3 \cos^3 \omega t,$$
$$V_2 = {}^1/_2\, BV_0^2 + (AV_0 + {}^3/_4\, CV_0^3) \cos \omega t + $$
$$+ {}^1/_2\, BV_0^2 \cos 2\,\omega t + {}^1/_4\, CV_0^3 \cos 3\,\omega t.$$

In most cases the relationship between V_2 and V_1 cannot be expressed by a simple mathematical formula, but can usually be deduced from the form of the curve $V_2(t)$ (cf. *fig. 1*). The following rules are useful in this connection.

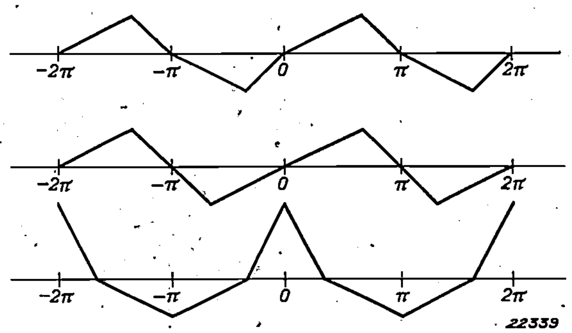


Fig. 1. *a)* Symmetrical function. *b)* Odd function. *c)* Even function.

a) A symmetrical function ($f(t) = -f(t+\pi)$) is composed of the fundamental wave and odd-numbered harmonics.

b) An odd function ($f(t) = -f(-t)$) is composed of sine functions only.

c) An even function ($f(t) = f(-t)$) is composed of cosine functions only.

The first rule in particular has an extremely useful application in amplifier technique for the purpose of reducing the distortion as far as possible. This will be considered in closer detail for a simple amplifying valve by studying the connection between the anode current and the grid voltage (*fig. 2*). The anode current produces a voltage drop at the resistance R which is proportional to the current. Take the point O as the working point. It is seen from the figure how the anode current,

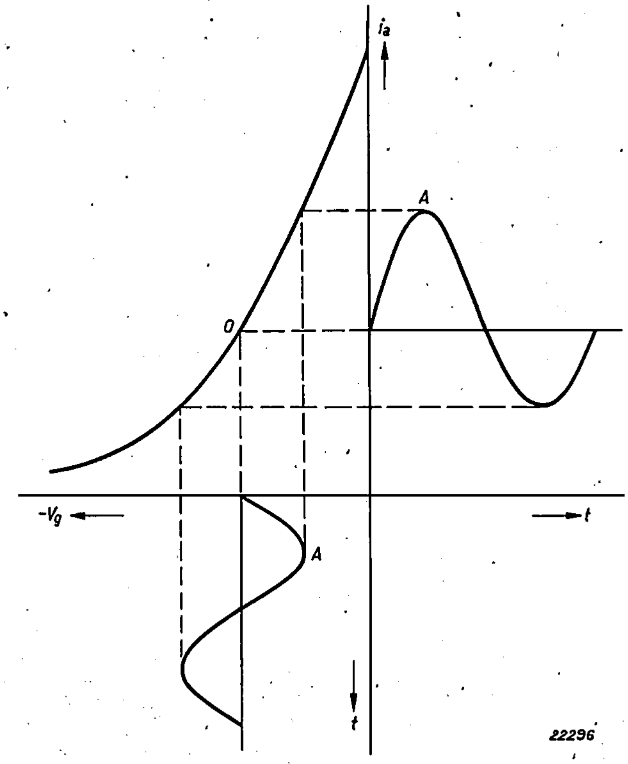
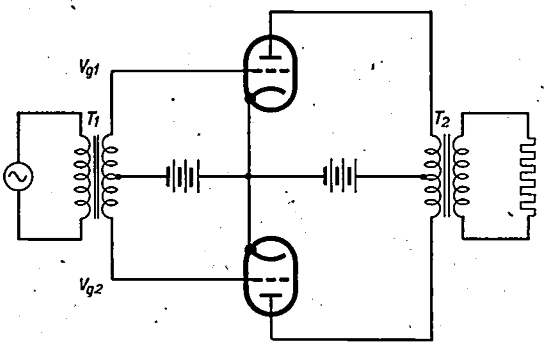


Fig. 2. Anode-current characteristic of an amplifying valve. The anode-current/grid-voltage characteristic is plotted in the top left-hand quadrant. O is the working point. The curve in the bottom left-hand quadrant shows the sinusoidal variation of the grid voltage. In the top right-hand quadrant the anode current is plotted against the time.

and with it the voltage drop at the resistance R, are dependent on the time, when grid voltage has a sinusoidal time function. It is evident that the curve is not symmetrical, for its curvature is much greater at low anode currents than at high currents. If zero time is taken at the point A, the anode current is found to be an even function of the time and on expansion gives a series of cosine functions only.

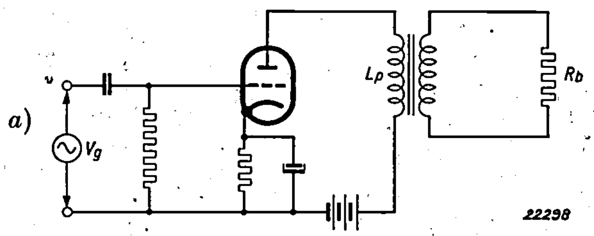
Now consider two valves connected in a circuit such that the anode current of one valve just reaches its maximum value when the current in the other valve is a minimum. This is the so-called push-pull circuit (see *fig. 3*). We must plot the common characteristic of the two valves. Without entering into details, it may be sufficient to indicate that a symmetrical arrangement is obtained by connecting the valves in push-pull, and in consequence all even-numbered harmonics disappear. In other words: the even-numbered harmonics, which are produced in each of the two valves, are brought into phase opposition so that they compensate each other. Since, in triodes, the second harmonic predominates throughout a wide range over all other harmonics, the distortion is very considerably reduced in this circuit.



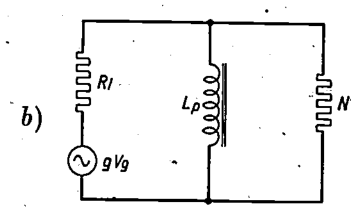
22297

Fig. 3. Push-pull circuit. The transformer T_1 is so wound that V_{g2} is a minimum when V_{g1} is a maximum. The anode currents in the two valves are thus in phase opposition; they are combined in the transformer T_2 in such a way that mutual amplification is obtained.

If the input voltage is not sinusoidal, analysis of the conditions becomes more complex. The signal must then be regarded as a sum of sinusoidal oscillations (Fourier series) of different frequencies and amplitudes, and the amplifier has to amplify this sum as a whole. New frequencies are then produced, not only harmonics but also combination tones, i.e. vibrations with a frequency of $m\omega_1 + m\omega_2$, where ω_1 and ω_2 are two input frequencies and m and n are integers. It is just these combination tones which reduce the quality of reproduction, but only in simple cases can their intensity be calculated. One of the principal difficulties is that with two frequencies the distortion is often not the same for both; however, this problem will not be discussed in the present paper.



22298



22299

Fig. 4. *a)* Circuit of an output stage. *b)* Equivalent circuit. The amplifying valve has been replaced by the internal resistance R_i in series with the generator gV_g (g = amplification factor of the valve). The transformer with the load resistance has been replaced by the primary self-inductance L_p with N^2 times the value of the load resistance in parallel thereto.

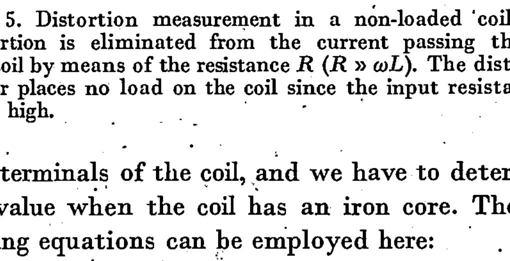
Magnetic Distortions

Those distortion effects will now be discussed in more detail which originate from transformers and choke coils with iron cores present in the circuit. Special reference will be made to choke coils, since transformers and choke coils have the same effect on the distortion phenomena under discussion.

To simplify analysis, we shall consider the distortion produced by an output transformer, and it will be assumed that the load on the amplifier is a resistance and not a loudspeaker as is usually the case (*fig. 4a*). We must first construct the equivalent circuit for this output stage. The amplifying valve can be represented by a current generator in series with the internal resistance of the valve (*fig. 4b*). Since only low frequencies are under consideration (the reasons for this will be discussed later) we can neglect the leakage and capacity characteristics of the transformer. The equivalent circuit of the transformer then consists of the primary self-inductance, which carries a load equal to N^2 times the load resistance (N = transformation ratio of transformer) (see fig. 4). For our purposes this circuit is still comparatively difficult to analyse, so that in the first place we will consider several simpler cases.

Distortion due to non-loaded Iron-core Coils

Undistorted current, i.e. pure sinusoidal current, is passed through the coil; this may be done by connecting a resistance in series with the coil the value of which is very large in comparison with the coil impedance (see *fig. 5*). An e.m.f. appears across



22340

Fig. 5. Distortion measurement in a non-loaded coil. The distortion is eliminated from the current passing through the coil by means of the resistance R ($R \gg \omega L$). The distortion meter places no load on the coil since the input resistance is very high.

the terminals of the coil, and we have to determine its value when the coil has an iron core. The following equations can be employed here:

$$H = \frac{0.4\pi n i}{l} \text{ oersted} \quad \cdots \quad (2)$$

$$V = n\,O\,\frac{\mathrm{d}B}{\mathrm{d}t}\,10^{-8}\,\text{volt/cm} \quad \cdots \quad (3)$$

where: H is the magnetic field intensity;
B is the magnetic flux density;
V is the voltage impressed on the coil;
i is the current through the coil;
n is the total number of turns;
l is the mean length of the lines of force, and
O is the cross-section in sq. cm.

Thus already we have an expression connecting the electrical voltage V and the magnetic flux B, and a second expression connecting the electric current i and the magnetic field intensity H. To find the relationship between V and i, we still lack an expression connecting B and H, to deduce which further information must be obtained concerning the magnetic properties of the core iron.

This subject has already been discussed in an article in a previous issue of this Review [1]).

We know that the relationship between B and H:
1) Is not linear,
2) But has an hysteretic character, in other words, a closed curve is traced in the BH plane, when H passes through a cycle (*fig. 6*). The hysteresis loop

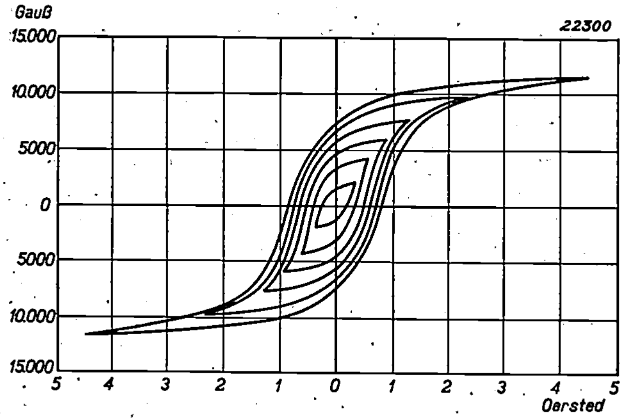


Fig. 6. Examples of hysteresis loops. The various loops relate to different values of B_{max}.

consists of two arms, a rising (lower) branch and a falling (upper) branch; when H increases the ascending curve is traced out and when H diminishes the descending curve is obtained. The form of the hysteresis loop is determined by variations in the magnetic flux density; at very large amplitudes saturation is obtained.

As yet it has not been possible to evolve a mathematical function for the hysteresis loop which is valid for all amplitudes. Only at low flux densities can a relatively simple expression be obtained and to which we shall confine our remarks for the moment. That the curve traced is a loop signifies that energy is lost in the core, these losses being the so-called hysteresis losses; the area enclosed by the loop is a measure of these losses. The form of the loop determines the distortion; if the loop were an ellipse no distortion would occur and merely a phase displacement of B with respect to H. But since the loops have cusps, distortion results. Assume a specific relationship connecting B and H, from which both the hysteresis losses and the distortion can be calculated. It is then found that for many materials the two arms of the loop can be represented by a quadratic equation in terms of the magnetic field intensity, i.e. if we take a new set of bh axes through the lower cusp of the hysteresis loop, the ascending arm is represented by the equation

$$b = \mu_0 \, h + \nu \, h^2 \quad \ldots \ldots \quad (4)$$

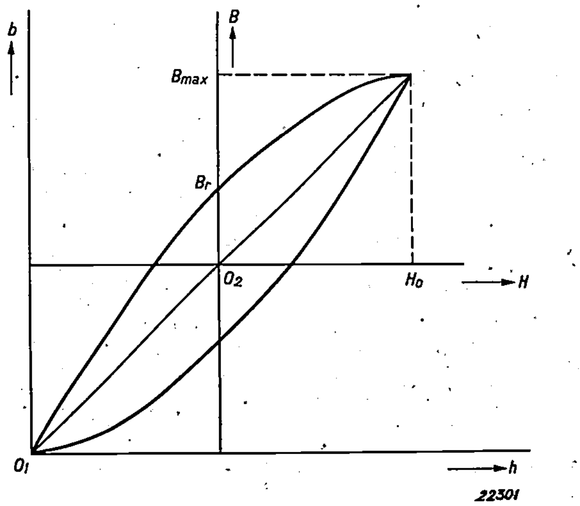


Fig. 7. Hysteresis loop at low amplitudes. The origin of the coordinate axes bO_1h is located at one cusp of the loop; the origin of the axes BO_2H is at the centre of the loop.

where μ and ν are constants of the material (μ_0 is the known initial permeability). This Rayleighs' "law" applies over a wide range about the point $B = H = 0$, i.e. one of the turning points of the loop. We can now deduce directly an expression for the whole of the hysteresis loop by locating the origin at the centre of the loop:

$$B = (\mu_0 + 2 \, \nu \, H_0) \, H \pm \nu \, (H_0^2 - H^2), \quad (5)$$

where $H_0 = H_{max}$, and the plus sign applies to the upper (descending) branch of the loop. From these formulae we get the expressions for the permeability and the remanence, thus:

$$\mu = \frac{B}{H_0} = \mu_0 + 2 \, \nu \, H_0 \quad \ldots \ldots \quad (6)$$

$$B_r = B_{(H=0)} = \nu \, H_0^2 \quad \ldots \ldots \quad (7)$$

1) For a detailed discussion, cf J. L. Snoek, Philips techn. Rev. 2, 77, 1937.

2) We here adopt the method of analysis due to H. Jordan, E.N.T, 1, 7, 1924.

From equation (5) connecting the field intensity and the flux density, and equations (2) and (3) connecting the electrical and magnetic magnitudes, we can calculate the voltage appearing across a coil when an undistorted, sinusoidal current: $i = i_0 \cos \omega t$; is passed through the coil. The magnetic field intensity is then, according to equation (2), proportional to i, thus:

$$H = H_0 \cos \omega t,$$

We thus get:

$$B = (\mu_0 + 2\,\nu\, H_0)\, H_0 \cos \omega t \pm H_0^2 \sin^2 \omega t, \quad \cdot \cdot \quad (8)$$

where the minus sign applies for $0 < \omega t < \pi$ and the plus sign for $\pi < \omega t < 2\pi$.

By Fourier analysis we can expand these expressions to a series of sine and cosine functions, thus:

$$B = (\mu_0 + 2\,\nu\, H_0)\, H_0 \cos \omega t + \frac{8}{3\pi}\,\nu\, H_0^2 \sin \omega t -$$

$$-\frac{8}{\pi}\nu\, H_0^2 \left\{\frac{\sin 3\omega t}{1\cdot 3\cdot 5} + \frac{\sin 5\omega t}{3\cdot 5\cdot 7} + \frac{\sin 7\omega t}{5\cdot 7\cdot 9} + \ldots\right\} \cdot \quad (9)$$

It is seen that odd-numbered harmonics alone appear in the magnetic induction. This was to be expected since we are dealing here with symmetrical conditions in which even-numbered harmonics disappear. The voltage applied to the coil is found from equation (3) to be:

$$\frac{V}{\omega\, n\, O\, 10^{-8}} =$$

$$= -(\mu_0 + 2\,\nu\, H_0)\, H_0 \sin \omega t + \frac{8}{3\pi}\,\nu\, H_0^2 \cos \omega t -$$

$$-\frac{8}{\pi}\nu\, H_0^2 \left\{\frac{\cos 3\omega t}{1\cdot 5} + \frac{\cos 5\omega t}{3\cdot 7} + \frac{\cos 7\omega t}{5\cdot 9} + \ldots\right\} \quad (10)$$

To express the voltage V in terms of the current i_0, we must substitute further in (10):

$$H_0 = \frac{0.4\,\pi\, n\, i_0}{l}.$$

Thus the voltage is made up of the following components:

1) A component with a phase displacement of 90 deg. with reference to the current,
2) A component in phase with the current, and
3) Harmonics.

1) Since $V = L\, di/dt$, where L is the self-inductance, we get for L:

$$L = \alpha\,(\mu_0 + 2\,\nu\, H_0), \quad \alpha = \frac{4\,\pi\, n^2\, O}{l}\, 10^{-9} \cdot \quad (11)$$

If the coil has no iron core and no resistance, this component alone is obtained.

2) Since the component is in phase with the current, energy is lost in the coil, viz: hysteresis loss, which may be expressed by saying that the coil has a loss resistance of R_h. This is given by the expression:

$$R_h = \alpha \frac{16}{3} f\, \nu\, H_0 = \beta \frac{16}{3} f\, \nu\, i_0, \quad \cdot \cdot \quad (12)$$

$$\beta = \frac{1.6\,\pi^2\, n^3\, O}{l^2}\, 10^{-9} \text{ and } f = \frac{\omega}{2\pi}\cdot$$

where f is the frequency.

3) This component is the distortion required. The amplitudes of the third, fifth and seventh harmonics are given by the expressions:

$$V_3 = \beta \frac{16}{5} f\, \nu\, i_0^2;\quad V_5 = \beta \frac{16}{21} f\, \nu\, i_0^2;$$

$$V_7 = \beta \frac{16}{45} f\, \nu\, i_0^2 \quad \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \quad (13)$$

It is seen that the amplitude of each harmonic increases with the square of the current. It now remains to compare the distortion voltage with the voltage of the fundamental wave.

If: $\nu\, H_0 \ll \mu_0$, we get:

$$\frac{V_3}{V_1} = \frac{0.2\, n}{l}\,\frac{16}{5}\,\frac{\nu}{\mu_0}\, i_0 \quad \cdot \cdot \cdot \cdot \quad (13a)$$

Hence the percentage distortion is independent of the frequency (the same loop is traced out at all frequencies) and is proportional to the current through the coil. Equation (10) gives for the ratio of the various harmonics:

$$V_3 : V_5 : V_7 = 1/5 : 1/21 : 1/45 \quad \cdot \quad (13b)$$

As an example the results of distortion measurements on a coil composed of nickel-iron sheets are represented in *fig. 8*. The equations satisfactorily

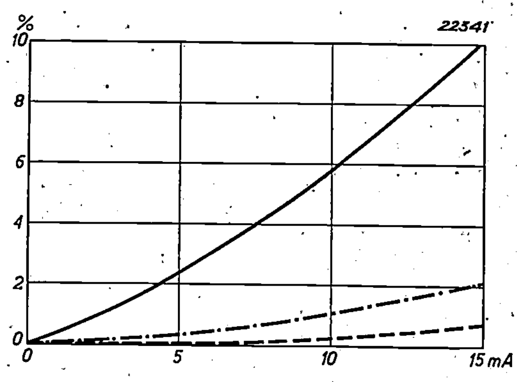


Fig. 8. Distortion in a non-loaded coil with nickel-iron core. ——— third harmonic —·—· fifth harmonic — — — — seventh harmonic.

represent the general shape of the curve, although systematic deviations are found to occur: The distortion is not absolutely proportional to the current, and the ratio of the various harmonics is not quite constant; the higher harmonics are not as high as found by calculation.

By combining equations (12) and (13) the relationship between the hysteresis resistance and the distortion is obtained, thus:

$$\frac{V_3/i_0^2}{R_h/i_0} = 0.6 \quad \cdots \cdots \quad (14)$$

With this expression the distortion of a coil can be calculated when the loss resistance is known. Equation (14), also, is not completely satisfied for this coil; the ratio is not constant, for the distortion does not vary linearly with the current and is smaller than demanded by this equation. On the other hand, it may be concluded from these experiments that in broad outline the theory gives a fairly accurate representation of the distortion at low flux densities.

Distortion in Loaded Coils

In the previous section we have discussed the distortion occurring in non-loaded coils, i.e. in coils which are not shunted by an external impedance. We will now examine the conditions when such an impedance is present, and assume that undistorted current again flows through the coil and impedance, and has such a large amplitude that the current through the coil is the same as in the absence of impedance (see *fig. 9*). Since the

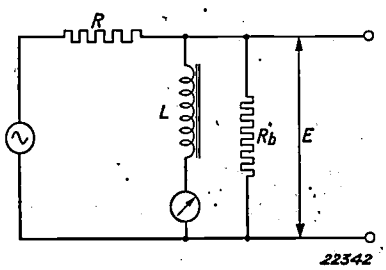


Fig. 9. Measuring the distortion in a loaded coil. By connecting the ammeter in series with the coil, distortion measurements are made for the same values of the fundamental wave as in fig. 4. With high distortion values a selective apparatus must be used, which gives an indication of the amplitude of the fundamental wave alone.

coil is now shunted, the undistorted e.m.f. across the coil will cause a current of the higher harmonic frequencies to flow in the shunt circuit formed by the coil and its external impedance. The flow of this current reduces the distortion voltage in the coil, and the quotient of this voltage drop and the distortion current may be regarded as an internal impedance with respect to the higher harmonics. At a first glance it might be assumed that this impedance has the same expression as applicable to the fundamental wave. But it must be remembered that the coil, which is traversed by both the distortion current and the fundamental wave, is not a linear impedance and in view of this these currents are not independent of each other. The detailed calculations cannot be given here; only the results obtained will be stated.

The impedance for the fundamental wave (i_1) is represented by the expression:

$$Z(i_1) = R_h(i_1) + j\omega L(i_1) \quad \cdots \quad (15)$$

where R_h and L are function of i_1; in addition R_h is proportional to the frequency.

For the internal impedance in respect of the third harmonic we then get:

$$Z(i_3) = 26/35\, R_h(i_1) + j \cdot 3\,\omega L(i_1) \cdot (16)$$

It should be pointed out that in this expression the terms R_h and L stand for values related to the amplitude of the fundamental wave. As long as the induction in the iron is low, R_h will remain small (R_h is proportional to the current), so that in the majority of cases we have to take the self-inductance alone into consideration. The same applies for the higher harmonics also. In this way we can calculate the distortion current flowing through the circuit, and hence also the voltage obtained in the internal impedance as well as that applied to the coil. To simplify matters, the external impedance is assumed to be a resistance. The distortion then becomes smaller when the coil carries the resistance as a load. Since the internal impedance increases with the order of the harmonics, the higher harmonics will be reduced to a greater extent than the lower harmonics. Hence, if we make the external resistance equal to ωL for the fundamental wave, the third, fifth, seventh and ninth harmonics will be reduced to:

$$\frac{1}{\sqrt{10}},\ \frac{1}{\sqrt{26}},\ \frac{1}{\sqrt{50}} \text{ and } \frac{1}{\sqrt{82}}$$

of their original values. In practice equation (16) is adequately satisfied. Although it has been deduced for low flux densities (Rayleigh's formula) it is evident that it represents quite satisfactorily the conditions at high flux densities also.

Nevertheless, in certain cases, the distortion may increase when the coil is shunted by an impedance. This may be demonstrated by using as load, a condenser with such a capacity that at the frequency of one of the harmonics it is in resonance with L. The circuit then constitutes a series resonance for this harmonic, and the current flowing through the circuit is limited solely by the loss resistance of the circuit, i.e. by the hysteresis resistance. Hence this current will be very high and in consequence a high voltage will be impressed on the condenser by this harmonic also.

Eddy-Current Coils

In our analysis we have not taken into consideration that eddy-current losses occur frequently in a coil with an iron core. To examine to what extent these currents will modify the results we have obtained, we must again refer to the equivalent circuit of the coil.

We will assume that the frequency is so low that no appreciable displacement of the lines of force as yet occurs as a result of eddy currents, so that the induction is uniformly distributed over the whole cross-section of the iron. In an equivalent circuit, the eddy currents can be represented as flowing in an eddy-current loss resistance. It may be asked whether it is more correct to put this resistance in series or in parallel with the coil. This can be determined by considering the origin of the eddy currents; it cannot be established experimentally, since each circuit can be transformed by calculation into the other. We shall analyse the two forms of circuit arising (see *fig. 10*):

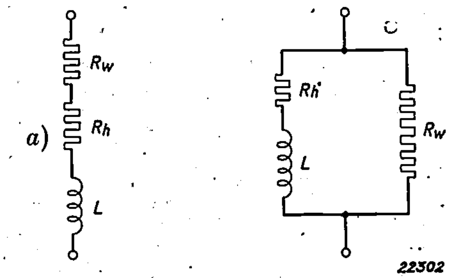


Fig. 10. *a)* Circuit with a coil in series with the eddy-current resistance. L = self-inductance, R_h = hysteresis resistance, R_w = eddy-current resistance.
b) The same circuit with eddy-current resistance in parallel.

1) Eddy-current Resistance in Series.

This circuit does not show that the eddy-current losses are due to currents flowing in the iron, for a circuit of this type generates a magnetic field which is opposed to the initial field. Eddy currents thus reduce the magnetic flux in the core at a given current. This is not indicated in the circuit diagram for the initial current continues to flow through the self-inductance.

2) Eddy-current Resistance in Parallel.

In this arrangement the formation of eddy currents is clearly illustrated. The eddy-current resistance constitutes a load on the coil; if the external current is kept constant the current in the self-inductance must drop. In this circuit, the eddy currents differ in no way from an ordinary load; the circuit is definitely better than the previous one. Furthermore it is an advantage that, contrary to a series resistance whose value increases with the square of the frequency, the parallel resistance is independent of the frequency at low loss values.

The effect of eddy currents on the distortion is now evident. Similar to a resistance due to an external load, these losses reduce the distortion. If the distortion is measured at such a high frequency that the eddy currents can affect the result (their effect increases with the frequency, since the resistance in parallel remains constant), then the value obtained is not the same as the distortion e.m.f. measured without an external load on the coil.

Distortion due to Transformers

After these preliminary remarks, we can return to a discussion of the distortion in an output stage. To make direct use of the above results, we shall make a small alteration in the equivalent circuit (fig. 4b). Previously we substituted for the amplifying valve a voltage source furnishing a voltage $g\ V_g$ and connected it in series with the internal resistance of the valve. The valve can also be replaced by a current source of intensity $S\ V_g$ (S = slope of the amplifying valve) which is connected in parallel with the internal resistance of the valve (*fig. 11*). The circuit is then identical to that shown

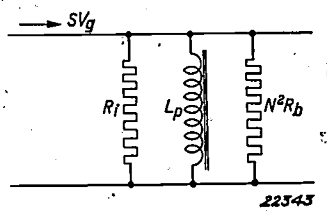


Fig. 11. Equivalent circuit for the output stage. The amplifying valve has been replaced here by a current source SV_g with R_i in parallel.

in fig. 9, when R_i and N^2R_b are combined to form a single resistance. The results arrived at above can then be directly applied; the magnitude of the distortion is determined by the current intensity and by the ratio of the impedance values of the transformer (at different harmonics) to the equivalent resistance of R_i and R_b. It is now seen why a marked distortion has to be expected at low frequencies only. At high frequencies the internal impedance rapidly becomes too great to permit large distortion currents, so that distortion practically disappears; moreover, the current passing through the coil diminishes steadily as the frequency increases.

An amplifier must not only be devoid of distortion but, at a specific low frequency, the amplifier gain must not diminish by more than a certain fraction. Again here the ratio of ωL to $\dfrac{R_iR_b}{R_i + R_b}$ is a determining factor, as may be seen from the following. The gain of the circuit in fig. 11 is given by:

$$V_2/V_1 = \frac{g\,\omega L\,R_b}{\sqrt{R_i^2\,R_b^2 + \omega^2 L^2\,(R_i + R_b)^2}}\,, \quad (17)$$

where V_2 is the voltage applied to R_b and $V_1 = V_g$. The amplification at high frequencies is therefore:

$$V_2/V_1 = \frac{g\,R_b}{R_i + R_b} \quad \dots \quad (18)$$

If p is the ratio of the gains at low and high frequencies we have:

$$\frac{V_2/V_1(\omega_1)}{V_2/V_1\,(\omega = \infty)} = p\,(\omega_1)\,,$$

so that from (17) and (18) we get:

$$\omega_1 L = \frac{R_i\,R_b}{R_i + R_b}\sqrt{\frac{p^2}{1-p^2}} \quad \dots \quad (19)$$

We can now calculate the magnitude of L for the amplifier gain to be reduced to a fraction p at a specific frequency ω_1. It is seen that when

$$\omega_1 L = \frac{R_i\,R_b}{R_i + R_b}$$

the gain drops to 70 per cent of its normal value. Therefore, to determine both the frequency characteristic and the magnitude of the distortion, we must compare the impedance of the transformer with the external load. In some cases the frequency characteristic will be satisfactory, while the distortion will be too high; in other cases the reverse will be found. This will depend entirely on the characteristics of the transformer iron. With $p = 0.7$ the internal impedance with respect to the third harmonic is already three time greater than the external impedance, so that the distortion at this frequency is already considerably reduced. At still lower frequencies, the distortion naturally becomes greater.

In the example dicussed here, a direct current also flows through the coil, which destroys the symmetry of the hysteresis loop, so that even-numbered harmonics also are produced. For the rest the same considerations as given above apply in this case.

Distortion at High Induction Values

The above analysis is only valid for low flux densities. As soon as these latter become large the problem is no longer susceptible to mathematical treatment. Fortunately, the phenomena discussed above then have a more or less similar form, so that it is unnecessary to measure the distortion either for all flux densities or for all practical loads. If it is known how the self-inductance values and the distortion vary with the amplitude in the case of the non-loaded coil, we can estimate the approximate distortion with different loads. However, in this case it is very difficult to take eddy currents into consideration, since in the majority of materials the permeability is dependent on the amplitude. It becomes necessary therefore, to avoid these eddy currents as far as possible, in other words to carry out measurements at low frequencies, especially as the distortion of amplifier coils is only serious at low frequencies.

Finally, *fig. 12* shows as a further example the

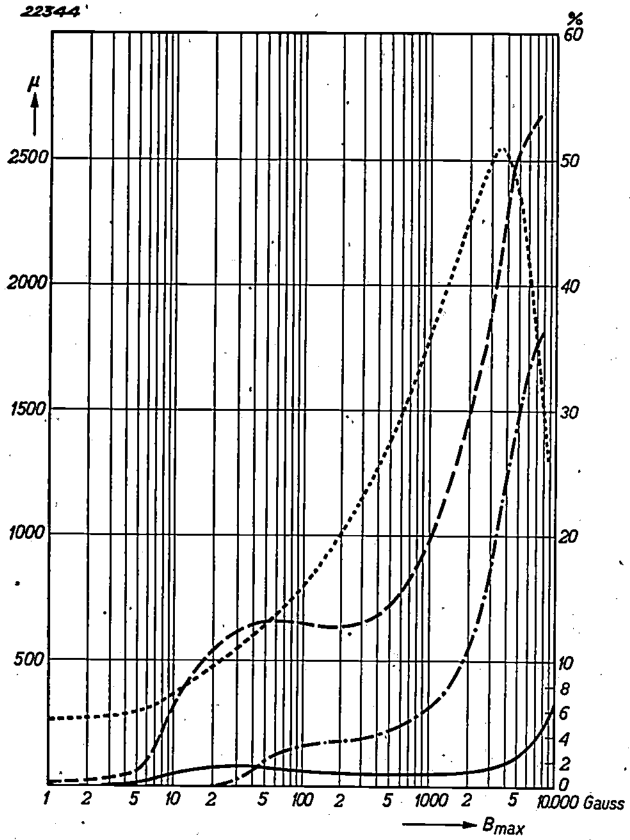


Fig. 12. Distortion of a silicon iron plotted in relation to B_{max}. ...... Permeability, — — — — — Third harmonic on no load. —·—·—·—·— Fifth harmonic on no load. ———— Third harmonic with a load equivalent to a resistance $R_b = \omega L$, where ω is the circular frequency of the fundamental wave.

distortion in silicon iron on no-load, as well as on load, these values being obtained at the fundamental frequency $\omega L = R_b$. It is seen particularly clearly that the internal impedance of the coil satisfactorily fulfils its purpose of reducing the distortion. The continuous curve indicates that these laminations are still effective at $B = 8000$ gauss, with a distortion of less than 5 per cent, and that on the whole the distortion is negligible as compared with that due to the amplifying valve itself.

HIGH-VACUUM GAUGES

by F. M. PENNING.

Summary. The various principles for measuring high vacua are briefly discussed. Some new designs of gauge are described in detail, particularly those which are based on the characteristics of electric gas discharges.

Definition of Gas Pressure; Units

The pressure of a gas is defined as the force applied by the impact of the gaseous molecules against unit surface of the walls of the enclosure containing the gas. The C.G.S. unit in which gas pressures are measured is 1 dyne per sq cm and is termed a microbar [1]). Low pressures are usually expressed in units of millimetres of mercury, a pressure of 1 mm. of Hg being termed the Torricelli or sometimes the Tor. The relationship between these units is:

1 mm Hg = 1 Tor = 1333 microbar = 1333 dynes per sq cm = 1/760 atmosphere.

In this paper, discussion will be confined to gas pressures below one atmosphere, and particularly to those below 1 mm Hg.

Principles of Pressure Measurements

Apart from a direct measurement of the gas pressure, every other property of a gas which varies with the pressure can theoretically be employed as a measure of that pressure. The associated properties which can be used for this purpose are:

1. Gas pressure,
2. Thermal conductivity of the gas,
3. Internal friction or frictional drag,
4. Radiometer effect,
5. Dependent gas discharge, and
6. Independent gas discharge.

Where high-vacuum gauges based on these principles have already been described in the literature (see bibliography), brief mention only will be made to them in the present article; while a more detailed description will be given of any individual new types which have been devised.

Gauges depending on a measurement of the Gas Pressure per se

Manometers or pressure gauges based on this principle obtain a pressure reading by measuring the displacement of the surface of a liquid (liquid gauges) or of an elastic (mechanical gauges). To magnify the sensitivity of measurement, the pressure can be multiplied before measurement, viz, by compression in a known ratio (MacLeod gauge). The surface, which is displaced under the action of the pressure, need not form part of the wall of the actual enclosure, but can be freely suspended. This method is adopted in the

Molecular-Jet gauge (*fig. 1*), which was

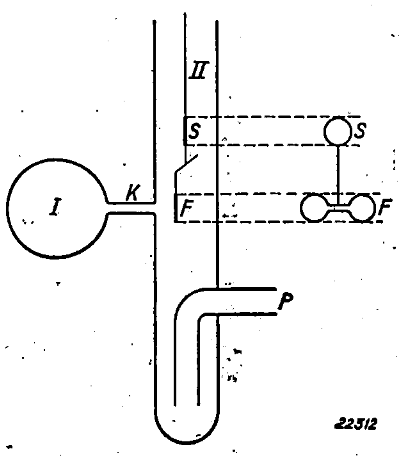


Fig. 1. Molecular-jet gauge for measuring vapour pressures (diagrammatic). The space *II* is connected to a pump at *P*; the vapour pressure of the substance in *I* deflects the vane *F* from its equilibrium position.

designed by Mayer for determining very low vapour pressures [2]). The substance whose vapour pressure is to be determined is contained in the bulb *I*, which is connected to the tubular vessel *II* through a capillary; *II* is highly evacuated through the wide connecting tube *P*. In vessel *II* a small, very light vane with two aluminium discs, 0.01 mm thick and 4 mm in diameter, is suspended; one of the discs is situated directly in front of the opening of *K*. The stream of molecules issuing from *K* displaces the small vane from its position of equilibrium, the magnitude of the angle of deflection being measured by means of a small mirror *S*. In the gauge constructed, this angle of deflection was found to be proportional to the pressure below a limit of 10^{-2} mm. The gauge must be calibrated with a substance of known vapour pressure.

[1]) A considerable divergence of opinion is found in the literature in regard to the units bar and microbar (barye). In Germany and the United States the C.G.S. unit has sometimes been termed the bar. At present the universal definition of these units is: 1 bar = 10^6 dynes per sq cm = 0.99 atmosphere and 1 barye = 1 microbar = 1 dyne per sq cm.

[2]) H. Mayer, Z. Physik, 67, 240, 1931.

Mayer obtained a sensitivity of 10^{-6} mm Hg per mm deflection. A serious drawback of this gauge is that it is very susceptible to tremor, and it must, therefore, be carefully set up to protect it against extraneous vibrations. By virtue of its construction this gauge cannot be used for measuring constant gas pressures.

Gauges depending on the Variation of Thermal Conductivity with Pressure

One of the most remarkable achievements of the kinetic theory of gases was to account for the fact that the thermal conductivity of a gas is independent of its pressure. From this it appears impossible to construct a pressure gauge depending for its action on variations in the thermal conductivity. But this conductivity is independent of the pressure only within a narrow range of pressures, for at high pressures the transmission of heat through a gas is a function of the pressure, since in addition to conduction convection of heat also takes place. Theoretically, therefore, a thermal-conductivity gauge is feasible within this range of pressures; a gauge of this type has indeed been described.

At low pressures, the thermal conductivity is independent of the pressure only for as long as the free path λ of the gaseous molecules is small compared with the distance d between the hot and the cold masses ($\lambda \ll d$). It is evident that at zero pressure ($\lambda \gg d$) the thermal conductivity also must be zero and that there must hence be a transition region in which the thermal conductivity varies with the pressure. If $d = 1$ cm, then $\lambda = d$ for a pressure of approximately 0.01 mm Hg and a transition stage becomes feasible between 10^{-1} and 10^{-3} mm Hg.

In this transition region the pressure can be measured as follows: A metal radiator to which a constant quantity of heat is supplied per second is fixed along the centre of a glass tube. The heat loss to the glass wall determines the temperature of the metal, which is hence a measure of the pressure. Gauges of this type require calibration before use, e.g. against a MacLeod gauge. The more common forms of this gauge are:

1) The Pirani Gauge.

In this gauge, a thin wire which is heated electrically and whose resistance is a measure of the temperature is generally used as a metal radiator (*fig. 2a*).

2) The thermo-couple gauge.

In addition to the resistance, other properties of a metal, such as the thermo-electric force, can also be employed for measuring the temperature. In this case a heated wire is fitted with a thermo-couple (*fig. 2b*), and instead of passing an electric

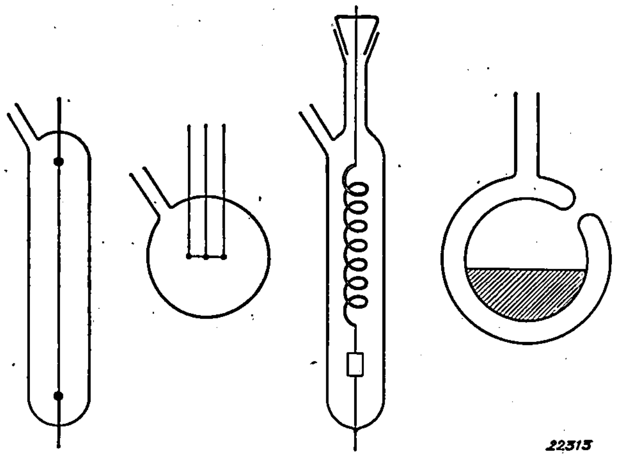


Fig. 2. Principle of the thermal-conductivity gauge.
a) Resistance (Pirani)
b) Thermo-couple (Voege and Rohn).
c) Bimetallic (Klumb and Haase)
d) Dewar flask (Herzog and Scherrer).

current through the wire, heat is supplied to it by irradiation of the thermo-couple with a light-source of constant intensity. In both arrangements the temperature to which the thermo-couple is raised is a measure of the pressure.

3) Bimetallic manometer.

Recently Klumb and Haase [4]) made use of the change in shape of a bimetallic strip (straight strip or helix) to measure the temperature. The application of the principle of a bimetallic helix to the construction of a gauge is shown in *fig. 2c*. The helix is here again heated by the electric current; the rise in temperature causes the small mirror attached to the helix to turn through a specific angle, such rotation being compensated by a torsion head. The number of degrees which the torsion head has to be turned is a measure of the pressure. Results obtained with this gauge for a number of gases are shown in *fig. 3*. The time taken to obtain a reading with the gauge is comparatively long, being about 1 minute.

4) Thermal-conductivity gauge depending on the Rate of Volatilisation of Solid or Liquid Substances.

A gauge constructed on this principle has been described by Herzog and Scherrer [5]). The vessel in which the pressure has to be measured communicates with the space between the walls of a

[3]) Th. Haase, G. Klages and H. Klumb, Phys. Z., **37**, 441, 1936.

[4]) H. Klumb and Th. Haase, Phys. Z., **37**, 27, 1936.

[5]) G. Herzog and P. Scherrer, Helv. Phys. Acta, **6**, 277, 1933.

double-walled sphere which is filled with solid carbon dioxide (*fig. 2d*). The sphere is thus a type

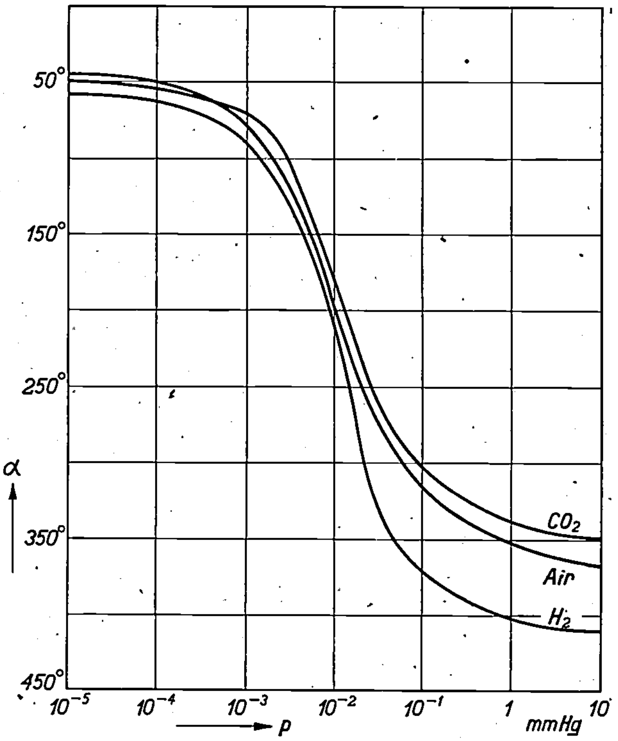


Fig. 3. Angle α through which the torsion head of the bimetallic manometer must be turned to compensate the rotation of the mirror.

of Dewar flask in which the carbon dioxide volatilises the slower the lower the pressure of the gas between the walls of the vessel, so that the rate of volatilisation of the carbon dioxide is a direct measure of the pressure. This gauge is not intended for accurate pressure measurements.

Gauges depending on the variation of Internal Friction with Pressure

Similarly to the thermal conductivity, the internal friction of a gas is also independent of the pressure as long as $\lambda \ll d$, but over a specific range of pressures this property also can be used for measuring high

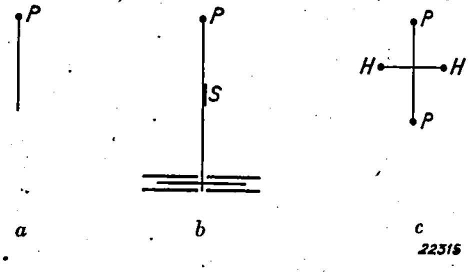


Fig. 4. Principles of the damping-type gauges.
a) Oscillating quartz fibre.
b) Rotatable disc with small mirror S.
c) Rotatable quartz cross *HH*; the surface of *H* is used as a convex mirror. The point of suspension is marked *P* in each case.

vacua. Gauges operating on this principle again require special calibration, while the quantitative readings obtained depend on the type of gas under measurement. There are two types of this gauge, viz:

1) Damping gauges, in which the damping of linear or circular oscillations is measured. Of the three forms which have been devised, the first has a thin fibre of glass or other material which is set in vibration by an impulse (*fig. 4a*), while in the second and third forms circular oscillations are imparted by small magnets to an oscillator, either a disc (*fig. 4b*) or a small quartz cross (*fig. 4c*), suspended by a fibre; in the latter case one of the small spheres on the cross acts as a mirror. In all three forms the damping of the motion is a measure of the pressure.

2) Molecular gauge of Langmuir and Dushman in which a rapidly-rotating aluminium disc *A* applies a torque to a thin mica disc *M* (*fig. 5*), this torque being directly proportional to the pressure of gas in the gauge.

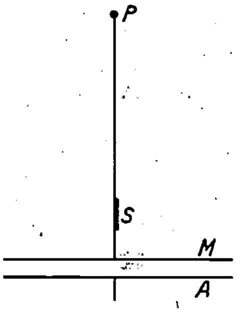


Fig. 5. Molecular gauge of Langmuir and Dushman. *A* - Rapidly-rotating aluminium disc; *M* - mica disc, whose deflection from the equilibrium position is read off by means of the small mirror *S*.

Gauges depending on the Radiometer Effect

These gauges are based on the principle that in very highly-rarefied gases (free path » dimensions of the containing vessel) two surfaces at different temperatures exert a mutual mechanical force. The rotation of the vanes in the Crookes' radiometer also is due to this force. Consider a flat plate *B* (*fig. 6*) at the absolute temperature *T*

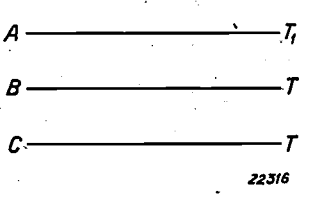


Fig. 6. Mechanical force applied to a plate *B* at the temperature *T* situated between a second plate at the same temperature and a third plate at a higher temperature T_1. Owing to the higher velocity of the molecules arriving from above, the pressure applied to the upper side of *B* is greater than that applied to the under side.

which is located between two other plates A and C at temperatures of T_1 and T respectively. At equilibrium the number of molecules striking against each sq cm of the three plates will be the same for each plate, viz, n. At the low gas pressures in question here nearly every molecule which strikes against B, will have registered its last impact at the surface of the opposite plate and will therefore be at the temperature of this plate. Molecules at a temperature of T_1 thus strike against the upper surface of B and molecules at a temperature of T against the lower surface; the molecules arriving from B are also at a temperature of T. The force applied to B per sq cm is, according to the kinetic theory of gases, $nm\ (v_a + v_b)$, where m is the molecular weight of the gas in question, v_a and v_b the average velocity components perpendicular to the surface with which the gaseous molecules either arrive at or leave the surface respectively, and n is the number of collisions per sq cm per sec. Since v is proportional to $\sqrt{T}$, i.e. $v = c\sqrt{T}$, the difference in pressure p between the top and bottom surface of B is:

$$\Delta p = m\,n\,c\,(\sqrt{T_1} - \sqrt{T}).$$

If the external wall of the enclosure is at the temperature T the pressure in the rest of the enclosure will be

$$p = 2\,m\,n\,c\,\sqrt{T}$$

and hence

$$\Delta p = \frac{p}{2}\left(\frac{\sqrt{T_1}}{\sqrt{T}} - 1\right).$$

The resultant force is, therefore, independent of the type of gas in the gauge.

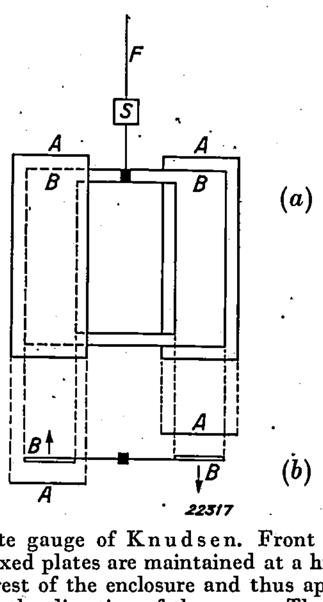


Fig. 7. Absolute gauge of Knudsen. Front view (a) and plan (b). The fixed plates are maintained at a higher temperature than the rest of the enclosure and thus apply a force on the vanes B in the direction of the arrow. The rotation of B is read off by means of the small mirror S.

1) Absolute gauge of Knudsen.

The absolute gauge of Knudsen consists of the type in fig. 6, but with a glass wall in place of the plate C. The two strips B (*fig. 7*) have been combined to form a small frame which is suspended by a fibre F. The fixed plates A are electrically heated to a temperature T_1, which is higher than the rest of the enclosure, and thus apply a pressure on B in the direction of arrow; the rotation of B which is read off by means of a small mirror attached to F, is a measure of the pressure. The absolute value of the pressure is obtained from the dimensions and the period of oscillation, etc., of the system using the equation given above. The Knudsen gauge is the only gauge which always give a direct reading of the absolute pressure of a gas; this is only possible with the Mac Leod gauge when the gases and vapours present obey Boyle's law.

A short time ago Du Mond and Pickels[6]) described a new form of this gauge, which is suitable for general purposes. In this design a scale is provided on the gauge, which permits a direct reading of the pressure between 10^{-4} and 10^{-6} mm Hg. This particular gauge also requires calibration, and still shares the same disadvantages as the original Knudsen manometer in that it needs careful handling, is difficult to make and can only be used when suitably protected against vibration.

2) Aluminium-Foil gauge of Knudsen.

This gauge is based still more closely on the principle illustrated in fig. 6. A vertically-suspended aluminium foil of the type used in electroscopes here serves as the plate B, its deflection from the equilibrium position being a measure of the pressure.

3) The Molecular vacuum-meter of Gaede[7])

A new gauge of this name and based on the principle under discussion here was recently placed on the market. Two strips HH mounted along the glass wall (*fig. 8*) which are raised to a suitable temperature by electrical heating elements here act as the hot plates. A small frame with two thin strips BB is suspended from a fibre. The equilibrium position which BB assumes under the action of the gas pressure lies in the plane EE and is made to coincide with the elastic equilibrium position. To displace the frame BB, e.g. to the position shown in fig. 8, a specific force which is directly proportional to the pressure must be applied to it. This deflection is obtained by passing a current i through

6) J. W. M. Du Mond and W. M. Pickels, Rev. Sci. Instr., **6**, 362, 1935.

7) W. Gaede, Z. techn. Phys., **15**, 664, 1934.

the coils S, thus causing a rotation of the small magnet M suspended from the fibre. The magnitude of i for a particular position of BB is, therefore, a measure of the pressure, which latter is read

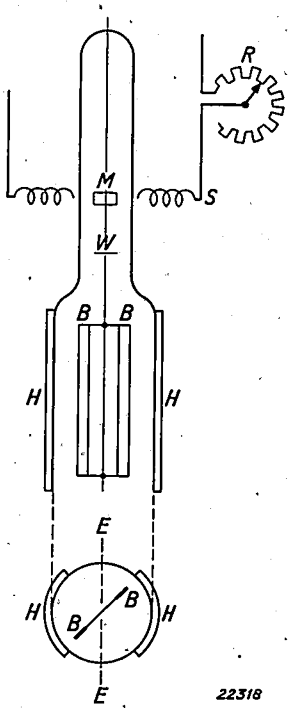


Fig. 8. Molecular vacuum-meter of Gaede. The strips H on the glass wall are raised to a temperature higher than the rest of the enclosure by means of heating elements placed in contact with them. A thermal directive force is applied to the small frame BB provided it is not kept in its equilibrium position EE by an elastic opposing force. By means of the small magnet M and the current i in the coils S, the frame is brought back to its initial position. i is proportional to the pressure in the tube. The magnitude of i which is read off from the adjustment of the variable resistance is a measure of the pressure.

off directly on the control knob of the variable R. The position of BB is adjusted by means of the pointer W. This gauge is calibrated empirically, although the deflection is independent of the type of gas used. Since the "thermal" directive force determines the period of oscillation of the system as well, the latter can also be taken as a measure of the pressure; the relative accuracy of this method is stated to be one per cent. The gauge can be used as a damping manometer also (see above), in which case readings are governed by the internal friction, i.e. to a first approximation by the molecular weight of the gas present. By combining the two methods, the molecular weight can be determined in addition to the pressure.

Gauges based on a Dependent Gas Discharge

In the case of a dependent gas discharge the primary carriers of electricity are nôt produced by the discharge itself, but derived from an external source, e.g. from a hot wire or by an actinic action. If the electrons, ions or radiation quanta of sufficient energy generated in this way pass through a gas the latter will be ionised; if the pressure of the gas is sufficiently low the number of ions and electrons produced will be proportional to the pressure and the path traversed by the ionised particles. *Fig. 9* shows the number of ionisations N

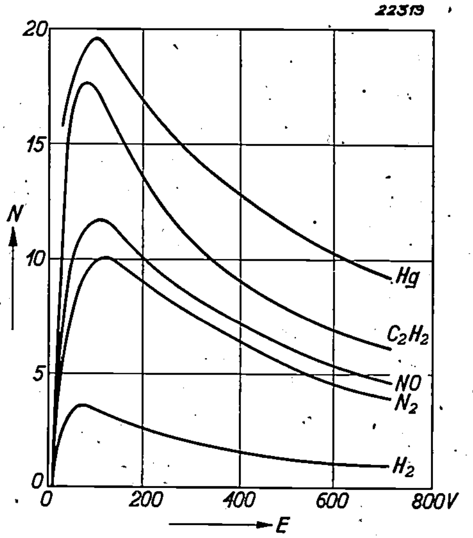


Fig. 9. Ionisation by electrons plotted against the accelerating potential E in volts. N = number of ionisation reactions per cm of path and per mm pressure.

per cm of path and per mm of pressure of the electrons in a variety of gases as a function of the electron energy. It is seen that for most gases N is a maximum at an energy of the order of 100 electron-volts, and that therefore for measuring low pressures a mean electron energy of this magnitude should be used. Since a much smaller number of ionisation reactions is produced by positive ions and radiation quanta of moderate energy, the importance of using electrons for ionisation is evident.

Ionisation Gauge.

In the gauge of this name, ionisation is produced by electrons emitted from a hot tungsten filament F (*fig. 10*). The apparatus is generally in the form of a triode, in which the grid G and the anode A are concentric with the heated filament F. One of the electrodes A or G is maintained at a negative potential with respect to the filament, so that ions alone flow towards it (current i_+) while the primary and secondary electrons can move only towards the third and positive electrode (current i_-). As long as the free path of the electrons is large compared to the path traversed, the pressure is then proportional to i_+/i_-. If i_- is made large, e.g. 10 milliamps, and a sensitive galvanometer is used for measuring i_+, pressures down to 10^{-6} mm can be measured. If desirable, the ion current also can be amplified to facilitate measurements.

The maximum sensitivity with the method is obtained when G is made positive, for part of the

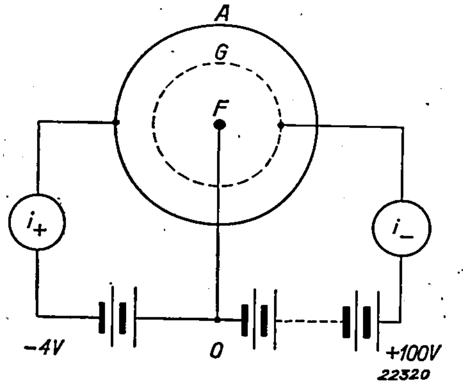


Fig. 10. Ionisation gauge. The ion current i_+ flows to A, and the electron current i_- to G. The pressure is proportional to i_+/i_-.

electrons will then shoot through the mesh of the grid G and describe longer paths before coming back to G, thus having more chance to cause ionisation. By means of a magnetic field the path of the electrons can be still further lengthened [7]).

It follows from fig. 9 that the readings of the ionisation gauge depend on the type of gas used. One of the first observations of Found and Dushmann was that the ionisation was proportional to the molecular weight of the gas used, but it is seen from fig. 9 that this is not the case, and more accurate measurements carried out later by Reynolds [8]) with ionisation gauges showed that this proportionality is not generally valid.

Instead of by direct measurement of the ion current, the pressure of a gas can be determined in still another way with the ionisation gauge. If A is taken as the anode and G as the negative electrode, the insertion of a resistance in the grid circuit will not affect the anode current in any way if the electrodes are located in a vacuum, since no grid current will flow. When gas is admitted to the gauge a positive ion current will flow towards the grid, and the insertion of a resistance in the path of this current will alter the grid voltage and in consequence the anode current also. The alteration in current may be used as a measure of the pressure [9]). Another feasible method is to use the ion current for charging a condenser which is connected either to the grid of another vacuum triode or to a relay-amplifying valve so that the anode current of the second valve is dependent on the condenser voltage [10]). The total rate of change of anode current of this second valve is then a measure of the pressure.

Finally, instead of deriving an electron saturation current from the cathode, the anode voltage can be made so small that the current in vacuo is limited by the electron space charge. If positive ions are then generated at low gas pressures, these are able to neutralise the space charge due to a larger number of electrons and thus allow the cathode current to increase considerably. In this method of measurement proposed by Spiwak and Ignatow [11]) the anode current itself is a measure of the pressure. Theoretically this method was already employed some years previously by G. Hertz [12]) for measuring the ionisation potential of gases.

Gauges based on an Independent Gas Discharge.

In an independent gas discharge, the particles responsible for carrying the charge are generated in the gas or at the electrodes by the discharge itself and without any external agency other than the applied voltage. Although in this case a number of factors, such as the current, current density, and the dimensions of the various components of the discharge, are dependent on the pressure, these have hitherto been used rarely for quantitative pressure measurements, since in general they are difficult to reproduce and also vary within wide limits with the characteristics of the gas present [13]). In order to obtain a rough idea of the vacuum in a glass vessel, a common method used in high-vacuum technique is to impose a high-frequency electrical field and observe the luminous effect resulting therefrom. Down to about 10^{-3} mm Hg the gas itself then appears luminous, while at lower pressures the glass walls alone exhibit a green fluorescence.

This form of gas discharge ceases at a gas pressure of approximately 10^{-3} mm, when using a direct voltage or low-frequency alternating voltage in a valve of moderate dimensions and at pressures of the order of 1000 volts, and for this reason will be quite useless for making measurements at low pressures. But by using a magnetic field and suitably disposing the electrodes, measurements

[8]) N. B. Reynolds, Physics, **1**, 182, 1931.

[9]) H. Teichmann, Z. techn. Phys., **9**, 22, 1928.

[10]) R. Sewig, Z. techn. Phys., **12**, 218, 1931.
A. Butschinsky, Techn. Phys. U.S.S.R., **3**, 223, 1936.

[11]) A. Spiwak and A. S. Ignatow, Sow. Phys., **6**, 53, 1934.

[12]) G. Hertz, Z. Phys., **18**, 307, 1923.

[13]) By using high alternating voltages (7000 volts eff.) Wellauer obtained good results in a range of pressures from $5 \cdot 10^{-3}$ to 10^{-1} mm Hg (Arch. f. Elektrot., **24**, 4, 1930).

can be made at much lower pressures and on this principle a simple gauge can be constructed for pressures down to 10^{-5} mm.

Philips' Vacuum-meter

The Philips' vacuum-meter is based on the principle just discussed and will be described in greater detail by reference to *fig. 11*. If two plates

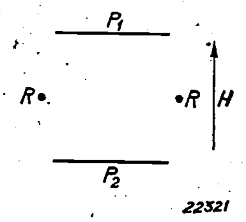


Fig. 11. Principle of the Philips' Vacuum-meter. The electrons at the plates P_1 and P_2 (connected together as a common cathode) do not move along straight lines to the ring *RR* (anode), but oscillate in spiral paths between P_1 and P_2 under the influence of the magnetic field.

P_1 and P_2 situated opposite to each other form the cathode for the discharge and a ring *R* acts as a anode, then in the absence of an applied magnetic field the electrons emitted from P_1 and P_2 will travel along curved paths to the ring *R*. But if a sufficiently powerful magnetic field is applied in the direction perpendicular to P_1, the electrons emitted from P_1 will travel in narrow spirals round the magnetic lines of force towards P_2, will be repelled towards P_1 by the retarding electric field and will thus oscillate many times between P_1 and P_2 before reaching the anode. In consequence, even at very low gas pressures the electrons will be able to collide with a sufficient number of gas molecules producing an adequate ionisation to permit an independent discharge through the gas.

The gauge constructed on this principle is shown in *fig. 12*. The plates *P* and the ring *R* are located

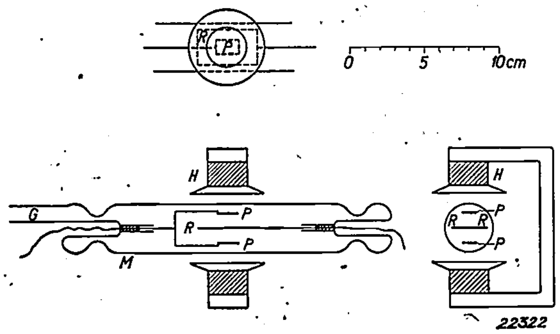


Fig. 12. Philips' Vacuum-meter with two plate-type cathodes P_1 and P_2 and a ring anode *R* in the field of a permanent magnet *H*.

in the field of a permanent magnet *H*, which midway under the pole pieces has a field intensity of approximately 370 oersted. The electrodes are connected up according to the circuit shown in *fig. 13*, viz, through a resistance of 1 megohm to a low current source of 2000 volts. The current *i* through the gauge

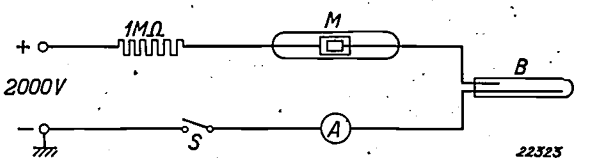


Fig. 13. Circuit of the Philips' Vacuum-meter with the tuning lamp *B* connected as a current measurer, and the micro-ammeter *A*.

valve *M* is hence a measure of the gas pressure, and can be read off on a micro-ammeter with resistances connected in parallel. In qualitative measurements, the instrument can be replaced by a small glow-discharge lamp, such as the Philips 4662 tuning lamp in which the length *l* of the glow discharge is a measure of the current flowing and hence also of the pressure in the valve *M*.

The relationship between *i* or *l* and the pressure *p* is shown in *fig. 14* for a pressure range from $2 \cdot 10^{-3}$ to 10^{-5} mm Hg. The curves are based on mean

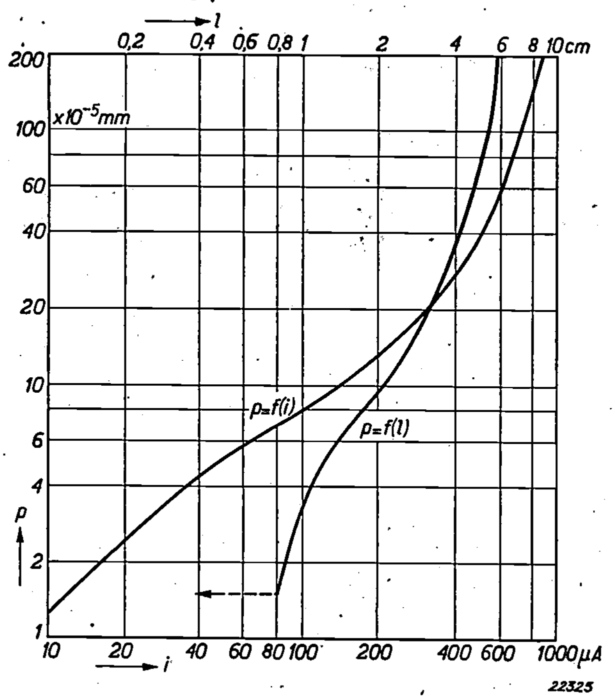


Fig. 14. Gas pressure *p* in the Philips' Vacuum-meter plotted as a function of the discharge current *i* or the length *l* of the glow-discharge in the tuning lamp (mean values for H_2, CO, A and air).

values for air, hydrogen, carbon oxide and argon, while the values of *p* are in agreement within a factor of 2. This gauge [14]) which has been described in detail in another paper [15]) has the advantage of being extremely simple in construction and does

[14]) Marketed by E. Leybold's Nachf. A.G., Cologne-Bayental.

[15]) F. M. Penning, Physica, 4, 71, 1937.

not require protection against tremor, while one reading made without preliminary adjustment is sufficient to measure a pressure value. It is therefore extremely suitable for industrial use and permits simultaneous and continuous supervision of a multiplicity of pumps. Moreover, it can be sealed to the apparatus of vessels being evacuated or in close proximity to them. Compared with the Mac Leod gauge, it has the advantage of giving continuous and instantaneous readings which are easily visible at a distance; it can be used also for measuring the pressure of condensible vapours and does not itself introduce a supplementary vapour pressure. The latter feature is important when the gauge is used for measurements in conjunction with oil high-vacuum pumps. In conclusion, it should be stated that in the first models of this gauge made the cathode plates P were of iron, but it was shown by experience that the sensitivity of the gauge was increased when the cathode was made of thorium or zirconium; moreover with these metals cathode sputtering was less than with iron.

Tube connecting Gauge and Gas Chamber

In gauges in which gas is continuously consumed, as for instance in those instruments designed on the electrical-conductivity principle, a pressure gradient is produced in the tube connecting the gauge and the gas chamber; this tube must therefore not be made too narrow.

Other potential errors in pressure measurement may be caused by inserting a liquid-air trap between the gauge and the gas chamber, when, in particular, the pressure of condensible gases and vapours, which are given off in the gas chamber, escapes measurement entirely or is only partially measured. A second source of error may be caused by the bore of the trap *not being the same where it enters or leaves the liquid air* [16]). In *fig. 15* the section between A and B represents a trap at a temperature T', with one end of diameter d and the other of diameter D; to the left of A and the right of B the temperature is T. The pressure to the left of A is p, and that to the right of B is P. Over the pressure range in which the conditions of flow conform with the Poiseuille formula (free path $\lambda \ll$ diameters of tubes d and D) we have:

$$\lambda \ll d,\ p = p' = P.$$

[16]) M. Rusch and O. Bunge, Z. techn. Phys., **13**, 77, 1932.

In the range covered by the laws of Knudsen ($\lambda \gg D$) we have:

$$\lambda \gg D,\ \frac{p}{p'} = \sqrt{\frac{T}{T'}} \text{ and } \frac{P}{p'} = \sqrt{\frac{T}{T'}},\ \text{also } p = P.$$

Although d and D are different and no errors are likely to accrue in the two boundary pressure ranges, this is no longer true for the transition range in which λ is approximately equal

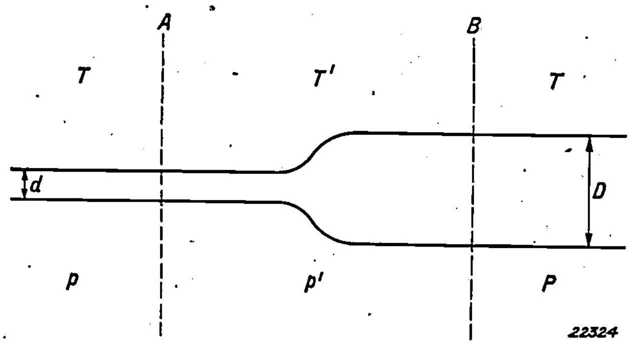


Fig. 15. Gas pressure in a tube of variable diameter at different temperatures.

to d. In this range p/p' and P/p' are determined by d and D respectively, and for $d \neq D$ we have also $p/P \neq 1$; errors of 10 per cent can readily accrue here. If a trap is necessary, then if accurate measurements are required one with a uniform cross-section should be used.

Synopsis

A synopsis of the more common high-vacuum gauges in use is given in Table I, together with the principal ranges of pressures for which they are suitable.

BIBLIOGRAPHY.

S. Dushman. Production and measurement of high vacua. Schenectady, 1922.

A. Goetz: Physik und Technik des Hochvakuums. Brunswick, 1926.

F. H. Newman: The production and measurement of low pressures. London, 1925.

G. W. C. Kaye: High vacua. London, 1927.

S. Dushman: Recent advances in the production and measurement of high vacua. Journ. Franklin Inst. **211**, 689, 1931.

Table I.

Type of Gauge	Range (mm Hg)										Pressure reading dependent on type of gas	Principle
	10^{2}	10^{1}	10^{0}	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}		
Hg gauge	—	—	—	—							No.	Pressure
Macleod gauge		—	—	—	—	—	—	—			Partially*)	Press. after preliminary compr.
Pirani gauge (fig. 2a)				—	—	—	—				Yes	Thermal conductivity.
Knudsen gauge (fig. 7)						—	—	—	—		No	Radiometer effect.
Molecular vacuum-meter (fig.8)		—	—	—	—	—	—	—	—	—	No	Radiometer effect
Ionisation gauge (fig. 10)						—	—	—	—		Yes	Ionisation by electrons
Philips' vacuum-meter (fig. 11)						—	—	—			Yes	Gas discharge in magnetic field

*) In the presence of condensable vapours.

THE USE OF AMPLIFIERS (REPEATERS) IN TELEPHONY

by W. SIX and H. MULDERS.

Summary. Following a discussion of the principle of two-wire and four-wire repeaters for amplifying speech currents in telephony, three different types of repeater are described, viz., a four-wire repeater equipped with triodes, a four-wire repeater equipped with pentodes and inverse feed back, and a two-wire repeater also with pentodes and inverse feed back.

In a previous article [1]) published in this Review it has already been pointed out that before the introduction of loading coils the range of transmission in telephony was fairly small. While the adoption of coil-loaded cables increased the range over which calls could be made, it did not entirely remove all restrictions to telephonic intercommunication. An outstanding advance in developing an international telephone service was in fact not made until 1920 when the general principles of amplification technology were worked out.

By inserting loading coils in telephone cables the attenuation can certainly be considerably reduced, but over very great distances the weakening of speech currents is still excessive. For transmission over long distances, it is imperative therefore to supply additional energy to the transmission line by introducing a type of relay action. In the first application of this principle the line was interrupted at a particular point, the incoming line connected to a telephone receiver and the outgoing line to a microphone, the diaphragms of the receiver and the microphone being linked mechanically (Brown relay). The signals reaching the receiver then cause the microphone diaphragm to vibrate, so that an amplified signal is transmitted through the outgoing line. The energy transmitted through the first part of the line is therefore used to operate the relay, while the energy passed through the second part is furnished by the microphone battery.

Nevertheless the general application of amplifiers of repeaters to telephone technique only became possible when the production of the triode amplifying valve provided an inertia-free and efficient relay, which also permitted a high amplification ratio to be obtained.

Two-Wire and Four-Wire Circuits

As stated above the incoming signal is utilised to operate the relay, which in the case of the triode (three-electrode valve) means that the incoming line is connected to the grid of the valve through a transformer. The outgoing line is connected to the anode circuit through another transformer. This circuit will naturally operate in one direction only; to make two-way traffic possible, two methods can be employed, either a two-wire or a four-wire circuit may be adopted (see *fig. 1*).

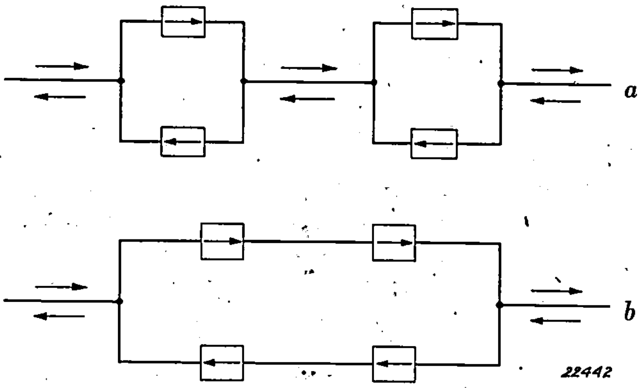


Fig. 1. *a*) General scheme of a two-wire circuit.
b) General scheme of a four-wire circuit.

In the two-wire circuit, two repeaters are connected to the line at the same time, the input of one repeater being in parallel with the output of the second repeater. The repeaters in the various repeater stations are thus linked through a twin-core conductor.

In the four-wire circuit, telephone traffic in the one direction is entirely distinct from that in the opposite direction. Hence two twin-conductors are required and at the terminals the input of one repeater is connected in parallel to the output of the other.

Paralleling the input and output, particularly with two-wire circuits, is not possible without further precautions. For the amplified signals would pass from one repeater to the input of the other repcater, where they could again be amplified, so that in fact the repeaters would commence to oscillate. This is prevented by means of so-called balancing network shown in *fig. 2*, which fundamentally is a bridge circuit. The four arms of the bridge are formed by the line impedance *1*, the artificial line N_2 and the two impedances between the points a - b and b - c; these impedances are determined

[1]) W. Six: The use of loading coils in telephony, Philips Techn. Rev. 1, 353, 1936.

by the input impedance of the repeater. The two transformer windings a - b and b - c are similar, so that the impedances between a - b and b - c also

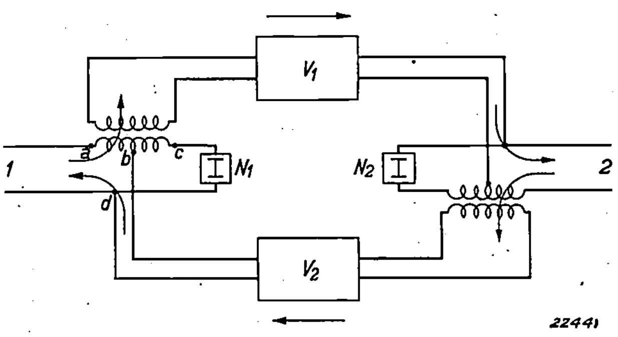


Fig. 2. Balancing network. The impedances of the artificial lines N_1 and N_2 are equal to the impedances of the lines *1* and *2*. In this way the output signal of V_2 produces no voltage at the input of V_1 and vice versa.

are equal. Furthermore, the impedance of the artificial line N_1 is made equal as far as possible to the line impedance for all frequencies. If a potential difference obtains between the points b and d (output voltage of V_2), the potential difference between a and c must be zero, since the bridge is balanced. No voltage is therefore applied to the input of V_1. On the other hand, a signal incoming over the line *1* must be passed to the input of V_1 and after amplification transmitted to the line *2*. But it is difficult to make the impedance of the artificial line exactly equal to the line impedance at all frequencies, since particularly with coil-loaded cables this impedance is non-uniform in a certain frequency range (see *fig. 4*), owing to the irregular intervals between the loading coils and the difference in their self-inductance values.

The propagation of electric waves through cables has already been discussed theoretically in this Review [2]).

From equation (2) of the first article, we can readily deduce that for a wave passing through a cable the ratio $V:I$, which is termed the impedance, is represented by:

$$Z = V : I = \sqrt{\frac{R + j\omega L}{G + j\omega C}}, \quad \ldots \ldots (1)$$

where R is the resistance, L the inductance, G the dielectric conductivity (insulation loss) and C the capacity per unit length of the cable. This expression also applies to coil-loaded cables provided that the frequency is low, i.e. the wavelength is long, compared to the intervals between the loading coils.

But if this is not the case, the fact that the self-inductances are not uniformly distributed over the cable must be taken into consideration. The cable is then more satisfactorily represented by the equivalent circuit shown in *fig. 3*, which is identical to the low-pass filter discussed in the second article. By analogy to the discussion in the series of articles dealing with electrical filters, it may be expected that the attenuation in loaded

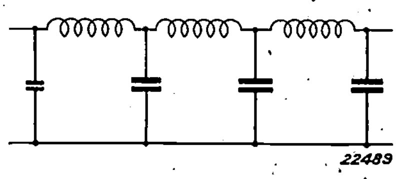


Fig. 3. Equivalent circuit for a coil-loaded cable. If the uniformly-distributed capacity is replaced by condensers, a system equivalent to a low-pass filter is obtained.

cables will remain small up to nearly a specific cut-off frequency of:

$$\nu_1 = \frac{1}{\pi\sqrt{LC}}$$

and will then increase rapidly. Where no losses occur, the impedance, which in the notation used in the second article quoted must be represented by Z_π', should also become infinite at the cut-off frequency and be expressed as a function of the frequency as follows:

$$Z = Z_\pi' = \frac{\sqrt{L/C}}{\sqrt{1 - \frac{\nu^2}{\nu_1^2}}}.$$

These equations for the low-pass filter cannot be applied to cables without some modification; since adequate consideration is not given to the fact that the capacity is distributed uniformly along the line. The qualitative deductions made therefrom are, however, correct, as may be seen from *fig. 4*, where the attenuation and the impedance are plotted

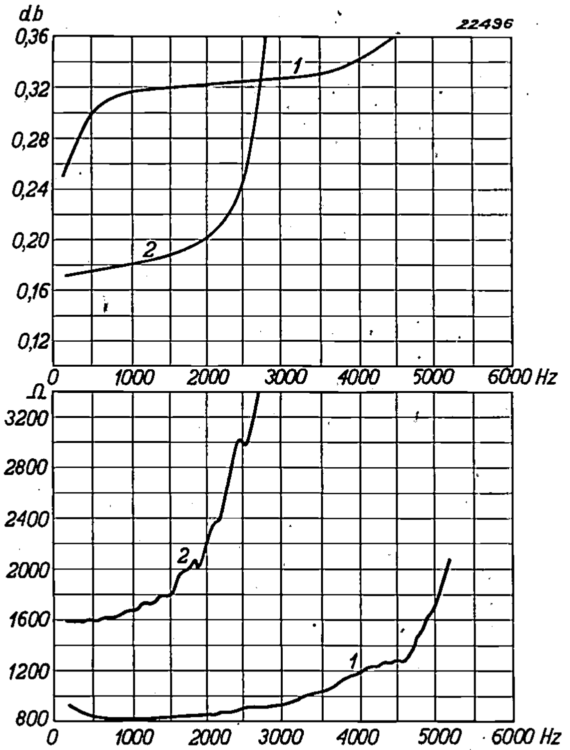


Fig. 4. Attenuation of an impedance in (*1*) a lightly-loaded cable, and (*2*) a heavily-loaded cable.
1 - 44 millihenries at intervals of 1830 m.
2 - 177 millihenries at intervals of 1830 m.

---

[2]) W. Six: The use of loading coils in telephony, Philips techn. Rev. **1**, 353, 1936, which is the first article; and Balth. van der Pol and Th. J. Weyers, Electrical Filters V, Philips techn. Rev. **1**, 367, 1936, which is the second article.

as functions of the frequency for a low-loaded cable and a high-loaded cable.

Close to the cut-off frequency, the impedance varies considerably even with very small changes in the self-inductance and in the intervals between the loading coils; it gives a wavy curve when plotted as a function of the frequency if the coils are not equally spaced, and for this reason it is difficult to keep the balancing network in balance in this frequency range.

If the bridge circuit is not fully balanced at a specific frequency, then although this circuit will still produce a certain attenuation, part of the signal originating from one repeater will be passed back to the input of the other repeater. As a result the gain is limited, for in the complete circuit which is made up of the two repeaters and the two balancing networks a certain attenuation must be present and must be greater than the total gain in the circuit; if this were not the case the amplifiers would start oscillating.

If the amplification ratio of V_1 and V_2 is, for instance, 20 decibels, then the attenuation in the balancing network must be greater than 20 decibels at all frequencies. But, as already indicated above, this is difficult to achieve close to the cut-off frequency of the cable. A filter is, therefore, connected in front of the amplifiers V_1 and V_2, and serves to produce a marked attenuation of the frequencies lying above the band required for satisfactory acoustic intelligibility. The amplification ratio at frequencies above this band will, in consequence, be much reduced, so that the attenuation of the balancing network can also be made much smaller.

Another important point in the case of two-wire repeaters is with regard to the input impedance, i.e. the impedance between the points *a* and *d*, which must be as nearly equal to the line impedance as possible. The greater the difference between these impedance values, the greater will be the fraction of the incoming signal which is reflected back to the preceding repeater, thus further reducing its stability. Reduction in the stability by reflection can also be described in another way, viz., by saying that the line impedance seen from the points *a* - *d* is determined by the termination of the circuit at the preceding repeater situated at the opposite end of the cable and that therefore the balancing network will only be equivalent when the line is terminated at that point by its impedance.

If, therefore, the impedance of the line is Z and the input impedance of the repeater is W, the coefficient of reflection will be:

$$F = \frac{Z - W}{Z + W}\ ^{3)}.$$

In addition to the input impedance, the impedance of the artificial line must also be equal to Z, as already stated above. *Fig. 5* shows diagram-

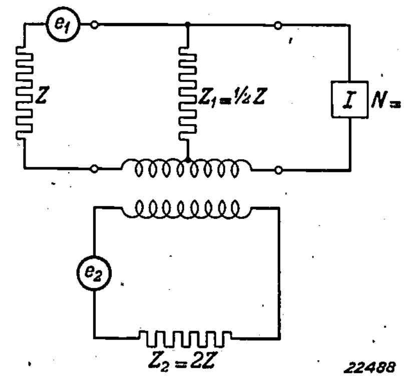


Fig. 5. With this choice of impedances in the balancing network, the bridge is not only balanced but at the same time the line is terminated by its impedance.

matically how these two requirements can be satisfied. In this diagram Z is the impedance of the line, Z_1 the input impedance of the repeater, Z_2 the output impedance of the other repeater and N the artificial line; e_1 is the voltage in the line and e_2 the voltage generated in the amplifying valve. If the impedance of N is equal to Z, the voltage e_2 will not cause current to flow in the branch Z, since the bridge is balanced. If $Z_1 = {}^1/_2\, Z$ and $Z_2 = 2\,Z$, it may be found by calculation that the voltage e_2 produces no current at all through the branch N and that the terminal impedance of the line is equal to Z.

With four-wire circuits, somewhat different conditions obtain. Naturally a certain attenuation must be present also in this circuit which is made up of repeaters, lines and balancing networks. But in this case, the attenuation of the telephone lines also is included in the circuit. If the repeater gain is made so great that the voltage is slightly lower at the end of the line than at the beginning, there can be no question of oscillation, even when a balancing network is not used. The only disadvantage without such a network is that the signal may return to the input of the line and thus give rise to echo phenomena. For this reason balancing networks are used in these cases also, but the artificial lines need not be given the same accuracy in simulation as is required with two-wire circuits.

The amplification ratio at the individual repeater

[3]) A cable with impedance can be substituted electrically by a voltage source in series with an impedance Z. If a voltage source e with an internal impedance Z is shorted by an impedance W, the voltage applied to W will be $eW/(Z+W)$. If $Z = W$, the voltage will be ${}^1/_2\, e$. This may also be regarded as a reflected voltage equal to ${}^1/_2\, e\,[1 - 2W/(Z+W)] = {}^1/_2\, e\,(Z - W)/(Z + W)$.

stations is in this case limited by cross-talk phenomena between the telephone channels as well as by the interference level. In practice a gain of 12 to 15 decibels can be obtained with two-wire repeaters and of 30 to 40 decibels with four-wire repeaters.

Frequency Characteristics of Repeaters

In general, the gain at all frequencies should be made equal to the attenuation produced by the line, i.e. within a frequency band from 300 to 2500 cycles, which as a rule is adequate for telephone channels. With coil-loaded conductors, the attenuation increases considerably at higher frequencies particularly in the neighbourhood of the cut-off frequency. In a heavily-loaded line, i.e. one equipped with coils with high self-inductance values and located close together, the attenuation in the speech band is on the average smaller, but on the other hand the cut-off freqeuncy is lower, while with lightly-loaded lines the attenuation is greater but the cut-off frequency is higher.

It is, therefore, necessary in the case of heavily-loaded lines to insert a wave filter or network between the lines and the repeaters or in the repeaters themselves, whose main purpose is to counteract the increase in attenuation close to the cut-off frequency and at the same time provide a slight balance for the drop in the attenuation at low frequencies. This measure is not necessary with lightly-loaded lines.

Choice of Repeater System

There is still a wide divergence of opinion in different countries with regard to the relative suitability of the various systems for producing a gain in transmission viz., two-wire or four-wire repeaters or light or heavy loading. In Holland the method which has been adopted generally during recent years consists in the use of four-wire repeaters and light loading [5]). Although double the number of conductors are required with the four-wire system, it yet allows a much higher gain to be obtained than the two-wire system, so that a much greater attenuation is permissible in the cables with the same number of repeaters, and in consequence a smaller copper cross-section can be used. Moreover, as already stated above, the four-wire repeater is much simpler to design, since in it the balancing networks with artificial lines are required only at the terminals of the line and need not be made with the same accuracy as in two-wire repeaters; in addition, rejector filters to cut out high frequencies, which are necessary with two-wire repeaters, can be dispensed with in the case of the four-wire type.

The method of loading adopted in Holland in which 65-millihenry coils are located at intervals of approximately 3.68 km, has a cut-off frequency of roughly 3400 cycles, which is sufficiently above the speech frequencies, so that the attenuation in the cable requires compensation only in the range from 300 to 800 cycles. The wide interval between the loading coils has resulted, moreover, in a marked saving in the cost of the coils.

Inverse Feed back

Inverse feed back or negative reaction has been employed in amplifier technique for many years, and consists in feeding back in phase opposition part of the output voltage to the input of the amplifier. In these circuits the gain is naturally reduced, but their advantage lies in the fact that at the same time the distortion is lower and the gain is made more stable, in other words it is made less dependent on the characteristics of the amplifying valve and the feed voltages. For some years inverse feed back has been used in all amplifiers employed in carrier-wave telephony in which an extremely low distortion is essential in view of cross-modulation between the different transmission channels. The Dutch Post Office authorities first employed this circuit in intermediate repeaters for low-frequency telephony, its chief advantage being not so much that the distortion is reduced but that the gain is constant. Reduced gain which is inseparable from inverse feed back is not really a drawback, since by using a valve with a high effective amplification (pentode) roughly the same gain can be obtained as was realised in the past with a triode.

Description of Three Types of Repeater

Three types of repeater have been developed in the Philips Laboratory, all three being designed for indirectly-heated amplifying valves. The use

[4]) For every section of line joining two repeaters, the average power at the output of one repeater is much greater than that at the input of the other repeater. This difference in level is equal to the gain of the repeater. For lines with opposing directions of speech, the level is high in one line and low in the other line. Cross-talk will therefore be the greater, the greater the difference in level, i.e. the greater the gain.

[5]) A new telephone system in Holland, by A. H. de Voogt, Post Office Electr. Engin. Journ., 25, 195, 1932-33.

[6]) C. J. van Loon: Improvements on radio receivers, Philips techn. Rev. 1, 264, 1936.

[7]) A new feed-back repeater, by G. H. Bast and E. H. Stieltjes, Post Office Electr. Engin. Journ. 28, 225, 1935-36.

of these valves considerably simplifies the circuit compared with triodes having directly-heated filaments as hitherto in use and in which the filaments of four valves were always connected in series. For these latter a separate battery for the negative grid bias was necessary, while at the same time complicated circuits had to be adopted to decouple the repeaters from each other so as to avoid cross-talk. The simplified circuit enables a very compact repeater unit to be obtained. These repeaters are all-mains fed from an alternating-current supply.

I) The first repeater developed was a four-wire repeater with triodes without feed back. A repeater unit, containing four repeaters of this type, is shown in *fig. 6*.

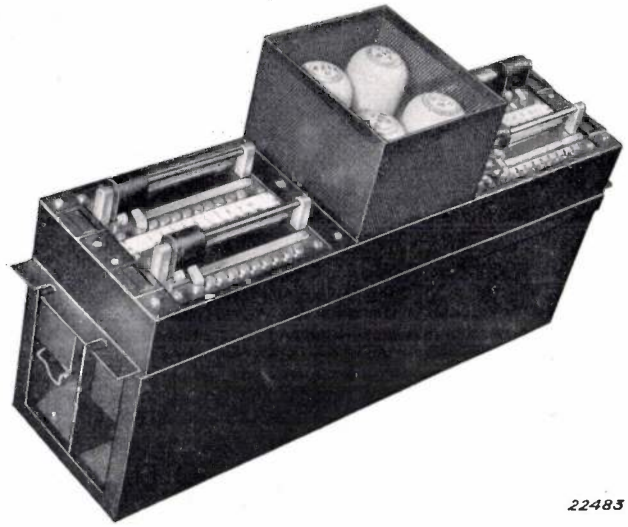


Fig. 6. Repeater unit with 4 four-wire repeaters equipped with triodes.

II) The fundamental circuit of the four-wire repeater with feed back is shown in *fig. 7*. In this repeater a pentode is used as an amplifier, and the feed back voltage is tapped from the rheostat *R*. This voltage is determined by the

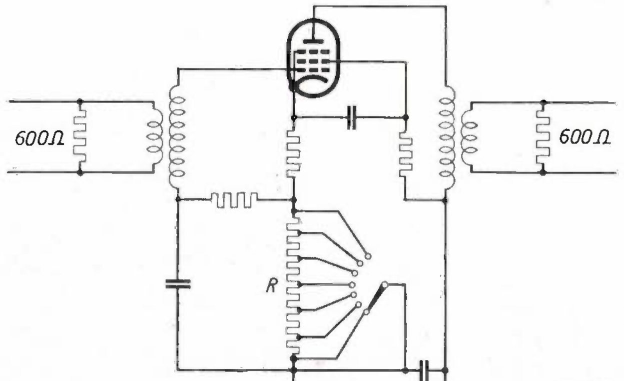


Fig. 7. Fundamental circuit of the Philips four-wire repeater with pentodes and inverse feed back.

resistance, so that the gain is controllable. Between 300 and 4000 cycles the characteristic of the repeater is a straight line within 2 decibels (see *fig. 8*).

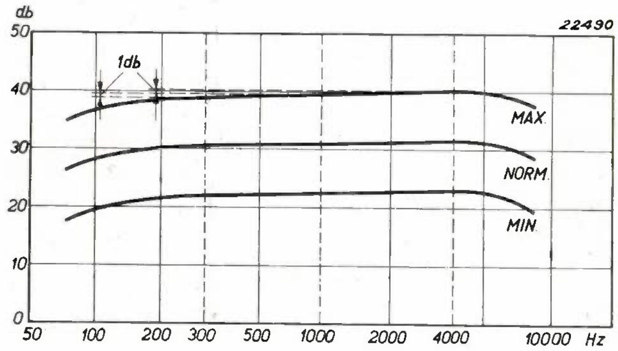


Fig. 8. Gain in the four-wire repeater with inverse feed back plotted against the frequency.

To obtain satisfactory matching to the line, the input and output transformers are terminated with resistances equal to the impedance of the telephone line, viz, 600 ohms. Strictly speaking, the terminal resistance together with the input or output impedance of the amplifier in parallel to it should be made equal to the impedance. The input and output impedances are, in this case, much greater than 600 ohms and hence practically without effect on the combined value. The non-linear distortion for an output of 50 milliwatts is shown in *fig. 9*

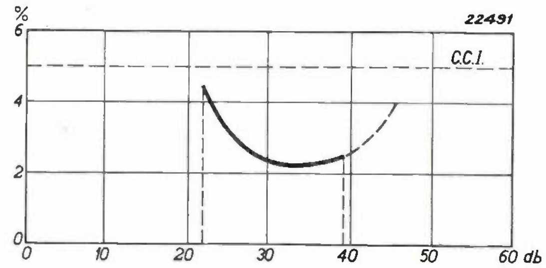


Fig. 9. Non-linear distortion in the inverse feed back repeater plotted against the gain at a power output of 50 milliwatts.

as a function of the amplification. Throughout the whole range this distortion is below the limit stated in the Standards of the Comité Consultatif International des lignes téléphoniques à grande distance (C.C.I). The principal advantage offered by this repeater, as compared with the preceding repeater which is equipped with triodes, is in its marked independence of the feed voltages, as may be seen from *fig. 10*. These curves in fact show that the pentode amplifier with negative reaction still satisfies the C.C.I. standards for lines with more than 12 repeaters when the anode and filament voltages drop to 60 per cent, while in the case of the triode amplifiers the voltage must not drop below 82 per cent if they are still to conform with these standards.

III) The fundamental layout of a two-wire

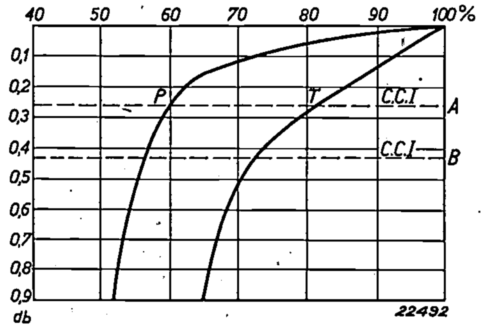


Fig. 10. Gain in decibels below the normal value plotted against the filament and anode voltages in percentages of the normal value. T - triode amplifier without negative reaction; P - pentode amplifier with negative reaction. A - C.C.I. limits for lines containing more than 12 repeaters. B. - C.C.I. limits for lines with 12 repeaters or less.

tional to the output voltage. The feeding back of the current raises the internal resistance of the repeater, while it is reduced by the voltage back-feeding [8]). By a suitable combination of these two reactions the required internal resistance can be obtained in the repeaters and at the same time the requisite degree of negative reaction realised. In this repeater the internal resistance has been made equal to the line impedance, so that the output transformer need not be terminated by a 600 ohm resistance as is necessary with the four-wire repeater. In the latter, half of the energy output was lost in the terminal resistance, while in the two-wire repeater the whole of the energy output is available for useful work.

In *fig. 12*, the impedance (magnitude $|Z|$ and phase angle φ) of this repeater is plotted against

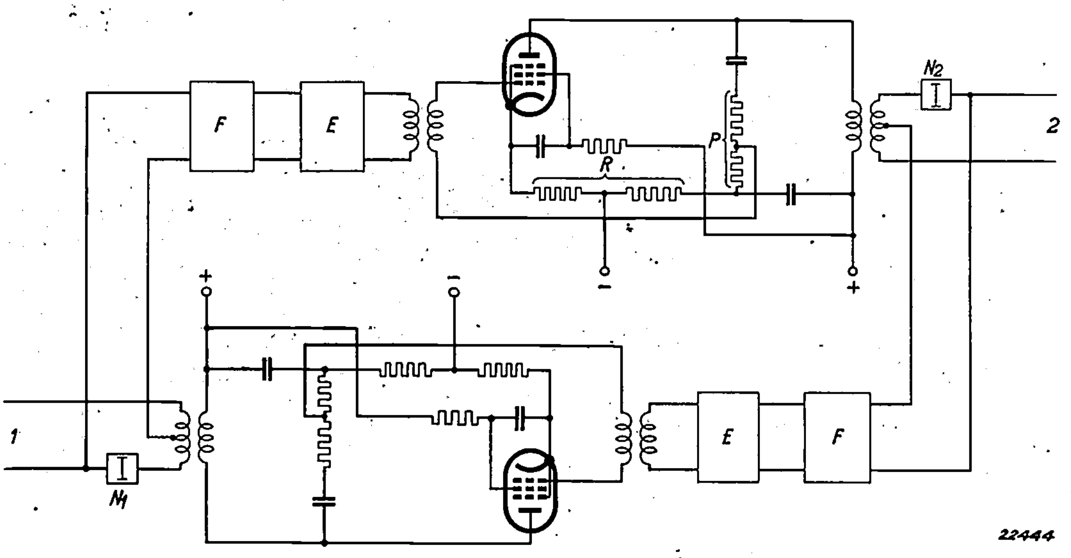


Fig. 11. General circuit diagram of Philips two-wire repeaters with pentodes and inverse feed back. N_1 and N_2 artificial lines. E — Balancing networks. F — Filters.

repeater with inverse feed back is shown in *fig. 11*. In this repeater, a combined current and voltage feed back has been used, in other words the negative-reaction voltage is made up of two components: one component is tapped from the resistance R and is hence proportional to the output current, and a second component taken from the potentiometer P and is hence propor-

the frequency. The broken line represents the im-

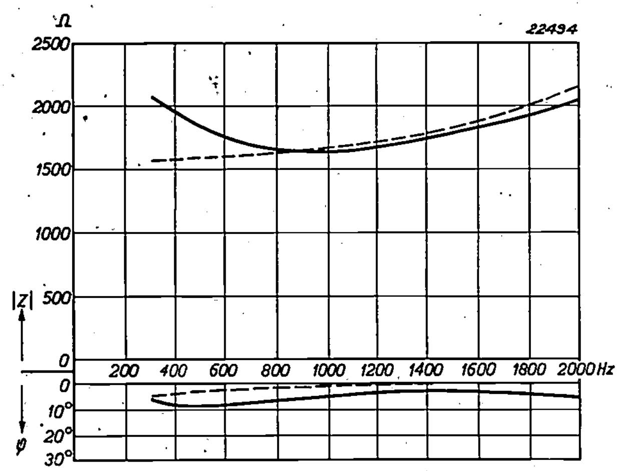


Fig. 12. Full lines: Impedance of the two-wire repeater (magnitude $|Z|$ and phase angle φ). Broken line: Impedance of a cable with 1.3 mm cores and loaded with 177-millihenry coils at intervals of 1830 m.

[8]) Negative reaction or inverse feed back consists in feeding back to the input the voltage variations occurring in one or other of the circuit components at the output side of an amplifying valve, so that these variations become reduced. If the component is so chosen that the voltage applied to it is proportional to the output voltage, the alteration in the output voltage, caused for example by altering the external resistance load connected to it, will be less, i.e. the effective internal resistance is reduced. If the feed-back voltage is made proportional to the output current, this current becomes less dependent on the value of the resistance load, and the internal resistance is increased.

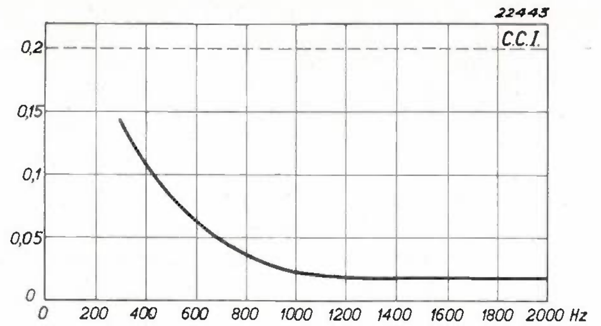


Fig. 13. Coefficient of reflection of the two-wire repeater plotted against the frequency, when using a cable for which this repeater has been designed.

pedance of the cable for which the repeater was designed (diameter of cores 1.3 mm; loading coils of 177 millihenry at intervals of 1830 m).

The reflection coefficient $F = Z - W/Z + W$ is plotted against the frequency in *fig. 13*. Throughout the whole frequency range this coefficient is considerably below the limiting value specified by the C.C.I.

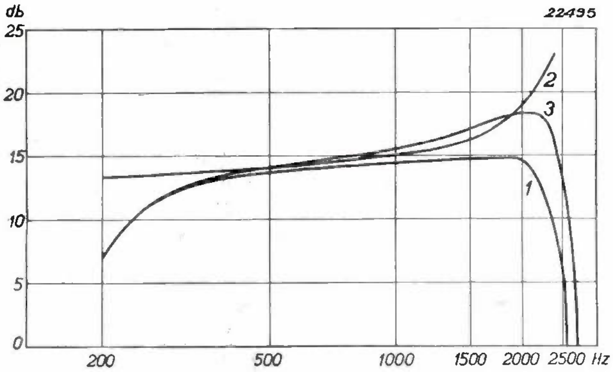


Fig. 14. Gain in two-wire repeater containing pentodes, plotted against the frequency. *1* - Repeater with rejector filter for high high frequencies. 2 - Attenuation of the associated cable. 3 - Repeater corrected by means of a balancing network.

Fig. 14 gives the frequency characteristic of this repeater. Curve *1* is the characteristic of the repeater with rejector filter for cutting out high frequencies. Curve *2* is the attenuation of the cable, while curve *3* is the characteristic of the repeater when using a balancing network which compensates for the increase in attenuation in the neighbourhood of the cut-off frequency of the cable.

Finally, a photograph of a repeater of this type is reproduced in *fig. 15*.

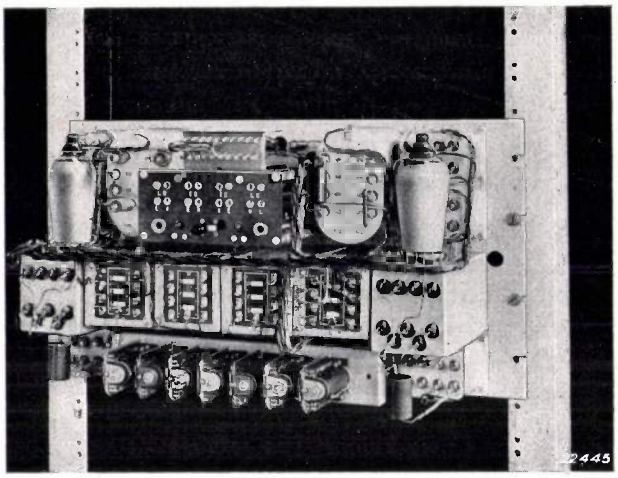


Fig. 15. Photograph of a Philips two-wire repeater with pentodes and inverse feed back.

A RECORDING FIELD-STRENGTH METER OF HIGH SENSITIVITY

by M. ZIEGLER.

Summary. An apparatus for measuring the signal strength of radio transmitters radiating on wave lengths between 10 and 2000 m is described. The sensitivity of the apparatus is calibrated during measurement by means of an oscillator incorporated in it. Field strengths from approximately 1 microvolt per m upwards can be measured with this apparatus.

Introduction

The measurement of the field or signal strength of radio transmitters is now a matter of routine in wireless engineering and may be carried out for a variety of practical purposes. A few applications in which the field-strength meter has proved indispensable may be quoted from the innumerable uses to which this instrument has already been put: The energy radiated from a transmitting aerial with a known input is a criterion for determining the efficiency of the aerial, while the distribution of field strength may be explored over the area surrounding a transmitter to ascertain the intensity of interference at a particular point, due either to an interfering transmitter or to an oscillating receiver. The directional characteristics of an array of transmitting aerials can be investigated, as well as the absolute minimum and maximum signal strengths due to fading ascertained. The purpose of a measurement of the field strength may have purely a scientific interest, but in the majority of cases measurements have primarily a technical value, such as those enumerated above. Practical data concerning the power radiated from a specific type of aerial, as well as a close knowledge of the effects of local geological and topographical conditions on the propagation of waves, are essential for planning and selecting the best site for a transmitting station which shall operate efficiently and be economical to construct and run.

The various problems entailed in a measurement of the field strength are very diverse, and the practical and theoretical requirements imposed on the measuring apparatus used are widely divergent. Firstly, very weak fields due to remotely-situated transmitters must be capable of measurement, and secondly, measurements must be feasible in the immediate vicinity of the transmitting aerial; in other cases again the apparatus must be used on sites to which it cannot be conveniently transported by vehicles. For still other purposes, it may be desirable to install the measuring arrangements permanently in order to carry out measurements at a particular spot of the variations in signal strength with time over periods of several weeks. Furthermore, on certain occasions measurements have to be carried out on long waves and on other occasions on ultra-short waves.

Every type of measurement of the field strength required in normal practice can be carried out directly with the apparatus described below. Before passing to a description of the apparatus, the principle on which the field-strength meter operates must be briefly outlined.

Principle of the Field-Strength Meter

A field-strength meter is composed of an aerial in which a certain electromotive force is induced by the field of the transmitter, and a voltmeter for measuring this e.m.f.

The effective height of the aerial is of particular interest in this connection, and is defined as follows: If the electric field strength for a particular wave is $e = E \sin \omega t$, then an e.m.f. of

$$V = h\,e \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

will be induced in an aerial of effective height h. If V is measured and the effective height is known, the signal strength can be calculated. The effective height of a vertical aerial with a high top capacity (*fig. 1a*) and whose length l is small as compared

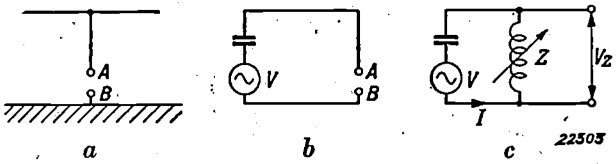


Fig. 1. *a*) Vertical aerial with high top capacity (capacity aerial). A vertical electric field induces a voltage V which is equal to the product of the field strength and the length of the aerial.
b) Equivalent circuit of the capacity aerial. The induced e.m.f. may be regarded as being connected in series with the capacity of the aerial.
c) By connecting a self-inductance Z in series with the aerial, a circuit tuned to the impressed signal can be obtained. At resonance the voltage V_Z can be many times V.

with the wave length is l for a vertically-directed electric vector; the e.m.f. induced by the wave will then be $V = l\,e$.

Adopting the above definition it is also possible to speak of effective height in the case of a frame aerial. The effective height of a frame aerial (flat coil, *fig. 2a*) of n turns, each enclosing a surface O whose dimensions are small as compared with the wave length, is

$$h = \frac{2\,\pi\,n\,O}{\lambda}\,\sin\,\alpha, \quad \cdots\cdots \quad (2)$$

where α is the angle between the plane of the frame and the magnetic vector. As is well known a frame aerial has directional characteristics. A frame aerial suitable for use with e.g. a portable field-strength meter has 16 turns of $0.4 \cdot 0.4 = 0.16$ sq. m surface, when for $\lambda = 300$ m and $\alpha = 90$ deg, we get $h = 0.054$ m. This height is considerably less than the effective height of a vertical aerial of several meters.

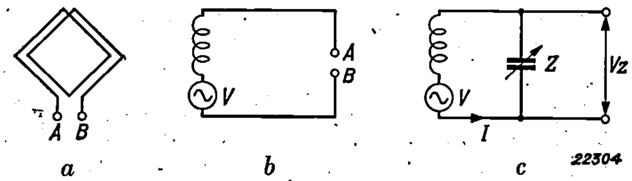


Fig. 2. *a*) Frame aerial (inductive aerial). A magnetic field of unit intensity, which is perpendicular to the plane of the frame aerial, induces a voltage $V = \frac{2\,\pi\,n\,O}{\lambda}$ where n is the number of turns, O the area of the frame, and λ the wave length.
b) Equivalent circuit of the inductive aerial. The induced e.m.f. may be regarded as connected in series with the self-inductance of the frame aerial.
c) By connecting a capacity Z in series with the frame aerial, a circuit tuned to the received signal can be obtained. At resonance the voltage V_Z can be many times V.

The equivalent circuits of the capacity aerial and the frame aerial with an assumed induced e.m.f. are represented in figs. 1*b* and 2*b* respectively.

The problem therefore resolves itself into a measurement of the voltages V indicated in figs. 1b and 2b. But only the terminals A and B are accessible. Theoretically, it is sufficient to measure the potential difference between A and B with a voltmeter, whose impedance is high compared to the reactance of the vertical aerial or the frame aerial. Although it is in fact not imperative, the resonance principle will be used in this case, and a circuit tuned to the signal under measurement formed by connecting an inductance or a condenser in series, the selective and amplifying characteristics of this circuit offering important advantages (figs. 1*c* and 2*c*). Either the current I which flows through the circuit can be measured, or the potential difference V at the series-connected impedance. If at the same time the resistance of the circuit is ascertained, V can be directly calculated, for:

$$V = Ir = V_Z\,\frac{r}{Z}\,.$$

What type of aerial is provided with the field-strength meter?

The requirements of an aerial for use in conjunction with the field-strength meter are that it shall have a specific ascertainable effective height and at the same time be transportable. A frame aerial automatically satisfies these requirements, while a capacity aerial would assume an unfavourable form, for instance those mounted on a short vertical pole fitted with a horizontal plate at the top would have a capacity with respect to earth which is high as compared with that between the vertical part and earth. It is thus reasonable to consider a condenser aerial, but the effective height of an aerial of this type cannot readily be made appreciably greater than that of a standard frame aerial.

A further difficulty with the capacity aerial is that it cannot be so easily tuned as a frame aerial, for a variable selfinductance is required for this purpose (fig. 1*c*) which cannot be realised to suit all cases. In the more normal circuit shown in *fig. 3* the gain in the voltage due to the field is

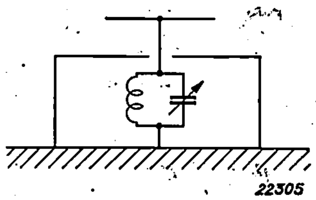


Fig. 3. A short vertical pole on which a horizontal plate is mounted, is an example of a mobile capacity aerial of well-defined characteristics. The effective height is equal to the distance between the plate and the top of the metal box enclosing the measuring apparatus.

much smaller than the resonance amplification obtainable with the frame aerial. Since, moreover, a frame aerial has directional characteristics as compared with a capacity aerial, which in general will prove an advantage, as a standard it is to be preferred to a condenser aerial.

However, where an aerial with a great effective height is essential for the purpose of measuring very weak signals, and if the directive action is either dispensable or even undesirable, one or other of the capacity aerials may be employed. Any aerial system, whose effective height is unknown and which is coupled in any manner to an amplifier with an indicating instrument (e.g. an aerial installed in a motor car), can always be calibrated as a whole:

1) By comparison with a standard aerial, and
2) By measuring the theoretically-known field of an accurately-calibrated testing transmitter.

Important requirements which the field-strength meter must meet

A field-strength meter has to meet the following requirements:

a) High sensitivity in order to measure weak signals,
b) High selectivity to permit weak signals to be measured even when interference is produced by more powerful stations on an adjoining frequency;
c) Constant characteristics, in order to avoid the need for frequent recalibration in the laboratory.

To meet requirements a) and b) a sensitive amplifier is used containing a number of amplifying valves in cascade and various tuned circuits, by means of which the induced voltage is amplified for measuring purposes. On the other hand the characteristics of the apparatus cannot be made constant over long periods by any simple and direct method, for the sensitivity varies with the age of the valves and is to a great extent determined by the accuracy of tuning, etc. For this reason, the apparatus has been so designed that the sensitivity can be adjusted to the required value as necessary.

The sensitivity of the apparatus can be checked very easily, viz, by observing the effect of a known e.m.f. which has the same frequency as the field under measurement and which is connected in series with the frame aerial as indicated in *fig. 4*. equal to the calibration voltage, the deflection of the voltmeter will be reduced or increased in the corresponding ratio. By this comparison method the reliability of measurements made is mainly

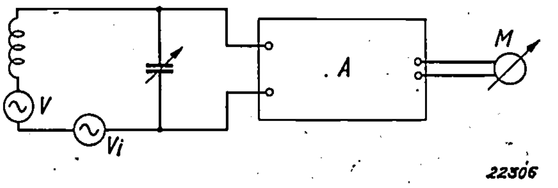


Fig. 4. Principle of the field-strength meter. The voltage V induced in the frame aerial is compared with the calibration voltage V_i. To measure low voltages, the voltage which is produced by V or V_i at the condenser is amplified by an amplifier A.

dependent on the accuracy of the calibration voltage.

Electrical Layout of the Apparatus

In fig. 4 the general principle of the Philips field-strength meter is illustrated and in *fig. 5* a detailed layout of the various component circuits. Referring to this diagram we shall first investigate how the signal under measurement produces a deflection on the measuring instrument.

The voltage of frequency ν which is induced at the terminals of the aerial circuit, as already described above, is passed through a selective amplifying stage H and then to the mixer valve M. The intermediate-frequency signal of frequency ν_m

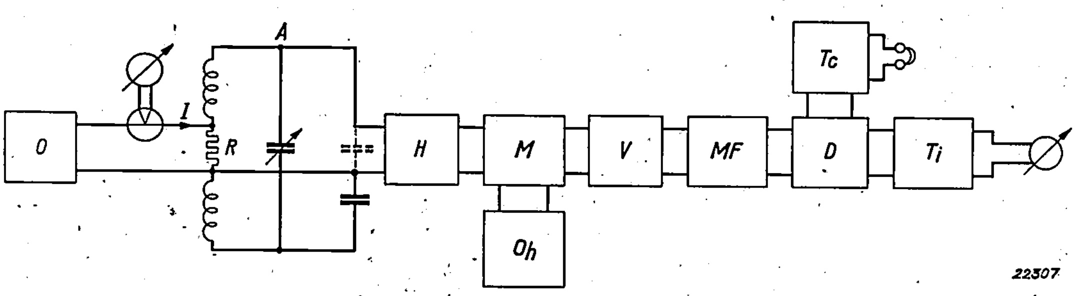


Fig. 5. Diagrammatic layout of the signal-strength meter. O Calibration oscillator to furnish the calibration voltage $V_i = IR$. A Aerial circuit; H High-frequency amplifier; M Mixing valve; O_h Heterodyne oscillator; V Attenuator; MF Intermediate frequency amplifier; D Detector; T_i Indicator triode, which furnishes the current for the instrument; T_c monitor triode for the current passed to the headphones.

Not only can the sensitivity of the voltmeter, which in this case includes the amplifier and the instrument, be checked by this method, but also the gain which occurs in the resonance circuit formed of the frame aerial and the condenser. *The calibration voltage produces exactly the same effect as an induced e.m.f. of the same magnitude* [1]), and will therefore also give the same deflection on the voltmeter. If the induced e.m.f. V under measurement is not which is produced by interaction with the auxiliary voltage of frequency $\nu + \nu_m$ originating from the oscillator O_h, is filtered out and is passed through a variable attenuator V of which further details are given below. Subsequently, the signal is amplified

[1]) This is not quite correct, since the frame aerial is never entirely devoid of capacity and the induced e.m.f. is not concentrated at a single point. But deviations from these ideal conditions can be neglected in practical measurements.

by two stages *MF* tuned to the frequency ν_m and then rectified in a diode detector *D*. The direct voltage now produced is passed to the grid of a triode T_i of which the anode current flows through a milliammeter. This triode, which we shall term the indicator valve, is so adjusted that its anode current is 5 milliamps when no signal is impressed on it. A rectified signal makes the grid of the triode negative with respect to the cathode, so that the anode current drops and produces a deflection of the pointer from the no-signal reading. The total amplification of the signal under measurement must be high enough to give a deflection easy to read, yet not so high that the anode current of the indicator valve becomes zero. The amplification stage is also controllable in addition to the attenuation stage, by means of the mixer valve and the two intermediate-frequency valves, so as to permit accurate adjustment.

The ratio between two signals of the same frequency can be directly read off on the instrument without altering the adjustment. This ratio cannot be determined if it is greater than 10, but with the attenuation stage, the amplification of one signal can be reduced or increased in a definite ratio as compared with the amplification of the other. The maximum attenuation is 1000, so that voltages in a ratio up to 1 : 3000 can still be compared with each other.

Naturally, at all practical adjustments, the amplification ratio must be independent of the amplitude. Both the maximum amplification in each stage and the method by which the amplification is varied are so calculated that the maximum deflection is obtained on the instrument before any other part of the apparatus becomes overloaded.

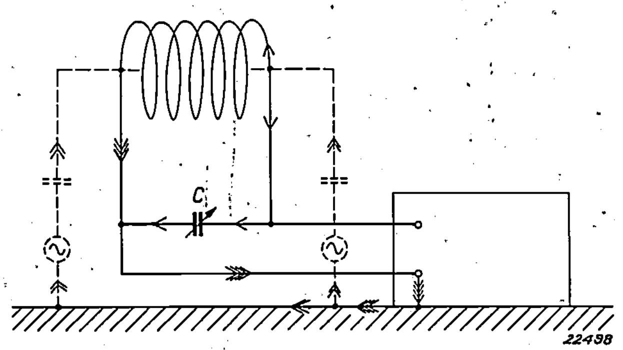


Fig. 6. Asymmetrical frame-aerial circuit. If one end of the aerial winding is earthed, the electric field will cause currents to flow, which, independently of the direction of the frame aerial, will produce a voltage at the terminals of the amplifier. The arrows indicate the currents which are produced by the vertical electric field between the frame aerial and earth. It is seen that part of these currents flow through the condenser C, and hence produce an alternating voltage at the input terminals of the amplifier.

We shall now give a description of the various components of the field-strength meter.

Aerial circuit: Frame aerial. If a frame aerial is connected up as shown diagrammatically in fig. 4, the capacity of the aerial and that of the apparatus with respect to earth (*fig. 6*) will, as a whole, exhibit the same characteristics as a capacity aerial, so that a different type of signal will be generated that would be deduced from equation (2). This follows from the fact that the signal does not fall to zero, even when the plane of the frame is parallel to the magnetic vector. This undesirable characteristic, the socalled "aerial effect", may be eliminated by mounting the frame symmetrically, when the currents generated by the capacity effect balance each other (*fig. 7*).

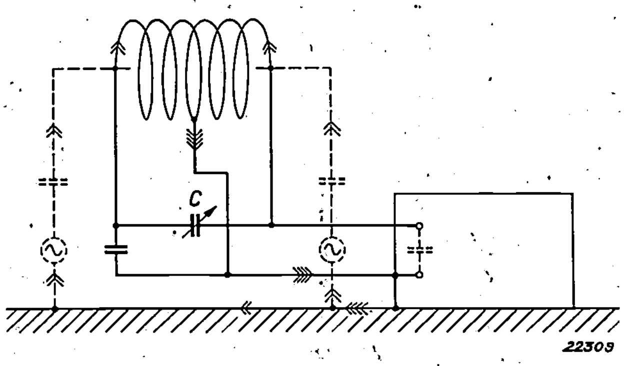


Fig. 7. Symmetrical frame-aerial circuit. Since the centre of the aerial winding is earthed, the currents induced by the electric field counteract each other. No current flows through the condenser *C*.

The frame aerial is tuned with a variable condenser *C*. The measuring equipment is provided with a series of interchangeable frame aerials.

Reception with capacity aerial. For reception with a capacity aerial, the rotatable frame holder can be replaced by coils which automatically make the necessary contacts, so that a circuit of the type shown in *fig. 8* is obtained. Each coil is fitted with

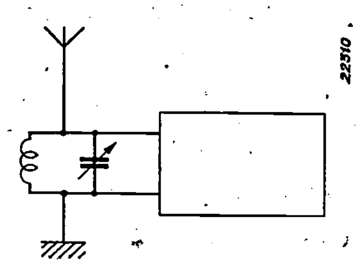


Fig 8. Connection of the aerial circuit when using the signal-strength meter with a capacity aerial.

a terminal to which the aerial is connected; the apparatus is earthed.

The calibration signal. The potential difference at the resistance *R* (fig. 5) is used for calibration and is produced by an alternating current I of the required frequency furnished by a special oscillator (O). As already indicated in fig. 4, the calibration voltage is in series with the frame aerial. The resistance is only of the order of 1/8 ohm, so that for $I = 8$ milliamps, $V_i = 1$ millivolt. The sensitivity of the thermo-couple which is used for current measurement and the resistance R are independent of the frequency, hence also the calibration voltage. The oscillator through its five adjustable ranges furnishes all frequencies between 150 and 30 000 kilocycles and is so rated that the maximum current flowing through the thermo-couple will never overload the couple. This maintains the characteristics of the thermo-couple constant and also raises the reliability of the apparatus as an absolute measuring instrument.

High-frequency Amplifier. The function of this amplifying stage, which consists of an amplifying valve with tuned anode circuit, is twofold, viz:

1) To raise the high-frequency selectivity in order to suppress those signals with a frequency $\nu + 2\,\nu_m$ which may give rise to a disturbing intermediate-frequency signal, and
2) To make the ratio of the signal strength to the interfering noises as satisfactory as possible. By using a high-frequency amplifying valve with very low mush the noise level can be reduced to the unavoidable thermo-electric fluctuations in the first circuit. Even when using a small frame aerial a signal due to a field of only a few microvolts per m will still be sufficiently audible above the mush.

Mixer Valve and Auxiliary Oscillator. An ordinary mixer hexode is used as a mixing valve, and a special triode which has five switched wave-length ranges similar to the calibration oscillator is connected up as an oscillator. The tuning condenser is coupled mechanically to the condenser in the high-frequency stage, so that for tuning the apparatus only one other knob in addition to that for the aerial condenser has to be adjusted. The gain in the mixer valve is regulated by means of the variable negative grid bias.

Attenuator. The attenuator V is coupled to the mixer valve by a pair of circuits tuned to the intermediate frequency signal. By means of a switch the attenuation factor can be adjusted as required to 1, 3, 10, 30, 100, 300 or 1000. The attenuator consisting of wound wire resistances which have been very carefully calibrated, is balanced so that the input and output resistances have exactly the same values at all adjustments. The attenuated signal passes to the first intermediate-frequency amplifier through two coupled tuning circuits.

Intermediate Frequency Amplifier. This amplifier has the following components and characteristics: Two amplifying valves, four tuned circuits, and controllable gain by variation of the negative grid bias. In addition, the grid bias of both valves can be automatically controlled by the impressed signal, such that strong signals are amplified less than weak signals, which may be desirable when recording signal strengths subject to marked fluctuation.

Rectification. The amplified intermediate frequency voltage is passed to the two diodes of a duo-diode-triode. One of the diodes serves exclusively for generating a direct voltage for the automatic gain control just referred to, the whole or part of this voltage being applied to the grid of the intermediate frequency amplifiers. Rectification by the other diode furnishes a low-frequency alternating voltage with a D.C. voltage component, the latter being passed to the indicator valve described above. The alternating voltage is passed to the triode grid of the duo-diode-triode. Reception can be followed audibly by means of headphones connected to the anode circuit of this triode.

Indicating Instrument. The milliammeter with a range of 5 milliamps, which gives readings of the anode current, is connected up in such a manner that the pointer is to the right of the scale when no current is flowing. When a current of 5 milliamps flows the pointer will give a reading to the left of the scale opposite the scale zero. In the absence of a signal, the anode current can be adjusted to the correct value by regulating the potential difference between the cathode and the grid. The scale has been calibrated by experiment in such a way that with a constant sensitivity, i.e. when not using the automatic gain control, the deflection on the instrument is proportional to the input signal strength.

This division of the scale is determined principally by the I_a-V_g characteristic of the triode, so that it is important for the operating characteristics of this valve to remain constant. The anode current is, therefore, maintained automatically constant by means of a neon lamp and is thus unaffected by any fluctuations in the voltage supply. A second instrument, e.g. a recorder, can be connected in series with the milliammeter incorporated in the apparatus.

The field-strength meter with frame aerial and battery box is shown in *fig. 9*.

Variable Selectivity. Although the meter has ten circuits, the selectivity has not been made excessively sharp, since the eight intermediate-frequency circuits are connected in pairs as band filters. The coupling of two pairs of these circuits has been made variable, so that the selectivity can be increased by making the coupling looser.

Current Supply. The apparatus is designed for connection to either a direct-voltage supply (batteries) or a 50-cycle A.C. mains supply.

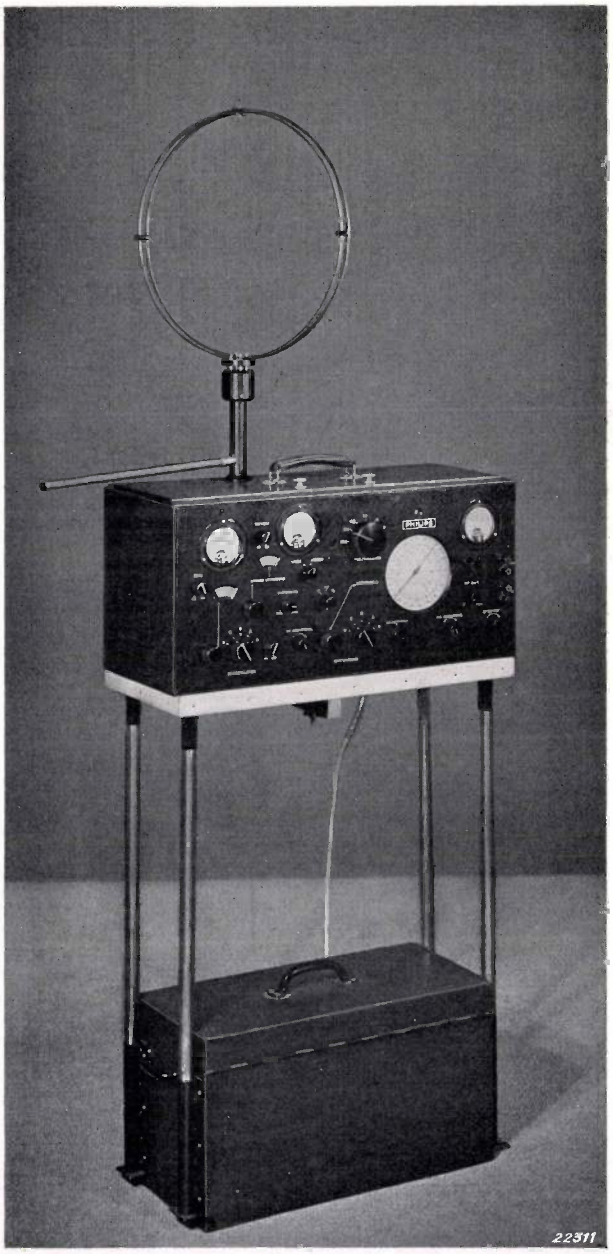


Fig. 9. Philips' field-strength meter with battery box. The apparatus is equipped with a frame aerial for wave lengths from 28 to 75 m. On the left of the front panel is the meter for the calibration voltage; on the right is the meter which indicates the output current. The large circular tuning scale is provided for frequency calibration in each of the five frequency ranges. Above and to the left of the tuning scale is the knob for the attenuator.

Exemples of various measurements

The application of the field-strength meter is illustrated by the following practical examples of its use;

a) *Field-Strength of a Local Transmitter.* One of the most important uses of the field-strength meter is for measuring the field distribution in the area about a transmitting station. The apparatus described here is particularly suitable for this purpose, since it is capable of measuring signal strengths over a very wide range of values.

Assume that the signal strength of the Hilversum station, transmitting on a wave length of 300 m with an aerial output of 15 kilowatts, has to be measured in the neighbourhood of Muiden.

As already stated above, the effective height of the frame aerial provided for this wave length, when suitably orientated, is 0.054 m at maximum signal strength. By accurate tuning and satisfactory orientation of the frame aerial, and using an attenuation of e.g. 1/30, without automatic gain control, the pointer can be adjusted to read 5 by regulating the amplification. The frame is now turned into the position corresponding to the minimum signal strength (there will remain a residue of e.g. 1 per cent), and the calibration oscillator is switched in. After tuning and adjusting the current at the low resistance to 8 milliamps (calibration voltage = 1 millivolt), a reading of 6.5 is obtained with an attenuation of 1/10, all other adjustments remaining unaltered. Hence the effective value of the signal originating from the field radiated by Hilversum is

$$V_{\mathrm{eff}} = 1 \cdot \frac{30 \cdot 5}{10 \cdot 6.5} = 2.3 \text{ mV}.$$

The effective value of the field strength is thus

$$\frac{V_{\mathrm{eff}}}{h} = \frac{2.3}{0.054} = 42.2 \text{ mV/m}.$$

Where the field strength is to be explored over a large area, the field-strength meter can be installed in a motor van. It is then an advantage to use, in place of the frame aerial, a vertical aerial 0.5 to 1 m in height, which has no directional characteristics at all and which does not have to be taken down when moving from place to place. The absolute calibration of the mobile field-strength meter is carried out with a field strength previously explored by a frame aerial by the method described above.

b) *Signals radiated from radio receivers.* It is essential to suppress as far as possible the radiation of signals from auxiliary oscillators in superheterodyne receivers in view of the mutual interference which they cause in reception. In a particular case, it was stipulated that the field radiated by a receiving set equipped with a particular aerial should not exceed 5 microvolts per m at a distance of 25 m.

To investigate whether a particular set satisfies this condition on a wave length of 30 m, a frame aerial is used which is adapted to this wave length and has two annular windings of 40 cm diameter. The effective height in the optimum position is 0.0525 m. The field radiated by the receiver under investigation is so small that the calibration signal is much more powerful than the induced signal, so that the frame aerial before calibration does not have to be turned into the position for which the received signal has a minimum strength.

The amplification is so adjusted that a reading of 10 is obtained with a calibration voltage of 500 microvolts and 1/1000 attenuation.

After switching off the calibration signal, a deflection of only 3.1 is found for the induced signal without attenuation. The signal strength is therefore

$$V_{\text{eff}} = 500 \cdot \frac{3.1 \cdot 1}{10 \cdot 1000} = 0.155\ \mu\text{V}.$$

and the effective value of the field strength is:

$$E_{\text{eff}} = \frac{0.155}{0.0525} = 2.95\ \mu\text{V/m},$$

c) *Recording the signal strength of a distant transmitter.*

The transmission of electromagnetic waves over very great distances has been the subject of systematic investigations during recent years. *Fig. 10* reproduces an oscillogram of the signal strength of the L.R.1. transmitter at Buenos Aires, which was recorded at Eindhoven during the night hours.

For this measurement a vertical aerial 12 m in height was used, which owing to the lack of top capacity had an effective height of only 6 m. The measuring equipment was supplemented by a recording milliammeter. As the automatic gain control was in circuit, the deflection of the pointer increased more slowly than the intensity of the received signal; this is shown in fig. 10. In cases where the field strength is expected to show very large variations, the amplitude can be controlled still more effectively than was done in the present case.

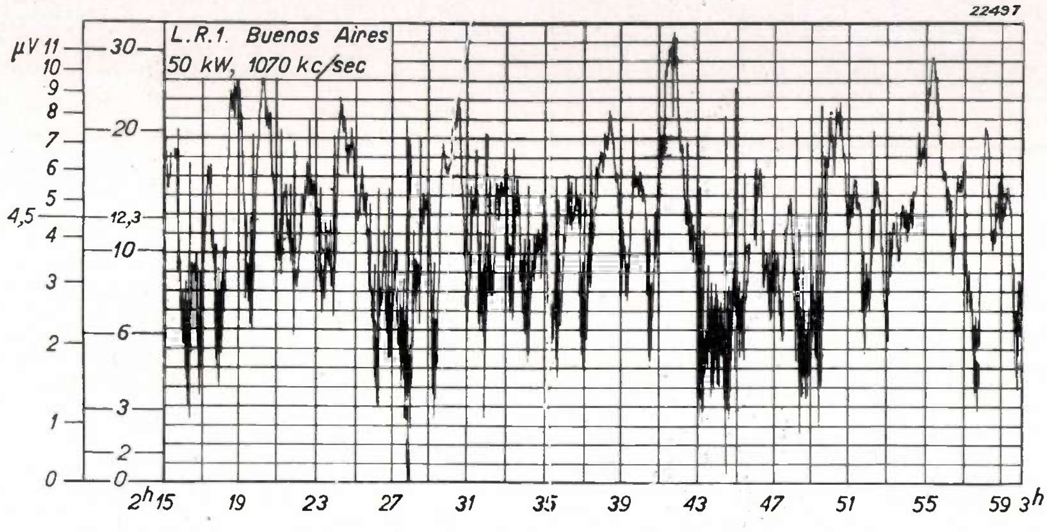


Fig. 10. Oscillogram of the field strengths of the L.R.1. transmitting at Buenos Aires, recorded on December 15, 1935, at Eindhoven. The scale 0 to 30 was obtained by plotting the deflection for calibration signals of 1 and 0.5 millivolt using different settings of the attenuator. A station with a field strength of 450 micro-volts per m and about the same frequency thus gives a deflection of 12.3 = 4.5 microvolts per m when using an amplification 100 times smaller than employed in the according; this deflection gives the absolute calibration of the scale.

The division of the relative scale from 0 to 30 was carried out with a constant calibration signal and the attenuator, by keeping the adjustment of the automatic gain control unaltered and varying the intermediate frequency signal in the ratios of 1, 3, 10, 30, etc. and then determining the deflection on the oscillogram. The scale value in microvolts per m was determined by picking up a station transmitting on approximately the same frequency and having a constant field strength, the latter being measured by the method described under a) above.

If no interference is experienced in reception, field strengths of less than 0.1 microvolt per m can be measured by this method. In the recording strip reproduced, the interference level lay between 1 and 2 microvolts per m.

Use of the Apparatus as a Voltmeter

In conclusion, reference should be made to an important use to which this meter can be put, viz., as a selective voltmeter with adjustable sensitivity and high aperiodic input impedance for measuring sinusoidal alternating voltages. After removing the frame aerial or aerial coil, the alternating voltage to be measured is applied to the grid and cathode of the first amplifying valve, using the special terminal provided for this purpose. With a switch the unknown voltage V_x can be compared to the voltage V_i of the calibration oscillator (*fig. 11*). The sensitivity of the instrument as voltmeter is approximately 1 millivolt.

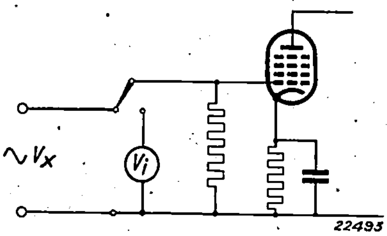


Fig. 11. Circuit using the meter as voltmeter. The unknown voltage V_x is compared to the voltage V_i of the calibration oscillator.

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

No. 1172: M. J. Druyvesteyn and N. Warmoltz: Ein neuer Dunkelraum in der Nähe einer Glühkathode in einer Bogenentladung (Physica 4, 51 - 68, Jan. 1937).

A new dark space round the oxide-coated cathode of an arc discharge, which has already been briefly described in Abstract No. 1117, is discussed in detail in this paper. This dark layer is about 20 times the thickness of the space charge sheath close to the cathode and occurs at current densities above 0.1 amp per sq cm and at pressures between 10^{-2} mm and 3 mm. At lower pressures the new dark sheath disappears and a bright sheath with the same dimensions appears in its place. Starting from the cathode, a sharp increase of the concentration and a decrease of the mean energy of the fast electrons occur at the boundary between this layer and the bright part of the discharge. This may be accounted for by a scattering of the fast electrons which leave the boundary of the space charge sheath close to the cathode. This scattering increases with increasing current. A dark or bright sheath is observed according as the increase in concentration of the fast electrons outside this layer or the reduction in mean energy of these electrons has the greater effect on the emission of light.

No. 1173: K. F. Niessen: Über die Wirkung eines vertikalen Dipolsenders auf ebener Erde in einem Entfernungsbereich von der Ordnung einer Wellenlänge (Ann. Phys., **28**, 209 - 224, Jan. 1937).

Formulae which are sufficiently accurate for application up to distances of the order of a wave length are deduced for the intensity of the horizontal component of the magnetic field and the vertical component of the electric field of a vertical dipole transmitter. These formulae are then simplified for types of soil and wave lengths for which the Sommerfeld numerical distance is small compared with unity.

No. 1174: H. C. Hamaker: A general theory of lyophobic colloids. II. Rec. Trav. chim. Pays Bas, **56**, 3 - 25, Jan. 1937).

Continuing the analysis reviewed in Abstract No. 1154, a formula is deduced which gives the total energy of interaction between two colloidal particles, expressed approximately as a function of the distance apart of the particles; it also contains as independent parameters the charge E of a colloidal particle and the electrolyte concentration c. In an E - c diagram the different states of a colloid can be represented as a function of E and c. A variety of phenomena, such as peptisation, reversible and irreversible flocculation and thixotropy, can be analysed in a clear and simple way by means of the curves in the E - c diagram.

No. 1175: J. A. M. van Liempt and P. Leydens: Die Farbenwiedergabe beim Photographieren mit Neonlicht (Rec. Trav. chim. Pays Bas, **56**, 26 - 28, Jan. 1937).

Colour reproduction in photography with neon light is investigated for different negative materials;

be measured by this method. In the recording strip reproduced, the interference level lay between 1 and 2 microvolts per m.

Use of the Apparatus as a Voltmeter

In conclusion, reference should be made to an important use to which this meter can be put, viz., as a selective voltmeter with adjustable sensitivity and high aperiodic input impedance for measuring sinusoidal alternating voltages. After removing the frame aerial or aerial coil, the alternating voltage to be measured is applied to the grid and cathode of the first amplifying valve, using the special terminal provided for this purpose. With a switch the unknown voltage V_x can be compared to the voltage V_i of the calibration oscillator (*fig. 11*). The sensitivity of the instrument as voltmeter is approximately 1 millivolt.

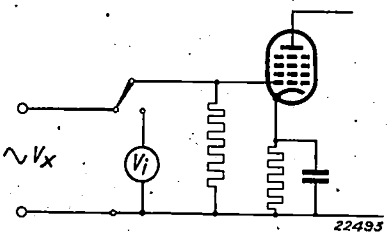


Fig. 11. Circuit using the meter as voltmeter. The unknown voltage V_x is compared to the voltage V_i of the calibration oscillator.

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

No. 1172: M. J. Druyvesteyn and N. Warmoltz: Ein neuer Dunkelraum in der Nähe einer Glühkathode in einer Bogenentladung (Physica 4, 51 - 68, Jan. 1937).

A new dark space round the oxide-coated cathode of an arc discharge, which has already been briefly described in Abstract No. **1117**, is discussed in detail in this paper. This dark layer is about 20 times the thickness of the space charge sheath close to the cathode and occurs at current densities above 0.1 amp per sq cm and at pressures between 10^{-2} mm and 3 mm. At lower pressures the new dark sheath disappears and a bright sheath with the same dimensions appears in its place. Starting from the cathode, a sharp increase of the concentration and a decrease of the mean energy of the fast electrons occur at the boundary between this layer and the bright part of the discharge. This may be accounted for by a scattering of the fast electrons which leave the boundary of the space charge sheath close to the cathode. This scattering increases with increasing current. A dark or bright sheath is observed according as the increase in concentration of the fast electrons outside this layer or the reduction in mean energy of these electrons has the greater effect on the emission of light.

No. 1173: K. F. Niessen: Über die Wirkung eines vertikalen Dipolsenders auf ebener Erde in einem Entfernungsbereich von der Ordnung einer Wellenlänge (Ann. Phys., **28**, 209 - 224, Jan. 1937).

Formulae which are sufficiently accurate for application up to distances of the order of a wave length are deduced for the intensity of the horizontal component of the magnetic field and the vertical component of the electric field of a vertical dipole transmitter. These formulae are then simplified for types of soil and wave lengths for which the Sommerfeld numerical distance is small compared with unity.

No. 1174: H. C. Hamaker: A general theory of lyophobic colloids. II. Rec. Trav. chim. Pays Bas, **56**, 3 - 25, Jan. 1937).

Continuing the analysis reviewed in Abstract No. **1154**, a formula is deduced which gives the total energy of interaction between two colloidal particles, expressed approximately as a function of the distance apart of the particles; it also contains as independent parameters the charge E of a colloidal particle and the electrolyte concentration c. In an E - c diagram the different states of a colloid can be represented as a function of E and c. A variety of phenomena, such as peptisation, reversible and irreversible flocculation and thixotropy, can be analysed in a clear and simple way by means of the curves in the E - c diagram.

No. 1175: J. A. M. van Liempt and P. Leydens: Die Farbenwiedergabe beim Photographieren mit Neonlicht (Rec. Trav. chim. Pays Bas, **56**, 26 - 28, Jan. 1937).

Colour reproduction in photography with neon light is investigated for different negative materials;

it is found that neon is a satisfactory source of light for orthochromatic materials. However, owing to the shorter exposure times, preference is usually shown for panchromatic material, with which good results may be obtained by using a mixture of neon and mercury lights. Nevertheless, the quality of this mixture is slightly less satisfactory than the mixture of sodium and mercury lights (cf. Philips Techn. Rev. **2**, 24, 1937).

No. 1176: J. A. M. van Liempt and J. A. de Vriend: Studien über das Verbrennungslicht einiger Metalle (Rec. Trav. chim. Pays Bas, **65**, 126 - 128, Jan. 1937).

The quantity of light emitted during combustion, as well as the comparative efficiencies of the combustion process is determined for tungsten, molybdenum, tantalum, cerium and carbon.

No. 1177: M. J. O. Strutt: Les performances de certains types des lampes changeuses de fréquence dans les récepteurs toutes ondes (Onde él. **16**, 29 - 44, Jan. 1937).

The various requirements which mixing valves for short waves have to satisfy are summarised under eleven headings. The most important characteristics of octodes operating on short waves are discussed and the means indicated for overcoming the difficulties encountered. The characteristics of octodes in general are discussed and the latest improvements in construction are briefly stated. In this connection, data are given of the behaviour of the latest types of valves for short waves, which show that they are improvements on the types in use hitherto.

No. 1178: J. H. Gisolf: The counting of electrons by means of a discharge tube (Physica, **4**, 69 - 70, Febr. 1937).

Different types of discharge tubes can be used as electron counters in a special circuit, which is described in detail in this paper.

No. 1179: F. M. Penning: Ein neues Manometer für niedrige Gasdrücke, insbesondere zwischen 10^{-3} und 10^{-5} mm. (Physica **4**, 71 - 75, Febr. 1937).

A glow-discharge tube in a magnetic field may be adapted for the measurement of gas pressures, especially between 10^{-5} and 10^{-3} mm. The current passing between two electrodes is a useful measure of the pressure. The length of the negative glow on the rod-shaped cathode of another glow-discharge tube is a very suitable indicator for this current. This new manometer, which is marketed by E. Leybolds' Nachfolger A.G., Cologne-Bayenthal, under the name of the Philips' Vacuummeter, is very simple in its use. It gives an uninterrupted, instantaneous and clearly visible indication of very low pressures, those of condensable vapours included. (Cf. also the description of this instrument on p. 201 of this issue).

No. 1180: J. Sack: The charge of electricity on metal drops which leave the welding rod during electric-arc welding (Physica, **4**, 104 - 106, Febr. 1937).

Liquid drops of metal, which during welding pass from the welding rod to the work, are caught before reaching the latter in order to measure their charge with a Wulf string-electrometer, and in addition their weight is determined. The drops from the coated iron welding rod PH-50 always carry a negative charge, both in A.C. and in D.C. welding, irrespective of whether the polarity of the welding rod is positive or negative. This negative charge of 2 to $3 \cdot 10^{-9}$ coulomb is relatively high for such small drops of metal of only 3 to 4 mm diameter, and can produce a voltage of 10 kV below earth. A satisfactory explanation of this phenomenon has not yet been given.

No. 1181: A. Bouwers, F. A. Heyn and A. Kuntke: A neutron generator (Physica **4**, 153 - 159, Febr. 1937).

In this paper, a tube is described by means of which high-speed ions and neutrons can be produced. A high-voltage generator used in connection with this tube is also discussed, as well as a new method for supplying the potential to the ion-source.

No. 1182: F. A. Heyn: The radio-activity of cobalt, nickel, copper and zinc induced by neutrons (Physica **4**, 160 - 165, Febr. 1937).

The radio-activity of cobalt, nickel, copper and zinc induced by neutrons of different energies is investigated. If copper and zinc are bombarded with fast neutrons, half-life values are obtained which suggest the existence of a new nuclear reaction in which two neutrons are emitted by the bombarded nuclei. Experiments to confirm this assumption are described.

VOL. 2. No. 8 AUGUST 1937

Philips Technical Review

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

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

SURGE PROTECTION OF LOW-TENSION OVERHEAD LINES

by J. W. G. MULDER.

Summary. In this article the requirements for an efficient protection of overhead low-tension transmission lines against excess voltages and surges are discussed. The construction and method of operation of the Philips lightning arrester are described.

Introduction

Overhead power lines which are used for transmission purposes in nearly all countries are comparatively frequently exposed to excess voltages created by electrical phenomena taking place in the atmosphere. The probability of a transmission line being directly struck by lighting is not very great; but it is by no means exceptional for these lines, which possess a certain capacity with respect to earth, to become charged by atmospheric effects to a potential under which the weakest parts of the insulation of the associated installations must necessarily fail. Since the insulators from which the lines are suspended are usually adequately rated to withstand high excess voltages, it is frequently the coils in the electricity meters and the insulation of the house connections, etc., which suffer when a surge is experienced.

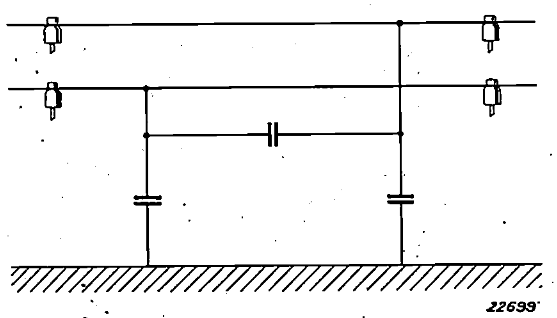


Fig. 1. Equivalent circuit of an overhead line.

Over-voltages may also be produced by electromagnetic induction when powerful surges occur in neighbouring conductors, such as metal masts and trees, when these are struck by lightning. Travelling waves are then produced which are liable to damage the insulation and the whole of the plant connected to the network.

In low-tension networks, over-voltage conditions can be initiated by a host of causes other than atmospheric discharges, as for instance, high short-circuit currents in adjoining high-tension lines, or by the impulses sustained on switching heavy loads, etc. But since excess voltages due to these causes are far less powerful than those due to atmospheric causes, they will not be discussed in this article, particularly as it may be assumed that the whole of the insulation has been suitably adapted to withstand these abnormalities of non-atmospheric origin.

Intensity of Lightning Flashes

The magnitude of the discharge currents occurring in lightning strokes has been investigated in some detail in reference to high-tension transmission networks. Various methods are used for measuring the intensity of these currents; thus the diameter of the puncture produced by the discharge current in a thin sheet of paper held between two metal plates can be measured and the corresponding current intensity found by calibration. Better results can be obtained with magnetisable steel needles; in this method the maximum magnetic field intensity which has been applied to a needle is determined by measuring the magnetism remanent in the needle after it has been left near a high-tension pole; this measurement is independent of the duration of the field. The maximum value of the discharge current can then be calculated for the particular arrangement used from the maximum magnetic field intensities measured at the location of the needle.

Mc Eachron and Mc Morris [1]) used this method for recording 156 positive and 255 negative discharge currents at 1225 different points. Of these 411 recorded discharge currents 0.73 per cent. were found to be above 15 000 amps, the maximum intensity measured being 17 000 amps. These investigators also calculated from their measurements the probability of lightning discharges occurring with intensities above a specific limiting value. Naturally this probability decreases very rapidly as the intensity increases, as may be seen from *fig. 2* which shows the time intervals at which a discharge above a certain current intensity may be expected in one needle within a given area.

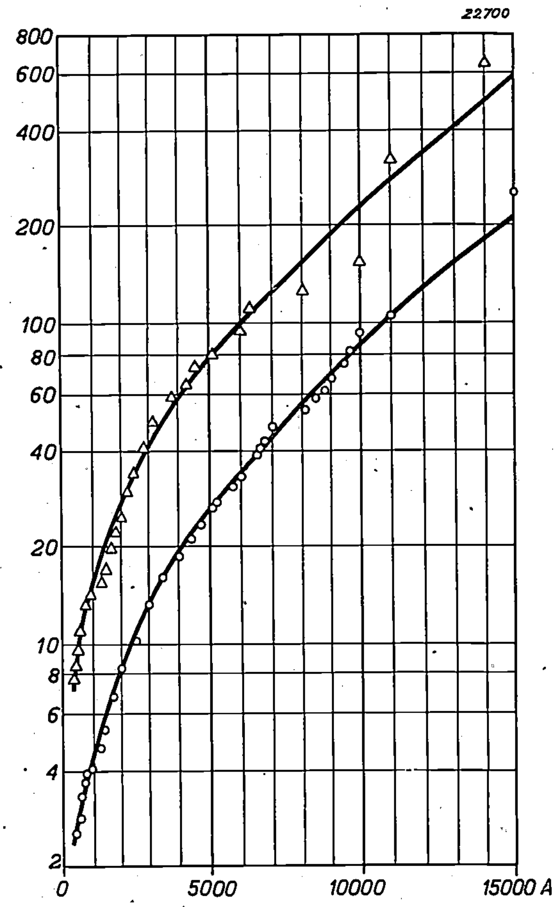


Fig. 2. Time in years in which a single discharge in one needle above a certain current intensity may be expected to occur, plotted against the current intensity (according to Mc Eachron and Mc Morris). The upper curve refers to an urban and the lower curve to a rural district.

A systematic investigation of atmospheric discharges is clearly both cumbersome and tedious, for the occurrence of these discharges depends largely on fortuitous circumstances. During the last few years, however, a mass of observational data has been collected and correlated both in the United States by the Franklin Institute and in Switzerland. It is found that the intensity of these discharge currents covers a very wide range and also that current intensities considerably higher than the values stated above are possible even in Europe. An intensity of 30 000 to 60 000 amps is by no means exceptional, and in isolated cases currents up to even 100 000 and 200 000 amps may be encountered. Under the auspices of the Dutch Standards Institution, a committee was set up to make recommendations regarding the precautions necessary as a protection against lightning. This committee will shortly publish the results of its investigations. The vast amount of data which has been collected and classified, considerably amplifies our understanding of the generation and effects of lightning discharges as compared with the state of knowledge of a few years ago. One of the tasks of the committee has been the preparation of a map showing the nature and distribution of lightning strokes in the Netherlands over a certain period. This map shows that the distribution of the strokes is by no means arbitrary, but that in certain regions the probability of lightning strokes causing damage is much greater than in others. It has thus been shown, in confirmation of everyday experience, that the danger of lightning strokes is particularly severe in the neighbourhood of rivers and in places where the subsoil water levels are high. According to the experience of the electricity authorities, an exceptionally high percentage of discharges on overhead lines takes place in just these areas. Various places, such as isolated farms and the terminations of overhead lines, frequently have the reputation that the associated electricity meters become damaged at regular intervals, for instance, monthly.

Precautionary Measures

For many years, attempts have been made to protect overhead lines by equipping them with devices for conducting the excess voltage to earth. These devices have to meet extremely severe specifications, and it was soon found that the horn arresters (*fig. 3*) originally used did not provide an altogether adequate safeguard. Although quite able to deal with surges of the intensities in question while their shape facilitated the extinction of the power arc, the initial breakdown voltage of these

[1]) Mc Eachron and Mc Morris: Discharge currents in distribution arresters. El. Eng., 54, 1395 - 1399, December, 1935.

[2]) Other protective devices such as condensers and damping circuits are not discussed in this article.

arresters yet proved to be too high. The distance apart of the two poles of the spark gap had in fact to be made so small that a flashover occurred at potentials which were below those liable to damage the connected meter coils. Moreover, any dirt accumulating could readily cause a short circuit between electrodes situated so close together, so that it became customary to make the gap between the electrodes much greater than was desirable for low initial breakdown. Since with the majority of

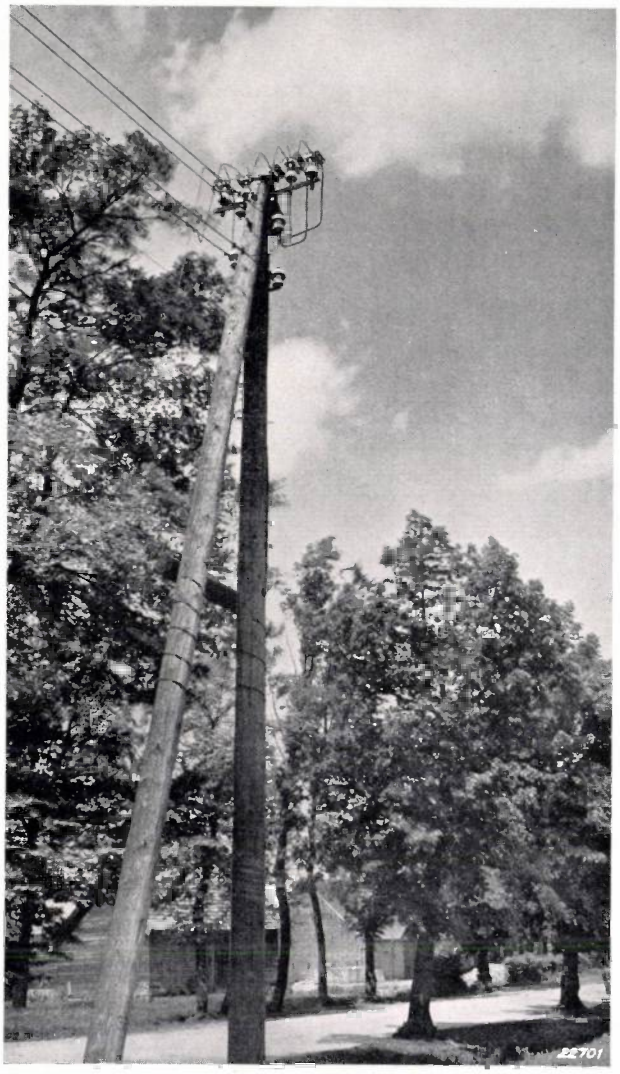


Fig. 3. Method of mounting horn arresters; these are always fixed above the pole to allow the arc to travel upwards out of the horns.

spark gaps the flashover takes place with a certain delay both in air and rarefied gases, the socalled dynamic breakdown voltage and not the static breakdown voltage has to be taken into consideration here.

The definition of the dynamic breakdown voltage implies that a certain time interval is required for the discharge to be built up. Hence, with a particular spark gap, the breakdown voltage is the higher the shorter the time the potential in question is applied to the electrodes. The lowest breakdown voltage is thus observed with direct voltages by the static method, while for potentials subject to rapid fluctuations, as with a steep-fronted surge wave, much higher values are encountered by the dynamic method. But a certain time must also elapse to puncture the insulation of an electrical installation. An essential characteristic of an efficient lightning arrester is therefore that the dynamic breakdown voltage of the arrester is lower than that of the insulation which has to be protected and which is in parallel to it (*fig. 4*).

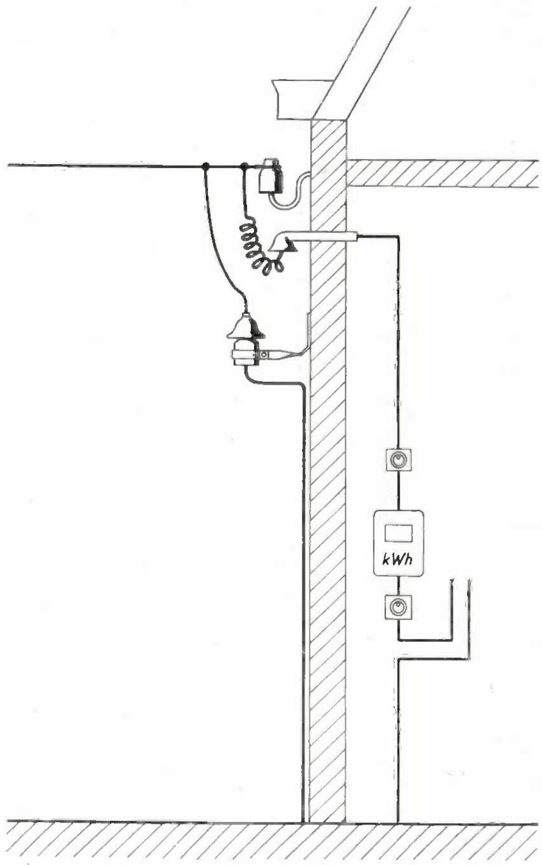


Fig. 4. Circuit of a house installation incorporating a lightning arrester.

But horn arresters operate with a considerable inertia and in fact only respond rapidly to very high voltages; the protection provided for the insulation of the connected installations and plant is thus inadequate.

Another requirement is that the discharge must be automatically extinguished in the lightning arrester as soon as the over-voltage wave disappears. The voltage at which the discharge in the arrester collapses, i.e. the extinction voltage, must, in consequence, be made as high as practicable, in any case much higher than the mains voltage.

The principal requirements of an efficient arrester are therefore:

a) Low dynamic breakdown voltage,
b) Capacity for handling large discharge currents, and
c) Sufficiently high extinction voltage.

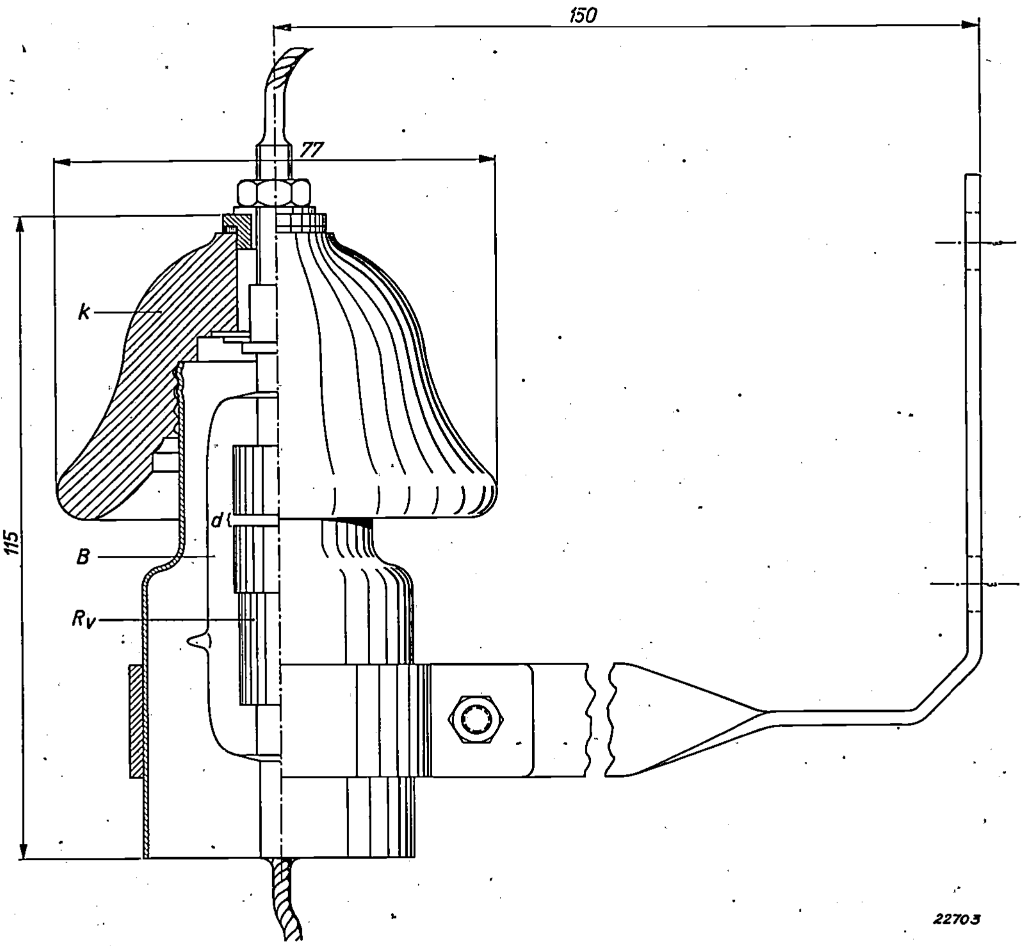


Fig. 5. Section through a Philips lightning arrester (type 4394) for low-tension networks, with dimensions given in mm. A hood K protects the discharge tube B against rain. The discharge gap d is located between two heavy iron electrodes. R_v is the series resistance.

Characteristics of Lightning Arresters

a) and b) are determined from the characteristic curves which can be plotted with an impulse-voltage generator and a cathode-ray oscillograph:

1) A current-voltage curve with a surge of specific wave front gradient and duration;
2) A socalled "protection characteristic" ($V=f(V_l)$) which indicates the limiting voltage V which is produced at a certain point of an arrester when a surge with a peak value V_l arrives from a distance which is large as compared with the length of the impulse. This voltage is determined for a certain impedance of the line (e.g. $Z = \sqrt{\frac{L}{C}} = 500$ ohms) and is usually measured at one end of the line; as a rule a steeper wave front is used than when plotting the current-voltage characteristic.

Permissible Loading

b) and c) are determined by loading tests in which the conditions of test are closely specified, and which *inter alia* must not appreciably alter the form of the characteristics in question. Since, as we shall see, extinction is determined by the temperature of the electrodes, a certain number of loading impulses must be applied at prescribed intervals; at the same time an alternating voltage of 50 cycles (1.2 to 1.5 times the mains voltage) is applied to the arrester. If required the loading impulses can be synchronised with the most dangerous point on the low-frequency voltage phase.

In addition to the above, the arrester has also to satisfy the following requirements:

d) The breakdown voltage at 50 cycles must be much higher than the normal voltage (e.g. more than twice the latter) and in fact so high that the arrester does not respond to the highest normal excess voltages.

e) The insulation of the arrester for low-tension networks must be able to sustain safely an over-voltage of 3 000 volts at 50 cycles, when the spark gap is removed and the insulator is exposed for 5 minutes to a water spray of 3 mm per minute impinging at an angle of 45 deg.

f) Its mechanical construction must be sufficiently robust to reduce as far as possible the danger of a permanent short or earth when the device is damaged and an overload is applied.

Official Test Specifications

Official specifications of tests for lightning arresters have not yet been formulated in any country. It is obvious from the data given above regarding the intensities which may occur with lightning discharges that devices capable of withstanding a direct stroke would prove far too costly for adoption on a large scale. Since many thousands of lightning arresters are required and must be located at comparatively short distance apart along the lines, no attempt has been made to provide adequate protection against a direct stroke.

In Germany the Verein deutscher Elektrotechniker (V.D.E.) and in the Netherlands the Keuringsbureau van Electrotechnische Materialen van de Vereeniging van Directeuren van Electriciteitsbedrijven ("inspection and testing bureau for electro-technical materials of the directors of the electrical power stations" or "Kema") have drafted specifications which at the present time are being examined as regards their practical suitability. The two drafts differ in a number of essential points, the principal differences being in regard to the extinction voltage required, whether synchronisation in the extinction test is desirable or not, the number of loading impulses required and the most suitable intervals between successive load tests, and the form of the impulse wave to be used. It should be pointed out that the results of a test depend largely on the form of the wave.

The three principal requirements, viz., of a low dynamic breakdown voltage, large discharge capacity and a high extinction voltage, are not simultaneously promoted to the same extent by the majority of design factors, so that the practical compromise which is arrived at depends on which of the three desiderata is regarded as the most important one.

The Philips lightning arrester for low-tension networks of the type described in this article has an adequate extinction voltage (1.25 × working voltage) and as low as possible a dynamic breakdown voltage (1 500 to 3 500 volts, with an average of 2 400 volts found in tests on 123 specimens).

Modern Arresters

Considerable advances have been made in recent years in the construction of efficient and suitable lightning arresters, and these are now gradually replacing the obsolete horn arresters. The experience gained with these modern types of arresters has been satisfactory, and there has been a marked reduction in the number of faults sustained in the overhead lines of low-tension networks equipped with such arresters. From this it may be concluded that these devices are reliable in service, and in consequence, they are being installed in steadily-growing numbers in feeder networks.

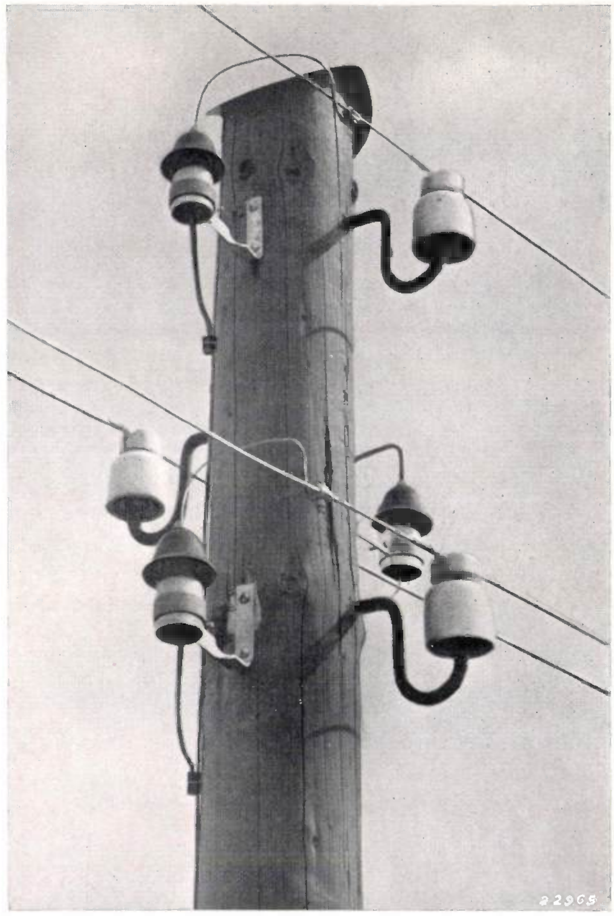


Fig. 6. Showing how the lightning arrester is mounted; it is no longer necessary to place it above the pole or mast, which thus makes the arrangement more compact and easier to inspect than with horn arresters.

A description of a lightning arrester manufactured by Philips for low-tension networks is given below, with detailed reference to its main characteristics.

Philips Lightning Arrester

A section through the Philips lightning arrester is shown in *fig. 5*. The porcelain hood K provides the necessary protection against rain to allow the arrester to be mounted in the open. Below this

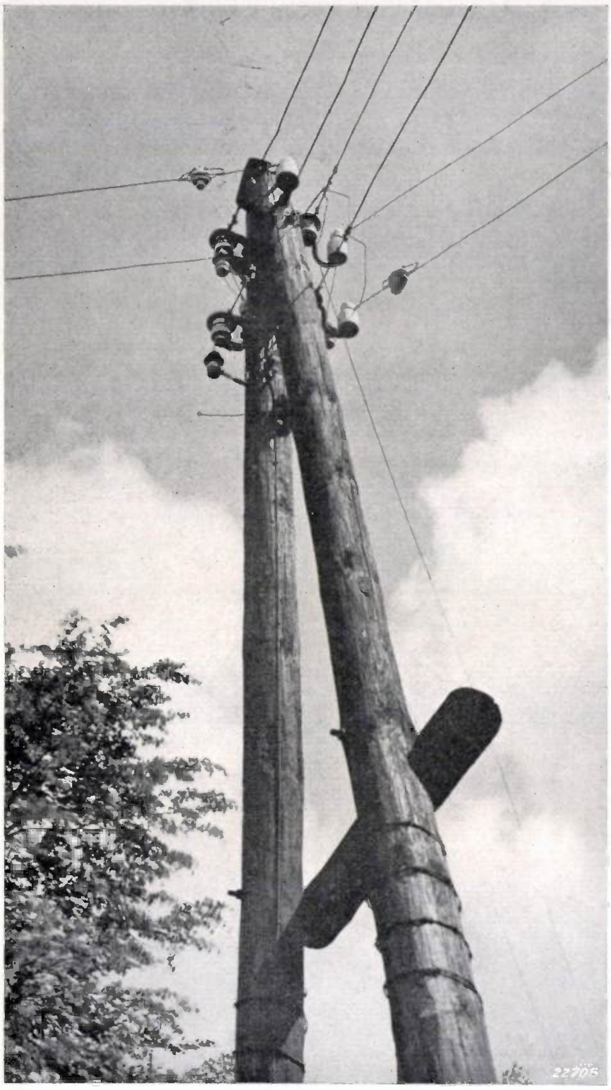


Fig. 7. Low-tension lines safeguarded with one of Philips lightning arresters. This terminal pole was exposed to an abnormal danger from lightning strokes, according to practical experience, before the arresters were installed. The arrester on the left under the insulators has failed; one of its electrodes is suspended by a wire and is visible from a considerable distance.

hood is a hermetically-sealed discharge tube B filled with gas at a low pressure and enclosing a discharge gap d. A resistance R_v is connected in series with G. Constancy of the discharge gap and the resistance are ensured, irrespective of changes in atmospheric conditions, by enclosing these two components in a sealed chamber. This is important, as the properties of these components to a large extent determine the dynamic breakdown voltage and the extinction voltage.

To avoid high self-inductances, loops and bends are avoided as far as possible in the leads connecting the arrester to the feeders and to the earth lead (cf. *fig. 6*). The resistance of contact of its earth connection should be made as small as possible by care in making joints.

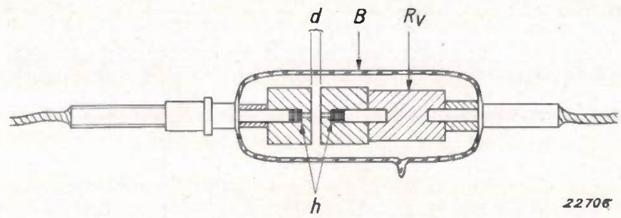


Fig. 8. Section through the discharge tube B. d is the discharge gap and R_v the series resistance. The recesses h contain substances from which alkali metals are liberated on heating.

A useful characteristic of this type of arrester, to which attention must be called, is that if the device is directly struck by lightning the glass tube is destroyed, which allows the lower electrode to drop and hang suspended from the connecting wire in a conspicuous position. This feature, as may be seen from *fig. 7*, considerably simplifies the location of damaged arresters.

Properties of the Discharge

A detailed discussion of the vital component of the arrester, viz., the discharge gap, will now be given. The discharge tube is shown in section in

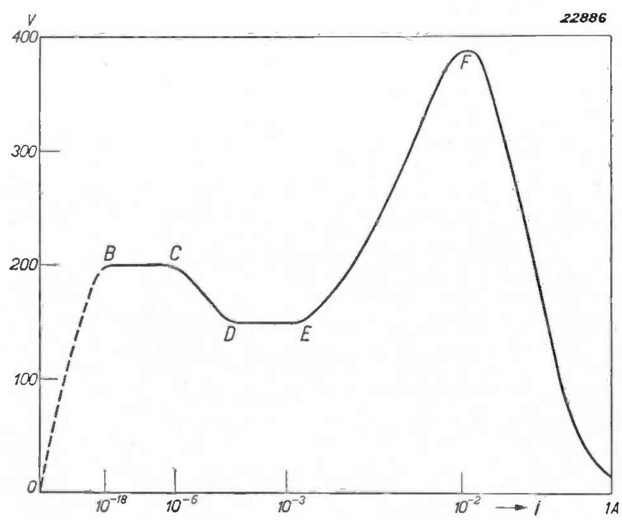


Fig. 9. Diagrammatic characteristic of voltage and current with a discharge between two parallel plates.

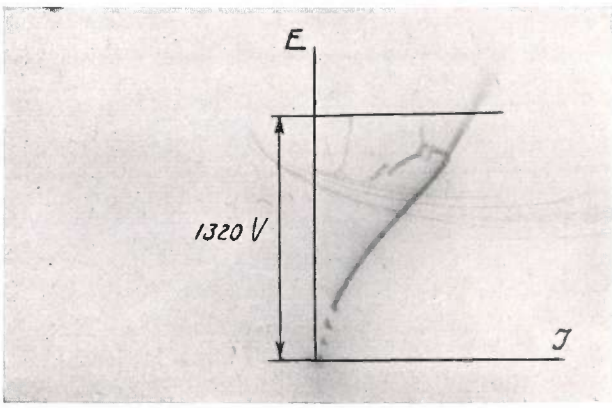
fig. 8. The electrodes are made of stout iron because a high thermal capacity is essential for the rapid extinction of the discharge. The recesses h contain

a mixture of substances from which alkali metals are liberated by chemical reaction, these metals coating the electrode surfaces. The reaction being progressive, this surface coating is regenerated during each discharge, to that it is always in a satisfactory condition, thus ensuring the requisite low dynamic breakdown voltage.

In the April, 1937, issue of Philips Technical Review, M. J. Druyvesteyn and the present author have discussed the general characteristics of discharges between parallel plane plates. The method of operation of the lighting arrester can be readily gathered from fig. 1 in that article, which has been reproduced as *fig. 9* here. In the present case, however, the voltage is drawn on a different scale, although the currents are approximately correct.

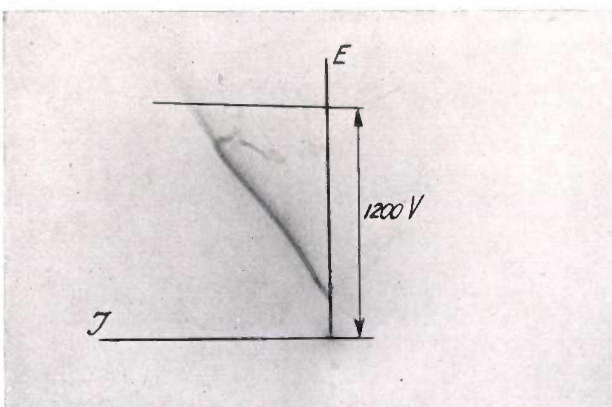
For the moment we shall neglect consideration of the delay in operation which in fact occurs only with an exceptionally steep voltage increase. It is apparent that the voltage must exceed a certain threshold value $B\ C$ for a fairly appreciable current

a



22961

b

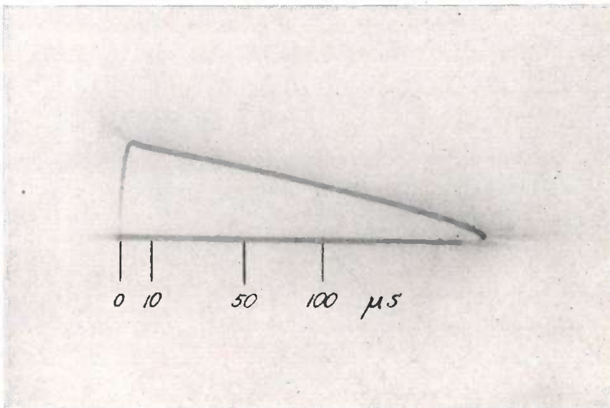


22962

Fig. 10. a) Current-voltage characteristic of a Philips lightning arrester for low-tension networks, type 4394, when a positive current impulse of 300 amps with a wave front of 30 kilovolts per microsecond is applied.
b) The same for a negative current impulse of the same form and intensity.

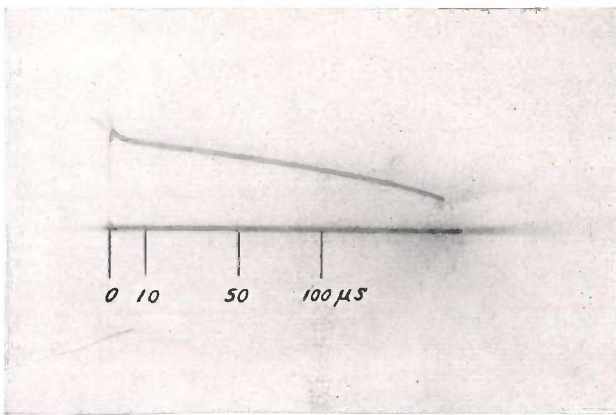
[3]) Philips techn. Rev., 2, 122, 1937.

to flow through the tube. Over the section $C\ D\ E\ F$ a corresponding current discharge takes place. The level of the threshold value $B\ C$ is determined



a

22963



b

22964

Fig. 11a). Variation of current and b) variation of voltage plotted against the time when a discharge impulse of 300 amps is passed through a Philips lightning arrester, type 4394. The peak values of the current and voltage are 300 amps and 1320 volts respectively.

by the number of electrons which happen to be present in the gas. If no electrons are present, the threshold value is not reproducible; but the potassium coating on the electrodes (fig. 8) is slightly radioactive and thus furnishes sufficient initial electrons for the discharge to be always initiated at a specific voltage.

The shape of the section $F\ G$ in which the discharge is transformed to an arc is determined by the temperature of the electrode (cathode), which increases from F to G. To enable high current intensities to be dissipated, the threshold voltage F must be exceeded, this voltage being considerably higher than $B\ C$. But if the temperature of the electrode rises too rapidly after reaching the point F, the discharge voltage along the branch $F\ G$ may drop below the normal supply voltages. This signifies that after the surge disappears the arrester no longer has an extinguishing action, since the extinction voltage is below the supply voltage. The arrester is then destroyed, for along the descending curve $F\,G$ the current can assume arbitrary high values until the material of the electrodes is fused.

But such a disastrously rapid increase in temperature is avoided by the high thermal capacity of the electrodes. Extinction is also promoted by the ease with which heat is conducted from the surface inwards as soon as the discharge current decays; the surface thus becomes rapidly cooled. It should be stated in this connection that with impulsive discharges the transient value of the current can be 1 000 to 100 000 times greater than shown in fig. 9 for raising the electrode to the requisite temperature, since this figure relates to a continuous static load.

The built-in series resistance R_v promotes regular extinction, for if the circuit has an adequate resistance a progressive increase of the current due to the mains voltage is prevented. But a resistance of this magnitude would make it impossible for the over-voltage impulse to assume considerably higher values than permissible with the steady current, as is in fact necessary. This difficulty may be overcome by making the resistance of a material whose specific resistance rises considerably with increase in voltage.

In this way, a rapid breakdown in the presence of an excess voltage as well as rapid extinction when the normal voltage is restored can be satisfactorily obtained.

In conclusion, reference must be made to some results of measurements made by the "Kema" on our behalf as regards the behaviour of the arrester with travelling waves. In these measurements, a voltage wave with a wave-front gradient of 30 kilovolts per micro-second and a duration of 100 microseconds was earthed through an arrester. The current-voltage characteristic was plotted for both polarities (*fig. 10*, *a* and *b*) for a current impulse of 300 amps; the variation of current and voltage are shown in the oscillograms reproduced in *fig. 11 a* and *b*. It may be concluded from these investigations that the lightning arrester operates so rapidly and at such a low voltage, and hence suppresses the voltage waves with such effectiveness, that it can with certainty prevent a breakdown of the insulation in electrical installations even when this insulation is of a low order.

MAGNET STEELS

by J. L. SNOEK.

Summary. The principal properties of a magnet steel are remanence and coercive force. How these properties are connected with the composition and the structure of the steel is discussed in this article on the basis of current theories of magnetism.

Introduction

Permanent magnets offer the important advantage, in comparison to electromagnets, that no supplementary source of energy is required for the purpose of maintaining the magnetic field. Where a constant magnetic field of given strength is required within a certain space, as in loudspeakers, gramophone pick-ups, ribbon microphones, small dynamos and measuring instruments, the permanent magnet is preferable to the electromagnet provided its use does not entail an increase in cost or weight.

Owing to the much improved magnetic properties of modern magnet steels, a very wide field has been opened up for the use of permanent magnets which in its turn has given the designer a variety of new problems to deal with. In this article, recent developments in the production of magnet steels will be discussed on the basis of current theories of magnetism.

Principles of Ferromagnetic Hysteresis

As a starting point, the hysteresis curve for an ordinary ferromagnetic material is shown in *fig. 1.*

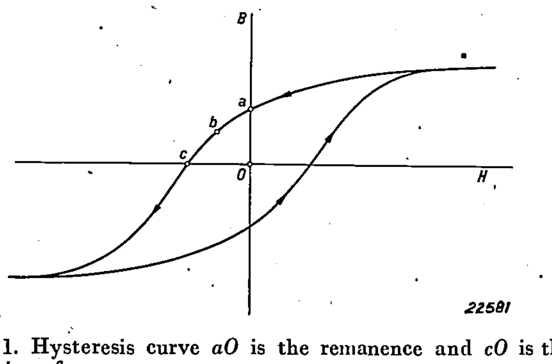


Fig. 1. Hysteresis curve *aO* is the remanence and *cO* is the coercive force.

The chief characteristic of this curve is that in strong magnetic fields the induced magnetism increases little, or not at all, with further increase in the field intensity, in other words with field intensities of the order of 1000 oersted a state of magnetic saturation is reached. The magnitude of this saturation is of interest in the course of our investigation, since as a rule the residual induction or remanence, which remains after the removal of the magnetic field, is roughly equal to half the magnetic saturation. If a high remanence is required, it is therefore important to use alloys having a high magnetic saturation.

If a magnetic field is applied in opposition to the direction of magnetisation and is steadily increased, the residual magnetism will become reduced as shown by the curve *a b c*, which has therefore been termed the "demagnetisation curve". The remanence eventually becomes zero at the point *c*. The field intensity applied when this point is reached is termed the coercivity or coercive force of the material and may be regarded as a measure of the hysteresis.

In determining the practical applications of a particular magnet steel, the remanence and the coercive force are not the only two properties which have to be taken into consideration, for the form of the demagnetisation curve is also an important factor. This curve must in fact be of such a shape that the product of the residual magnetism and the opposing demagnetising field can be made as large as possible. But since the shape of this curve, when reduced to equivalent values of the remanence and the coercive force, is roughly the same for all magnet steels, provided that flaws in casting and heterogenities in structure are avoided, it is apparent that the magnetic characteristics of a magnet steel are to a first approximation determined by the magnetic saturation and the coercive force. The various factors which influence these two magnitudes will be discussed individually below.

---

[1]) For further details of the many points raised in this article, reference should be made to the standard work of W. S. Messkin and A. Kussmann: Die ferromagnetischen Legierungen, (Berlin, 1932), which is cited as M.K. in the text.

[2]) In this respect, the magnetic steels are sharply differentiated from magnetically softer materials in which very considerable deviations may obtain. This marked difference may be accounted for theoretically by the very special character of the internal strains, as found for instance in a rolled nickel-iron strip (cf. Philips techn. Rev., 2, 77, 1937), and the fact that these strains cannot occur in magnet steels in view of their treatment during manufacture.

Magnetic Saturation

At the beginning of this century the nature of the components of a material responsible for ferromagnetism was still very obscure. Nevertheless there was some justification for the assumption that in a magnetic substance the atoms themselves could be regarded as small elementary magnets. Ordinarily these small magnets have a random orientation owing to thermal rotational motions; but when an external magnetic field is applied to the substance they take up a definite configuration and assume a direction coinciding with that of the applied field, so that the magnetised condition becomes macroscopically apparent.

This theory provides a very satisfactory explanation of the behaviour of socalled paramagnetic materials, but on the other hand, it could not at the outset account for the behaviour of ferromagnetics, which differ from paramagnetics in that they can be magnetised with considerably lower field intensities and that after the removal of the field a certain residual magnetism remains.

Pierre Weiss was the first to point out that ferromagnetic phenomena could be explained on the hypothesis that a very powerful mutual action existed between the elementary magnets which tended to bring the latter into a mutually parallel configuration. According to Weiss the existence of this powerful intra-coupling of the elementary magnets was the sole difference between ferromagnetics and paramagnetics. These intra-molecular forces are so powerful that the material is always in a condition of spontaneous saturation. To account for the fact that a ferromagnetic can nevertheless be absolutely non-magnetic, Weiss assumed that the material was divided up into socalled elementary areas, within each of which the direction of the (spontaneous) magnetisation was entirely independent of that in any of the adjoining areas.

We now know that the carriers of the elementary magnetism of an atom are the electrons which are predestined for this duty by the nature of their linkage in the atom. The application of the modern electronic theory, particularly by Heisenberg and Bloch, has confirmed the basic assumptions of Weiss's theory and has enabled this theory to be further developed.

A theoretical explanation has indeed been forthcoming for the intramolecular forces demanded by the Weiss theory, but it appears that these forces are effective over very short distances only. The energy required to produce a marked deviation from parallelism is hence so great that large deviations between neighbouring particles will occur only rarely. Where the particles are separated by great distances, the coupling is in fact so weak as to be ineffective.

A weak magnetic field will suffice to produce a uniformly parallel configuration of the particles throughout a particular area, but thermal rotational motion remains a disturbing factor and its effect will be the greater the weaker the intramolecular forces [3]).

It is, in fact, possible to interpret quantitatively the variation of saturation with different series of alloys in terms of their composition by taking the above theory as a basis.

In general, the magnetic saturation of a series of solid solutions of two ferromagnetic elements varies, to a first approximation, linearly with the concentration. Where there is a marked deviation from this rule (cf. M.K., p. 125), these deviations may be explained by the forces operating between the elementary magnets being so weak at a certain concentration that thermal rotational motion assumes the upper hand. This is indicated by the marked depression of the Curie point.

Also in other cases, where heat treatment is found to have a marked effect on the magnetic saturation (nickel-manganese, cf. M.K. p. 134), the observed behaviour may be accounted for by deviations in the intramolecular forces as a result of a modification in the arrangement of the atoms.

On adding a non-magnetic material to an alloy the mutual forces are as a rule very rapidly reduced in magnitude.

In accord with these observations, the search for materials with a high saturation has been directed to the production of solid solutions composed of three ferromagnetic elements, viz., iron, cobalt and nickel, with small additions of other elements.

All alloys suitable for commercial uses are largely composed of these three elements, although very high coercive forces are also found with alloys containing little or even no iron alloys (e.g., the Heusler alloys, cf. M.K. pp. 146 - 146), but the remanence is so small that these alloys are of little practical use.

For the same reason, all other possible iron compounds, mainly mixtures of oxides isomorphous with magnetite (Fe_3O_4), are also unsuitable and valueless; these will therefore not be discussed in this article.

3) The intensity of this mutual action can be determined by measuring the critical temperature (the so-called Curie point), above which ferromagnetism disappears, giving place to ordinary paramagnetism.

Coercive Force

To discuss the question of the coercive force of permanent magnets, the theory of magnetism must be considered in greater detail.

In the past an attempt was made to explain magnetism by assuming that the electrons circulated round the nucleus with a high velocity in certain plane orbits. These "circulating currents" created a magnetic field according to known laws. The main argument in favour of this hypothesis was that no other cause was conceivalble at the time. But from spectral investigations it appeared that each electron had its own magnetic moment, which gave rise to the conception that the electricity within the electron itself was circulating about a certain axis with a high velocity. From that time onward two possible causes for paramagnetism and ferromagnetism have thus been recognised.

Further investigations revealed that in nearly all cases the latter cause is alone responsible for the magnetic phenomena observed. The circulating currents compensate each other because just as many electrons spin in one direction as in the opposite direction. To a certain extent the same also applies to the magnetic moment of the electron itself, where again the axes of rotation of the different electrons are in the main opposed to one another. A small number of electrons (one or two per atom) is however not subject to this limitation in motion and it is just these electrons with their magnetic moments whose properties we have to consider.

The axial directions of the magnetic moments are on the whole easily rotatable, but in many cases this degree of freedom is restricted by a variety of factors.

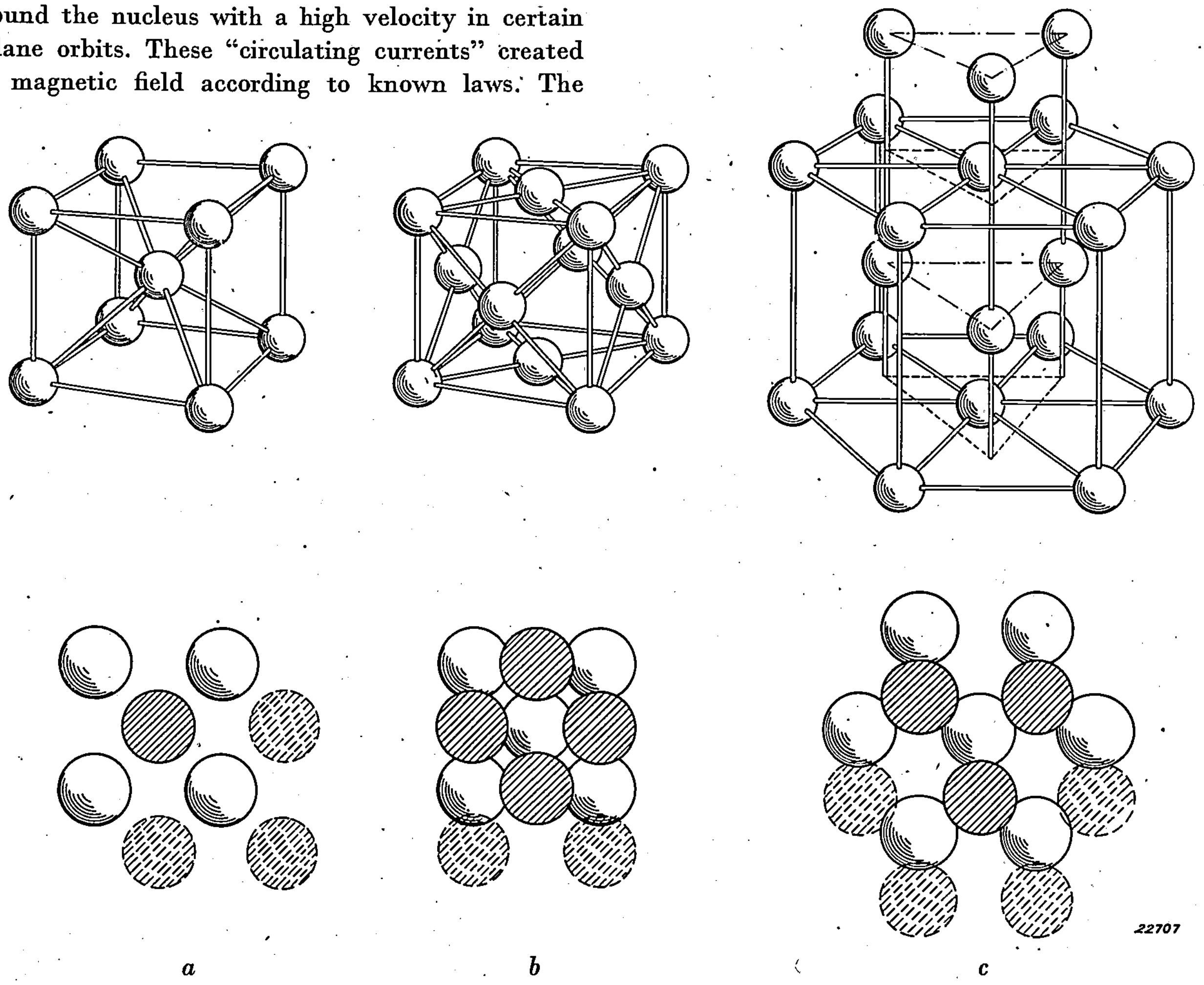


Fig. 2. Internal structure of the atom.
a) In a cubic space-centred lattice (alpha structure).
b) In a cubic body-centred lattice (gamma structure).
c) In a hexagonal lattice.
The diagrammatic plans below the structural models show the two lowest atomic layers (the second is hatched). The third layer then coincides with the first, the fourth with the second, and so on. The second atomic layer is shown shaded in order to indicate that it coincides with the first as regards the mutual arrangement of the atoms.

We have already stated that the small magnets composing a ferromagnetic substance are subject to mutual forces which tend to move neighbouring magnets parallel to each other. The common axis will then still remain freely rotatable, which is obviously not the case with a material having a coercive force, so that additionel forces have to be assumed to obtain a restricted rotation. These forces are provided by a mutual action between the electrons and the ionic lattice, which so to speak forms the framework on which the material is built up. As a result of this mutual action there are preferential directions in the crystal for the common axis of the specific moments to occupy, and a certain amount of energy, the socalled crystal energy, is necessary to rotate the magnetisation from a preferential direction into a less favourable one.

These forces cannot be calculated in advance, but it is evident from the above that their nature is closely dependent on the form of the ionic lattice. The three chief forms of lattice found in ferromagnetic metals are shown in *fig. 2*. With a lattice structure of pronounced asymetry, as shown in fig. 2*c*, the inherent moments exhibit a definite preferential direction. To serve as an example *fig. 3* shows the magnetisation curve of hexagonal

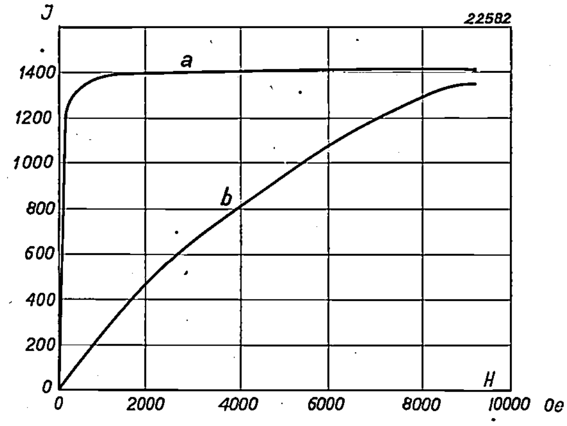


Fig. 3. Magnetisation curve of cobalt: a) in the direction of the hexagonal axis, b) in a direction perpendicular thereto. To obtain magnetic saturation in the second case a much higher field intensity is required than in the first case.

cobalt, curve *a* applying to magnetisation in the direction of the hexagonal axis and curve *b* to magnetisation in a direction perpendicular thereto. It is seen that the inherent moments tend to become parallel to the hexagonal axis, since a very high intensity is necessary to obtain magnetic saturation perpendicular to this axis. In cases where the symmetry is of a higher order, e.g. in cubic crystals, such as nickel and iron, the coupling is much looser; this follows from the fact that the three principal axes are equivalent from reasons of symmetry.

Now the majority of ferromagnetic alloys have a cubic symmetry, and in order to augment the linkage to the ionic lattice it is first necessary to reduce the high degree of symmetry obtaining. This can be done quite easily by applying a unilateral compressive or tensile stress to the material, when firmer bonds are produced almost exclusively, the existence of these bonds being exhibited by the fact that magnetisation is entirely different in the principal direction of deformation than perpendicular thereto.

Some materials, such as nickel, behave in the same way as cobalt when pressure is applied to them, viz., the inherent moments are located in preference parallel with the principal axis, while in others, such as iron, these moments are mainly perpendicular to the principal axis. When a tensile force is substituted for a compressive stress, the reverse behaviour is observed. It is thus possible to modify the positions of the elementary magnets relative to the ionic lattice and also their bonding forces by applying external forces.

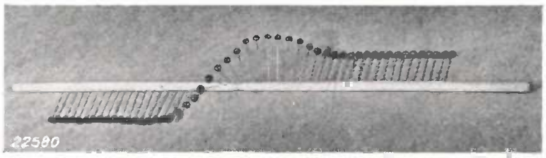
Unfortunately, the existence of preferential directions of magnetisation is in itself insufficient for obtaining a high coercive force. The hysteresis loop of, for instance, a single crystal of cobalt is so small that it cannot be drawn in fig. 3. The same behaviour is observed with nickel under compression and iron under tension.

This failure is probably associated with the fact that a material, which is magnetised in a certain direction, always contains nuclei which are arranged in opposition to this direction. We will consider the effects of these nuclei using the model shown in *fig. 4*. Assume that the material has been magnetised principally in one direction *A*, as shown on the left of fig. 4, which at the same time is a preferential direction in the crystal. The part *B* shown on the right of the figure is a nucleus which has been magnetised in the opposite direction, this magnetisation being just as favourable as *A* in regard to the crystal energy. Between these two parts there will be a gradual transition, as indicated in *fig. 4*.

We thus see that in spite of the crystal energy no work is necessary to alter the state of magnetisation in the model shown in fig. 4; to effect this change it is only necessary to allow the nucleus *B* to grow at the expense of *A*. This will change the magnetic moments to the left of the transition zone from a favourable to an unfavourable configuration, while to the right of this zone exactly the same number of moments will be transformed from an unfavourable to a favourable arrangement.

In the case of a crystal lattice, the corresponding conditions are rather more complicated than those shown in the model in fig. 4, and the coercive force will depend on the shape of the nuclei present. But

B



A

Fig. 4. This model shows how the magnetisation gradually changes between two areas with the elementary magnets in opposite configurations. By displacing the transition layer, magnetisation *A* (left) can be readily converted to magnetisation *B* (right).

again here the field intensity required to induce the growth of the nuclei will be smaller than is necessary in the absence of these nuclei for reversing the direction of magnetisation.

Hence, to produce a high coercive force, not only must magnetically-preferential directions be produced but the growth of nuclei must also be counteracted. Experiments have shown that this may result from marked changes in the direction and magnitude of the local strains.

These non-uniform strains always occur in commercially-produced materials owing to the presence of "disturbing" (i.e. insoluble) impurities, which also account for the fact that very low coercive forces were obtained in the case of iron only when this element could be prepared with a high degree of purity. On the other hand, soluble admixtures forming solid solutions with the iron have no effect on the coercive force, since very low coercive forces are found in the nickel-iron group of alloys. The following practical solution is thus arrived at for creating maximum strain abnormalities in iron by the intentional introduction of disturbing impurities: As much as possible of the foreign element must first be dissolved in the molten metal, so that by solution a uniform distribution of this element is obtained throughout the metal. During cooling segregation of the disturbing impurities occurs and the distribution of strain necessary in the alloy to obtain the requisite coercive force can be arrived at by regulating the conditions of cooling.

To illustrate how sharply the coercive force reacts to segregation, the results obtained in an investigation of rapidly-cooled mixtures of ferrosoferric oxide (Fe_3O_4) with mangano-manganic oxide (Mn_3O_4) are reproduced in *fig. 5*. It is seen that at a certain concentration limit the coercive force rises rapidly; at this point segregation evidently takes place during cooling.

Survey of Magnet Steels: The Oxide Magnet.

The oldest known permanent magnet, magnetic iron ore or loadstone, which is widely distributed

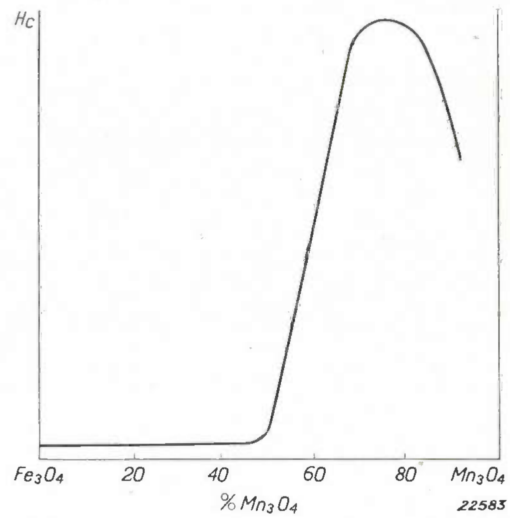


Fig. 5. Variation of the coercive force in rapidly- cooled mixtures of Fe_3O_4 and Mn_3O_4. At high temperatures, the formation of a solid solution is complete, but during cooling a slight segregation takes place which is distinctly revealed by the change in coercive force. Mn_3O_4 itself is non-magnetic.

in Nature, is the classic example of an oxide magnet of the type just referred to. The coercive force and particularly the remanence of this magnet are very unsatisfactory judged by present standards. During recent years the oxide magnet has been the subject of close research, and higher coercive forces have been obtained with it, although its low retentivity remains a serious disadvantage which has not yet been overcome.

Martensitic Steels

The first artificially-prepared permanent magnets used commercially were obtained by hardening steel. In addition to plain carbon steels, a distinction is drawn between chromium steels, tungsten steels and cobalt steels on the basis of the proportion of the predominating alloying element. A summary of the characteristics of these magnet steels, taking in each case the most reliable values, is given in the table at the end of this article. A marked preference is shown for cobalt steel, although the high price of cobalt prohibits its general use.

The hardening of steel does not exactly follow along the lines of the simple process described above. Yet the higher solubility of carbon, as the element responsible for the hardening of steel, at high temperatures than at low ones is used, although this increase in solubility is mainly due to the

[4]) J. L. Snoek, Physica, 3, 463, 1936.

iron having a crystal structure (austenitic or gamma structure) different at the high temperatures in question from that at low temperatures, when the structure is ferritic or of the alpha type (see fig. 2). With excess carbon and sufficiently rapid cooling, a transitional structure, martensite, is obtained which possesses a high mechanical and magnetic hardness. In modern steels, the use of carbon as an alloying element has been superseded, so that martensite is no longer an important component of magnet steels.

Non-Martensitic Steels

These steels may be classified in two groups. In the first group a metal, such as titanium, tungsten or molybdenum, is precipitated from the matrix as a compound, the best results being obtained with molybdenum dissolved in an iron-cobalt alloy. The description of the hardening process given above applies *in toto* to this type of alloy, and searching experiments have shown that here the increase in the coercive force takes place simultaneously with segregation, so that this segregation is fundamentally responsible for hardening.

In the second and technically more important group, the mechanism of the process follows along somewhat different lines. During segregation a compound is not directly precipitated from the matrix, which here has an alpha-structure, but an equilibrium is set up between the alpha phase referred to above and the gamma phase produced by cooling [5]). Owing to the marked difference in composition between these two phases, a separation of the homogeneous component in two phases is associated with an exchange of material between those regions in which the alpha phase persists and those where the gamma phase is produced. Segregation is therefore preceded by diffusion, during which a marked change in composition occurs locally in the metal before the formation of the gamma phase is initiated. This preparatory reaction, which is thus closely associated with the presence of a supersaturated solution, is the true cause responsible for magnetic hardening [6]).

It is probably not fortuitous that the best results have been obtained with just these steels, which include the Ni-Al and Ni-Al-Co-Ti alloys. As already indicated above, the maximum attainable deformation of the metal should be produced; during this transformation the limit of elasticity of the material is usually reached. If a crude process is employed resulting in the appearance of discontinuous transition phases, the mutual bonds between the particles will be disturbed so that the original structure will be partially reconstituted. In fact, it has been observed during transformation from the alpha to the gamma phase that in a certain transition state the metal is mechanically very weak. This is in complete agreement with the observation that in the case of Fe-Ni-Al steel this transformation is not initiated at the same time as the coercive force increases, but that on the contrary the transformation is a sign of an almost abrupt decrease which is probably connected with a partial "slipping" of the material. On the other hand, during diffusion preceding slipping the strains are to a certain extent built up step by step, so that cohesion between the atoms is never for a moment lost.

The table below correlates the magnetic properties of tbe principal magnet steels.

Table I.

Material	Remanence in gauss	Coercivity in oersted
Carbon steel (plain)	9 000 - 8 000	50 - 60
Tungsten steel	12 000 - 11 000	56 - 67
Chromium steel	10 800 - 9 700	58 - 66
Chrome-tungsten steel	11 000	76
Cobaltsteel	9 000	300
Ni-Al steel	6 000 - 5 000	400 - 600
Ni-Al-Co-Ti steel	8 000 - 6 000	600 - 900

[5]) W. G. Burgers and J. L. Snoek, Physica, 2, 1063, 1935.

[6]) The same has been found for duralumin which was hardened in the same way.

TECHNICAL CONSIDERATIONS IN THE LIGHTING OF COUNTRY ROADS

by G. B. VAN DE WERFHORST.

Summary. In this article the standards underlying the assessment of lighting conditions on highways are discussed. Originally the intensity of illumination measured horizontally was taken as a criterion for this purpose; later the brightness of the road surface was favoured, while at the present time the contrast between the object and the background is taken as a standard of comparison. The coefficients of reflection of the object and the background, the comparative values of specular or regular reflection of the road surface, the screening of the emitted beam which is necessary to avoid dazzle, as well as the effect of the area of the radiating surface of lamps, are also discussed.

In this article road or highway lighting embraces the illumination of highways and roads situated outside built-up areas. The principles which underlie the planning of a new lighting installation and which have to be considered in assessing the efficiency of existing installations have already been discussed in a previous article [1]). The subject matter of that article may be briefly summarised as follows: Fixed lights which are provided in the public interest on all class I country roads and on certain class II roads must afford such conditions of visibility that the traffic on these roads can circulate with all reasonable safety.

A speed of 50 m.p.h. should be reasonable and permissible under the conditions of illumination provided. Aesthetic considerations regarding the colour of the light do not come into question; disturbing effects originating from light sources outside the road boundaries must be eliminated.

The driver of a car must be able to perceive immediately the presence of any obstacle situated at a distance exceeding 300 yards. Moreover within this distance, it must also be possible for him to recognise the exact nature of the object. At the same time objects within this distance and approaching from a lateral direction must produce such a peripheral stimulus that the driver is able to react directly to this stimulus. The speed with which an object can be perceived and recognised is thus of paramount importance in the present case. Supplementary lighting of the edges of the road and of strips skirting it is essential for satisfactory perception.

Far removed from the application of the above basic principles, it was believed a few years ago that the lighting of a highway could be assessed exclusively from the amount of light incident on the road surface in a horizontal direction (the horizontal intensity of illumination in lux). It is evident that this gave a one-sided and erronous scale of measurement, since we are never able to see directly the light falling on a surface or on a solid object, but only that part of the incident light which is reflected by the surface in the direction of vision. We see the brightness of the illuminated surface in the direction of vision and not the actual intensity of illumination. In recent years, closer attention has therefore been given to the brightness of road surfaces and various attempts were made to increase it. Cohu [2]) has investigated the reflection properties of road surfaces in order to determine from the intensity of illumination the brightness apparent to the road user. Paterson, Waldram [3]) and others have studied means for increasing the brightness of road surfaces and methods for measuring the brightness values. Yet the conditions of perception required to make traffic safer are not determined merely by the brightness of a highway, for perception demands the existence of a contrast, and in fact a contrast between the object viewed and its background. In the present article, the contrast is defined as the ratio of two brightnesses, of which one, e.g. that of the road surface, is greater than the other, viz., that of the object. In regard to our perceptive abilities, this ratio is not of the nature of a constant, since the contrast sensitivity of the eye is not the same at all brightness levels. The contrast sensitivity becomes reduced with diminution in brightness, and at low brightness levels only comparatively heavy contrasts can be differentiated [4]). Since from considerations of cost, it is necessary to make do with low intensities of illumination for lighting country roads, we are compelled to take into account the lower contrast

[1]) Philips techn. Rev., 2, 110, 1937.

[2]) M. Cohu. Revue génerale de l'électricité, 37, 755 - 767, 1935.

[3]) C. C. Paterson, Modern Street Lighting, J. Record Trans. 45, part 4, 1935, J. M. Waldram, Road Surface Reflection, Ill. Eng., 27, 305 - 313 and 339 - 351, 1934; 28, 34, 1935.

[4]) Cf. also P. J. Bouma, Philips techn. Rev., 1, 166, 1936

sensitivity of the eye at these brightness levels. We must therefore make the fullest use of all properties of the highway and of the objects on it in so far as they affect the conditions of reflection, so as to arrive at the maximum possible effect [5]). The already low contrast sensitivity of the eye must also be prevented from becoming further reduced by the presence of disturbing influences, such as dazzling light sources.

Owing to the diversity of conditions arising, it is only possible to discuss in broad outline the reflection characteristics of a highway, including the kerbs and pathways skirting it, and those of the objects on the roadway.

It should be noted at the outset of this discussion, that an illuminated surface or object never reflects all the light incident on it; a greater or lesser fraction of this light is always absorbed, while the reflected component can be reflected in a variety of ways. To understand exactly what occurs, we must examine in closer detail the nature of reflection.

A distinction is drawn between diffuse and regular or specular reflection. A surface with diffuse reflection reflects the incident light uniformly in all directions irrespective of the angle of incidence of the light, so that the surface appears equally bright when viewed from any direction, e.g. a matt stucco ceiling, a sheet of matt white drawing paper or of grey drawing paper, which however reflects a much smaller part of the incident light. The apparent brightness of a surface with diffuse reflection depends, *inter alia*, on the quantity of light incident on it and hence on the distance between the light source and the illuminated surface; the brightness of the surface is always much lower than that of the light source.

On the other hand, a surface with regular-reflecting properties reflects the incident light in one direction only as determined by the direction of incidence of this light. On looking at the surface in a direction coinciding with that of the reflected light rays, the image of the light source is seen. Since with a "mirror" of this type the absorption loss is small, the brightness of the image is roughly equal to that of the light source itself, and in fact quite independent of the distance between the light source and the illuminated surface. But if the surface is regarded in a direction other than in that in which the light is reflected, it will appear dark, irrespective of the amount of light falling on it. Examples of this ordinary mirrors and smooth surfaces of water.

Perfect diffuse and perfect regular reflection are extreme concepts, which are obviously opposites in almost every particular. The mistake is frequently made in practice of taking one or other of these extremes as the only standard; objects and surfaces which conform wholly or almost wholly to one or other of these extremes of reflection are exceptions. Very frequently diffuse and regular reflection occur simultaneously, and often one or other predominates. Thus for instance ordinary glazed paper has mainly diffuse reflection, although the smooth surface also produces a secondary regular reflection. A mirror with a thin film of dust exhibits principally regular reflection, although the dust particles on the surface also give diffuse reflection.

In such cases the reflecting surface may be roughly regarded as homogeneous in so far as reflection is concerned. On the other hand, there are many practical cases which are not homogeneous of which some give diffuse reflection, others regular reflection and still others both types of reflection. This heterogeneity of reflection is found, for instance, with a host of small mica chips and in fact in all kinds of crystalline formations in natural rocks. Every stone pavement and many tarmacadam road surfaces exhibit this combination of reflections. A section through a road surface of this type is shown in *fig. 1*, where the surfaces A_d and B_s

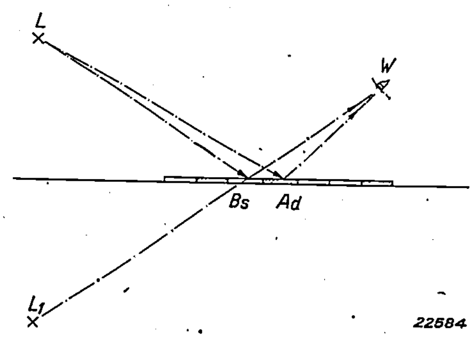


Fig. 1. Road surface with areas of regular reflection (B_s) and of diffuse reflection (A_d).

reflect the light rays emanating from a light source L by diffuse and regular reflection respectively. The observer W views the surface A_d with a certain brightness at a distance WA_d, while in surface B_s he sees the light source L specularly reflected at L_1 and at a distance WL_1. Hence, the observer when looking at a point on the road surface must accommodate his eye to two very different distances, which is naturally very difficult. Without being conscious of this fact, the observer may frequently

[5]) Where contrast is referred to in this article, only one type of contrast is implied, as for instance that obtained by illumination with socalled "white light" (light sources with a continuous spectrum: sun, gaslight, incandescent lamps, etc.).

find it difficult to judge the distance of a certain point under these conditions, since he cannot see the exact position of the point distinctly [6]).

Not only must the regular-reflecting parts of the road surface be taken into account in this connection, but it has to be remembered that many areas giving diffuse reflection when dry become converted to areas of regular reflection in wet weather. A road surface which in dry weather is almost wholly diffuse-reflecting may exhibit a combination of both types of reflection when wet, e.g. brick paving. In fact in wet weather regular reflection may predominate.

In the cases discussed here, it has always been assumed that a given reflection can be resolved into two components, of which one may be regarded as regular reflection and the other as diffuse reflection. In very many cases, this resolution is however not possible, and reflection is more of an intermediate nature which cannot be resolved as a combination of diffuse and regular reflection, but requires for description a notation of its own which is still lacking. Hence, in analysing road lighting conditions, we are compelled to make do with the terms "diffuse reflection" and "regular reflection", and later when discussing the conclusions arrived at we can speculate on the possible significance attaching to the occurence of an intermediate form of reflection of the type referred to.

After this digression we can return to a consideration of the road and of the objects present on it. The objects on a highway may be divided into two main groups: Individuals and vehicles. Individuals include pedestrians, cyclists and motor-cyclists. With all individuals, including motor-cyclists, the reflection properties of their clothes are the determining characteristics, and by a sheer accident introduce no difficulty in dealing with this problem. We can in fact start from the assumption that **clothing is always diffuse-reflecting**. Bouma carried out measurements of the different coefficients of reflection and examined the frequency of their occurrence. His results are correlated in the form of curves in *fig. 2*. It is seen that in most cases the coefficient of reflection of overcoats is below 1 per cent and of other garments below 2 per cent, and that coefficients of 6 per cent occur only rarely. If we therefore take a mean coefficient of 5 per cent for individuals, we are erring on the right side.

In the second group comprising vehicles, practically only power-driven vehicles, some with trailers, have to be considered in the case of country roads under discussion here. A recent census showed that 80 per cent of these vehicles are dark in colour

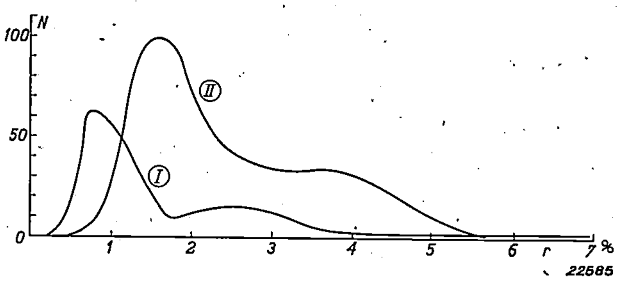


Fig. 2. Coefficients of reflection of clothing. Curve *I*: Overcoats, and curve *II*: Other outer garments. The coefficient of reflection r is plotted in percentages along the abscissa, and the frequency with which a garment of a given coefficient of reflection occurs relative to another is plotted along the ordinate.

2 per cent very bright, while the colour of the remaining 18 per cent conforms roughly to the average hue of the road surface. To investigate how sharply these vehicles stand out against the background, we will neglect the regular reflection of the vehicles, since it is only important when the light of a street lamp incident on the vehicle is reflected directly in the direction of the eye of a person on the road; it is obvious that this will only result in an accidental flash of light being received. The liability of regular reflected light rays of this type falling on the eye of a person on the road may be augmented by the three following circumstances:

a) Most private motor cars have curved surfaces of a variety of shapes: The different areas on a vehicle of this type will therefore reflect the incident rays in many directions. On the other hand, most commercial vehicles have few specularly-reflecting surfaces and parts, while their contours are also much flatter. The liability of fortuitous flashes in the direction of another road user is hence very small with these vehicles.
b) As a rule the person on the road is himself moving and is continually changing his position with reference to the vehicle seen under the light of a lamp; if the vehicle gives any regular reflection a reflected light ray will then strike his eye for a brief period in each direction of regular reflection.
c) The vehicle seen under the light of a lamp is itself moving. In this case, in conjunction with the conditions referred to under b), the danger of a road user receiving an accidental flash will be much increased. If, on the other hand, the vehicle is stationary, this liability is limited to the two cases a) and b), which must obviously

[6]) As long as his field of vision contains a sufficient number of well-defined and recognisable characteristics, the observer on the road will hardly be aware of this lack of visibility, particularly as he does not focus the eye on to the road surface. But if the conditions on the road surface closely approach those under consideration here, which may occur on a wet road during the evening, the surface will appear blurred to the road user and will not exhibit sufficient characteristics.

not be over-estimated in the case of non-specularly-reflecting vehicles, such as the majority of commercial vehicles. For this reason stationary vehicles are very dangerous. If the background and the object appear to the road user to have the same brightness and no difference in colour (which is in fact more frequently the case with existing road-lighting systems than generally assumed, the road user will barely be able to perceive the object owing to the absence of a contrast. Moreover, the motion of an object contributes considerably to visibility.

However the problem is regarded, the feasibility of perception, which in fact demands something more than that light rays shall accidentally impinge on the eye for a very brief period, remains an uncertain factor which is incapable of calculation. Regular reflection of the vehicles must therefore not be taken as a basis in planning a lighting installation, and it should be regarded merely as a contribution to the general effect produced by the lighting. In regard to the diffuse reflection of vehicles, a coefficient of reflection not exceeding 5 per cent applies for vehicles with a dark finish, and one of 50 per cent for the very small group of vehicles with a very bright finish. For the intermediate group, which as regards brightness and colour is comparable to an average road surface, the accidental flashing of reflected light rays already referred to must provide us with a means of safety.

Considering vehicles and individuals in a single group, a coefficient of reflection of a maximum of 5 per cent can be taken as a reasonably accurate basis.

The reflection of the road surface is less uniform: Stone pavements (many types of stone), brick paving, asphalt surfaces (in great variety) and concrete surfacings exhibit considerable differences. Moreover, many road surfaces have an intermediate form of reflection as referred to above. It is, therefore, practically impossible to ascribe specific coefficients of reflection to a road surface, which can be taken as a reliable basis for calculations.

In contradistinction to these indefinite factors, we have however the following well-defined criteria:

Assume that we are standing during the day on a long straight country road, whose right-hand traffic lane has a surface of grey asphalt, while the left hand lane (a subsequent widening of the original road) has a greyish-yellow asphalt surface (rolled gravel); along one side the road is skirted by a pavement of Swedish granite and along the other side by brick pavement.

On looking at the different road surfaces, a definite brightness stimulus is obtained from each half of the road. The greyish-yellow asphalt appears brighter than the grey, and the Swedish granite pavement brighter than the greyish-yellow asphalt. A marked difference is moreover apparent between the brightnesses of the Swedish granite and brick pavements. Each part of the road surface gives its own general brightness stimulus and these stimuli merge to a uniform brightness only at a point miles along the road. The brightness stimuli from the different road surfaces are so characteristic, both when viewed against each other and individually, that it must be found possible to express these differences in terms of specific coefficients of reflection (cf. fig. 3).

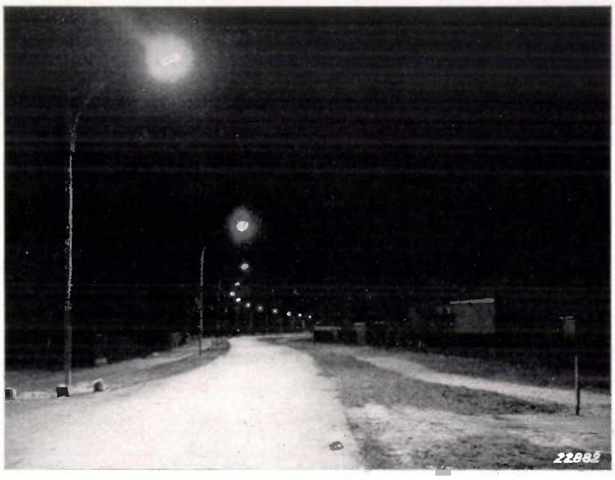


Fig. 3. Example showing the importance of illuminating the areas skirting country roads in addition to the road itself, so as to obtain a sufficiently extensive background. It may be seen how each of the traffic lanes produces its own particular brightness stimulus.

We might be inclined to regard these pronounced differences to be due to colour stimuli. But if we look at the same road under monochromatic light from sodium lamps the same stimuli are experienced, so that the observed differences must be considered as due to differences in brightness.

If, now, we take a series of samples of diffuse-reflecting paper with different tones of grey, whose coefficients of reflection have been determined by careful calibration and which differ by only 2 per cent from each other, and place these samples on the different parts of the road referred to above, we can determine by direct observation which sample on each part of the road has the same brightness as the road surface itself, and cannot then be readily distinguished. It is also possible to determine which samples are brighter and which darker than the surface; this can be done so accurately that the coefficient of reflection can be precisely determined to within 1 per cent.

We can thus express the above-mentioned pronounced total brightness stimuli in terms of a moment consider only diffuse reflection (regarding regular reflection, see below), it is evident that under an equivalent illumination an object with a coefficient of reflection below 5 per cent will appear

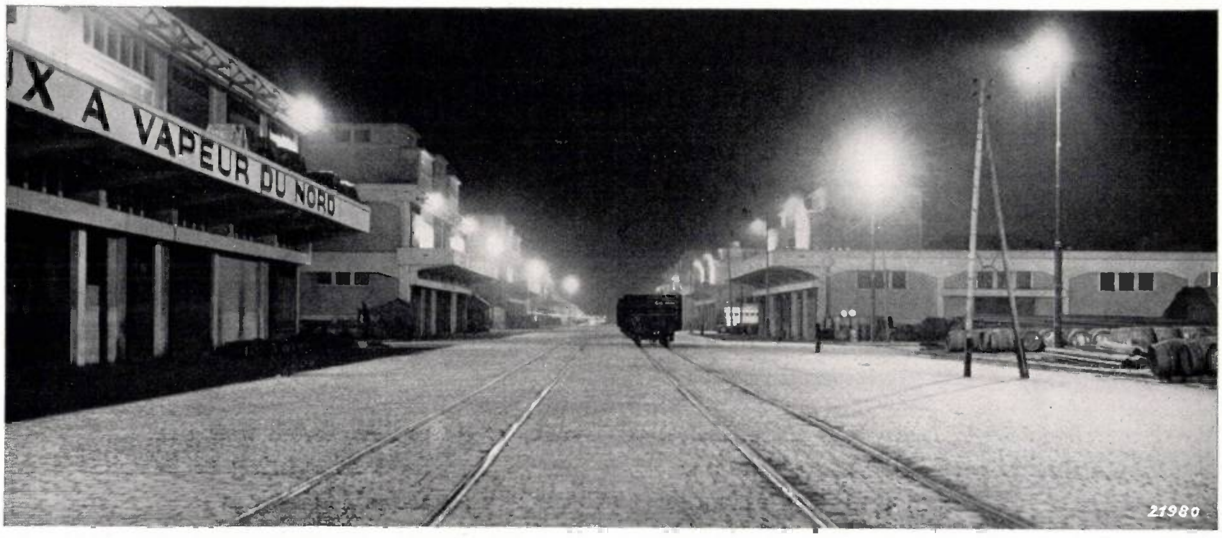


Fig. 4. The object on the roadway stands out dark against the bright background.

Fig. 5. In daylight also, the object appears dark against the background, except the ladies with very bright summer dresses cycling along the side track.

coefficient of diffuse reflection of these samples. In this way we can deduce that the coefficients of reflection of the different parts of the road lie between 5 and 35 per cent, except for isolated areas of (black) tarred asphalt. If, furthermore, we assume a value of 10 per cent for the average road surface, pathways and skirting fields, and for the dark against the brighter background. This result, which may be frequently observed in daylight (cf. *figs. 4* and 5), must therefore be utilised to the full when artificial illumination is provided [7]). To

[7]) It may be noted in this connection how few colour contrasts, even during the day, play a part in the perception of objects on the road.

the road user, the road and its boundaries appear as horizontal surfaces and the object mainly as a projection on a vertical plane. To obtain the greatest contrast it would, therefore, be ideal to make the bright horizontal surface as bright as possible and in particular leave the dark, vertical surfaces as dark as possible; this would require a system of illumination which throws the whole of its light vertically downward. An ideal system of this character cannot be realised in practice, and we are compelled to adopt a compromise in the form of a number of separate lighting units located at definite distances apart and mounted at a certain height, and at the same time cut out all non-vertical rays. Various problems have then to be considered, such as the best type of lighting unit, the optimum distances apart and the most suitable mounting heights, as well as the most favourable distribution of light from the lamps. If both the object and the road surface may be assumed to give only diffuse reflection, the optimum characteristics could be determined to a fair degree of accuracy by taking as a basis the coefficients referred to above. The brightnesses are then proportional to the illumination intensities and could be deduced directly from the latter. While diffuse reflection alone comes into consideration with objects on the road, regular reflection also must be taken into account when analysing the conditions of illumination on the road surface. As already stated above, different road surfaces have very different regular-reflection values which moreover vary with the angle with which the road user views the road surface and with the angle of incidence of the light [8]). Since, as already pointed out, the light from the lamps is not radiated exclusively in a vertical direction and all possible angles of incidence below 90 deg. have therefore to be taken into account, calculation is rendered very difficult by these highly variable factors which are moreover by no means easy to assess quantitatively. Laboratory measurements with small samples of road surface materials may give very interesting data for the theoretical assessment of the brightness stimuli, but the surface of an open road will be found to give entirely different brightness values. The reason for this is that, in the first place, our perceptions produce a total stimulus which is by no means equal to the sum of the component stimuli as determined in laboratory tests. Moreover, in practice the brightness stimulus is affected by the colour of the light and the more or less marked dazzle due to roadside and vehicle lamps. If all these influences could be taken into consideration when planning a lighting scheme, it would still be impossible to assess what effect each of these factors has on the visibility on the highway in the case of an existing installation. A most important advance in dealing with this problem has therefore been made with the appearance of the Philips visibility meter which was evolved by Holst and Bouma and provides means for the direct measurement of the visibility on the road itself together with all associated effects [9]). From a large number of measurements made with this visibility meter on various highways a minimum visibility value for the contrast on the road has been found which must be satisfied by the conditions of lighting to permit traffic to travel with reasonable safety at a speed of 50 m.p.h. It has been shown that to achieve this end the object and background must have an average minimum contrast of 3 : 4 [10]).

It has been frequently recommended that the best way to improve the contrast is to make use of the regular reflection of the road surface. Since the regular reflection of every surface increases the smaller the angle of incidence of the light rays, it has been assumed that this property could be used with advantage in road lighting and the light distributed with very low angles of incidence. Moreover it is claimed that this also has the advantage of permitting the lamps to be placed at greater distances apart and at lower mounting heights. The nature of this problem is illustrated in greater detail in *figs. 6* to *9*, where AB is a much magnified part of the road surface as seen by the road user W between the points A and B. This surface is illuminated by rays from the light source L impinging in the direction *1*. Since the surface AB is assumed to be very small, the incident rays may be regarded as forming a parallel beam. In all figures only that part of the light subject to regular reflection is shown. In *fig. 6* the incident rays make an angle α with the horizontal; part of these rays, viz., *1*, *2*, *3*, *4*, *5*, *6*, strike the surface at points where they are reflected in the direction of vision of the observer. On the other hand, rays *7*, *8* and *9* are, for example, reflected in such directions that they do not reach

[8]) The angle of incidence is here defined as the angle which the incident ray of light makes with the horizontal road surface, and not as the angle between the ray and the perpendicular drawn to this surface.

[9]) Cf. Philips techn. Rev., 1, 349, 1936.

[10]) Actually, we have here introduced a new basis for the assessment of the quality of a lighting system, viz., „the feasibility of travelling with safety at a speed of 50 m.p.h."

the observer and thus do not contribute to this brightness stimulus. Other rays again, e.g. *10* and *11*, although reflected in the direction of the observer are prevented from reaching him by inequalities of the at all is reflected in the direction of the observer, who thus sees AD as a dark patch.

In wet weather, the depressions in the road surface are filled with small puddles, as shown in

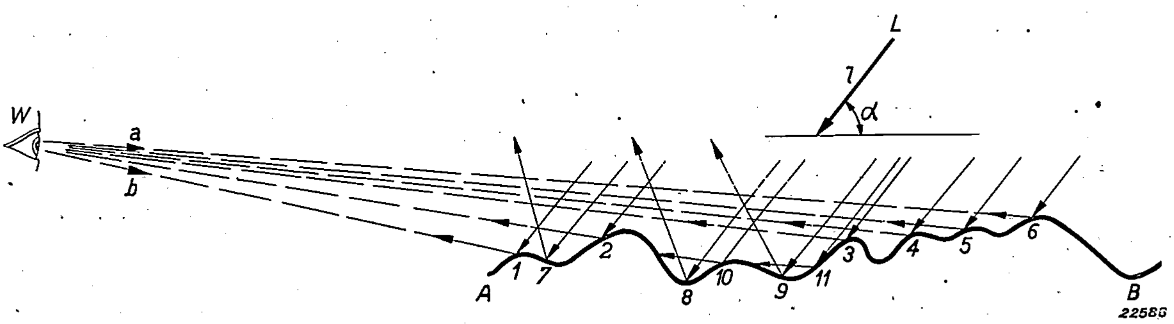


Fig. 6. Regular reflection at a highly-magnified area of a road illuminated at an angle α; part of the light is reflected towards the observer and part in an entirely different direction, while still another part is absorbed by inequalities on the surface.

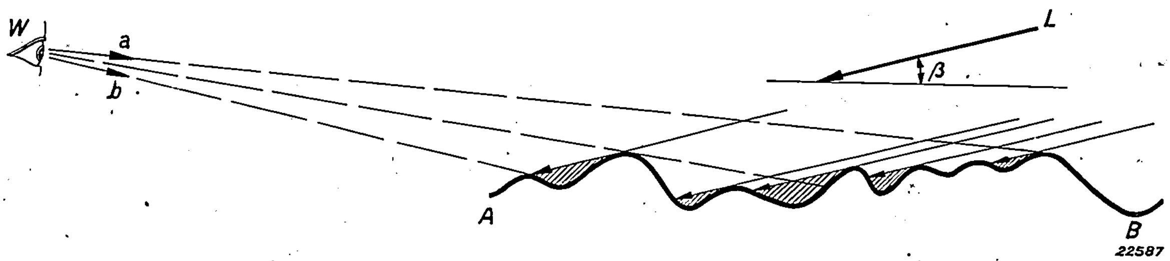


Fig. 7. Regular reflection at a highly-magnified area of a road surface illuminated at an angle β, with the inequalities of the surface casting shadows.

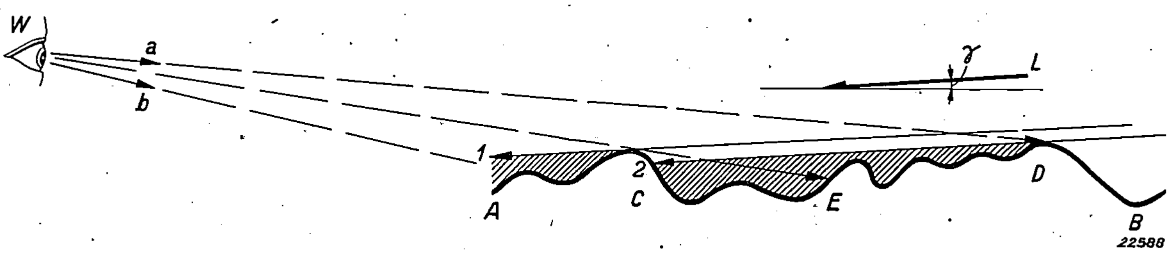


Fig. 8. Regular reflection at a highly-magnified area of a road surface illuminated at an angle γ. The surface inequalities cast long shadows; the areas in shadow also cannot give diffuse reflection.

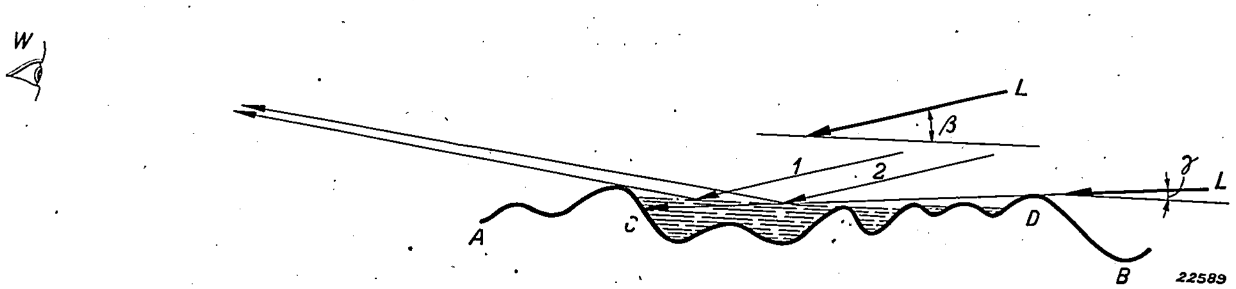


Fig. 9. Regular reflection at a highly-magnified area of a road surface in wet weather when illuminated at an angle β or γ; with an angle of incidence of γ the effect of the long shadows on the whole persists.

road surface. If a surface with similar contours is illuminated at an angle β as shown in *fig. 7*, the observer already sees fairly large areas of deep shadow on the road. If the angle of incidence is still further reduced, to γ as in *fig. 8*, the area from A to D remains in deep shadow, so that no light *fig. 9*. With an angle of incidence of β regular reflection may indeed be greater in the direction W than shown in fig. 7, although with an angle of incidence of γ the area CD remains in shadow and only the regular reflection occurring at the highset points of the surface is visible.

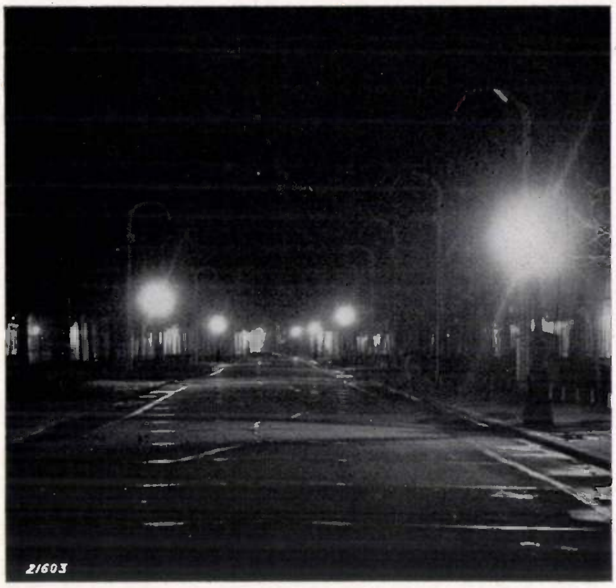


Fig. 10. With a very acute angle of incidence, only the highest points of the wet surface reflect the light specularly.

It must be remembered that even the smoothest road never has a mathematically smooth surface, particularly as it is frequently made rough to prevent skidding. Experiments have shown how with small angles of incidence inequalities of the order of a few millimetres appear to the observer to be depressions several centimetres deep. Owing to the presence of long shadows, there is just as little likelihood of diffuse reflection with small angles of incidence, since the areas *CD* even if giving perfect diffuse reflection receive no light at all and cannot therefore give a reflection. A depression of 9 mm will in fact give a shadow 10 cm in length with an angle of incidence of 5 deg. The moderate degree of reflection obtained with these angles of incidence is apparent if we consider the inequalities due to ruts in asphalt road surfaces carrying a heavy traffic. Only the peaks give regular reflection along the line joining the observer and the light source; beyond this line everything appears dark to the observer and narrow streaks of light are in consequence produced on the dark road surface (cf. *fig. 10*). In practice it is useless to attempt to increase the brightness of the road surface by using angles of incidence less than about 10 deg., in other words to use that part of the rays emanating from the light source L which make an angle β with the area AB (see *fig. 11*). There is, however, another reason which is a serious obstacle to this kind of illumination. Assume that in fig. 11 W_1, W_2, W_3, W_4 are points on the level of the eye of a road user moving in a forward direction. If the light source L emits rays in the whole of the region below the horizontal, light rays will fall on the observer's eye to the left of W_2 when these make angles of less than 10 deg. with his direction of vision assumed to be straight ahead horizontally, and to the left of W_0 when these angles are smaller than 15 deg. Nevertheless, it is found that dazzle, i.e. the reduction in the contrast sensitivity, is a maximum when the angle of incidence of the light rays on the eye is small [11]). Thus Bouma, among others, found that with white light impinging at $\beta = 2.5$ deg. and under conditions normally occurring on a roadway the contrast sensitivity is reduced to 43 per cent of the value obtained in the absence of dazzle, and at $\beta = 5$ deg. to 62 per cent [12]). Bouma and other investigators [13]) observed that the contrast sensitivity is reduced to the minimum permissible values only when the angle of incidence exceeds 15 deg. (Bouma considers 15 to 20 deg. as desirable). Direct rays from the light source L must not impinge on the eye before the observer reaches the point W_3, so that all light beyond this angle of $\beta = 15$ deg. must be screened off.

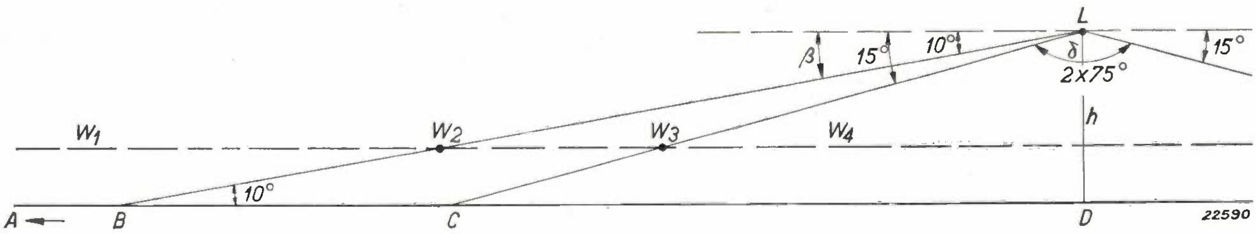


Fig. 11. Vertical section through a road, illuminated by the light source L, and with a moving observer W whose eye travels along the line W_1 W_2 W_3 W_4. The light radiated from L is screened to form a beam with an angle δ (2 × 75 deg.).

The permissible angle of radiation δ of the lamp must therefore be: $\delta = 2 \times 75$ deg.; for a given mounting height h of the lamp, this angle determines the range and hence also the interval a

[11]) The angle of incidence is here the angle which the light rays make with the direction of vision of the observer.

[12]) P. J. Bouma, De Ingenieur, **49 A**, 290 - 294, 1934.

[13]) Philips techn. Rev., **1**, 225, 1936, and the references given in that article.

between lamps; conversely with a given interval a between the lamps the mounting height h is fixed. It now only remains to determine the distribution of light from the lamps with particular light sources.

Within the angle of radiation of the lamps found above, regular reflection of the road surface, particularly in wet weather and with most surface materials, may become an important consideration. Taking this type of reflection as a basis, definite specifications for the design of the lamps and the characteristics of the light sources can be deduced. In *fig. 12* the conditions of illumination on a road

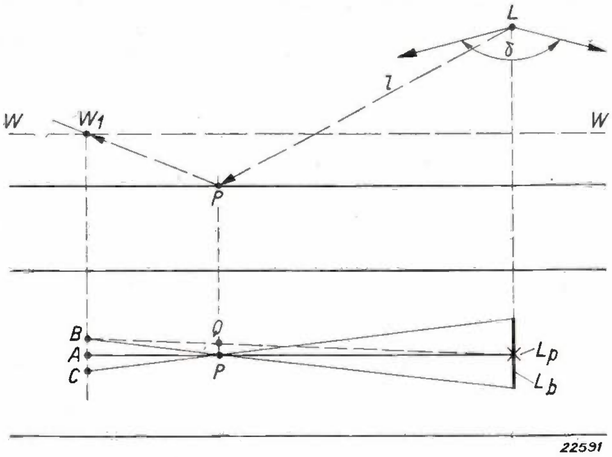


Fig. 12. Plan and elevation of a road illuminated from a light source L above it, showing the effect of regular reflection on the observer W, with a punctiform source L_p and with a transverse source L_b.

illuminated by the light source L and with the observer's eye at a level WW are shown in plan and elevation (for the sake of simplicity the beam of light and the direction of vision of the observer are assumed to be straight along the road). First, regard the light source L as punctiform L_p, e.g. an electric lamp with a clear bulb or a lamp with prismatic glass so that the lamp appears to radiate light in all directions as a point source. The observer at A sees the point P of the road surface illuminated with a brightness of the same order as that of the light source (regular reflection alone is being considered here). If the observer is at B, he will see the point Q similarly illuminated, while the point P will appear dark. If we now take a light source L_b giving the same luminous flux, but with a larger radiating surface in a direction transverse to the road, the observer will then see the point P on the road surface brightly illuminated not only from his position A but also to the side of A as far as B and C. It will always appear with a lower brightness than in the former case, since the light source L_b owing to its greater radiating surface will have a lower brightness than the light source L_p assuming the total luminosities are equivalent. By using a light source L_b placed perpendicular to the length of the road, regular reflection may be obtained also at inequalities of the road surface, embracing the above-mentioned areas reflecting in all directions, such reflection being apparent no

Fig. 13. Transverse light sources with large radiating surfaces produce wide luminous bands with a comparatively low brightness on a wet road surface.

longer as narrow bands of dazzling brightness, but as broad bands of lower brightness (see *fig. 13*), so that the troublesome dazzling reflection obtained with a punctiform light source L is avoided. Furthermore, by a judicious arrangement of light sources, these wide luminous bands may be brought contiguous to each other so that also in wet weather the road surface appears bright overall.

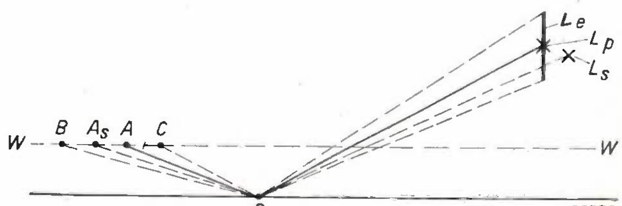


Fig. 14. Vertical plane to a road with an elevated light source L, showing the effect of regular reflection on an observer W with the eye at a level WW, with a punctiform light source L_p, with a vertically-expanded light source L_e and with a specular reflector $L_p + L_s$.

A similar effect is produced when the light source is expanded in a vertical direction. It may be seen from *fig. 14* that with a punctiform source of light L_p illuminating a horizontal specularly-reflecting part of the road surface, the observer sees the point P brightly illuminated only from A. But if the lamp is given a greater radiating surface in a vertical direction, e.g. equal to the apparent surface of an enamelled reflector, so that it acts as a light source L_e, the observer will see the point P brightly illuminated on moving from B to C, and with a lower brightness exactly as in the previous case. If the enamelled reflector is replaced by a specular reflector, the latter will give a mirror image for each point, e.g. L_s for the point P. The observer will then however see the point P illuminated not only at A, but also at A_s, although again with a much higher, usually dazzling, brightness; moreover the distance from which he sees P illuminated is much smaller than with the light source L_e. Owing to the unevenness of the road surface, which as already stated includes surfaces giving regular reflection in all directions the point by point reflection of a punctiform light source L_p and its mirror image L_s will also be frequently visible to the observer as an elongated illuminated band. This is also the case with a light source L_e elongated in the vertical plane, although the contiguous points of this bright band will now have a lower brightness in the direction of the observer. A lighting unit consisting of a combination of lamp and light source expanded both laterally and vertically will therefore distribute the perceptible reflection over larger surfaces on a road surface with many areas giving regular reflection, and at the same time reduce the otherwise dazzling brightness.

The sharply-defined and trying reflection of long bright bands of pronounced brightness obtained in wet weather, beside which the rest of the traffic lane appears a deep black, should therefore be avoided as far as possible when the light source and the lamp together form a large radiating surface. This desirable effect is, however, not obtained when a vertical extension of the radiating surface is attempted by means of a specular reflector, and which produces points of intense brightness in the direction of vision. It is evident that when using a specular reflector the distribution of light and the path of the rays with a reflector of this type must be extremely carefully calculated and executed, and that furthermore a reflector of optimum design requires the most careful mounting in order to exclude disturbing points of brightness in the direction of vision of the road user. On the other hand, when selecting a diffusing reflector to give a specified light distribution, the fact that the above advantages must be foregone will be found in practice to be of comparatively small importance.

THE ENLARGED PROJECTION OF TELEVISION PICTURES

By M. WOLF.

Summary. An arrangement is described for the reproduction of television pictures in which the image on the fluorescent screen of a small cathode ray tube is projected on a ground glass screen measuring up to 1 × 1.20 m.

Introduction

The reproduction of television pictures by means of standard cathode-ray tubes has already been described in several previous articles which have appeared in this Review [1]). The pictures reproduced in this way on the fluorescent screen are, however, so small that it is difficult for an audience of more than a very few persons to view them conveniently at the same time. To facilitate viewing, larger tubes have been devised with a screen diameter of 40 cm, but in arriving at dimensions of this order a variety of difficulties are met with in the construction of the tubes, which become almost insurmountable when the dimensions of the tube are still further increased.

In particular, the screen end of the tube has to be given an increasing curvature as the diameter is itself increased, so as to prevent the tube collapsing under the pressure of the external atmosphere. But by increasing the curvature the edges of the television picture become distorted, and since this distortion increases with the distance from the axis of the tube, the area on the screen over which a television picture can be reproduced satisfactorily is limited by the distortion regarded as permissible in the picture. The flattest screen surfaces which can be produced technically without requiring an excessive thickness of glass are shown in *fig. 1* for screens of different diameters. Since it is only desired to bring out the difference in curvature of the screen ends of the tube, the tubes although having different diameters are here shown of the same size.

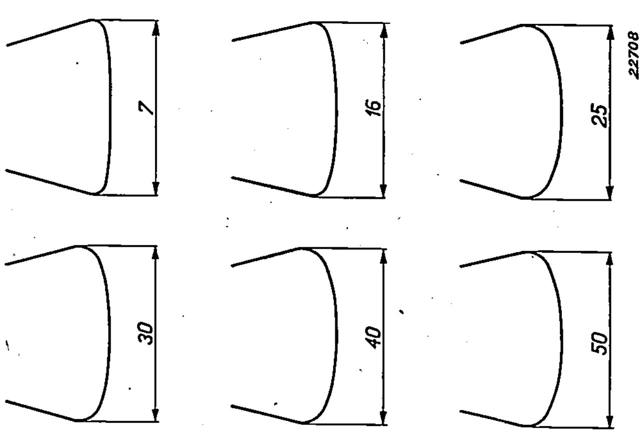


Fig. 1. Contours of several cathode ray tubes of different diameters showing the least curvature of the screens which can be tolerated from the standpoint of glass technology.

The adoption of very large tube-diameters is further limited by the fact that the precautions which have to be taken to obtain a robust tube not liable to collapse become rapidly more onerous as the diameter is increased.

In this connection, it may also be noted that the cost of making a tube increases so rapidly with the diameter of the screen that a large cathode-ray tube becomes altogether too expensive for use in television receivers.

The disadvantage of the extremely small picture normally obtained on the fluorescent screen can be remedied by making an enlarged projection of this picture. Projected television pictures of satisfactory definition and brightness measuring up to about 100 by 120 cm can be obtained by using small high-power television tubes in conjunction with a suitable optical system for enlargement. In ordinary rooms where the space available for an audience is comparatively small, the size of the projected picture should not exceed 40 by 50 cm. The best distance from which to view the projection screen is between five and ten times the width of the picture. Larger pictures are liable to be too large considering the space available in the average room.

A brief description is given in this article of the television tubes designed for projection purposes and of the optical means employed to make the best use of the radiation output of the television tube.

Television Tubes for Picture Projection

An optical system with the largest practicable relative aperture is required for the projection of television pictures. To avoid very expensive optical arrangements, the diameter of the picture must be made small and in practice the diameter of the image on the fluorescent screen should not exceed 8 cm.

It follows, therefore, that the sharpness of the focal spot of the electron beam has to conform to exceptionally severe requirements. With a screen raster of 405 lines the maximum diameter of the

---

[1]) Philips techn. Rev., 1, 16 and 321, 1936; 2, 33, 1937.

spot must not exceed 0.1 mm. This necessitates high anode voltages with which a focal spot of small diameter is much easier to obtain [2]) and which in particular allow a considerable amount of energy to be converted to light on the fluorescent screen. The Philips projection tubes are run on a voltage of between 20 to 25 kilovolts.

Electrostatic or magnetic lens systems can be used with equal effect for focusing the electron beam. Experience has shown that with magnetic lenses great sharpness can be obtained more easily than with the electrostatic system, because *inter alia*, a smaller number of electrodes is needed. This is an important advantage, since a system of electrodes which must be capable of dealing with a working voltage of 25 kilovolts has to be accommodated in a relatively confined space. For these reasons magnetic focusing has been adopted in the projection tubes developed by our laboratory.

The construction of a projection tube of this type is shown diagrammatically in *fig. 2*. On the front

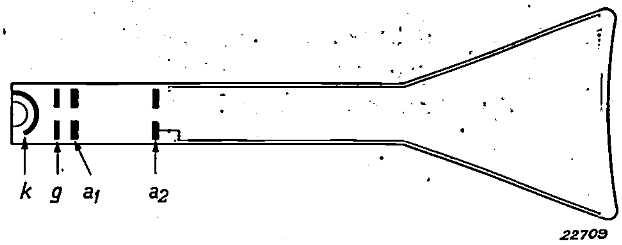


Fig. 2. Diagrammatic section of the projection tube. k - Cathode; g - grid; a_1 - first anode; a_2 - accelerating electrode.

surface of the grid g there is a small circular aperture behind which is the flat surface of the cathode k which is coated with an emissive material. The electrons are drawn out of the cathode by the first anode a_1, which also has a small aperture co-axial with the aperture in the grid; the final velocity is imparted to the electrons by the accelerating field between the first and the second anodes. This second anode a_2 is connected with a conducting surface which completely covers the inner wall of the tube between the anode a_2 and the fluorescent screen. No electric field exists in the space between a_2 and the conducting layer, apart from the comparatively small voltage drop produced at the screen by bombardment with the beam electrons. Between the anode a_2 and the conical part of the bulb is situated the magnetic focusing field, which is generated by an ironclad coil provided with an air gap. Owing to this air gap in the iron sheath the external magnetic field is restricted to a short region located just behind a_2. The magnetic fields for deflecting the electron beam are located also in this section of the tube.

A projection tube for use with a television receiver is shown in *fig. 3*.

The current intensity of the beam is mainly determined by the voltages at the grid and at the first anode, the field of the second anode penetrating only very slightly to the cathode. The system comprising the cathode, grid and first anode operates on exactly the same lines as that in a triode. Normally the voltage of the first anode is 250 volts with respect to the cathode; with a negative grid bias of about 40 to 50 volts the electron beam is then completely suppressed, while the current intensity of the beam at zero grid potential ($V_g = 0$) is 400 to 800 milliamps.

The current characteristic of the beam for a projection tube with 250 volts first anode voltage and 20 kilovolts at the last anode is shown in *fig. 4*.

The Optical System

A) Projection Lens

To project the image obtained on the fluorescent screen, it is desirable to use an optical system with the largest practicable relative aperture and also from reasons of cost to make the focal length of the projection lens as small as possible. But reduction of the focal length of the lens is limited, since the image on the screen to be projected by the lens is 8 cm in diameter, and the projected picture must have sharp definition right up to the edges.

To enlarge the area of sharp definition, an artifice was adopted in designing the projector tube. With lens systems of large aperture, the area of sharp definition of a flat object is principally limited by curvature of the image surface and only to a small extent by other optical defects due to using large angles of incidence. To produce a plane image on the projection screen, the fluorescent screen with its image is given a curvature corresponding with the curvature of that part of the lens image surface covered by the picture; in this way it is possible to reproduce a much larger area with high definition than when using a flat fluorescent screen. In general the screen is given such a curvature that the centre of curvature is located on the same side as the lens. For this reason the base of the projection tube has been made concave inwards, which is just the opposite to the shape adopted in standard cathode-ray tubes. Owing to the comparatively small diameter of the screen end a bulb of this type can be quite readily made without any loss in mechanical strength.

[2]) Philips techn. Rev., 1, 33, 1936.

Fig. 3 shows clearly the concave shape of the screen end of the tube.

By using tubes with screens of the correct shape television images of 48 by 55 mm have been projected with satisfactory definition over the whole surface on to a flat screen of 40 by 50 cm, using a projection lens with a relative aperture of 1/1.9. With plane surfaces of projection, pictures of not more than 45 mm in diameter could be projected with sharp definition using this optical system.

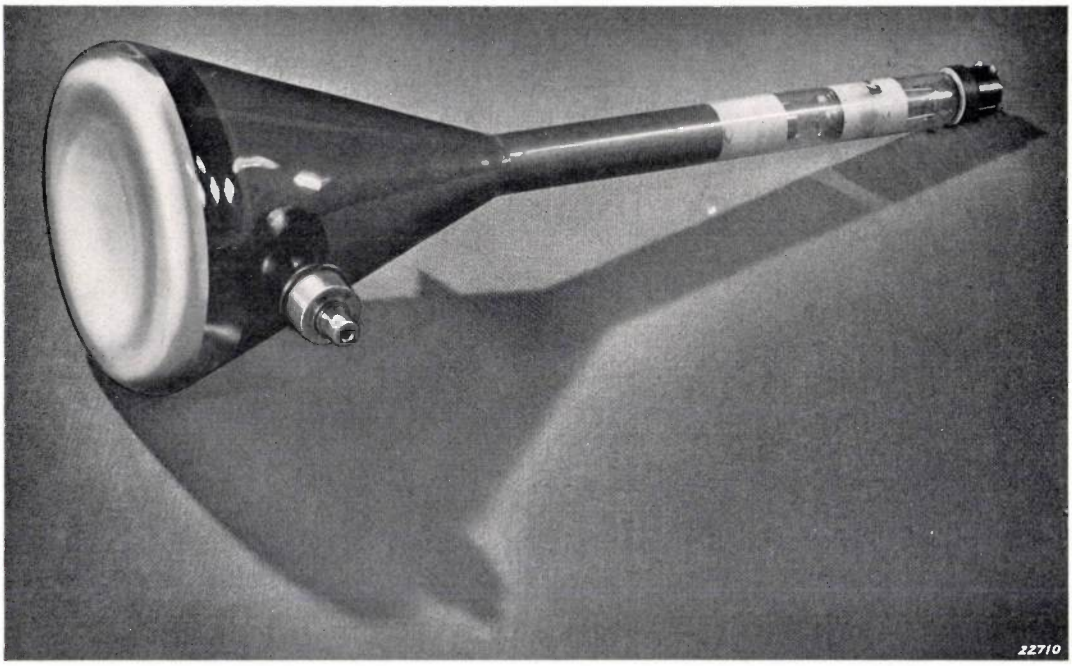


Fig. 3. Projection tube with concave screen surface.

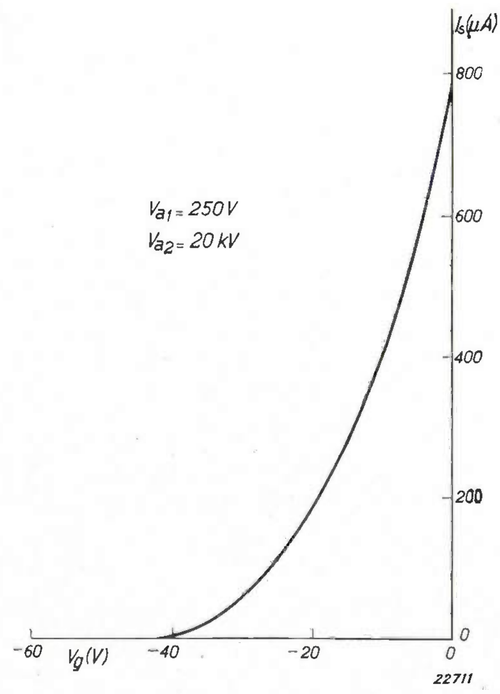


Fig. 4. Beam current plotted against grid voltage.

B) Projection Screen

The quantity of light transmitted in projection by a lens of this type is however, very small. If the efficiency of the optical system is defined as the ratio of the flux concentrated by the lens over a specific element of the picture surface to the total flux emitted by the corresponding surface element of the object, then analysis of the above-mentioned optical system would show an efficiency of only about 4 per cent. This result can be readily checked by calculation, if it is remembered that the radiation of the fluorescent layer in its immediate neighbourhood obeys Lambert's law and if the absorption and reflection losses in the optical system are also taken into consideration.

It is evident that the amount of light transmitted by the optical system must be utilised to the greatest effect. If the picture is projected on a diffuse-reflecting wall obeying Lambert's law,the brightness of the projected image with a tenfold enlargement may only be 0.0004 times the brightness on the fluorescent screen. Since the illumination of the fluorescent screen is of the order of 10^4 to $2 \cdot 10^4$ lux [3]), the brightness of the projected image will only be about 4 to 8 lux. To increase this value, it is essential to use screens having more or less specularly reflecting or translucent surfaces which possess lower dispersion values than represented by

[3]) The illumination in lux is defined as the illumination obtained on a white screen with reflecting properties obeying Lambert's law when it is illuminated with an intensity of the same number of lux.

Lambert's law. The use of specularly reflecting screens is impracticable in view of the comparatively short distance separating the optical system and the screen (short focal length): there would then be hardly room to accommodate the audience between the projection lens and the viewing screen. The second method, however, offers various important advantages. In this case the short distance separating the projection lens and the screen is an advantage since it enables the cathode-ray tube, the lens and the projection screen to be incorporated in a single housing, so that the total path of the light rays between the object and the image is confined to the interior of the receiver. Apart from the apparatus itself, no other components have then to be set up separately and the space available on the viewing side of the screen is unrestricted.

A ground glass sheet is used as the translucent screen. If a narrow pencil of parallel rays falls on one side of a screen of this type, the light will be dispersed by the matt surface so that on the other side of the screen rays will be projected on the viewer's eye not only when the viewer is located in the same straight line as the incident beam, but also when the line joining the viewer's eye with the point of incidence of the light ray on the screen is inclined to the normal. The dispersion produced by a particular screen can be represented diagrammatically by indicating the brightness in each direction by the length of an arrow drawn in that direction. The line joining the heads of the arrows then represents the dispersion curve of the particular screen. The curves for two ground glass screens with different degrees of roughness are reproduced in *fig. 5* drawn to the same scale.

It is obvious that with a small dispersion curve as shown in fig. 5*a* the brightness dispersed directly to the front with equivalent illuminations will be greater than with a dispersion curve of the type shown in fig. 5*b*, since under the conditions shown in curve *a* the energy incident on the rear side is mainly projected straight ahead. Dispersive screens can therefore be characterised by the intensity of light transmitted directly to the front. It is interesting to compare this intensity with that obtained in dispersion according to Lambert's law and to define the ratio of these two intensities as the intensification of the dispersive screen. With the majority of ground glass screens this intensification is very high, and in the cases illustrated in fig. 5 it is 9.9 for *a* and 4.6 for *b*.

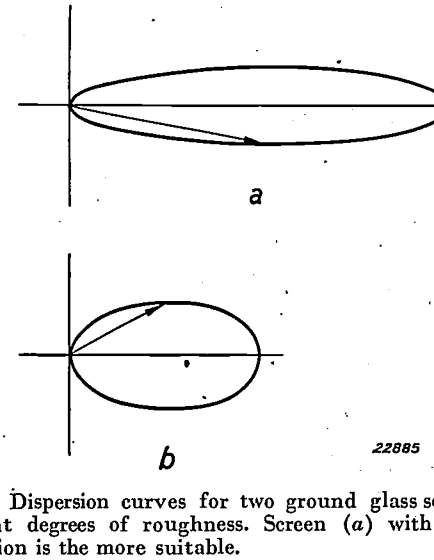


Fig. 5. Dispersion curves for two ground glass screens with different degrees of roughness. Screen (*a*) with the lower dispersion is the more suitable.

The use of screens with a minimum dispersion curve is limited by the fact that the middle and the edge of the projected image must not reveal excessive differences in brightness to an observer looking along the axis of projection, and also since the brightness must not diminish too much when the viewer moves out of the axis of projection, otherwise the area within which the picture can be viewed distinctly and comfortably will be too small.

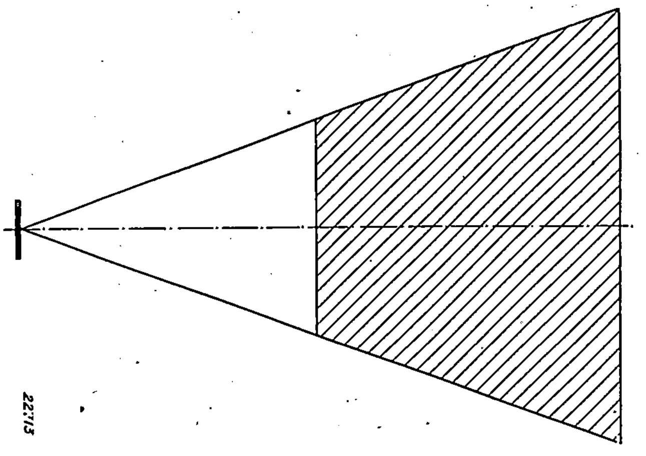


Fig. 6. Effective area (shown shaded) for satisfactory viewing of television pictures.

A serviceable middle course must therefore be found between these adverse and favourable features, and experience has shown that the screen whose dispersion curve is of the type given in fig. 5*a* provides a satisfactory compromise. Practical experience in the projection of films indicates that the permissible differences in brightness are fairly high: brightness differences of 50 per cent are only rarely obtained. This accounts for the satisfactory results obtained with the comparatively very small dispersion curve shown in fig. 5*a*. The area within which a good view of the projected picture is obtained is roughly bounded by two lines drawn from the middle of the screen at an angle of about 20 deg.

to the axis of projection and by the above-mentioned maximum or minimum distances from the screen, thus giving a trapezoidal area of about 8 sq. m with a picture measuring 40 by 50 cm.

picture is about 30 to 60 lux, which is quite adequate for television reception in rooms of moderate illumination. Two photographs of television pictures projected on ground glass screens are reproduced in *fig. 7*; these may be compared with those already published in Philips techn. Rev. **1**, 325, 1936, which show the images thrown on the fluorescent screen of a standard cathode ray tube.

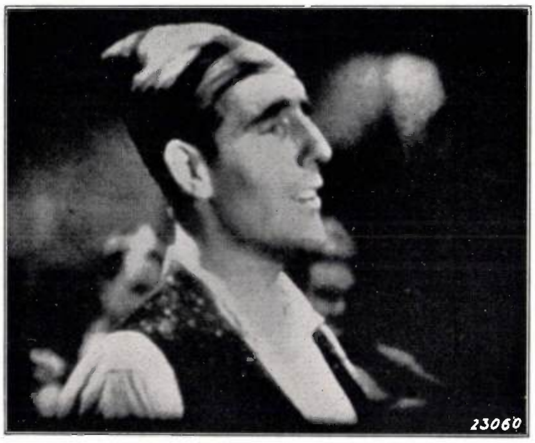


Fig. 7. Television pictures thrown on to ground-glass projection screens, 40 by 50 cm in size; total lines 405 and interlaced scanning.

In *fig. 6* the viewing area is indicated by shading. Outside this area viewing is also quite satisfactory although under slightly less ideal conditions. The intensity of illumination over the bright areas of the

PRACTICAL APPLICATIONS OF X-RAYS TO THE EXAMINATION OF MATERIALS. XII

By W. G. BURGERS.

24. Non-Ideal Crystals

Example: The Diamond

It has been frequently stated that the diffraction of X-rays produced by crystals can be interpretated as a reflection of the rays at the surfaces formed by the atoms composing the crystal, the radiation being reflected in exactly the same way as light rays by an ordinary mirror.

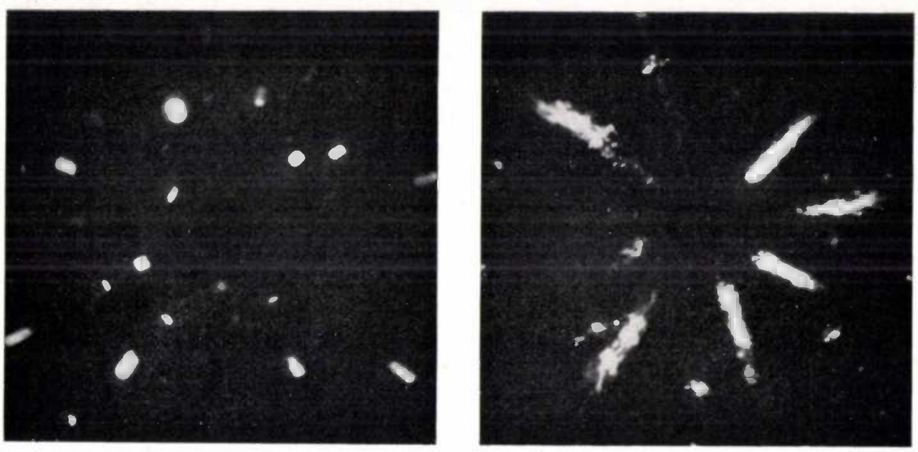
It is evident that the reflected rays are confined in one direction of reflection only when the reflecting lattice surfaces are plane; this is actually the case with undistorted crystal lattices. But in many cases the lattices are not in this ideal condition.

In the first place, naturally-grown crystals are frequently not ideal but built up from a large number of blocks in different orientations varying by not more than a few degrees; these arrangements are referred to as mosaic crystals. The divergence of these crystals from an ideal structure may be the result of the conditions of growth; for instance a truly parallel growth of the crystal surfaces may be rendered difficult by crystallisation taking place too rapidly.

Secondly, deviations from an ideal lattice structure may be due to mechanical strains. Owing to the resulting internal strains, equivalent lattice surfaces will then no longer be spaced at exactly the same intervals apart throughout a crystal or in the different crystallites from which the substance is built up; in some cases these surfaces may be more or less distorted.

It is obvious that all such deviations in the lattice must cause the X-rays to be reflected in directions which will differ somewhat from the direction of specular reflection obtained with normal and undistorted crystals. In the X-ray photogram this will lead to either a broadening or a weakening of the reflected beam which in specific cases may be shown in very different ways in the diffraction patterns. If the preparation under examination consists of a single crystal, which whould give a pattern with a comparatively small number of spots (Laue pattern), these spots in the case of a distorted crystal will become expanded to bands of different lengths, similar to the distorted image obtained in a curved mirror. If the material under examination has a fine-crystalline structure whose interference pattern consists of a large number of small spots which coalesce to form continuous haloes if the crystallites are sufficiently fine (Debye-Scherrer pattern), these haloes will become blurred or broadened when the individual crystallites are deformed by mechanical forces or from any other cause.

a



b

21763

Fig. 1. Diffraction patterns of natural diamond.
a. Colourless stone, with undistorted crystal lattice.
b. Cloudy stone, with distorted lattice built up of a large number of units with slightly-different mutual orientations (mosaic crystal).

Examples of these phenomena in fine-crystalline materials have already been given in article No. VIII of this series [1]). *Fig. 1* in the present article reproduces two X-ray photograms obtained with natural diamonds, *a* being that of a colourless stone and *b* that of a cloudy stone. The difference in character of the spots obtained by reflection of the X-rays at the crystal surfaces is evident. The expanded bands in *b*, which are composed of separate small spots, indicate that the cloudy stone in comparison with the colourless one, is built up of a mosaic of not perfectly parallel crystal units. It is clear that this difference in structure will be apparent also in other properties of the stone (e.g. brilliancy and brittleness).

25. Detection of Very Small Crystals. Example of Zinc Sulphide Crystal-Size Variation

In addition to the above-described cause of line broadening in X-ray photograms, another cause may also exist. This may be due to the fact that the diffraction of X-rays, which is essentially an interference phenomenon, will be confined to a narrow range of directions only when the effects produced by a large number of atoms are in concert. Accordingly, even with an entirely undistorted crystal lattice, reflection will be confined to the normal direction of specular reflection only when the crystal in question has a definite minimum size (about 0.1 μ). If the preparation under examination is composed of smaller crystals the rays will also be diffracted in directions which deviate considerably from that of specular reflection. As a result, the reflected rays will be included within a certain solid angle which gives rise to broadened lines on the photogram.

As an example of these conditions, two photograms are reproduced in *fig. 2* for a zinc sulphide preparation which was precipitated from a solution of a zinc salt. In order to follow the change in certain physical properties, radiographs were made of the preparation after two different heat treatments, viz., after drying at 100 deg. (fig.2*a*) and after ignition for an hour at 500 deg. (fig. 2*b*). The lines in fig. *a* are distinctly widened, which indicates that the crystals of the deposited and dried zinc sulphide are smaller than about 0.1 micron, while the sharp lines in fig. *b* show that by strong heating to 500 deg. these small crystals have coalesced to form larger ones. The latter must, however, be less than 10 microns since at this dimension the continuous interference lines are resolved into separate spots, as may be seen from fig. 2*b* for recrystallised nickel tubes in article No. IX of this series.

The occurrence of blurred interference lines can, therefore, be due either to the presence of very small crystals or to crystals which are distorted by internal stresses, so that in each particular case both possibilities must be given due weight. It will not always be easy to determine with certainty which of these conditions (or even if both together) are responsible for the widening of the lines. A very difficult problem in this respect is presented by very finely-ground powder, since on grinding not only are the crystals reduced to very small dimensions but a deformation of the small crystals may also take place as a result of the mechanical forces applied during grinding. Crystals having a more or less abnormal lattice structure can also be readily produced when substances are rapidly precipitated, for example by rapid sublimation from the vapour phase; this factor must be taken into consideration when dealing with the precipitated zinc sulphide referred to above.

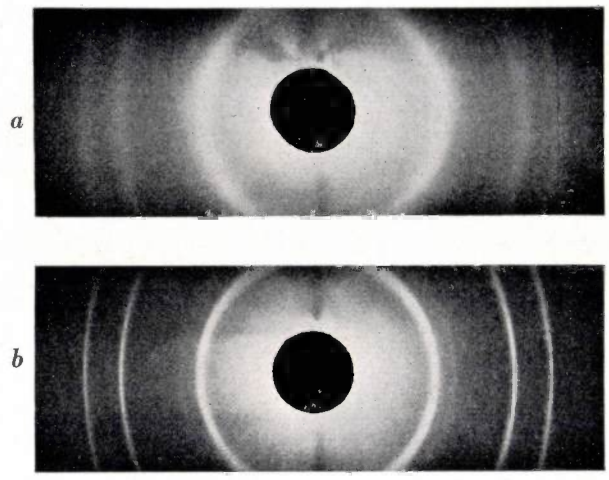


Fig. 2. Diffraction patterns of zinc-sulphide precipitates.
a. After precipitation and drying at 100 deg. C.
b. After one hour's ignition at 500 deg. C.
The broadened lines in *a* indicate the presence of small crystals with dimensions of 1 micron. By strongly heating to 500 deg. these small crystals coalesce to larger crystals, as indicated by the sharpness of the lines.

26. Amorphous Bodies

To conclude this series of articles, brief reference must also be made to the diffraction of X-rays produced by non-crystalline or socalled amorphous bodies, such as glass, artificial resins, etc. It appears probable that in general no directional interference would take place with these bodies, since in amor-

[1]) Philips techn. Rev., **1**, 373, 1936.

[2]) Philips techn. Rev., **2**, 29, 1937.

phous substances the molecules are not arranged in a crystal lattice in the same way as in crystals. Interference patterns can, however, be obtained even with amorphous bodies, because also in these bodies, exactly as in liquids, the molecules are not as a rule entirely free from some kind of regimentation, but exhibit certain characteristics of a mutual orientation. In consequence, an amorphous body, insofar as the dispersion of an incident radiation is concerned, may be regarded broadly as a highly-deformed lattice in which a regular orientation is confined to very small groups each composed of only a few molecules.

From the above considerations regarding the effects of lattice distortion and of the extremely small dimensions of the crystals, it is not surprising to find that the diffraction patterns obtained with amorphous bodies and liquids consist of a single or a few comparatively diffuse interference haloes. In *fig. 3a* to *d*, four patterns of this type have been reproduced, viz., for two phenol and cresol-formaldehyde artificial resins. It is seen from figs. *c* and *d* that resins obtained by the polymerisation of different original molecules give distinctive patterns with marked differences in the diameters and definitions of the haloes. This is due to the difference in the size of the combining molecules and hence to their distances apart, as well as to their orientation in the polymerised product. Thus to a certain extent, substances with such ill-defined structures as the resins can be differentiated by their X-ray diffraction properties. The diffraction patterns in figs. *a* and *b* were obtained with the same resins in different stages of polymerisation. This difference is not brought out in the patterns-this is not surprising if it is remembered that as polymerisation proceeds the original molecules coalesce in increasing numbers, but always on the same plan, to form the polymerised structure; the orientation of and distances between the molecules are not altered during this building-up process.

The marked difference in the interference patterns obtained with amorphous and crystalline substances enables a specific preparation or material

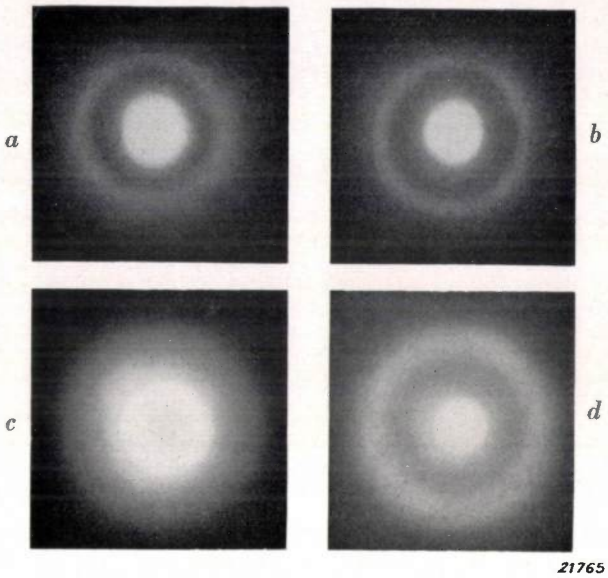


Fig. 3. Interference patterns of artificial resins.
a. Phenol-formaldehyde resin, slightly polymerised.
b. Phenol-formaldehyde resin, highly polymerised.
c. o-cresol formaldehyde resin.
d. p-cresol formaldehyde resin.

to be distinguished as crystalline or amorphous; this problem has an important application in practice, as for instance for determining whether certain precious stones are genuine or not. Stones prepared from glass may be readily distinguished by the appearance of one or more haloes in the interference pattern instead of the sharply-defined spots obtained with real stones, see e.g. the pattern in fig. 1*a* for diamond.

VOL. 2. No. 9 SEPTEMBER 1937

Philips Technical Review

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

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

FIVE-ELECTRODE TRANSMITTING VALVES (PENTODES)

by J. P. HEYBOER.

Summary. Following a general analysis of high-frequency amplification in transmitters, the characteristics of three-electrode, four-electrode and five-electrode valves are discussed, and the advantages of pentodes compared to four-electrode and three-electrode valves enumerated.

Introduction

High-vacuum amplifying valves are used in radio transmitters for various purposes, which may be classified under the following heads:
a) As low-frequency amplifiers,
b) As high-frequency amplifiers, and
c) As oscillators.
The first of these groups is used for the amplification of speech and music which are superposed as a modulation on the high-frequency carrier wave; for this reason these valves are sometimes also described as "modulating amplifiers".

The valves in groups b) and c) form components of the high-frequency part of the transmitter, while the oscillator is used for generating the required high-frequency oscillation which is then amplified in the succeeding high-frequency amplifiers.

The present article is limited to a discussion of high-frequency amplifiers.

A, B and C Amplification

If, for a particular amplifying valve, the anode current is measured as a function of the control-grid voltage, the potentials of the other electrodes being kept constant, a characteristic is obtained of the type shown diagrammatically in *fig. 1.*

When the valve is used as an amplifier, a certain negative bias is applied to the grid, which is superposed on the alternating voltage to be amplified.

Three types of amplification are distinguished according to the magnitude of the negative bias, viz.:

Class A amplification,
Class B amplification, and
Class C amplification.

Class A amplification is obtained when the negative bias is such that anode current still flows when the alternating grid voltage is zero and does not disappear even with the maximum negative

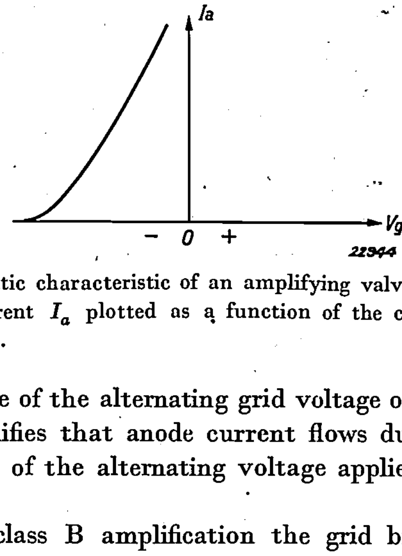


Fig. 1. Static characteristic of an amplifying valve with the anode current I_a plotted as a function of the control-grid voltage V_g.

amplitude of the alternating grid voltage occurring. This signifies that anode current flows during the full cycle of the alternating voltage applied to the grid.

With class B amplification the grid bias is so adjusted that the flow of anode current in the absence of a signal is nearly zero. If the bias voltage is superposed on the alternating voltage to be amplified, anode current will flow during half the cycle only.

In class C amplification the grid bias is more negative still than in the previous case; in the absence of a signal the anode current is zero, and when an alternating voltage is impressed on the grid, current will flow during a part of a half-cycle only.

If the alternating voltage at the grid varies sinusoidally with the time, the anode current can be readily determined, on the basis of fig. 1, as a function of the time for each of the three classes of amplification. The results obtained in this way are shown in *fig. 2*.

The curves in fig. 2 show that in class A ampli-

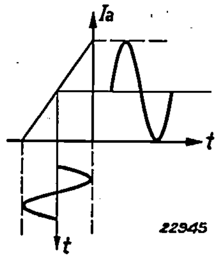
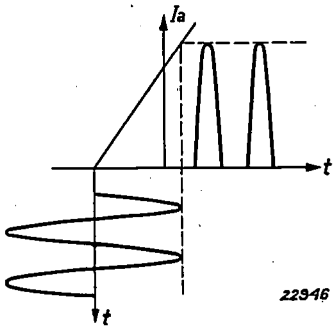
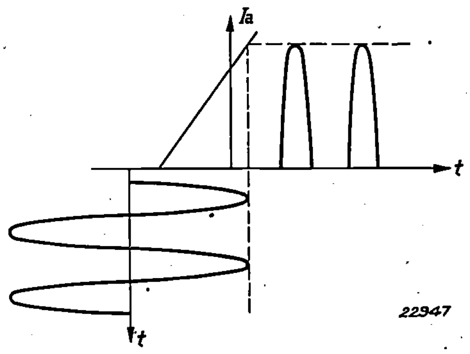


Fig. 2. Anode current plotted as a function of the time, on excitation of the control grid with a sinusoidal voltage, with *a)* Class A amplification



b) Class B amplification



c) Class C amplification

fication the anode current has almost the same curve as the grid voltage; this is not the case with class B and class C amplification because in the negative half-cycle no current flows. Class A amplification has, therefore, been favoured for low-frequency amplifiers in which undistorted reproduction of speech and music is a first consideration.

Class A amplifiers have, however, the disadvantage that a feed current continues to flow through the anode even in the absence of a signal voltage at the grid, the power carried by this current being dissipated in the form of heat. Since during operation of the valve the anode current fluctuates about this residual value, the efficiency is somewhat reduced. The maximum efficiency found with class A amplifiers is actually only of the order of 50 per cent, in other words 50 per cent of the D.C. power absorbed is transformed to useful work, while the remainder is dissipated at the anode as heat.

While this dissipation in the form of heat has frequently to be taken into account with receiving valves, its consideration is imperative in the case of transmitting valves owing to the heavy powers with which these valves have to deal, and the large monetary loss entailed in the loss of power. It is extremely important, therefore, to increase the efficiency of transmitting valves as far as practicable in order to obtain the maximum power output with a specified permissible anode dissipation. This result may be obtained in class B and class C amplification, as will be described in further detail below.

Power Output and Efficiency with Class B and Class C Amplification

The standard circuit of a high-frequency amplifier is shown in *fig. 3*. The grid is given such a high

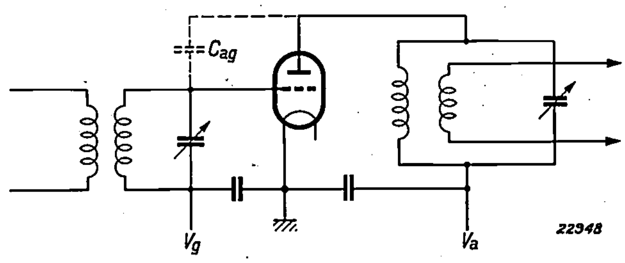


Fig. 3. Circuit using a three-electrode valve as a high-frequency amplifier. Owing to the capacity C_{ag} between the grid and anode, part of the alternating anode current is passed through the input circuit; when this circuit has an inductive reactance this may give rise to self-excitation.

negative bias that the anode current is zero (class C) or nearly zero (class B) in the absence of a signal. This negative grid bias is superposed on the high-frequency alternating voltage, which is furnished by the preceding stage in the transmitter and which is termed the exitation voltage.

A circuit is inserted in the anode lead of the valve having a resistance as load (e.g. the radiation resistance of an aerial) and which is tuned to the frequency of the excitation voltage.

Plotted as a function of the time, the anode current gives a curve of the form shown diagrammatically in figs. 2*b* (class B) and 2*c* (class C).

This current curve can be resolved into a D.C. component and various high-frequency components, as shown in *fig. 4* for a specific curve obtained with Class C amplification.

One of the high-frequency components has the same frequency as the excitation voltage (the fundamental frequency), while the other components have frequencies which are multiples of the fundamental frequency.

The anode circuit is tuned to the fundamental frequency; this implies that only those current

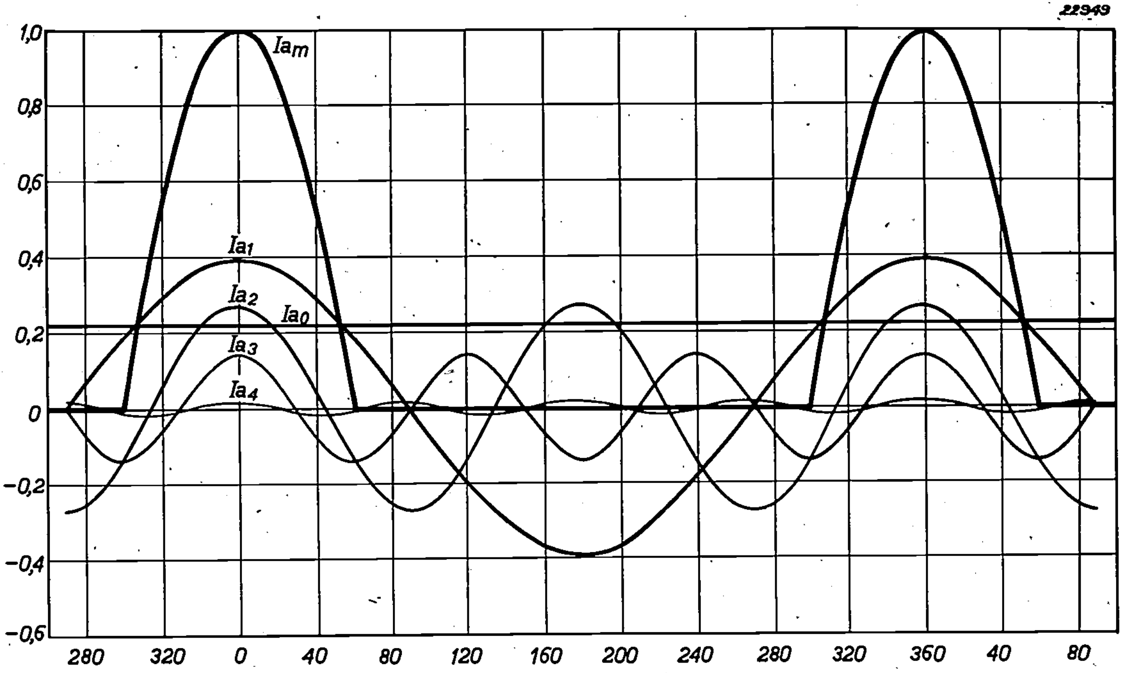


Fig. 4. Resolution of the anode-current curve of a class C amplifier into D.C. components and harmonics, when the current-carrying period is one third of the full cycle and the valve characteristic is linear.

components having the fundamental frequency (the first harmonic) encounter an impedance, which is moreover ohmic in character and which we shall term R_a, while no impedance is provided for the other harmonics. Only the first harmonic of the anode current, whose amplitude will be denoted by I_{a1}, will therefore produce a voltage drop across the resistance, this voltage drop being sinusoidal with an amplitude which will be denoted by $V_{a\sim}$. It is clear that I_{a1} and $V_{a\sim}$ are connected by the following expression:

$$V_{a\sim} = I_{a1} R_a \quad \cdots \cdots \quad (1)$$

From fig. 4 it is evident that the maximum of the first harmonic occurs simultaneously with the maximum I_{a_m} on the total anode-current curve. Since the maximum anode current naturally coincides with the maximum grid voltage $V_{g\sim}$, I_{a1} and $V_{g\sim}$ are cophasal. It follows from the fact that $V_{a\sim}$ is a voltage drop in R_a that $V_{a\sim}$ is in phase opposition to $V_{g\sim}$. The variations of the currents and voltages with time are shown in *fig. 5*.

The power passing from the valve to the anode resistance R_a is:

$$P_0 = {}^1\!/_2\, I_{a1}\, V_{a\sim}, \quad \cdots \cdots \quad (2)$$

while the D.C. power absorbed is:

$$P_i = V_a\, I_{a0}, \quad \cdots \cdots \quad (3)$$

where V_a is the direct anode voltage and I_{a0} the direct anode current.

The efficiency is therefore:

$$\eta = \frac{P_0}{P_i} = {}^1\!/_2\, \frac{I_{a1}}{I_{a0}} \cdot \frac{V_{a\sim}}{V_a}; \quad \cdots \quad (4)$$

and is thus determined by the ratios I_{a1}/I_{a0} and $V_{a\sim}/V_a$, the latter ratio being termed the voltage

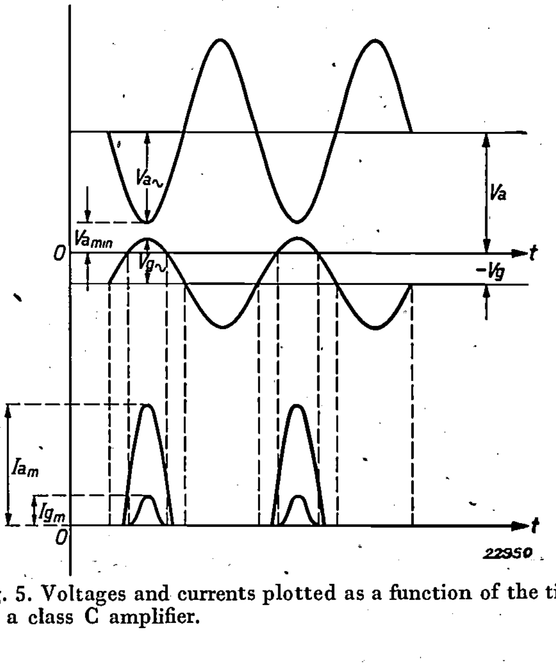


Fig. 5. Voltages and currents plotted as a function of the time for a class C amplifier.

amplification factor. A higher efficiency is therefore obtained when the ratios I_{a1}/I_{a0} and $V_{a\sim}/V_a$ are made as large as possible.

The ratio I_{a1}/I_{a0} is determined by the shape of the anode current curve and for a given form of curve may be found by expanding in a Fourier series.

If the curve is composed of semi-sinusoidal components (half-waves), as is nearly the case with Class B amplification, the ratio $I_{a1}/I_{a0} = {}^1/_2\,\pi = 1.57$; with class C amplification this ratio has a greater value and is usually between 1.7 and 1.8.

The voltage amplification factor $V_{a\sim}/V_a$ cannot in practice exceed unity and is usually about 0.9. (With four-electrode valves the conditions are somewhat different, as will be shown below). Why the amplification factor has this value may be explained as follows:

When the grid bias passes through its maximum value, the anode current is also a maximum and the anode voltage is a minimum. The residual anode voltage is then:

$$V_{a_{\min}} = V_a - V_{a\sim} \quad \cdots \quad (5)$$

(see fig. 5).

When the ratio $V_{a\sim}/V_a$ approaches unity, $V_{a_{\min}}$ becomes steadily smaller, finally becoming zero, and even negative when $V_{a\sim}/V_a$ is slightly greater than 1. If, however, $V_{a_{\min}}$ is zero or negative, the anode current must be zero, since the electrons are now no longer accelerated in the direction of the anode, but on the other hand are thrown back towards the grid after having passed through it. Instead of a maximum, the anode current then passes through a minimum value, so that the current curve is of the form shown diagrammatically in *fig. 6*, i.e. at the point where a maximum was

Fig. 6. Variation of the anode current of a three-electrode valve with sinusoidal excitation voltage (increasing from left to right). If the excitation voltage is so high that the anode voltage drops below the grid voltage at the peak value of the anode current, the greater part of the emission current passes to the grid at the expense of the anode current. This produces a trough in the current curve of such character that the first harmonic of the anode current no longer increases above a certain excitation voltage.

previously obtained the curve is now troughed, the trough being the deeper the smaller $V_{a_{\min}}$, in other words the greater $V_{a\sim}$ becomes.

The presence of this trough imposes a certain limit on all components of the anode current, as for instance on the first harmonic I_{a1}, since its maximum happens to be located just at the trough. As equation (1) is valid throughout, $V_{a\sim}$ also has a limit, which according to experience is about 0.9 of the direct anode voltage. We thus find that the alternating anode voltage is limited by the direct anode voltage.

Since the voltage and current amplification factors are now approximately known, the same also applies to the efficiency; for class B amplifiers the efficiency is about 70 per cent and for class C amplifiers 75 to 80 per cent.

It is seen that the efficiency is considerably higher than with Class A amplifiers.

Three-Electrode Valves as High-Frequency Amplifiers

If a three-electrode valve is used as a high-frequency amplifier in the manner described, a high power output as well as a satisfactory efficiency is obtained.

There are, however, two important disadvantages in this use of the three-electrode valve, viz.:

1) The control grid current is comparatively high, and
2) The valve is liable to become self-exciting.

Regarding the control grid current, it should be noted that the magnitude of this current is closely associated with the voltage amplification factor. We have seen (fig. 5) that the anode voltage is a minimum at the same time as the grid voltage is a maximum, and also that the alternating anode voltage must be made as high as possible in order to obtain a satisfactory efficiency. But the increase in the voltage amplification factor will be limited merely to making the minimum anode voltage equal to the maximum grid voltage. If, for instance, the grid voltage were higher than the anode voltage, then during the time in which the grid is at this potential a large part of the emission current would flow from the cathode to the grid at the expense of the anode current. This would naturally reduce the power output, since the trough in the anode current curve is then produced as already described, while the grid current curve would exhibit a high peak.

The minimum anode voltage must, therefore, never be smaller than the maximum grid voltage, but at most equal to it; with larger valves, especially those of the water-cooled type, the minimum anode voltage is in fact a multiple of the maximum grid potential.

Nevertheless, even when this condition is met, the grid current will still be fairly high, since owing to the periodically low anode voltage the accelerating action of the anode is small and hence the probability of electrons reaching the grid is comparatively high.

The grid current power, the socalled excitation power, is hence relatively high with a three-electrode valve and must be furnished by the preceding stage. It will be seen below that screen grid valves have more satisfactory characteristics in this direction.

The liability to become self-exciting is due to the capacity between the grid and anode, shown as C_{ag} in fig. 3, because the alternating anode voltage induces a current in the series circuit made up of C_{ag} and the grid circuit, whereby an alternating voltage is applied to the latter. If the impedance in the grid circuit is inductive, this feed-back voltage will be in phase opposition to the alternating anode voltage and hence just cophasal with the initial excitation voltage. In certain circumstances, the feed-back voltage at a given alternating anode voltage can thus become equal to the excitation voltage required to produce this alternating anode voltage; self-excitation then results.

This disadvantage of high-frequency three-electrode amplifiers can be eliminated in one of two ways:

1) By neutrodyning, i.e. applying a voltage to the grid which is equal to the feed-back voltage but in phase opposition to it.
2) Inserting an auxiliary grid between the anode and control grid, which is earthed on the high-frequency side, so that a screening action is produced between the anode and the control grid, and C_{ag} is thereby made very small.

Four-Electrode Valves

The latter method is used in four-electrode and five-electrode valves which are fitted with a screen grid g_2 between the control grid g_1 and the anode a. Experience has shown that neutrodyning can be dispensed with in these amplifying valves, as C_{ag_1} is very small. The values of C_{ag_1} are given below for typical four-electrode valves, as well as the corresponding values for two-three-electrode valves for purposes of comparison.

Four-electrode valves	QC 05/15	$C_{ag_1} = 0.004$ μμF
	QB 2/75	$C_{ag_1} = 0.02$ μμF
	QB 3/500	$C_{ag_1} = 0.01$ μμF
Three-electrode valves	TC 04/10-I	$C_{ag} = 6.8$ μμF
	TC 1/75	$C_{ag} = 10.4$ μμF

The introduction of the screen grid alters the shape of the anode-current/anode-voltage characteristics, so that these curves differ from those for three-electrode valves. This is shown by the characteristics for a TC 04/10 three-electrode valve and a QC 05/15 four-electrode valve in *figs.* 7 and 8,

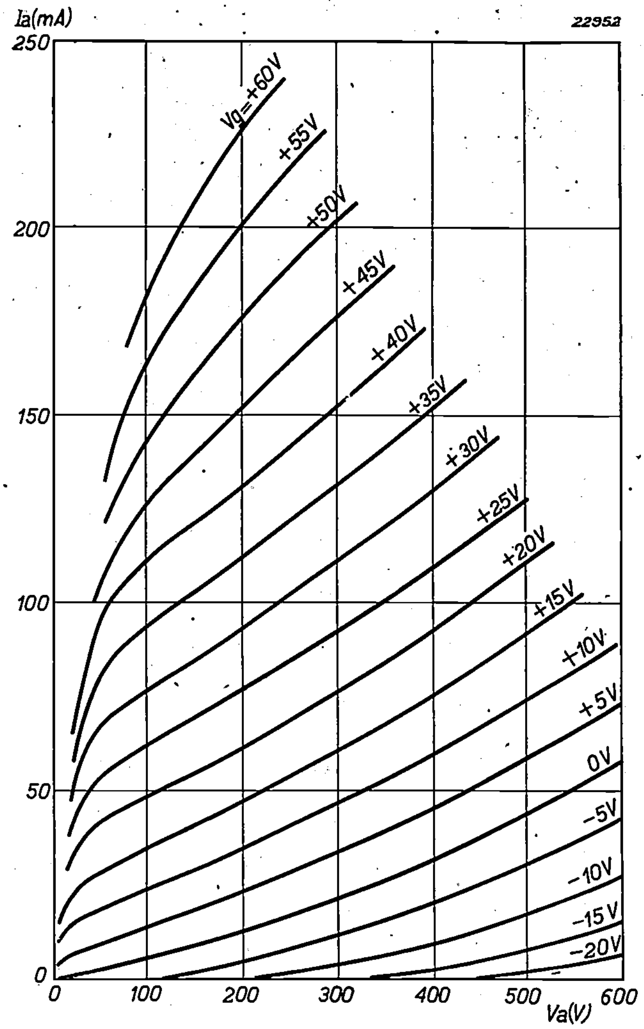


Fig. 7. Static I_a - V_a characteristic for the TC 04/10-I three-electrode valve. The anode current increases sharply with the anode voltage.

where by far the greater parts of the characteristics for the four-electrode valve are nearly parallel to the V_a axis, in other words the anode current is practically independent of the anode voltage. That this must in fact be the case is self-evident, for the screening effect produced by the screen grid practically eliminates the influence of the anode voltage on the anode current.

It may be seen from fig. 8 that the I_a-V_a characteristics of a four-electrode valve rapidly drop to zero, when the anode voltage is roughly equal to the screen grid voltage. As a result, the alternating anode voltage during operation can at most become equal to the difference between the direct voltages at the anode and screen grid, since the anode-current curve becomes troughed in the way already described when the alternating voltage increases. The voltage amplification factor cannot here reach unity but is approximately equal to

$$\frac{V_{a\sim}}{V_a} = \frac{V_a - V_{g_2}}{V_a} = 1 - \frac{V_{g_2}}{V_a} \quad . \; . \; . \quad (6)$$

If V_{g_2} is large in comparison with V_a, as occurs for instance in the QC 05/15 valve, where V_a = 500 volts and V_{g_2} = 125 volts, the amplification factor will be much less than 1 (e.g. in the QC 05/15 valve about 0.75). The power output and the efficiency are therefore both reduced.

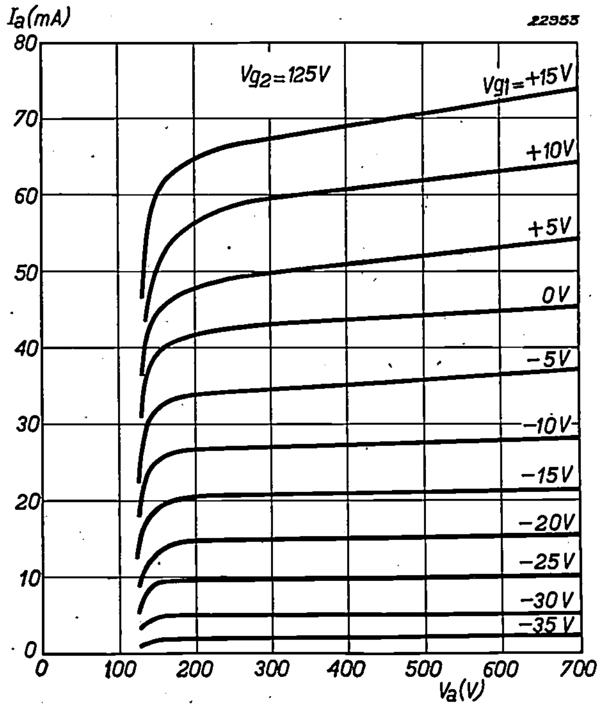


Fig. 8. Static I_a - V_a characteristic of the QC 05/15 four-electrode valve with 125 volts on the screen-grid. The anode current is nearly independent of the anode voltage, provided the latter is higher than the screen-grid voltage.

There is, however, the advantage, apart from the very low value of C_{ag_1}, that the control grid current, and hence the excitation power, are much smaller than in a three-electrode valve when run under otherwise equivalent conditions, so that the control stage can be given a much smaller rating and is therefore cheaper to make.

The fact that the control grid current is low is due to the presence of the screen grid to which a constant high potential is applied, while a similarly high voltage is not applied to the control grid during the working of the valve. The accelerating force applied to the electrons after passing through the control grid is, in consequence, greater than in a three-electrode valve; so that relatively less electrons pass to this grid and the grid current is lower.

Secondary Emission

The bend in the I_a-V_a characteristics of a four-electrode valve at the point $V_a = V_{g_2}$ is due to secondary emission at the anode and screen grid.

Of the electrons penetrating the control grid a part will strike the screen grid and others will pass to the anode. By the impact of the (primary) electrons secondary electrons are emitted from the screen grid, which travel towards the anode as long as its potential is higher than of this grid. If, on the other hand, the anode voltage is made equal to the screen grid voltage or slightly higher than the latter, practically no accelerating force towards the anode will be applied to the secondary electrons emitted from the screen grid and these will therefore not reach the anode. Should the anode voltage be lower than the screen grid voltage, the secondary electrons liberated from the anode will on the contrary pass to the screen grid, with the result that the anode current may even become negative.

Five-Electrode Valves (Pentodes)

To suppress secondary emission, a third or suppressor grid connected to the cathode has been interposed between the screen grid and the anode in pentodes. The secondary electrons from the screen grid are now unable to reach the anode owing to their low velocity and the repulsion applied by the suppressor grid. There is just a little opportunity for secondary electrons from the anode to travel to the screen grid, when V_a and V_{g_2} become smaller. This also eliminates the marked reduction of I_a in the neighbourhood of $V_a = V_{g_2}$. In fact the I_a-V_a characteristics of pentodes are nearly parallel to the V_a axis up to almost $V_a = 0$, as is shown in *fig. 9* for the PE 05/15 pentode.

Only when the anode voltage approaches zero does a bend appear in the characteristics, because then the primary electrons which have passed through the screen grid are no longer accelerated and the anode current hence drops to zero.

Thus when using a pentode as a high-frequency amplifier, the maximum value reached by the alternating anode voltage is roughly equal to the direct anode voltage, so that the voltage amplification factor again approaches unity and is hence higher than with a four-electrode valve other conditions being equal. The power output and efficiency are therefore also higher.

The advantages of the pentode may be summarised as follows:

1) Neutrodyning can be dispensed with, which is an important advantage particularly with transmitters transmitting on different wave lengths, and in high-frequency amplification.

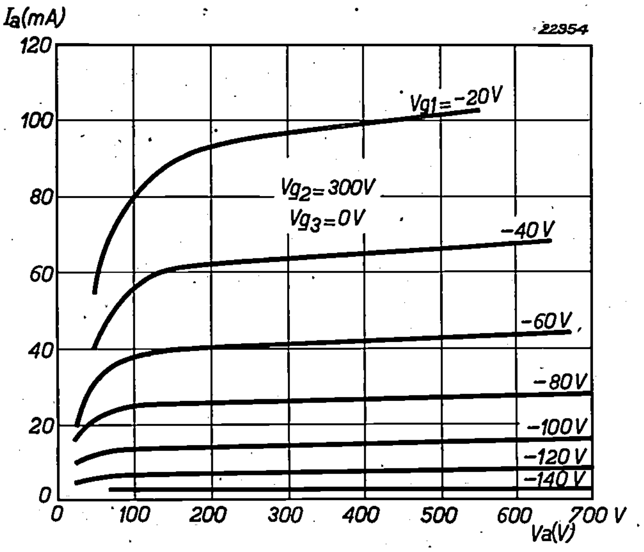


Fig. 9. Static I_a - V_a characteristic for the PE 05/15 five-electrode valve with a screen-grid voltage of 300 volts and zero suppressor-grid voltage. The anode current is independent of the anode voltage, except at very low values of this voltage.

2) Power output and efficiency are very satisfactory.

3) The excitation voltage is very small, or in other words power amplification is very high. A transmitter for a given power output therefore contains less stages when equipped with five-electrode valves than with three-electrode valves.

Use of Pentodes in Transmitters

In the use of pentodes in transmitters, a distinction may be drawn between the following applications:

a) High-frequency telegraphy amplification,

b) High-frequency telephony amplification (class B amplification),

c) Modulation of high-frequency oscillations, which are amplified by the valve.

a) When used for high-frequency amplification in telegraph circuits, the adjustment of the valve must be such that the maximum power output is obtained at the optimum efficiency. Class C conditions are therefore desirable in this case. The limits of adjustment are due to the maximum permissible dissipation in the anode and the screen grid and the maximum direct cathode current.

b) When used as a high-frequency telephony amplifier the excitation voltage is modulated on a low-frequency carrier, which means that modulation in one of the preceding stages of the transmitter is necessary. Adjustment must naturally be such that the modulation of the excitation voltage is preserved without distortion in the aerial current.

For this purpose the negative bias at the control grid must be made so large that with the screen grid voltage used the anode current is nearly zero (class B amplification).

That there is then a linear relationship between the aerial current and the amplitude of the excitation voltage may be seen from the following:

It is shown in fig. 2*b* that the anode current curve is composed of half-waves at intervals of half a cycle, i.e. the time during which the current flows is always half of a full cycle.

For a current curve of this type the ratio between the maximum current I_{a_m} and the amplitude of the first harmonic I_{a1} is constant and equal to $^1/_2\,\pi$, it is thus independent of the value of I_{a_m}.

As the static characteristic of the valve is nearly straight I_{a_m} is proportional to $V_{g_1\sim}$, from which it follows that I_{a1} is proportional to $V_{g_1\sim}$. From equations (1) and (2) we get furthermore:

$$P_0 = {}^1/_2\, I_{a1}\, V_{a\sim} = {}^1/_2\, I_{a1}^{\,2}\, R_a \quad \cdot \cdot \quad (7)$$

This power reaches the aerial, and if I_{ant} is the effective value of the aerial current and R_{ant} the aerial resistance, we have:

$$P_0 = I_{\text{ant}}^2\, R_{\text{ant}} \quad \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \quad (8)$$

We get from equations (7) and (8):

$$I_{\text{ant}} = I_{a1} \sqrt{\frac{R}{2\,R_{\text{ant}}}} = \text{const.} \cdot I_{a1}, \quad \cdot \quad (9)$$

i.e. I_{ant} is proportional to I_{a1} and according to the above is also proportional to $V_{g_1\sim}$.

c) Finally, the high-frequency oscillation which is amplified by the valve can be modulated with a low-frequency voltage by impressing this low-frequency voltage on the direct voltage of one of the electrodes. This can be done in the following ways:

1) Anode modulation,
2) Suppressor-grid modulation,
3) Screen-grid modulation,
4) Control-grid modulation,
5) Simultaneous modulation at the anode and screen grid.

In all cases the valve is used as a class-C amplifier.

In each of these methods of modulation there must be a straight-line relationship between the aerial current and the direct voltage at the electrode

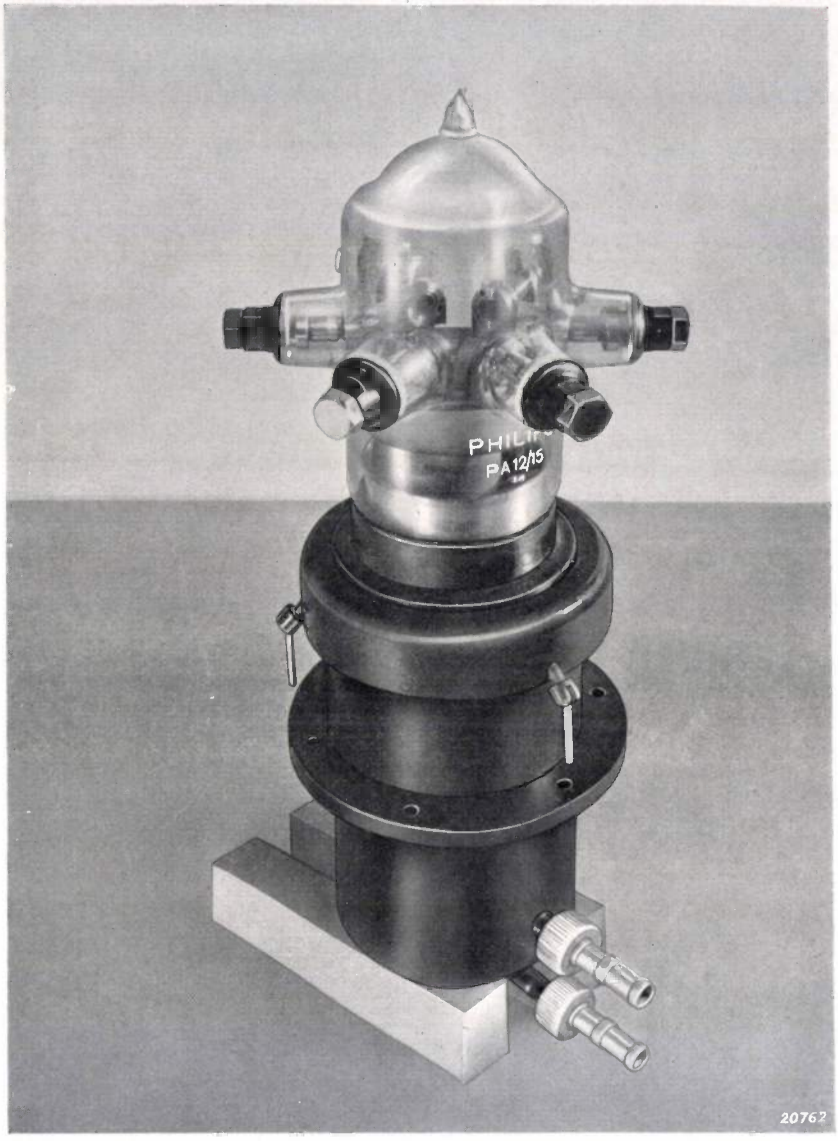


Fig. 10. The PA 12/15 five-electrode valve has a tungsten cathode and is water-cooled. It has an output exceeding 15 kilowatts and has been designed for wave-lengths of about 6 m, being thus particularly suitable for short-wave radio and telephone transmitters and for television transmitters.

on which the modulation voltage is impressed, since the modulation must be reproduced without distortion. It has been found in practice that a satisfactory linear relationship is readily obtainable by methods 1), 2) and 5), but that this is difficult with 3) and 4).

In modulation at the suppressor grid, the fact is used that the anode current is the smaller the higher the negative voltage applied to the suppressor grid. This is due to the partial rejection by the negative suppressor grid of the slower electrons, so that not all the electrons reach the anode.

The relationship between the aerial current and the suppressor grid voltage has been found satisfactorily linear so that good modulation may be obtained. A particular advantage found with modulation at the suppressor grid is that no power for modulation is required, for since the suppressor grid remains negative throughout modulation no current flows to this grid. The modulator can therefore be made small even with the largest of the five-electrode valves, e.g. the PC 3/1000.

Properties of the Philips Transmitting Pentode

Philips manufacture the following five-electrode transmitting valves:
PE 05/15, OC 05/15, PC 1/50, PE 1/80, PC 1.5/100, PC 3/1000 and PA 12/15. The power outputs of these valves lie between 15 watts with the first valve over 15 kilowatts with the last valve. Of these valves PE 05/15 and PE 1/80 have indirectly-heated cathodes, PA 12/15 has a tungsten cathode and the other valves a directly-heated oxide cathode. The PA 12/15 transmitting valve is water-cooled and is shown in *fig. 10.*

It has already been seen that the control grid current in these valves is very small as compared with that in three-electrode valves, and that the excitation power also is hence much smaller.

A comparison is made in the following table between the TC 1/75 three-electrode valve and the PC 1.5/100 five-electrode valve, the data relating to the so-called telegraph adjustment, i.e. at which the valves give their highest output.

		TC 1/75	PC1.5/100
Anode voltage	V_a	1500 V	1500 V
Control-grid voltage	V_{g1}	—160 V	—200 V
Screen-grid voltage	V_{g2}	—	300 V
Suppressor-grid voltage	V_{g3}	—	0 V
Anode current	I_a	123 mA	145 mA
Control-grid current	I_{g1}	12 mA	3.5 mA
Screen-grid current	I_{g2}	—	50 mA
Excitation voltage	V_{g1}	240 V	280 V
Excitation power	P_{hf}	2.9 W	1.0 W
Anode power input	P_{ia}	185 W	218 W
Anode dissipation	P_a	70 W	77 W
Screen-grid dissipation	P_{g2}	—	15 W
High-frequency power output	P_0	115 W	141 W
Anode efficiency	η_a	62.2 %	65 %
Ratio $\frac{\text{power output}}{\text{excitation power}}$	$\frac{P_0}{P_{hf}}$	40	141
Capacity between anode and control grid	C_{ag1}	10.4 μμF	0.03 μμF

It is seen from this table that the ratio of the power output to the excitation power is 141 with the five-electrode valve and 40 with the three-electrode valve, the five-electrode valve thus having a marked advantage in this direction.

THE RELATIONSHIP BETWEEN FORTISSIMO AND PIANISSIMO

By R. VERMEULEN.

The vigour which should characterise well-reproduced music is dependent upon adequate reconstitution of the original contrasts between fortissimo and pianissimo. It is well known that acoustical contrasts in the reproduction of a musical item are less pronounced than in the original rendering.

A check is imposed on the fortissimo by the restricted output of the amplifier and the loudspeaker; this limitation is, however, not of a fundamental character, since in the present state of technology the output can always be raised to the economic level justified in each particular case. The range of the contrasts which can be reproduced is also limited by the characteristics of the transmission members, which include not only the cables and the wiring in a radio plant, by which reproduction is transmitted to a distant point, but also the gramophone disc, the sound film and the "Philimil strip" [1]) which all permit the reproduction of music at times subsequent to their actual execution. All these reproduction and transmission units possess a characteristic maximum-permissible depth of modulation being determined either by cross-talk interference from other cores in the same cables, by the carrier wave used in broadcasting, by the mutual distance between the grooves in a disc, or by the width and optical transmission properties of the sound track on the film. This maximum modulation value must reproduce the fortissimo. In this connection it should be noted that the actual intensity of the maximum signal obtained is of secondary importance, since it can be magnified to any level by amplification, and its limitation in the fortissimo direction does not therefore of itself constitute a restriction of the brilliance.

It is, however, found that in the other direction, viz., towards the pianissimi, there is also a limit of response and that it is not practicable to use arbitrary low depths of modulation. The reproduction of the pianissimi is restricted by the interference phenomena which occur with every transmission system and which are usually referred to as "mush" or background noise. These interference phenomena may be due to a variety of very different causes. They are usually of such weak intensity that at a superficial glance they appear to be negligible, yet owing to the very wide range of sound intensities which are audible to the ear, even the slightest interference noise may still be detectible. With gramophone records, the surface inequalities of microscopic dimensions are the principal source of this interference; in the sound film individual silver grains of the light-sensitive emulsion and various other particles on the film, as well as the general deterioration of the surface, which determines the amount of light transmitted, may be responsible for interference.

If the interference is very weak, but of frequent occurrence, as for instance with inequalities in the disc grooves and with grains in the light-sensitive emulsion, impulses of very short duration will be produced during reproduction and these will occur at random and with an indeterminate frequency. The ear will detect these sources of interference as "mush" of roughly constant intensity and uniform timbre without any particular pitch. The best example of this is the shot effect, which has been discussed in a previous article in this Review [2]).

With the "Philimil" strips, the intensity of this background noise is very much reduced owing to the absence of grains in the coating and the sharpness of the freshly-inscribed surface. Nevertheless a residual mush remains also with these strips. Dust particles which can never be entirely excluded become audible as a low mumbling sound and impose a new, although lower, limit to the reproduction of weak sounds.

The height of the interference level, which is taken to include all types of unwanted and spurious noises, including mush and mumbling, also determines the maximum ratio between the fortissimi and pianissimi which can be reproduced or transmitted without distortion. It has already been pointed out that this ratio is too small for the reproduction in correct proportion of all the sound intensities obtained with a large symphony orchestra. It devolves, therefore, on the engineers operating and controlling the microphone potentiometer to maintain the frequency sweep recorded between the limits imposed on efficient reproduction, but without appreciably detracting from the general effect produced. To obtain this result the usual policy adopted is to leave the crescendi and decrescendi unaltered and to regulate only very slowly in the more constant parts. From the score

[1]) See Philips techn. Rev., 1, 107, 135, 211 and 230, 1936.

[2]) M. Ziegler: The causes of noise in amplifiers, Philips techn. Rev., 2, 136, 1937.

those passages are located where considerable differences in sound are to be expected as well as those where regulation will not spoil the general musical effect; these passages are marked in the score.

As an example of the variety of sound intensities which normally occur, there are reproduced below records of the sound intensities registered in the auditorium of the Philips Works' Theatre at Eindhoven during a concert given by the Amsterdam "Concertgebouworkest" under the direction of Prof. Dr. Willem Mengelberg. To record these sounds, they were picked up by a microphone and the microphone voltage measured, after amplification, by means of a recording voltmeter with logarithmic scale. The amplification was naturally carefully fixed at the start and not altered in the course of each measurement. At the same time a sound track was made on a "Philimil strip" in which the amplification was adapted to the limits of modulation using the method already described.

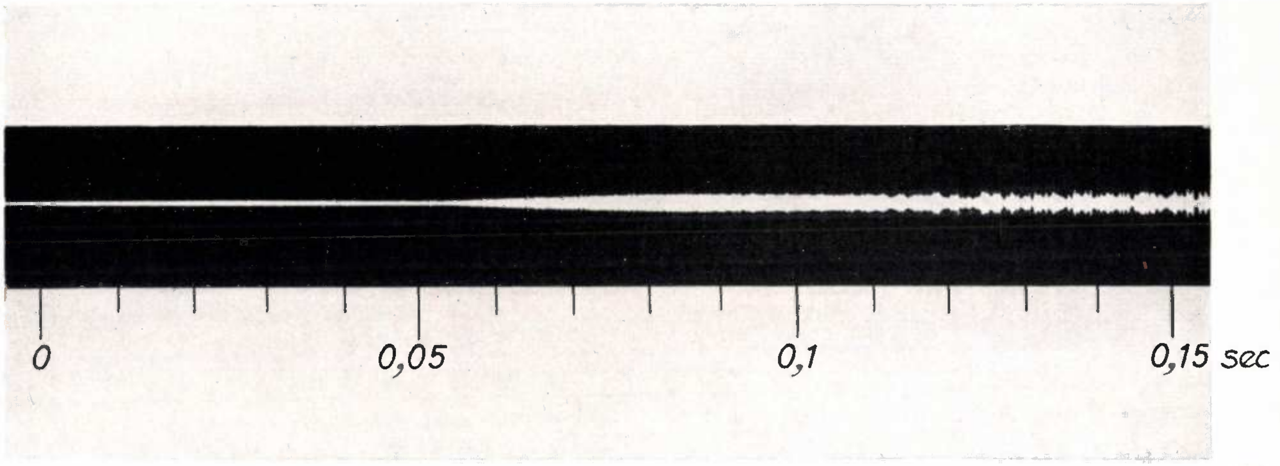


Fig. 2. Registration on the "Philimil strip" of the 160—161 bars where the sound intensity suddenly increases. This strip has been enlarged three times and the sound recorded thus covers only 0.15 of a sec. It is seen that one of the instruments started playing 0.05 sec before the others.

Fig. 1 reproduces the record made of the opening bars of the first part of the Sixth Symphony of Tschaikowsky, Opus 74. The top strip is the record made in the auditorium itself. The beginning of the record at a level of 35 decibels represents the noise in the auditorium at the end of the interval; this noise gradually subsided as the audience took their seats. The conductor on entering was greeted with applause of 55 decibels intensity for a period of 15 secs. The noise in the auditorium

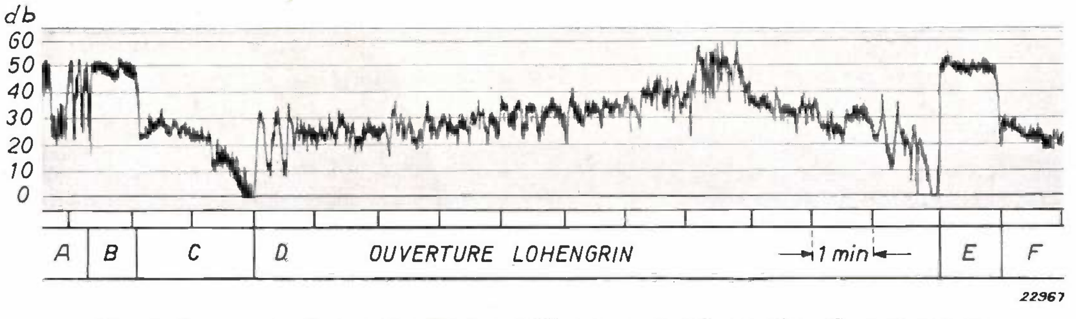


Fig. 3. Overture to Lohengrin, Richard Wagner, played by the "Concertgebouworkest" under the direction of Prof. Dr. Willem Mengelberg at the Philips Works' Theatre in Eindhoven. The letters denote:

A Last chord of the Eighth Symphony of Beethoven, which preceded the Overture.
B Applause. The return of the conductor is indicated by the renewed applause.
C Noises in the auditorium between the two pieces. The noises subside at the beginning of the overture.
D Registration of the Lohengrin Overture; the opening chords, the peak sound at the 54th bar and the closing chords can be clearly distinguished. The closing bars are played *pp*, only four violins playing, and the sound intensity then drops below the threshold of registration of the recorder. The range of sound intensities thus exceeded 60 decibels.
E Applause lasting 1 minute.
E Noises in the auditorium during the interval. As the audience left the auditorium during the interval these noises gradually subsided.

then quickly subsided, a few sharp peaks being however produced by coughing in the audience. The adagio was played by the bass-viols only, and which played *pp*; the bassoons then came in with the theme playing loud crescendi and diminuendi of nearly 15 decibels. It is not proposed to discuss in this article the rendering of each bar of the symphony, but for the guidance of those interested

Fig. 1. Explanation at the bottom of page 269.

and who have a score of this symphony available, the times of certain characteristic passages are indicated below the audiogram. The sharpest contrast is obtained at the *ff* start of the allegro vivo passage at the 161st bar following an organ point (a prolonged muted tone) of the first bassoon marked *pppppp* in the score. The difference in the sound intensities is 45 decibels; to illustrate this more clearly the corresponding page from the score has been printed below the sound records, with this difficult point marked with an arrow and two strokes. A section of the "Philimil" strip with the sound track of this passage is shown enlarged three times in *fig. 2*.

Under the sound-intensity record made in the auditorium, the record is reproduced of the voltage delivered by the photocell amplifier of the Philips Miller reproduction apparatus and which was registered on the following day. This record shows the effect of potentiometer regulation, and it is evident that the difference between the maximum and minimum sound intensities has been reduced, and that in particular the extreme maxima have been lowered and the minima not allowed to drop to the previous level as these are always limited by the interference level. The start of the record shows the background noise of the "Philimil strip", although after half a minute the potentiometer was stepped up and the noise in the auditorium recorded at a lower level. Just before the applause greeting the conductor, the potentiometer was stepped down in order to deal with the contrast in sound of the applause. During the interval of comparative quiet before the symphony commenced the amplification was again increased, and so on. It may be concluded from the measurements that the difference in sound intensity was about 55 decibels; to arrive at this figure the points at which the orchestra was not playing (e.g. bar 6 and bar 89) must naturally not be included. On the "Philimil strip" this difference has been reduced to 38 decibels, during which the extreme limits of modulation were not reached at any point in order to maintain the efficiency of registration.

Fig. 3 shows another registration made during the same concert, which included the overture to Lohengrin by Richard Wagner. Further details of this sound record are given in the caption under fig. 3.

It is interesting to note that, in spite of the successive crescendi and decrescendi, the intensity level on the average gradually increased, reached a maximum and then gradually diminished again.

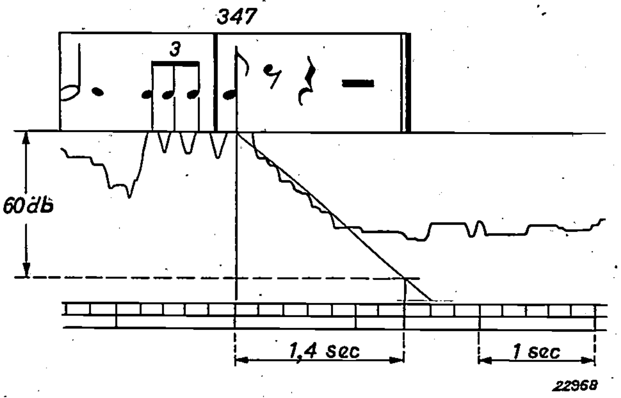


Fig. 4. Reverberation measurement. The time scale in this diagram is about 12 times greater than that used in fig. 3. Registration was made during the closing bars of the third part of the Sixth Symphony of Tschaikowsky. At the top edge, the sound intensity exceeded the range of the apparatus. The individual sounds can be clearly distinguished; the actual notes of these two bars are shown at the top of the figure.
After the closing chord the sound in the auditorium decayed to the prevailing noise level. Although there are irregularities in the course of the curve due to interference caused by various reflections, a mean line can easily be drawn. This line indicates that the rate of decay of the sound was 60 decibels in 1.4 sec, so that the mean time of reverberation of the auditorium was 1.4 sec.

Fig. 4 shows with a highly-magnified time scale a few bars from the end of the third part of the Sixth Symphony of Tschaikowsky already referred to. The orchestra finished this passage sharply with a fortissimo, after which the sound gradually died away by reverberation in the auditorium. In this way [3]) the mean reverberation in a full auditorium can be measured without distracting the audience with whistles or other obnoxious sounds.

[3]) This method was first used by E. Meyer and V. Jordan, E.N.T., 12, 213, 1935.

Fig. 1. Sixth Symphony part 1, P. Tschaikowsky, Opus 74, played by the Concertgebouworkest under the direction of Prof. Dr. Willem Mengelberg in the Philips Works' Theatre at Eindhoven. The axis of reference of the registered waves has been arbitrarily chosen so that only differences in sound intensity should be read off; the absolute sound level is unknown.
The same time scale is used as in fig. 3, viz., 8 mm = 1 min.
The upper strip is the record of the sound intensity in the auditorium. The figures given under this record are the numbers of the bars at characteristic points. The letters below the numbers represent:

A Noise in auditorium during the interval.
B Applause as the conductor enters.
C Beginning of orchestral music.
D Start of allegro vivo passage which has its greatest contrasts.

The times of its various passages are also given. The lower strip is the record of the same bars when reproduced with the Philips Miller apparatus. Below is the score in which the arrow points to the place corresponding to *D* in the records.

A SIMPLE ELECTRICAL MEASURING BRIDGE

Summary. In this article are described the arrangement and applications of a simple electrical measuring bridge designed for an alternating current supply, the "Philoscop, type G.M. 4140". The indicator consists of a five-electrode valve with a simple form of cathode-ray tube or "electron-ray" tuning indicator. With this bridge, resistances from 0.1 ohm to 10 megohms and capacities from 1 $\mu\mu$F to 10 μF, as well as self-inductances and composite impedances, can be measured. The apparatus is also suitable for other purposes, such as the tracing of interference, the testing of faulty insulation and for the measurement of low voltages.

Introduction

A simple measuring apparatus enabling rapid and reasonably accurate measurements of resistances, capacities and self-inductances to be made is frequently required in electrical work. The apparatus described in this article is a Wheatstone bridge designed for direct connection to an A.C. mains supply and suitable for the direct measurement of resistances between 0.1 ohm and 10 megohms and of capacities between 1 $\mu\mu$F and 10 μF. The ratio between the highest and lowest values which can be measured with this apparatus is $10^8 : 1$ in the case of resistances and $10^7 : 1$ with capacities. Furthermore by connecting suitable external resistance, capacity and self-inductance standards or combinations of these to the bridge, a large number of comparative measurements can be made on the basis of a bridge circuit.

In designing this apparatus, greater importance was attached to obtaining the widest possible range of applications for the bridge, than to a high degree of accuracy. The average accuracy with direct readings is just over one per cent, while for the comparative measurements referred to it is one per thousand. This relatively low accuracy is, however, adequately made up for by the rapidity with which measurements can be carried out, both in the high and low ranges, and by the fact that the apparatus does not require a battery as it is fed directly from the mains.

Sensitivity of the Measuring Bridge

One of the most interesting components of the apparatus is the indicating arrangement, which has a very high impedance and thus absorbs practically no current. In consequence, the "sensitivity" of measurement, i.e. the deflection of the instrument when the contact on the slide wire is displaced by a definite amount from the position of balance, is independent of the magnitude of the resistance under measurement. Usually the resistance of the indicating galvanometer is not higher than the bridge resistances, so that the sensitivity of measurement decreases rapidly as the resistance values increase.

This may be explained more clearly with the aid of the fundamental circuit of the Whaetstone bridge shown in *fig. 1.* If the difference in potential

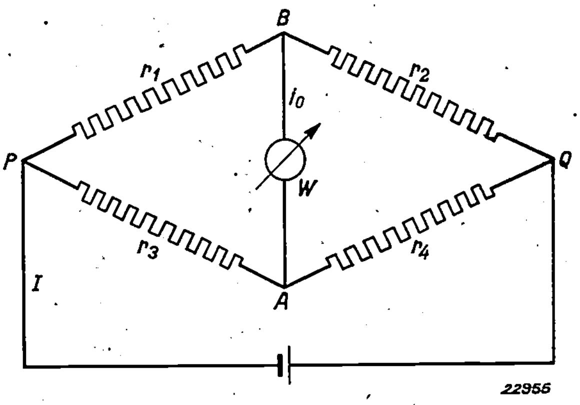


Fig. 1. Fundamental circuit of the Wheatstone bridge. $r_1 \dots r_4$ are the bridge resistances. w is the resistance of the galvanometer circuit.

between P and Q is equal to V_P-V_Q, then from Kirchhoff's laws, according to which the algebraic sum of the current intensities meeting at any point of a conducting network is zero and also in each closed circuit the algebraic sum of the products of the currents and the resistances is equal to the sum of the electromotive forces, we can calculate that:

$$V_A - V_B = \frac{(V_P - V_Q)\, w\, (r_2 r_3 - r_1 r_4)}{w\,(r_1 + r_2)(r_3 + r_4) + r_1 r_2 (r_3 + r_4) + r_3 r_4 (r_1 + r_2)}$$

The bridge is balanced, when $r_2 r_3 - r_1 r_4 = 0$; the difference in potential $V_A - V_B$ is then zero. In most cases, and also in the present apparatus, two of the bridge resistances, which are connected in series, e.g. r_1 and r_2, are made in the form of a slide wire. Their sum is therefore constant, but by displacing the point of contact B along the wire their ratio can be altered, until $V_A - V_B = 0$. r_4 is then the resistance required and r_3 the known resistance, and in the balanced condition we have:

$$r_4 = r_3 \cdot \frac{r_2}{r_1}.$$

If the bridge is brought out of balance, for instance by increasing r_1 by a small amount Δr, which at the same time decreases r_2 by Δr, the alteration in potential between A and B will be:

$$\Delta\,(V_A - V_B) = -\frac{(V_P - V_Q)\cdot\Delta r}{r_1 + r_2}\cdot\frac{1}{1 + \dfrac{r_1\,r_2}{r_1 + r_2}\cdot\dfrac{1}{w} + \dfrac{r_3\,r_4}{r_3 + r_4}\cdot\dfrac{1}{w}}.$$

The resistance of the slide wire $r_1 + r_2$ is usually small compared with the resistance w of the galvanometer which is used to indicate when the bridge current is zero. Most modern galvanometers are moving-coil instruments with a resistance not exceeding 1000 ohms, the resistance of the slide wire is usually however much lower, so that $\dfrac{r_1\,r_2}{w\,(r_1 + r_2)}$ is small and may be neglected compared with unity. It is evident that the other term in the denominator $\dfrac{r_3\,r_4}{(r_3 + r_4)\,w}$ can not be neglected, especially when measuring high resistances, as it determines the sensitivity in these measurements.

The instrument however used in the bridge described here has a *very high* resistance w, in fact higher than the maximum resistance to be measured: $\Delta\,(V_A - V_B)$, which is proportional to the reading obtained is hence:

$$\Delta\,(V_A - V_B) = -\frac{(V_P - V_Q)\,\Delta r}{r_1 + r_2},$$

and is therefore constant and independent of the value of the resistance under measurement. This high resistance, in conjunction with an adequate voltage sensitivity of the galvanometer, has been realised by connecting the points A and B to the cathode and the grid of an amplifying valve, as will be described in more detail when dealing with the indicating component.

The Bridge

If resistance of 0.1 ohm and 10^6 ohms are measured with the same bridge incorporating a single standard resistance of, for example, 1000 ohms, by merely altering the slide-wire ratio, this ratio will be small for 0.1 ohm and high for 10^6 ohms. These maximum and minimum ratios are, however, not accurately ascertainable, because at a given absolute accuracy in reading for both arms of the slide wire the percentage accuracy with which the smaller arm can be ascertained is much less than that in the case of the longer arm. A ratio which is either too high or too low is therefore undesirable. In practice the slide wire is used only for ratios between 0.1 and 10. This limitation in the ratio is obtained by connecting in series with each end of the slide wire, which in this case has a resistance of 1000 ohms, a resistance R which restricts the range of adjustment. This arrangement is shown diagrammatically in *fig. 2*. The resistance $r_1 + r_2$

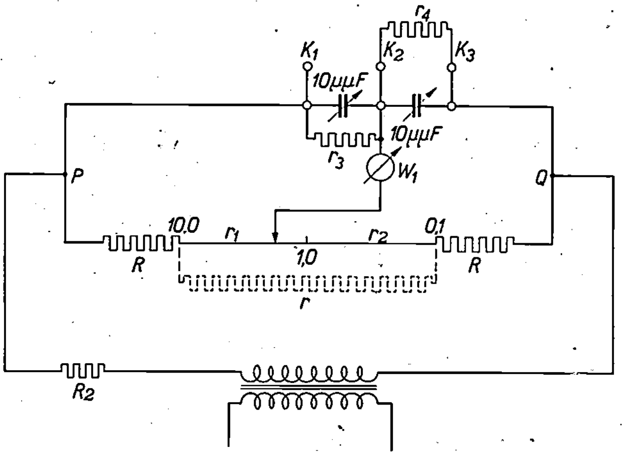


Fig. 2. Arrangement of the bridge with a slide wire in place of the resistances r_1 and r_2. The unknown resistances are connected across K_2 und K_3. The two resistances R in series with the slide wire limit the ratio between the arms of the slide wire to a range from 0.1 to 10. If r is shunted across the slide wire, this range can be reduced to form 0.8 to 1.25 (percentage scale), thus raising the accuracy. The variable capacities between K_1 - K_2 and K_2 - K_3 are used for adjusting the capacity of the wiring to 10 $\mu\mu$F. The transformator winding for feeding the bridge and the series resistance R_2 for protecting the winding also are shown.

of the slide wire and the series resistances R have such ratings that $\dfrac{R}{r_1 + r_2 + R} = 0.1$. It is thus not possible to measure resistance below 0.1 r_3 or above 10 r_3. Four resistances have therefore been incorporated in the bridge for r_3, which differ from each other in steps of 100. In this way, the interval from 0.1 ohm to 10 megohms is covered by four measuring ranges. The knob A_1 (*fig. 3*), which can be turned into ten different positions, is used for changing from one measuring range to another: four of these settings are used for the measurement of resistances (*fig. 4*). The position of the contact on the slide wire is adjusted and read off on the outer circumference of the circular scale R_1 (fig. 3), for which purpose the scale circle has been subdivided from 0.1 to 10, with the point 1 at which $r_1 = r_2$ situated at the centre. The division of the scale allows a direct reading of the ratio $(r_1 + R) : (r_2 + R)$.

It is evident from the above that a higher accuracy can be obtained if the ratio $(r_1 + R) : (r_2 + R)$ is not extended over a range from 0.1 to 10, but is restricted to a narrower range. This may be done by a resistance r which is shunted across the slide wire in a particular setting of the knob A_1. This

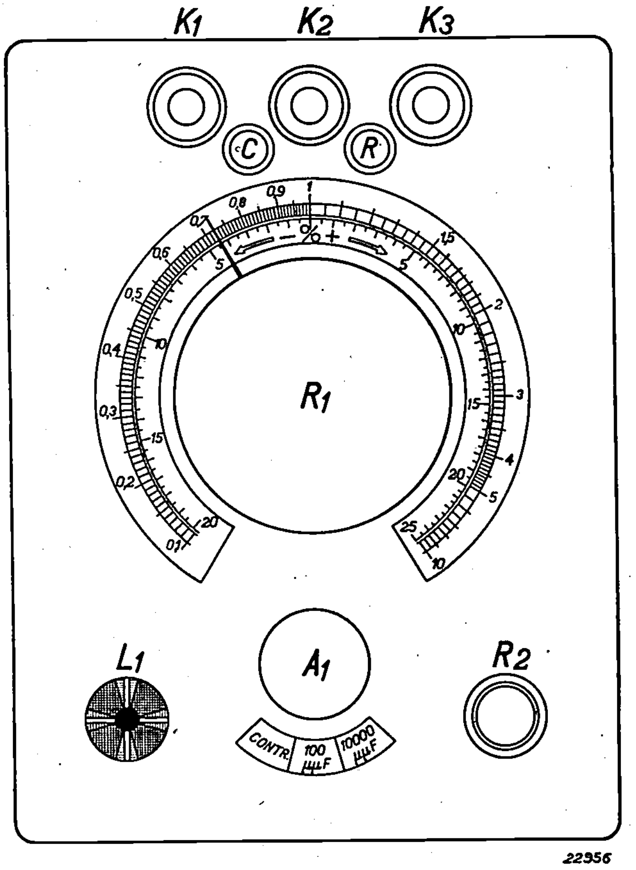


Fig. 3. Plan of apparatus. R_1 has a percentage scale and a decimal scale from 0.1 to 10, as shown. The knob A_1 serves for selection of ten different bridge settings (see also fig. 4). The "electron ray" indicator L_1 is viewed through a window. R_2 regulates the sensitivity by means of the grid leak of the "electron ray" indicator, K_1, K_2 und K_3 are terminals for connecting the resistances R and the capacities C. The earth terminal is not visible being on one of the sides.

resistance shunt is shown by broken lines in fig. 2. The resistance $\frac{(r_1 + r_2)\, r}{r_1 + r_2 + r} < r_1 + r_2$ is then connected across the ends of the slide wire between which regulation is possible. The resistance r is so adjusted that a complete rotation of the knob R_1 over the scale corresponds to a range of ratios from 0.8 to 1.25.

This higher degree of accuracy is used for measuring the percentage deviation of a resistance from a standard of nearly the same value. The corresponding percentage scale is marked along the inner circumference of the circular scale (see fig. 3). In this arrangement the bridge contains no standard resistance, so that this together with the unknown resistance, has to be connected up externally. This setting is termed the "open bridge" setting.

In another position of A_1 two exactly equal resistances are connected to the bridge in place of r_3 and r_4. In this test setting, R_1 should read 1.0 when the bridge is balanced. If this is not the case,

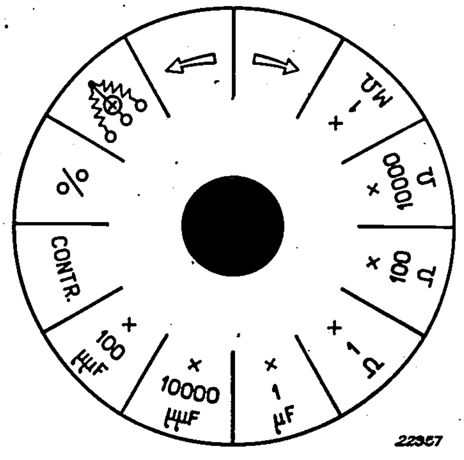


Fig. 4. Scale A_1, which turns together with its knob.

the adjustment can be corrected by slightly displacing the knob R_1 on its spindle. Readjustment of the bridge is, however, rarely necessary.

Since the bridge is connected to an A.C. supply, it is also suitable for *capacity measurements*, when r_3 and r_4 are replaced by capacities C_3 and C_4. The impedance of a capacity C is $1/j\,\omega\, C$ and the condition of balance of the bridge when using resistances: $r_2\, r_3 - r_1\, r_4 = 0$ then becomes $r_2\, C_3 - r_1\, C_3 = 0$ or $C_3 = r_2/r_1\, C_4$. To enable the same scale to be used as with resistance measurements, the standard capacity is connected up in place of the unknown resistance and *vice versa*.

For capacity measurements, the knob A_1 has three further positions for connecting up, as required, one of three capacities with ratios of $1 : 10^2 : 10^4$ to the bridge in place of r_4. The unknown capacity is connected to the bridge at two terminals which link up with a point in the bridge to which r_3 was connected in resistance measurements.

If the capacity under measurement C_3 is not a pure capacity, the currents flowing through C_3 and C_4 will not be in phase, which is in fact usually the case. The phase displacement of the voltage against the current is not strictly equal to —90 deg with condensers owing to dielectric losses, and in resistances is not usually exactly zero owing to their capacity or self-inductance. The condition $r_2\, C_4 - r_1\, C_3 = 0$ is then alone not sufficient; a second condition must also be fulfilled, viz., $\varphi_1 - \varphi_2 = \varphi_3 - \varphi_4$ between the phase angles of the four branches of the A.C. bridge, where φ_1 and φ_2 are the phase angles of r_1 and r_2, and φ_3 and φ_4 those of C_3 and C_4.

This result is readily obtained by writing:

$$r_1 = r_{10}\, e^{-i\varphi_1};\ r_2 = r_{20}\, e^{-i\varphi_2};\ r_3 = r_{30}\, e^{-i\varphi_3}$$
$$\text{and } r_4 = r_{40}\, e^{-i\varphi_4};$$

where $r_{10} \dots r_{40}$ are the moduli of $r_1 \dots r_4$ and $\varphi \dots \varphi_4$ are the phase angles. The condition of balance of the bridge $r_1 r_4 - r_2 r_3$ or

$$r_{10}\, r_{40}\; e^{-i(\varphi_1 + \varphi_4)} = r_{20}\, r_{30}\; e^{-i(\varphi_2 + \varphi_3)},$$

is then satisfied when:

$$r_{10}\, r_{40} - r_{20}\, r_{30} = 0$$
$$\text{and } \varphi_1 - \varphi_2 = \varphi_3 - \varphi_4.$$

Still another difference occurs between resistance and capacity measurements, due to the capacity of the bridge wiring. This capacity, which during measurements is in parallel with the standard and the unknown capacities, is adjusted to 10 μμF by small regulating condensers inserted between the terminals. The capacity of the wiring must be taken into account when calibrating the standard capacity and 10 μμF always be subtracted from the result obtained.

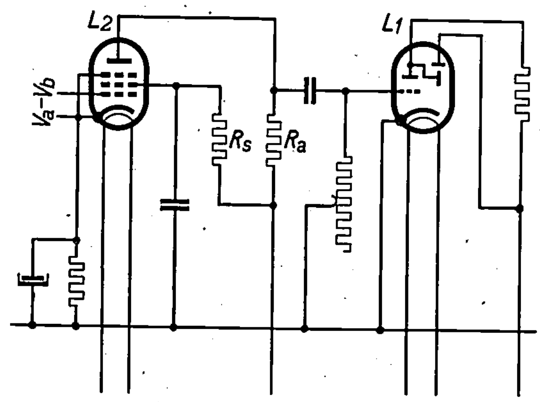
The capacity of the wiring is employed to useful purpose in measuring small capacities between 0 and 90 μμF. In the open bridge setting referred to above, R_1 is adjusted to the point 1 on the R_1 scale when no resistances are connected to the bridge The load on the two arms is then 10 μμF. An unknown condenser C_x will then give a value of 10 μμF + C_x when measuring for C_3. The limits between which C_x can be measured are determined by the condition: $10\,\mu\mu F < 10\,\mu\mu F + C_x < 100\,\mu\mu F$. The range between 0 and 90 μμF obtained in this way thus covers only the right half of the scale.

The Indicating Instrument

The indicating instrument constitutes the principal difference between the measuring apparatus described here and other standard types of Wheatstone bridge. As stated above, an amplifying valve, viz., the five-electrode valve EF 6 (L_2 in *fig. 5*), is used here in place of the usual galvanometer or telephone. The amplified voltage is indicated by the electron ray tuning indicator L_1. The bridge voltage between A and B in fig. 1 is applied across the cathode and the first grid of the penthode. The indicating instrument is thus made up of the penthode and the tuning indicator. The input impedance of the penthode is hence incorporated in the bridge, and is given by:

$$C_g = C_{kg} + C_{ag} \cdot \frac{gR_u}{R_i + R_u};$$

where R_i and R_u are the internal and external resistances of the five-electrode valve, g is the amplification factor, and C_{kg} and C_{ag} are the grid-cathode and the grid-anode capacities respectively.



22969

Fig. 5. Circuit diagram of the electron ray indicator L_1 with the pentode L_2, which acts as an amplifier. The bridge potential across A and B (see fig. 1) is connected to V_a- V_b. R_s supplies the screen grid voltage for the five-electrode valve. R_a is the anode resistance in the anode circuit of the five-electrode valve.

Thus C_g is approximately 5 μμF and $1/\omega C_g$ over 10^8 ohms at mains frequency. It can thus be claimed that the instrument has an adequately high internal resistance which satisfies the conditions set forth above.

Resistances with a blocking condenser are used for coupling the pentode and the tuning indicator in the usual way. The characteristics and method of operation of the tuning indicator are described in detail below.

In this valve a vertical, indirectly-heated cathode (*fig. 6*) emits electrons. The intensity of the electron stream can be controlled by a grid in the usual way, which in this design of valve surrounds only the lower part of the cathode. That part of the current which can be regulated in this way strikes an anode of the same length as the grid and assembled concentrically with the latter. A resistance of 2 megohms is connected to the anode, a constant voltage being applied across valve and resistance. A change in the p.d. between the cathode and grid therefore alters the voltage drop across this resistance.

The top of the cathode is surrounded by a second anode shaped like the generating surface of an inverted conical frustrum. The cone is co-axial with the cathode. From above a view is obtained of the inner surface of the cone, which is coated with a material which fluoresces under the impact of the electrons. The cone is joined directly to the anode voltage supply. In the field between the cathode and the cone, four vertical wires are arranged symmetrically and parallel to the cathode; two of these wires (B) are shown in fig. 6. They are connected to the anode a inside the valve, and when a voltage is applied to them, produce a controllable variation of the field distribution between the cathode and the conical frustrum.

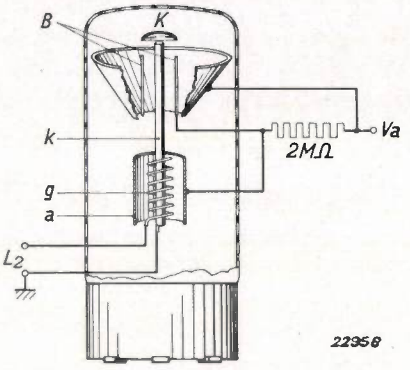


Fig. 6. Diagrammatic sketch of the electron ray tuning indicator. k is the common cathode for the three-electrode component (below) and the indicator component (at the top). a and g - Anode and grid of the three-electrode component. L_2 - Connection of five-electrode valve. B - Wires whose potential modifies the field distribution between the cathode and the cone. The luminous cross is produced by fluorescence on the inner side of the cone. K - Cap for screening the direct cathode light. The four wires B, of which one pair only is shown, are connected inside the tube to a, the anode of the three-electrode component. The voltage drop accross the 2-megohm resistance determines the voltage applied to the wires B.

If now, by altering the bridge voltage which is amplified by L_2 the potential difference between the cathode and the grid is altered, the resulting change in anode current in the lower system will produce a variable voltage drop in the 2-megohm resistance. This will alter the voltage applied to the vertical wires, so that the arms of the luminous cross, which is produced on the inner side of the cone by the modification of the field due to the vertical wires, will become broader or narrower. A variable broadening and constriction is thus superposed on the mean width at the mains frequency, but these rapid changes cannot be followed by the eye, so that the maximum width of the luminous arms is all that is visible. This maximum width, which subtends an angle Θ represents the reading of the tuning indicator. If no voltage is applied to the pentode, this angle will be a minimum. Adjustment is therefore always made to the minimum width in all cases in which the voltage has to be balanced to zero. The grid resistance of the indicator is controllable, and this alows the sensitivity, i.e. the variation of the angle Θ, to be regulated for a given change in bridge voltage.

Current Supply

A transformer of the usual type, which has separate secondary windings, furnishes the current necessary for the bridge, the anode voltages and the requisite heating voltage for the cathodes of the pentode, the tuning indicator and the rectifiers, the latter furnishing the anode voltage by double-wave rectification. A tapping is provided at the centre of the anode winding, while a filter circuit smoothes the anode voltage. A resistance in the cathode lead ensures that the grid of the five-electrode valve has an adequate negative bias with reference to the cathode. The bridge is supplied with 2 volts, applied through a resistance to protect the transformer windings in case of a short of the bridge.

The supply transformer can be changed over to one of two voltage ranges: 100—150 volts and 170—230 volts. Although the apparatus is designed principally for mains connection, the transformer can be used over a frequency range of 40 to 10 000 c/s.

Procedure of Measurements

A variety of measurements can be made with the apparatus described here and shown in *fig*. 7. The simplest measurements are those on resistances between 0.1 ohm and 10 megohms and on capacities between 1 $\mu\mu$F and 10 μF. In both cases the required measuring range is selected with the switch A_1, by means of which the requisite comparison resistances and capacities can be connected to the bridge. After adjusting to the value corresponding most nearly to the value under measurement, the indicator is adjusted to the minimum width by means of the knob R_1. The product of the readings

Fig. 7. General view of the bridge.

R_1 and A_1 is the value required. To maintain this simple proportionality with both resistances and capacities, the unknown resistance is connected across K_1 and K_2 and the unknown capacity across K_2 and K_3. In capacity measurements, 10 $\mu\mu$F for the wiring must then be subtracted from the reading made.

Frequently, the resistances under measurement are subject to certain tolerances, and it is useful to express the difference between a resistance, capacity or self-inductance conforming with certain specifications, as a percentage of the rating of another component of the same design. These measurements are provided for by the percentage scale of R_1. A_1 is then turned to setting $^0/_0$, and with this open bridge setting in which the resistance r is in parallel with the slide wire, the standard resistance, capacity or self-inductance is connected across K_1 and K_2 and the unknown component across K_2 and K_3. The bridge is then balanced with the aid of R_1, and differences from —20 to +25 per cent can be directly read off on the percentage scale.

With the knob A_1 in position *1*, a resistance, capacity or self-inductance can be similarly measured within a range of $1/10\,x$ and $10\,x$ when using an external standard x. This position of A_1 also corresponds to an open bridge, but in which no resistance is shunted across the slide wire. The standard and unknown component are again connected across K_1 - K_2 and K_2 - K_3. The value of the standard multiplied by the reading on R_1 at which the bridge is balanced is the required impedance.

The various settings of the bridge obtained with A_1 give the following measuring circuits:

1) Open bridge; ratio between the two arms of the slide wire is variable from 0.1 to 10. Comparison standards and unknown components are connected externally.

2) Open bridge ($^0/_0$), with increased sensitivity, i.e. the ratio range of the two arms of the slide wire is reduced to 0.8 to 1.25; percentage scale; external connection of the comparison standard and the unknown component.

3) Testing position for readjustment of the bridge.

4,5 and 6) Capacity measurements with three capacities from 10 $\mu\mu$F to 10 μF incorporated in the bridge. The capacity under measurement is connected externally.

7, 8, 9 and 10) Resistance measurements with four resistances incorporated in the bridge. Four-stage measuring range from 0.1 ohm to 10 megohms. The resistance under measurement is connected externally.

The amplifier with tuning indicator incorporated in the apparatus can also be used as a separate

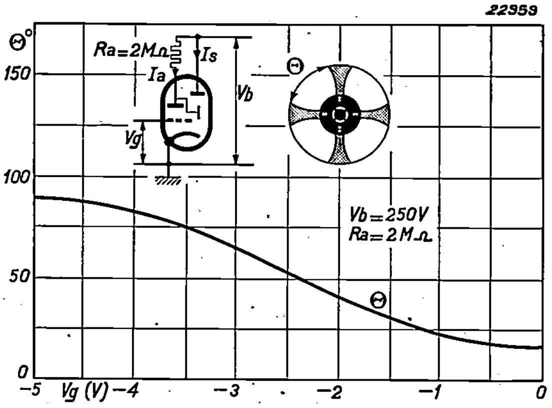


Fig. 8. Angle Θ between two arms of the luminous cross plotted against V_g between the cathode and grid of the three-electrode component of the tuning indicator. The maximum sensitivity is 25 degrees per volt. With a 2000-fold amplification in the five-electrode valve, the sensitivity of the combination L_1 and L_2 is 0.02 millivolt per degree.

voltage indicator. For this measurement the two points whose difference in potential is to be measured are connected across K_2 and the earth terminal. The broadening of the cathode ray tuning indicator is then a measure of the applied voltage. Similarly the apparatus can be used for the location of interference sources, while the output voltage of amplifiers or radio receivers can also be measured. In *fig. 8* the grid voltage of the tuning indicator is plotted against the angle Θ between the boundaries of the electron star, when the instrument is used for this purpose.

Contributed by P. G. CATH.

EQUIVALENT NETWORKS WITH HIGHLY-SATURATED IRON CORES WITH SPECIAL REFERENCE TO THEIR USE IN THE DESIGN OF STABILISERS [1])

By H. A. W. KLINKHAMER.

Summary. A design method is evolved which enables electrical equipment to be designed whose action depends on the characteristic shape of the *BH* curve of the iron in a choke coil or in a transformer with a closed magnetic circuit, without the need for complex calculations or making a large number of experimental models.

A characteristic or an oscillograph record obtained with a single specimen of the type of unit in question is shown to apply to similar designs but based on different scale, and such curve can, therefore, furnish valuable data of the properties of a large number of other designs.

A general proof of the theorem is given on the basis of arbitrary networks which contain, apart from the transformer referred to above, only constant magnitudes (resistances, capacities, self and mutual inductance coefficients, and constants of rotating commutator motors and amplifying valves). The network may also contain rectifying valves. The conditions of operation of large electrical equipment can, therefore, be studied on small-scale specimens, so that a single testing equipment can with advantage be devised for this purpose with which the most varied types of designs can be investigated.

To illustrate the general analysis of the problem, a stabliser for compensating fluctuations in mains voltage is discussed. A number of formulae are deduced for this device in which only the values of the required input and output voltages and the current intensity have to be substituted in order to obtain all the constructional data necessary.

Introduction

Devices whose action depends on the deviations from linearity in the relationship between the magnetic induction B and the field intensity H of iron have been used for a great variety of purposes from the very beginnings of electrical engineering and, in fact, form the subject of a large number of patents. Arrangements for frequency conversion and for controlling choking coils by means of initial magnetisation with direct current, and stabilisers for compensating fluctuations in mains voltage are but a few typical examples of the many devices in this group. There can be little doubt that these devices would be in more general employment if their use were not hampered by two difficulties:

Firstly, little definite information is usually available of the form of the *BH* curve, for the standard delivery specifications relating to transformer and dynamo sheets require only one and at most two points on this curve to be given and then only in reference to the lower limit.

Secondly, advance calculations are very difficult. Since sinusoidal currents are not under consideration, A.C. vectors can only be employed to give a rough approximation. Moreover, the actual performance frequently differs from that previously deduced theoretically by means of A.C. vector analysis. Analysis fails to provide a more accurate treatment applicable to practical conditions, in spite of many attempts which have been made in this direction.

The first difficulty may be overcome by the determination of the characteristics of each piece of equipment separately, for instance by altering the height of the core, although this method is impracticable for mass production. Moreover, the equipment then acquires larger dimensions and is more costly to construct, since in every direction the unfavourable deviation of the *BH* curve has to be taken into consideration.

The second difficulty need not hamper the manufacturer, for to draft a design of a particular piece of equipment it is by no means necessary to make a large number of complex calculations; a design can be based entirely on the results of direct measurement, without having to build a series of experimental models of the particular equipment to be designed.

How this can be done forms the subject of the present article.

Equivalent Designs

To analyse the problem generally, consider an arbitrary network of which one branch contains a choke coil with a closed iron core. We will limit

[1]) This article is a summary of the contribution of the author: "Over het ontwerpen van inrichtingen met verzadigde ijzerkern" to the Prof. Feldmann memorial volume: Electrotechnische Opstellen (Waltman, Delft, 1937).

our discussion to cores of a shape where the cross-sectional area of the core is the same throughout the magnetic circuit.

Let the various constants of the network be as follows:

N Number of turns on the core.
l Length of iron circuit.
F Cross-sectional area of iron.
R, L, C Resistance, self-inductance and capacity.

Let the variables of the network be as follows:

t Time.
Q Charge or quantity of electricity passed.
$I = dQ/dt$, current, for $t = 0$, $Q = Q_0$ and $I = I_0$.
U External e.m.f.
E Reactive e.m.f. of the elements forming the network.
B Inductance of the iron core.
H Field intensity of the core.

Starting from given initial conditions such as:

$$Q = Q_0 \text{ and } I = I_0$$

the values of the variables will be determined completely by the following set of equations:

For every branch except that containing an iron choke coil

$$\begin{cases} U = U(t), & (1) \\ I = \dfrac{dQ}{dt}, & (2) \\ -E = L\dfrac{dI}{dt} + RI + \dfrac{Q}{C} & (3) \end{cases}$$

For the iron-cored circuit

$$\begin{cases} -E = NF\dfrac{dB}{dt}, & (4) \\ 4\pi NI = Hl, & (5) \\ B = f(H), & (6) \end{cases}$$

For each mesh $\Sigma(E + U) = 0$, (7)

For each junction point $\Sigma I = 0$ (8)

The function f in equation (6) is determined by the BH curve of the material of the core, which may be assumed to be known. The functions $U(t)$ which represent the external e.m.f.'s in terms of the time are also known.

Make the following substitutions in equations (1) to (8):

$$t = \frac{1}{k}t', \qquad (9)$$

$$E = k\frac{F N}{F'N'}E', \qquad (10)$$

$$U = k\frac{F N}{F'N'}U', \qquad (11)$$

$$Q = \frac{1}{k}\frac{N' l}{N l'}Q', \qquad (12)$$

$$I = \frac{N' l}{N l'}I', \qquad (13)$$

$$L = \frac{F l'}{F' l}\left(\frac{N}{N'}\right)^2 L', \qquad (14)$$

$$R = k\frac{F l'}{F' l}\left(\frac{N}{N'}\right)^2 R', \qquad (15)$$

$$C = \frac{1}{k^2}\frac{F' l}{F l'}\left(\frac{N'}{N}\right)^2 C', \qquad (16)$$

$$B = B', \qquad (17)$$

$$H = H'. \qquad (18)$$

F', l', N' and k are here constants which can have arbitrary values. According to equations (9) to (18), the new magnitudes t', E', U', Q', f', L', R', C', B' and H' differ from the original magnitudes only by constant factors.

As a result of the above substitutions, equations (1) to (8) become:

For each branch except that containing the iron core

$$\begin{cases} U' = \dfrac{1}{k}\dfrac{F'N'}{F N}U\left(\dfrac{t'}{k}\right) = U'(t'); & (1a) \\ I' = \dfrac{dQ'}{dt'}, & (2a) \\ -E' = L'\dfrac{dI'}{dt'} + R'I' + \dfrac{Q'}{C'}, & (3a) \end{cases}$$

For the iron cored circuit

$$\begin{cases} -E' = N' F'\dfrac{dB'}{dt'}, & (4a) \\ 4\pi N' I' = H'l' & (5a) \\ B' = f(H'), & (6a) \end{cases}$$

For each mesh $\Sigma(E' + U') = 0$, (7a)

For each junction point $\Sigma I' = 0$ (8a)

The only difference between this new set of equations and the original set is in the addition of an accent to all letters, except to the functional symbol f in equation (6).

The new equations describe the distribution of current and voltage in another network composed of the same circuit and the same core material, but with different network constants (R', L' and C' in place of R, L and C), with other core constants (l', F' and N' in place of l, F and N) and with other external e.m.f's ($U' = U'(t')$ in place of $U = U(t)$). The only factor which has not been changed by this transformation is the magnetic state of the core material at corresponding times.

Since the new set of equations has been derived with the aid of equations (9) to (18), the proportionality between the original and new variables as determined by equations (9), (10), (12) and (13) is in all circumstances maintained. We can, therefore, derive the unknowns of the new set of equations from those in the original set by using equations (9), (10), (12) and (13), provided the unknows are known for these initial equations, e.g. as a result of measurement.

The fact that the magnitudes in the original and new sets of equations are connected by equations (9) to (18) may be briefly described by saying that the new set is "equivalent" to the original set.

Since the variations of current and voltage are completely determined by the differential equations together with the initial conditions, it is sufficient to obtain complete equivalence between the two sets of conditions that:

1. The new network constants are connected with the original constants by equations (14), (15) and (16);
2. The new initial currents and charges are connected to the original ones by equations (12) and (13);
3. The new external e.m.f's are connected with the original values by equations (11) and (1a).

The relationship determined by the transformation equations will then be maintained at equivalent times, which implies that the variables in the one case are equivalent to those of the other case except for the constant reduction factors according to equations (9) and (13).

It follows, furthermore, from (10), (16), and (13), (14) that:

$$\frac{{}^1\!/_2\, C'\, E'^2}{{}^1\!/_2\, C\; E^2} = \frac{{}^1\!/_2\, L'\, I'^2}{{}^1\!/_2\, L\; I^2} = \frac{l'\; F'}{l\;\; F},$$

i.e. corresponding energies are in the same ratios as the volumes of the iron core belonging to the network.

As the BH curve does not lend itself to analytical treatment, the properties of the circuit must be represented by curves or characteristics. Where the characteristics have to be determined for an iron core not yet made, it is sufficient to insert any other core (with arbitrary l', F' and N') in the equivalent network and to plot the characteristics for this equivalent arrangement. According to equations (9), (11) and (17), the characteristics will then apply also to the design under calculation but will be plotted on a different scale. The oscillograph record obtained will also apply to all cases which are equivalent to the actual case investigated, but again be on a different scale.

If, in particular, the external e.m.f's are given by the expression:

$$U\;(t) = p \sin \omega t$$

then it follows from equation (1a) that the external e.m.f's of equivalent designs will be:

$$U'\;(t') = \frac{\omega'}{\omega}\,\frac{F'N'}{F\;N}\; p \sin \omega' t',$$

where $\omega' = \omega/k$ has been substituted.

No attention need be given to the initial conditions, if only the steady state is under investigation. The conditions arising during switching on have then entirely disappeared.

If $U(t) = 0$, i.e. we are dealing with an independent network, no attention need be given, on the other hand, to the external e.m.f's. According to (1a), $U'\;(t') = 0$ also in this case, independent of the value of k. The time scale will here be a different one according to the value of k, which is inserted after choosing R', C', Q'_0 and I'_0 in accordance with equations (15), (16), (12) and (13).

The above analysis can be readily applied to cores with more than one winding. In the transformation equations, the number of turns on that winding belonging to the same part of the network as the brach under consideration is then inserted for N and N'.

The mutual inductances, which can then be obtained between branches of the various sections of the network or between branches of the same section, can also be included in equation (2).

In the first case the reduction equation becomes:

$$M_{12} = \frac{l'\,F\,N_1\,N_2}{l\,F'N_1'N_2'}\,M_{12}', \quad \cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot \quad (19)$$

where the indices relate to the section under consideration and N denotes the number of turns on the corresponding core windings. The second case is a particular case of the first where $N_1' = N_2'$ and $N_1 = N_2$.

In an analogous manner, the transformation equations may be deduced for the constants R_{12} (here $R_{12} \neq R_{21}$) applying to rotating electrical machinery. If the rotational e.m.f., E_1, which is induced in an armature winding belonging to section 1 as a result of a current I_2 flowing through a field winding belonging to section 2, is given by the expression:

$$E_1 = R_{12} I_2$$

we get for R_{12} a reduction factor which is k times the reduction factor for M_{12} in (19).

Transformation equations can also be deduced in certain circumstances for the constants of amplifying valves.

The characteristics for a series of designs of various sizes and shapes can be reduced to a form permitting a direct comparison, viz., by applying equations (9) to (20) so as to reduce them to the equivalent case with $l' = 1$, $F' = 1$, $N_1' = 1$ and $N_2' = 1$. The arrangement is then converted to an imaginary equivalent one with a cubical iron core of 1 cm length of side, where it may be assumed that the iron circuit is closed by a yoke of infinitely high permeability. Each of the windings of this imaginary iron consists of one turn only. The rating of each condenser and each choke coil of the reduced network is equal to that of the equivalent member of the original network, divided by the volume of the original iron core.

Advance Testing Equipment

It appears worth while to construct a permanent equipment with which all types of equivalent circuits can be set up and tested to assist in arriving at a suitable design of equipment to meet a specific practical purpose. The results of measurement obtained with this equipment can then be converted to the particular case under consideration using the method discussed above. The reduction formulae are, in fact, further simplified by employing this procedure, since l', F', N_1' and N_2' of the testing equipment can be included once and for all in the constant factors.

Furthermore, the transformer of the testing arrangement can be given the most suitable design to simplify measurements as far as practicable. Since, with equivalent networks and when using the same frequency, which should preferably be made equal to 50 cycles, the rating of all branches is proportional to the volume of the iron core, it is advisable to make the core with not too large a volume. This will avoid the voltage, current or power values exceeding the measuring range of the measuring instruments available as well as the need for making the leads too thick or too stiff. Moreover, in this way the ratings of the rheostats, regulating chokes and condensers required in the equipment can be maintained within reasonable limits. The number of turns on the windings can be given any arbitrary value, and should merely be adapted to the characteristics of the instruments.

The materials used for the iron core of the testing arrangement must naturally be the same as that used for the core of the actual piece of equipment being designed. In practice, the same material is, however, always used, viz., high-alloy transformer iron, since its BH curve is extremely suitable for the duties in question [2]), and this material is not too expensive and is also easily obtainable.

A large variety of designs can thus be investigated with a single core of high-alloy transformer iron incorporated in the testing equipment.

[2]) The BH curve of high-alloy transformer iron has in fact a marked turning point, which is necessary for the reliable operation of equipment of this type. As mentioned in the introduction, the BH curves of different consignments of the same sheets are not always exactly similar and this must be taken into consideration in the practical application of the result obtained with small-scale models. In the example discussed below this point is considered.

By using the cathode-ray oscillograph, which is always available for immediate use and requires as little preparation before taking readings as a voltmeter or ammeter, the testing procedure is much simplified, since there is no restriction on the number of turns on the core windings and these can therefore be adapted to the sensitivity of the oscillograph. In consequence, an amplifier is not required in any measurement.

To permit the measurement of non-periodic phenomena with the cathode-ray oscillograph, these can be converted into periodic phenomena, for instance by periodical repetition using a contact disc revolving in synchronism.

With the testing equipment described, the optimum values for all types of designs can be rapidly found for all unknown magnitudes, by taking various combinations of these values in succession and reading off their effect directly on the instruments.

Stabiliser for Compensating fluctuations in mains voltage

A device designed for compensating fluctuations in mains voltage will now be discussed to serve as a concrete example of the general analysis made above.

The circuit diagram of a device of this type is shown in *fig. 1.* On an alteration in the mains

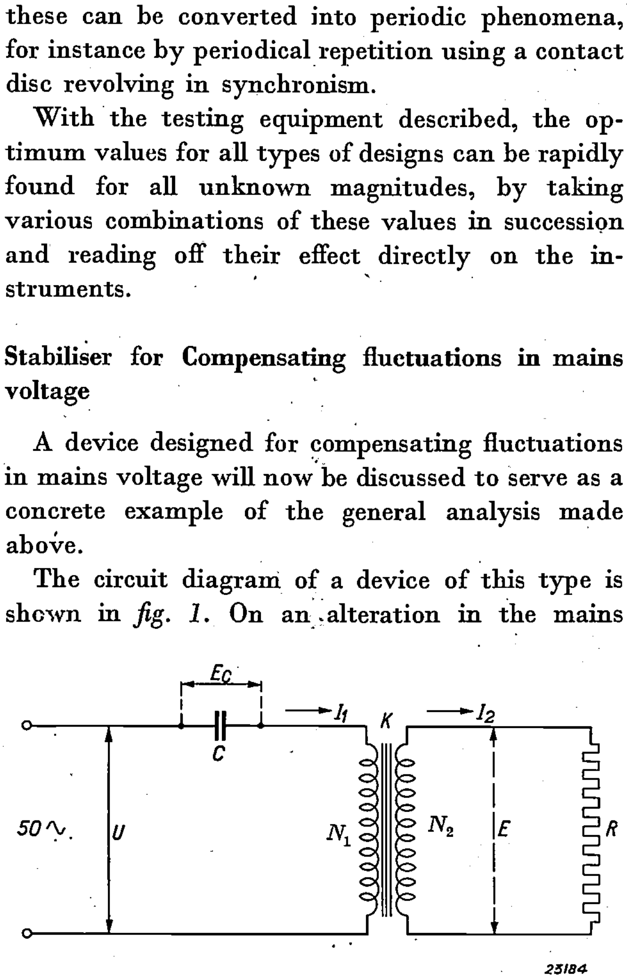


Fig. 1. Circuit diagram of a stabiliser with load resistance. K is the closed iron core; N_1 and N_2 are the number of turns of the coils in the windings. R is the load resistance, and C is a condenser with a potential E_C. U and E are the input and output voltages respectively; I_1 and I_2 are the current intensities.

voltage U, a corresponding change takes place in the primary current, although the magnitude of the iron field remains nearly constant owing to state of saturation of the iron; the same also applies for the output voltage E which is induced by the field in the secondary winding.

The efficiency of the device in relation to its proposed purpose is expressed by the factor of compensation, i.e. the ratio of the percentage variation in the input voltage to that in the output voltage.

For a device of this type, the output voltage E is plotted in *fig. 2* as a function of the input voltage U with constant R.

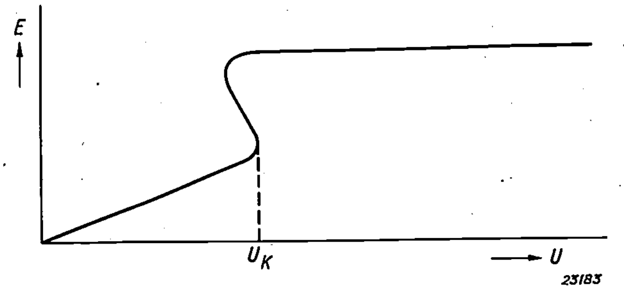


Fig. 2. Characteristic of a voltage stabiliser. U is the effective value of the input voltage; E is the effective value of the output voltage with a constant-resistance load. U_k is the effective value of the output voltage at the lower turning point of the curve. The upper, nearly horizontal, branch of the curve represents the working range.

The device functions as a stabiliser along the nearly horizontal part of the curve. The greater the value of U, the higher must be the ratings of the transformer and the condenser, and hence the higher will be the cost of the equipment. The operating range should, therefore, be located as far to the left as possible, although there is a limit at the point where $U = U_k$, i.e. where the ordinate touches the lower turning point, otherwise the working point could remain fixed in the lower branch of the curve when the mains voltage is applied.

If compensation is to be provided for possible deviations of the BH curve and for fluctuations in mains voltage of 10 per cent in either direction, the working range should be located between about 1.15 U_k and 1.40 U_k, with a nominal mains voltage U of 1.27 U_k.

It is required to arrive at the optimum shape of the E-U curve by selecting a suitable iron core, the optimum number of turns and the best constants of the network, so that the compensation factor will be sufficiently high and the ratings of the condenser and the transformer as low as possible, since these ratings determine the size and cost of the equipment.

The problem appears a very complex one owing to the large number of choices available, and hence the large number of combinations which have to be tested. It appears that tests are required with a series of iron cores of different shapes and sizes, as well as with different numbers of turns in conjunction with a wide range of different capacities.

But with the aid of the general analysis given above it is possible to simplify the problem considerably, because all designs can be referred to the same standard of comparison, viz., by reducing them all to the transformer in the testing equipment. Since as a rule networks of 50 cycles have alone to be considered, the question of a mathematical transformation of frequency can be neglected for the moment and we can hence put $k = 1$. While the original curve is determined by C, R, N_1, N_2, l and F, the number of parameters in the reduced curve has been diminished to two. The circuit shown in fig. 1 shows this to be the case, for if the transformer shown in fig. 1 is always the same with the same values of l', F', N', and N_2', only R' and C' remain variables.

By applying this reduction the problem is thus reduced to the following:

It is required to determine the values of R' and C' such that:

1) With the value of U' lying between 1.15 U'_k and 1.40 U'_k, the factor of compensation shall be sufficiently high, and

2) P_C'/P_R' and P_T'/P_R' shall be as small as possible for $U' = 1.27\ U_k'$. P_C' and P_T' are here the apparent powers of the condenser and the transformer, while P_R' is the watt consumption of the load resistance.

A solution of this problem can be arrived at without calculation, viz., by direct experiments with the testing equipment. For different values of R' and C' the E'-U' curve is plotted and therefrom the values of the three magnitudes in 1) and 2) above are determined. As to be expected, condition 1) is in opposition to that under 2). Hence according to the proposed use of the equipment in question, a compromise must be made between a satisfactory factor of compensation and a reasonably low cost of construction.

When suitable values have finally been selected for R' and C', the corresponding values of I_1', I_2', U', E' and E_C' are found by measurement. Together with the values of U, E and I_2, which are laid down in the specification, the measured values obtained provide all the data required for the design of the equipment. If the power output is, for instance, $E\,I_2 = P_R$ or $E'\,I_2' = P_R'$ respectively, we get:

From equations (10) and (13):

$$F = k_1 \frac{P_R}{l} \quad \ldots \quad (21), \text{ where } k_1 = \frac{l'\,F'}{P_R'} \quad \ldots \quad (21a)$$

From equations (10), (11) and (13)

$$N_1 = k_2 \frac{U\,l}{P_R} \cdots (22), \text{ where } k_2 = \frac{P_R' N_1'}{U'\,l'} \quad (22a)$$

From equation (13):

$$w_2 = k_3 \frac{l}{I_2} \cdots (23), \quad \text{,,} \quad k_3 = \frac{I_2' N_2'}{l'} \cdot (23a)$$

From equations (10), (11), (13) and (16):

$$C = k_4 \frac{P_R}{U^2} \cdots (24), \quad \text{,,} \quad k_4 = \frac{U'^2}{P_R'} C' \quad (24a)$$

From equations (13) and (22):

$$I_1 = k_5 \frac{P_R}{U} \cdots (25), \quad \text{,,} \quad k_5 = \frac{U' I_1'}{P_R'} \cdot (25a)$$

From equation (11):

$$E_C = k_6\, U \cdots (26), \quad \text{,,} \quad k_6 = \frac{E_C'}{U'} \cdots (26a)$$

Since the magnitudes on the right hand side of equations (21a) to (26a) are given by the testing equipment as the optimum combination of values for obtaining a specified factor of compensation, the constants k_1 to k_3 are known for all cases. The working out of the design is thus reduced to arriving at the values of U, P_R and I_2 in equations (21) to (26) to meet the specification. The length of the iron path l can be arbitrarily chosen and, for instance, adapted to the length of the iron path in an available stamping.

If a complete design has not to be worked out, but merely a rough estimate of the cost is required, it is sufficient to determine the apparent power of the condenser $P_C = I_1 E_C$ and the weight of the core $G = 7.8\, lF$. From equations (25), (26) and (21), we then get:

For the apparent power of the condenser:

$$P_C = k_7\, P_R \cdots (27), \text{ where } k_7 = \frac{E_C'\, I_1'}{P_R'} \quad (27a)$$

For the weight of the core in grammes:

$$G = k_8\, P_R \cdots (28), \text{ where } k_8 = \frac{7.8\, l'F'}{P_R'} \quad (28a)$$

These two magnitudes are thus expressed in terms of the rated power P_R which the stabiliser must furnish.

We have limited ourselves to an example in which a resistance and a capacity alone appear as network constants. The advantage of the method of calculation described is, in general, the greater, the less a particular case is susceptible to direct calculation. Thus this method has been found exceptionally useful in cases where pre-magnetised choke coils were used in conjunction with valve tubes.

THE RADIATION INTENSITY OF THE PHILIPS FASTNESS-TO-LIGHT METER

By J. F. H. CUSTERS.

Summary. Comparative measurements of the fading of samples on irradiation with sunlight and in the Philips fastness-to-light meter show that the radiation intensity in the fastness-to-light meter is about 50 times greater than the maximum solar radiation obtainable on a horizontal surface at a latitude of 52 deg.

A new apparatus for testing the fastness-to-light of all kinds of materials has already been described in this review [1]). In that article particular attention was called to the very high intensity of the light to which the sample was exposed and which enabled the fastness-to-light of a material to be determined in a very short time. The present article deals in greater detail with the determination of the ratio of the irradiation times required to obtain the same degree of fading on irradiation with sunlight and with the meter in question. The following experiments were necessary for the determination of this ratio, which has a value of approximately 50:

1) Irradiation for specific periods of time of one or more samples with sunlight and with the light given by the meter.
2) Registration of the amount of sunlight and meter light which is required to obtain a specific degree of fading.
3) Measurement of the fading itself, so as to permit direct comparison of the fading produced by sunlight and in the meter.

During exposure to the sun's rays the samples were placed on a horizontal surface covered with black velvet, T in *fig. 1*. Half of each sample was covered with a blackened metal strip. At a distance of 5 cm from the samples a sheet of ordinary window glass G, 2 mm thick, was placed, so that the air had free access to the samples.

A circular opening, 4 cm in diameter, was cut in the velvet-covered surface through which the sunlight passed into a cylinder, 15 cm wide and 20cm long, whose internal surface was painted a matt-white. Another circular opening, 2 cm in diameter, was cut in the base of the cylinder, through which a considerable part of the sunlight entering the cylinder was transmitted after repeated reflections at the inner white wall. To prevent sunlight falling directly on this aperture a mattwhite circular disc, 4 cm in diameter, was placed horizontally across the middle of the cylinder. A photo-electric cell was located under the base of the cylinder with its window against the lower opening in the cylinder; the cell was carefully screened against all extraneous light other than that reaching it through the opening. The weak electric current flowing through the cell, which is proportional to the intensity of

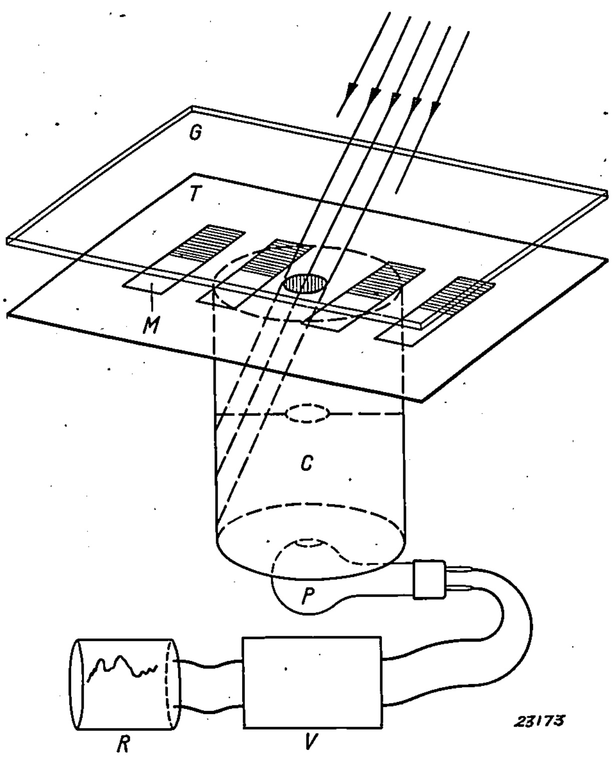


Fig. 1. Arrangement of apparatus for irradiation of different samples with sunlight and for recording the effective amount of solar radiation. T - Table, on which the samples were placed. M - Sample. G - Sheet of window glass. C - Cylinder, painted a dull white in the interior. P - Photo-electric celle. V - Amplifier. R - Recording milliammeter.

the incident light, was amplified about 500 times and then registered by a recording milliammeter with a revolving drum. As in the Ulbricht sphere, this current intensity is a measure of the light passing through the aperture, i.e. of the intensity of the illumination falling on the sample, irrespcctive of the direction of the sun's rays.

An arrangement of the type just described had to be used because if the light falls directly on the cell different current intensities are measured according to the angle of incidence. This is due to the fact that the sensitivity of the light-sensitive surface of the cell is not constant over the whole

[1]) J. F. H. Custers, A new apparatus for testing the fastness-to-light of materials. Philips techn. Rev., 1, 277, 1936. J. F. H. Custers. Ein neuer Lichtechtheitsprüfer. Die Chemische Fabrik 8, 103 - 105, März 1935.

of its area. By inserting the cylinder between the sample and the cell, a uniform illumination is always obtained over the whole surface of the cell, because the sun's rays are diffusely reflected one or many times at the cylinder wall before reaching the cell. Rotation of the apparatus about the vertical axis of the cylinder thus produces no alteration whatsoever in the current intensity through the cell, so that measurements can be made irrespective of the altitude of the sun and also with a moderately cloudy sky. It is evident that the whole of the light from the sky, i.e. from the sun and the sky generally, falls on the sample. A number of cell-current records are reproduced in *fig. 2* which were registered during two sunny days in June, the majority between 9.a.m. and 4 p.m. Record *A* was obtained

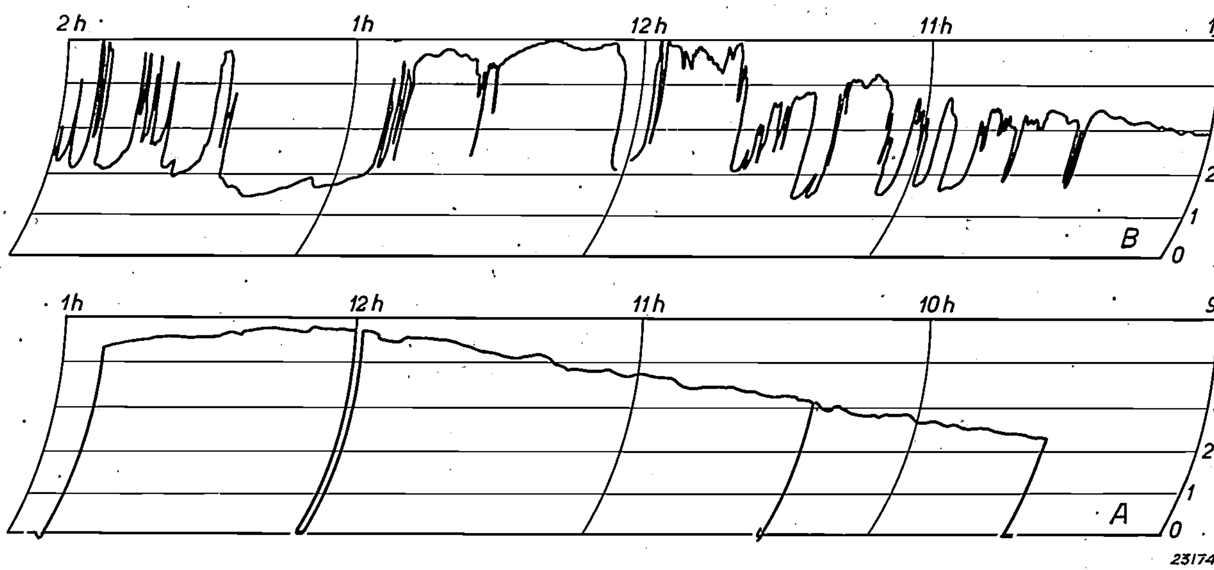


Fig. 2. Records of total radiation falling on unit surface of the sample in unit time. The time was measured in hours. The unit of intensity is arbitrary. *A* was recorded on June 9 and *B* on June 8.

on a unusually fine day with a perfectly clear sky and shows that the intensity of illumination increases progressively, apart from very small fluctuations, and diminishes again after midday. (In this article the intensity of illumination is taken as including both the visible and invisible radiation falling in all directions on unit surface of the sample in unit time). Record *B* was also made on a fine, although cloudy, day, the presence of clouds being shown by the marked fluctuations in the curve. Thc sudden drop in illumination to zero, which occurs at several points of the curve, was due to the top aperture of the cylinder being completely covered for short intervals for the purpose of testing the zero adjustment of the amplifier.

As already stated the deflection on the recorded curves is equal to the intensity of illumination of the samples, multiplied by a constant factor. For the sake of simplicity this factor will be taken as being unity. The area enclosed by the curve and the axis on the one hand and by the ordinates corresponding to two particular times on the other, then gives the amount of radiation which has fallen on unit surface of the samples during the corresponding interval of time; this value can be expressed in ergs per sq cm or in watt-seconds per sq cm.

The samples were irradiated by sunlight on June, 8, 9, 13 and 14, all very sunny days. If the area of each recorded curve is divided by the time of irradiation in hours, the mean intensity of illumination on each of these days is obtained; expressed in arbitrary units this average was 1.43, 1.89, 1.35 and 1.15 on June 8, 9, 13 and 14 respectively and is thus seen to vary considerably [2]). The total radiation falling on the samples on these four days was 27.5 during an aggregate time of irradiation of 19 hrs 35 mins. The maximum intensity of illumination was recorded at 11.45 a.m. on June 9 with an exceptionally clear sky, and was 2.4. With a constant intensity of this maximum value, 27.5/2.4 = 11.5 hrs exposure would have sufficed to produce the same effect as the 19 hrs 35 mins actually taken on the days concerned, i.e. assuming that the effect produced is determined exclusively by the product of the intensity of illumination and the time of irradiation; this is approximately true if the time of irradiation is not too long.

[2]) It should be noted that irradiation did not take place during the same hours on each of these days, so that the average values given are not directly comparable. On the other hand, the values 1.89 and 1.15 are comparable, since on June 9 and 14 irradiation took place roughly at the same hours and for the same length of time.

Similar samples were also irradiated in the fastness-to-light meter. As with this instrument the intensity of illumination is constant, it was sufficient to determine the time of irradiation required to produce a change in colour. As will be shown later, an irradiation time of 15.6 mins is required in this instrument to obtain a fading equivalent to that produced by 11.5 hrs exposure to sunlight, provided that the solar radiation is continuously maintained at its maximum value of 2.4, although this optimum condition cannot be realised in practice. The ratio required is therefore $\frac{11.5 \cdot 60}{15.6} = 45$.

To express the fading of the various samples on a numerical basis, reflection measurements were carried out on both the irradiated and original samples at different wave-lengths in the spectrum. A description of the reflection meter used for determining the reflection coefficients cannot be included here, but it should be stated that the direction of the incident light made an angle of about 30 deg. with the normal to the plane of the sample under examination and that the light which was diffusely-reflected along the normal was measured. A piece of white glazed board was taken as a standard of comparison, and its coefficient of reflection assumed to be unity at each wavelength; all samples were referred to this standard.

The reflection curves for the first four samples (*I*: 0.6 per cent Brilliant wool blue FFR extra; *II*: 0.6 per cent Wool blue N extra (825); *III*: 1 per cent Brilliant indocyanine 6 B; and *IV*: 1.5 per cent Fast wool blue GL (974)) chosen from the series of eight types of fast blues issued by the "Deutsche Echtheits Kommission", are given in *fig. 3*. As to be expected the coefficient of reflection R_λ is greater in the blue than in the orange and red. A number of reflection curves of sample *I* after irradiation for various periods in the fastness-to-light meter are given in *fig. 4*. With increase in the time of irradiation, the reflection becomes progressively more uniform throughout the spectrum, i.e. the colour of the sample changes from a saturated blue to a very unsaturated light blue, and the sample bleaches nearly a pure white. Better curves are obtained when, on the basis of the definition of the blackening of a photographic plate, the socalled "reflection-blackness D_λ" is plotted and which is defined by the expression:

$$D_\lambda = \log 1/R_\lambda.$$

If the coefficient of reflection is small (e.g. 0.01), the reflection blackness will be large (e.g. equal to 2) and *vice versa*. D_λ is regarded as a blackening in reflected light. The reflection-blackness curves for sample *I* are plotted in *fig. 5* and can be calculated directly from fig. 4. While the original sample has a low blackness value in the blue and a high blackness value in the red, the blackness value is much smaller in the red and greater in the blue after 8.5 hrs irradiation. These curves are very important because the diminution of D_λ at the maximum of the reflection-blackness curve is to a first approximation proportional to the time of

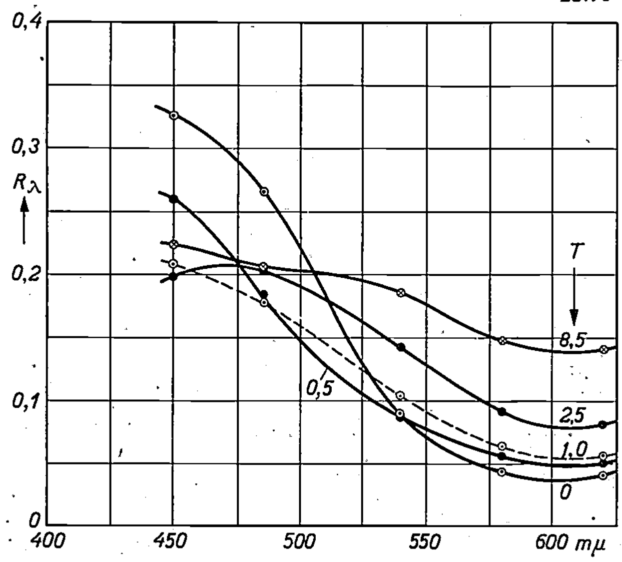


Fig. 4. Reflection curves for sample *I* after irradiation for different times in the fastness-to-light meter. The figures under the letter *T* denote the time of irradiation in hours.

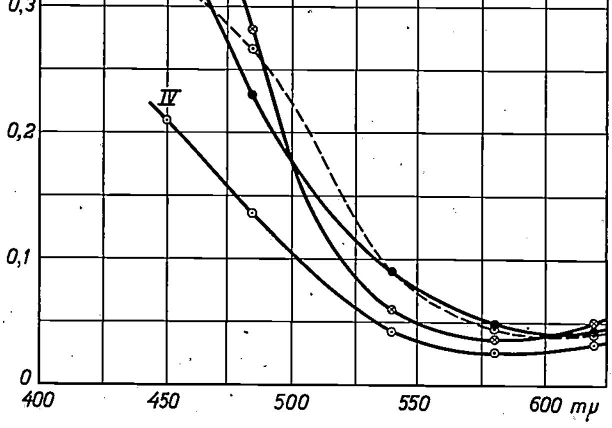


Fig. 3. Reflection curves for four different samples of blue (*I*, *II*, *III* and *IV*).

irradiation, provided the latter is not too long. In *fig. 6* ΔD_λ is plotted against the time of irradiation for a wave-length of 620 mμ, and on this curve we made the determination of the irradiation ratio referred to above.

Fig. 7 gives the blackening curves of the original sample *I* and the curve after exposure of the same sample to sunlight. At 620 mμ, D_λ has dropped from 1.382 to 1.334, so that $\Delta D_\lambda = 0.048$. To obtain this alteration in D_λ, irradiation was required for

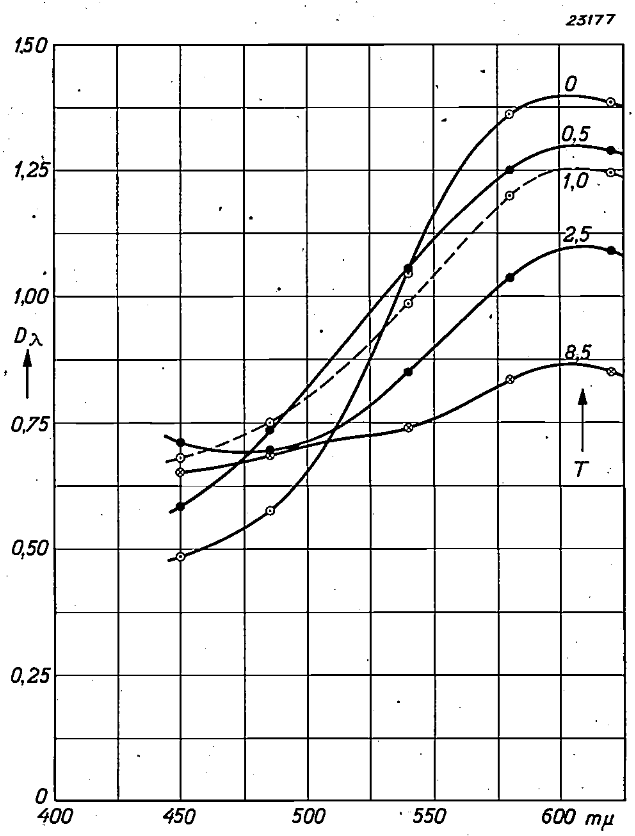


Fig. 5. Reflection-blackening curves for sample *I* for different periods of irradiation in the meter. The figures under the letter *T* denote the time of irradiation in hours.

11.5 hrs at the maximum intensity of 2.4 On irradiation on the fastness-to-light meter D_λ was reduced from 1.382 to 1.290 in 30 minutes, so that $\Delta D_\lambda = 0.092$; thus, in $\frac{0.048}{0.092} \times 30 = 15.6$ mins D_λ has been reduced by 0.048. It follows from this that 15.6 mins irradiation in the meter is equivalent to an irradiation with sunlight at maximum intensity of 11.5 hrs. The irradiation factor for various samples was determined in this way. For some samples it was found to be on the high side and for others on the low side; the average value was 50 with a maximum deviation of $\pm$ 15.

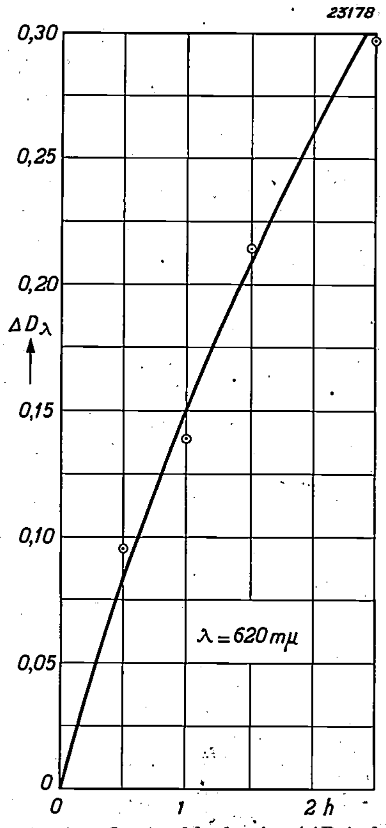


Fig. 6. Reduction in reflection blackening (ΔD_λ) plotted against the time of irradiation in hours at a wave length of 620 mμ.

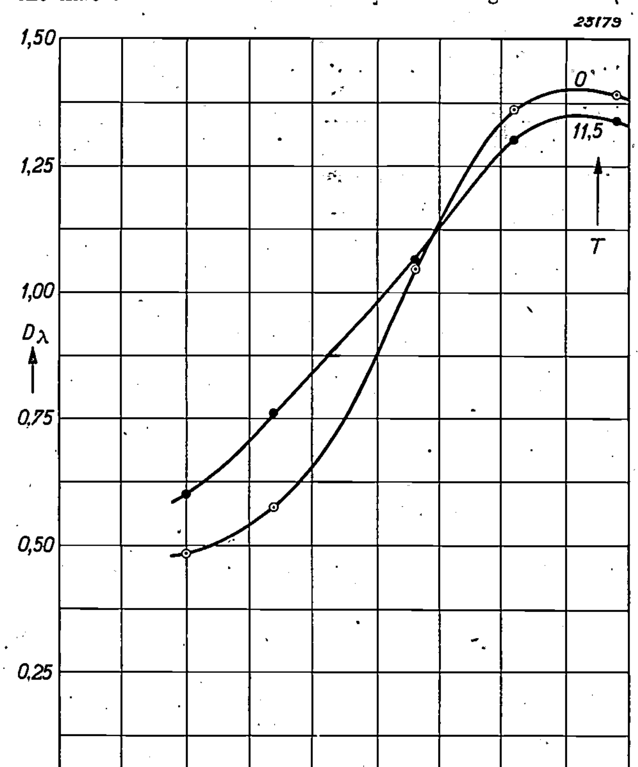


Fig. 7. Reflection blackening curves for sample *I* before irradiation and after 11.5 hours solar irradiation (with maximum solar intensity).

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

No. 1183: A. Bouwers: Convergence of electrons by means of magnetic coils (Physica 4, 200 - 206, March 1937).

The focussing effect of magnetic coils with respect to electron beams is investigated in this paper. The lens effect of a short coil is explained in an elementary manner and the path followed by the electrons is represented. For short and long coils, formulae are derived for the focal distance and for the angle of rotation about the coils. Finally, a number of special cases are considered, such as the borderline case between short and long coils and the effect of a non-homogeneous strong field, by means of which enlarged or reduced images can be obtained.

No. 1184: J. van Niekerk: Einige pharmakologische Untersuchungen mit Salzen von reinem Zirkonium und reinem Hafnium (Naunijn-Schmiedebergs Arch. exp. Pathol. Pharmakol. **184**, 686 - 693, March 1937).

The results of investigations on the pharmacological action of zirconium and hafnium oxychlorides are discussed. Their action in equimolecular quantities appears to be the same with corresponding test objects. The pharmacological effect of zirconium oxychloride is found to be identical with that of the group of zirconium compounds investigated by Marui.

No. 1185: E. M. H. Lips: Over de bewerkbaarheid van perlietisch gietijzer (Metaalbewerking 4, 21 - 26, March 1937).

The machinability of pearlitic cast iron can be determined without the need for specially-prepared specimens, since the structure already indicates that the machinability is mainly determined by: 1) The hardness of the pearlitic ground mass, which can be determined with a micro-scelerometer, 2) the amount of free graphite, which can be established by chemical analysis, and 3) the distribution of free graphite, which can be studied on an unetched preparation.

No. 1186: P. J. Bouma: Is geel licht minder verblindend dan wit licht? (Wegen, **13**, 132 - 135, March 1937).

In respect of simultaneous glare, yellow and white lights are roughly equivalent, but as regards successive glare and fatigue yellow light is very much better (cf. Philips techn. Rev. **1**, 225, 1936).

No. 1186*: J. W. M. Roodenburg: Der Einfluss der Tageslänge im Zusammenhang mit der künstlichen Pflanzenbeleuchtung im Winter (Ber. deutsch. botan. Ges. **55**, Heft 1, Febr. 1937).

The growth of leaves, which is directly associated with the assimilation of carbon dioxide, is found to be proportional to the quantity of incident light (duration times intensity) for an average irradiation period and intensity. But the growth of blossoms and other similar actino-periodical processes can be promoted by additional illumination with such low intensities that the assimilation of carbon dioxide is barely increased as a result thereof. The length of the day has, of course, a marked effect on the shooting of leaves and peduncles, and thus has an action similar to that of cell-activating hormones. It is probable that the duration of diurnal irradiation affects in some way or other the condition of the growth-promoting substance.

No. 1187: J. L. Snoek and M. W. Louwerse: Magnetic powder experiments on rolled nickel iron (II) (Physica 4, 257 - 266, Apr. 1937).

The Bitter patterns, which can be obtained from magnetic powders on thin-rolled sheets of nickel iron in the magnetic maximum-anisotropic state when the lines of force issue perpendicularly, are submitted to closer examination, especially along the narrow side of the strip material. On the long sides of the strip, remarkable zigzag patterns are observed with comparatively weak fields, which on a reversal of the field are transformed to complementary patterns, i.e. those in which the lines do not intersect those of the initial patterns. On the rolling plane complex patterns are obtained, in addition to the known parallel strips, but which on reversing the field are not transformed into the

---

*) There is not a sufficient number of reprints available for distribution of the publications marked *. Reprints of the other publications will on request gladly be supplied by the Administratie van het Natuurkundig Laboratorium, Kastanjelaan, Eindhoven.

corresponding complementary patterns. It is definitely established that the lines on field reversal do intersect. No satisfactory explanation for these phenomena has been advanced.

No. 1188: H. Bruining, J. H. de Boer and W. G. Burgers: Secondary electron emission of soot in valves with oxide cathode (Physica 4, 267-275, Apr. 1937).

When a carbon surface is covered with barium atoms, the capability for secondary emission is increased. However, soot precipitated from a flame, which is composed of very small particles, was found to exhibit a different behaviour. It may be deduced from the change of the secondary emission with time that the barium atoms precipitated on this carbon modification disappear from the surface, with even a moderate electron bombardment. Owing to local heating the barium migrates to soot particles at a greater depth. This migration of the barium is less marked with soot which has been sprayed from a suspension, since in this case the small particles are agglomerated to form large ones, as is confirmed by electron diffraction experiments.

No. 1189: W. Elenbaas: Der Gradient der Überhochdruck-Quecksilberentladung (Physica 4, 279-284, Apr. 1937).

In extra-high pressure mercury discharges, the potential drop per unit of length, i.e. the gradient, is greater than that which would be expected from measurements at one atmosphere pressure on the basis of previously-derived reduction formulae. Some possible causes for this are discussed.

No. 1190: J. H. de Boer: Über die Natur der Farbzentren in Alkalihalogenid-Kristallen (Rec. Trav. chim. Pays-Bas, **56,** 301-309, March, 1937).

The equilibrium concentrations of the colour centres found by Mollwo cannot be explained on the basis of the early theory of colour centres which assumed these to be neutral alkali atoms absorbed at the inner surfaces of lattices. According to Schottky, ionic spaces in the lattice and their associated clusters, such as play a part in ionic conduction and diffusion in alkali halides, may also have a bearing on the formation of colour centres. The energy levels of the occupied and unoccupied electron levels in the lattice and of the alkali and chlorine ions which are situated next to the spaces, are such that the colour centres produced by the absorption of light and the penetration of electrons may be regarded as neutral alkali atoms situated at a halogen ionic space. This provides an explanations for the fact that the absorption bands are reproducible and also accounts for the existence of an equilibrium concentration of the colour centres. Defects in the lattice can likewise be places where colour centres are formed, when the absorption bands are broadened or displaced and the equilibrium concentration is exceeded.

No. 1191: J. A. M. van Liempt and W. van Wijk: Die Mikrogasanalyse von Argon-Stickstoff-Mischungen (Rec. Trav. chim. Pays-Bas, **56,** 310-314, March 1937).

In this paper a arrangement is described for the quantitative analysis of small quantities of an argon-nitrogen mixture.

No. 1192: M. J. O. Strutt and A. van der Ziel: Erweiterung der bisherigen Messungen der Admittanzen von Hochfrequenz-Verstarkerröhren bis 300 Megahertz (El. Nachr. Techn., **14,** 75-80, March, 1937).

Methods are given whereby an "acorn" diode voltmeter can be calibrated with sufficient accuracy for measuring impedances. The various precautions are also discussed which must be taken into consideration in the design of transmitters required to operate efficiently up to 300 megacycles, as well as the adequate screening of the various components of the measuring apparatus. Impedance measurements carried out on shortwave valves, particularly "acorn" pentodes, up to 300 megacycles are also described. It was found that voltage amplification is possible up to 300 megacycles with these valves.

No. 1193: M. J. O. Strutt: Frequency changers in all-wave receivers. The performance of some types (Wireless Eng., **14,** 184-192, April, 1937).

This article is an English translation of the paper summarised in Abstract No. 1177.

No. 1194: W. G. Burgers: Optische demonstratie van eenige verschijnselen, welke optreden bij de verstrooiing van Röntgen- en electronenstralen door kristallen (Ned. T. Natuurk., **4**, 1 - 14, 33 - 38, 1937).

This article is a report of papers dealing with the demonstration apparatus described in Abstract No. **1149**. It was shown, *inter alia*, how this apparatus could be used for demonstrating complex electron-diffraction patterns.

No. 1195: J. Alfter and B. Visman: Grundsätzliches zur Automatisierung der Röntgenapparate (Röntgenpraxis, **9**, 244 - 247, April, 1937).

The main purpose of automatic control is to obtain radiographs of optimum quality by ensuring that the X-ray tube is always run at the limiting load when using the best tube for the particular purpose in view. The following advantages accrue:

a) Reproducibility of the quality of the radiation and the time of exposure;
b) Protection of the focal spot against overloads, and safeguards against switching on the apparatus when wrongly adjusted;
c) Manipulation of the apparatus is facilitated.

No. 1196: W. Elenbaas: Ontladungen in Kwikdamp van hooge druk (Ned. T. Natuurk. **4**, 65 - 87, April, 1937).

This paper, which was written for the 1936 Vacation Course in Illumination Technology, presents a survey of discharges through high-pressure mercury vapour.

No. 1197: J. A. M. van Liempt and J. A. de Vriend: The flash time characteristics of indirectly-ignited photographic flash lamps (Physica, **4**, 353 - 360, May, 1937).

Flux-time curves for indirectly-ignited photographic flash lamps with aluminium foil are given. From the standpoint of the practical applications of these lamps those factors are discussed which bear on the use of socalled synchronisers for igniting the lamp and opening the camera shutter after a specific interval.

No. 1198: F. A. Heyn: Radioactivity induced by fast neutrons according to the (n, 2n) reaction (Nature, **139**, 842, May, 1937).

Continuing the investigations referred to in Abstract No. **1141** in which copper and zinc were bombarded with fast neutrons which caused the expulsion of two neutrons from the nucleus owing to the absorption of one fast neutron, the author has extended these experiments to various other elements. Half-life periods were found for the induced radio-active products obtained in this way, which are in close agreement with those found by Bothe and Gentner when bombarding the same elements with gamma rays, a neutron being expelled in each case. In addition, a number of new radioactive isotopes was found.

VOL. 2 No. 10 OCTOBER 1937

Philips Technical Review

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

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

INVERSE FEED-BACK

by B. D. H. TELLEGEN.

Summary. In this article the influence of inverse feed-back on the properties of an amplifier is examined. After a discussion of the decrease of the non-linear distortion by inverse feed-back, the stability of an inverse feed-back amplifier is studied. In addition it is shown that the influence on the amplification by changes in the supply voltages or in the characteristics of the receiving valve can be decreased by inverse feed-back, and that the internal resistance of an amplifier can be either increased or decreased by inverse feed-back.

Principle of inverse feed-back

If part of the output voltage of an amplifier is fed back to the input, many properties of the amplifier will be modified. If the amplification is increased, we speak of direct feed-back, if on the other hand it is reduced we speak of inverse feed-back. This latter was investigated in 1928 in this laboratory by K. Posthumus[1]) and in the Bell Laboratories by H. S. Black[2]).

Several properties of inverse feed-back have been described previously in this periodical[3]). We propose in the present article to submit it to a closer investigation.

If inverse feed-back is applied to an amplifier, as shown in *fig. 1*, we must distinguish between the signal voltage E_s and the input voltage E_i of the amplifier itself. The difference between these two is the inverse feed-back voltage E_t. If the voltages are calculated as positive when they act in the directions indicated by the arrows, the following holds:

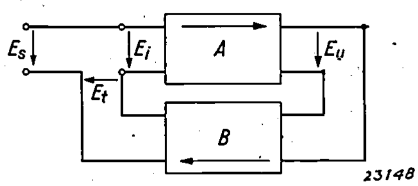


Fig. 1. Diagram of inverse feed-back, A amplifier, B inverse feed-back circuit. A portion E_t of the output voltage E_u is led back to the input side of the amplifier.

$$E_i = E_s - E_t \quad \ldots \ldots \quad (1)$$

Influence on the amplification and on the distortion

In order to calculate the change in amplification caused by inverse feed-back, we imagine a sinusoidal signal voltage of a definite frequency to be applied. Then, if α is the amplification without inverse feed-back and β is the portion of the output voltage E_u which is employed for inverse feed-back:

$$E_u = \alpha E_i,$$
$$E_t = \beta E_u, \text{ and therefore}$$
$$E_t = \alpha \beta E_i.$$

If this is substituted in (1) it gives:

$$E_i = \frac{1}{1+\alpha\beta} E_s \text{ and therefore}$$

$$E_u = \frac{\alpha}{1+\alpha\beta} E_s \quad \ldots \ldots \quad (2)$$

$E_u = \alpha E_i$ will in general differ from E_i, not only in size, but also in phase, and the same is true of the feed-back voltage $E_t = \alpha\beta E_i$. The factors α and $\alpha\beta$ are therefore not pure numbers, but quantities which include a phase angle, and which are represented in the vectorial notation of alternating current theory by complex numbers.

---

1) K. Posthumus, Brit. Pat. 323823.
2) H. S. Black, Stabilized feed-back amplifiers, Bell System techn. Journal 13, 1 (1934), published also in El. Eng. 53, 114 (1934).
3) C. J. van Loon, Philips techn. Rev. 1, 264, 1936.

If $\alpha\beta$ is real and positive, the amplification is decreased. If $\alpha\beta$ is real, negative and less than one in absolute value, the amplification is increased and we are concerned with direct feed-back. If $\alpha\beta = -1$, the amplification becomes infinite, that is to say, the system is no longer stable. In this case it is in fact possible for an undamped sinusoidal oscillation to be maintained in the system without a signal, because the fact that $\alpha\beta = -1$ means that at a given amplitude E_u the feed-back voltage E_t has just the magnitude and phase necessary to create an input voltage which generates E_u again. If $\alpha\beta$ is real, negative and greater than one in absolute value, one might think that the system is always unstable. Closer examination of the stability, to which we shall return later, has however shown that this is not necessarily always the case. In general, if the system is stable, it is the absolute value $|1 + \alpha\beta|$ which determines the change in amplification. When this absolute value is large with respect to one, $1 + \alpha\beta$ approaches $\alpha\beta$ and (2) is simplified to

$$E_u = \frac{1}{\beta} E_s \quad \ldots \ldots \ldots (3)$$

From this we see that the quantity α no longer occurs in (3). Now α, the value of the amplification without feed-back, will in general be a function of the frequency, to which it is not always easy to give the desired form, and moreover α depends upon the characteristics of the valves, in particular on their slopes, which may vary because of several causes. All valves of the same type have not exactly the same slope, and the slope varies to a certain extent during the life of the valves, and also depends upon the voltage of the mains or of the batteries, which is not always constant. When a strong inverse feed-back is applied, the amplification depends only on β which is determined by the inverse feed-back circuit. In general this circuit will include no valves and may more easily be constructed in such a way that β will depend on the frequency in the desired relation. To what degree the amplification depends less upon variations in α when inverse feed-back is applied, than without inverse feed-back, may be given simply for the case when $\alpha\beta$ is real, that is, when the inverse feed-back is accurately in phase. In order to determine the corresponding percentage change in $\frac{\alpha}{1+\alpha\beta}$ at a definite percentage change in α, we must calculate the quantity

$$\frac{\mathrm{d}\,\frac{\alpha}{1+\alpha\beta} : \frac{\alpha}{1+\alpha\beta}}{\mathrm{d}\,\alpha : \alpha}$$

which is found to be $\frac{1}{1+\alpha\beta}$ that is equal to the decrease is amplification.

Equation (3) suggests another advantage of inverse feed-back. An amplifier is in general not entirely linear, due to curvatures of the valve characteristics, and therefore harmonic and beat frequencies may appear in it. The inverse feed-back circuit may consist of linear elements, it may therefore be expected that, when the amplification depends on β alone, no harmonic or combination frequencies will appear in the output voltage E_u.

The fact that this assumption is correct, and to what degree the harmonics are suppressed was shown in a previous article in this periodical [4]). In the following the calculation of the influence on the second harmonic by inverse feed-back will be repeated in a somewhat more general form, which is not subject to the limitations mentioned in footnote (4). We assume that the signal voltage E_s is truly sinusoidal. If there is a second harmonic of a certain amount P present in the output voltage, there will be an amount of second harmonic βP present in the inverse feed-back voltage. Since the signal itself contains no second harmonics, according to (1) the input voltage will contain an amount $-\beta P$ of second harmonics. This amount $-\beta P$ leads to an amount $-\alpha\beta P$ of second harmonics at the output of the amplifier. Moreover, an amount Q of second harmonics is generated in the amplifier by the non-linear elements in it. In the case of small distortion Q depends only on the strength of the fundamental frequency in the input voltage, and is not influenced by the amount $-\beta P$ of second harmonics present. Thus Q is the amount of second harmonics in the output voltage of the amplifier when no inverse feed-back is applied, and when a sinusoidal signal voltage of such a size is applied that the fundamental component of the output voltage is just as large as in the case just considered with inverse feed-back. Since we began with an amount P of second harmonics in the output voltage, we arrive at:

$$P = -\alpha\beta P + Q,$$

Therefore

$$P = \frac{Q}{1+\alpha\beta} \quad \ldots \ldots \ldots (4)$$

It follows from the derivation that α and β in

4) C. J. van Loon, loc. cit. In this article the influence of inverse feed-back on the second and third harmonics is calculated. The derivation is limited to the case where α and β are real and independent of the frequency.

(4) must be taken at the frequency of the second harmonic, and not at the fundamental frequency. The amplitude of the second harmonic, therefore, at a given output voltage is decreased due to inverse feed-back by the same factor as is the amplification of a signal having the frequency of that harmonic. When the distortion becomes larger, and higher harmonics appear, the relations are no longer so simple [5]), but the fact remains true that the distortion at a given output voltage is decreased by strong inverse feed-back.

If a interference originates in the amplifier, due for instance to mains supply, then it may be shown in an analogous way to that used above for the second harmonic, that this is decreased in the output voltage of the amplifier by means of inverse feed-back in the same ratio, as is the amplification of a signal having the frequency of the disturbance.

Stability

The considerations which led to equation (2) have as a basis the tacit assumption that the inverse feed-back system is stable. In order to determine this fact we must examine the manner in which the system will continue to oscillate if it is left to itself after the application of a slight disturbance, that is to say, we must examine the free oscillations of the system. If these are all damped, the system is stable; if one or more of the free oscillations grows with time, the system is unstable. The borderline between these two conditions is that condition when one of the free oscillations is an undamped sinusoidal oscillation. As has already been shown, this limiting case will exist when $\alpha\beta = -1$. We may now determine $\alpha\beta$ by measurement or calculation as to magnitude and phase as a function of the frequency, and plot $\alpha\beta$ as radius vector for each frequency. The length r of the radius (see *figs. 2* and *3*) thus indicates the relation between the voltage E_t fed back and the input voltage E_i. The angle φ gives the amount by which the phase of E_t is in advance of that of E_i. The connecting line from the end points of the vectors which correspond to different frequencies will form a curve, which we call the vector diagram of the inverse feed-back, and which will usually be closed, since $\alpha\beta$ is generally zero not only for the frequency zero, but also for the frequency ∞. If the amplifier consists for instance of a single stage of resistance amplification in which a constant portion of the output voltage is employed for inverse feed-back (fig. 2), then C_1 is generally so large, and C_2 so small, that they have no in-

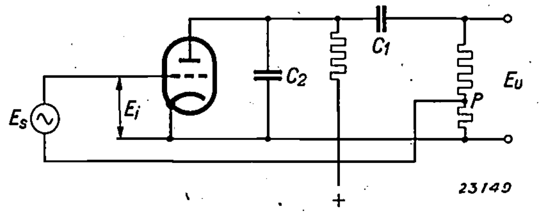


Fig. 2. Resistance amplifier with inverse feed-back. The voltage taken off at point P is in phase with E_i for average values of the frequency; for low frequencies the phase is 90° in advance, for high frequencies 90° behind.

fluence on the amplification over a large part of the frequency range to be amplified. In this range therefore $\alpha\beta$ is practically constant, and has a small phase shift, which changes very slowly with the frequency. At the lowest frequencies the amplification will however decrease because of the presence of C_1, and at the highest frequencies due to C_2. This change in amplification is accompanied by a phase change of the amplification, which is 90° at the frequencies zero and infinity. This has as result that the vector diagram of $\alpha\beta$ has about the form of a circle (*fig. 3*).

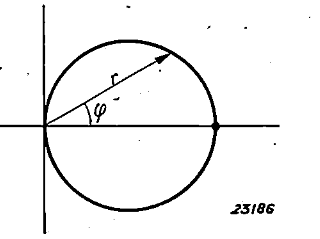


Fig. 3. Vector diagram of one-stage resistance amplification (Fig. 2) with inverse feed-back. The phase changes from 90° ahead at low frequency to 90° behind at high frequency.

If the amplifier consists of two stages of resistance amplification in cascade arrangement, in which a constant portion of the output is used for inverse feed-back, then the phase change of the amplification for the frequencies zero and infinity will be 180°, and we obtain a vector diagram of the form given in *fig. 4*. If the amplifier or the inverse feed-

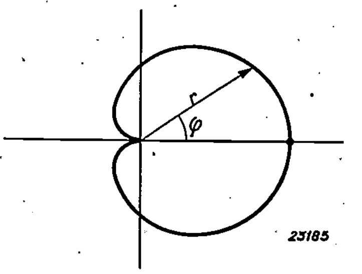


Fig. 4. Vector diagram of two-stage resistance amplification with inverse feed-back. The phase changes twice as fast as with one stage. Thus from 180° ahead to 180° behind.

---

[5]) R. Feldtkeller, Die 3te Teilschwingung in Verstärkern mit Gegenkopplung, Telegr. und Fernsprechtechn. 25, 217, 1936.

back circuit contains resonances, $\alpha\,\beta$ may be very much changed in the neighbourhood of those resonance frequencies as regards magnitude and phase, so that vector diagrams of complicated forms may occur.

In order to investigate the stability for all cases, we assume that we gradually increase the inverse feed-back from zero to the desired value, that is, that we substitute $n\,\beta$ for β, where n is a positive number independent of the frequency, which we allow to increase starting from zero. The vector diagram will then change congruently with itself. If n is small, the vector diagram is also small, and the point —1 on the real axis lies outside the diagram. The system will always be stable for small values of n, and we desire to know what happens to the stability when n increases from zero. We mentioned already that on the threshold of instability, one of the free oscillations has become an undamped sinusoidal oscillation. If this is the case, $\alpha\,\beta$ will equal —1. At the transition from stable to labile the vector diagram therefore passes through the point —1 on the real axis. Upon increase of n the system will thus remain stable as long as the point —1 on the real axis lies outside the vector diagram. If this point falls within the diagram, the system becomes labile. It may however happen that upon further increase of n the vector diagram passes the point —1 on the real axis for the second time. Then there are two possible cases: either the free oscillation which had been increasing with the time is again damped and the system therefore again stable, or a second free oscillation begins to increase instead of being damped and the system remains labile. Nyquist[6]) has shown that these cases may be distinguished by the forms of the vector diagrams. If the diagram has the form of *fig. 5a* for example, then the system upon increase of n is first stable, then labile and finally stable again. In this last state therefore we have a system for which $\alpha\,\beta$ is, for two frequencies, real, negative and of an absolute value greater than 1, while the system is still stable. If however the diagram has the form of fig. 5b for example, then the system upon increase of n changes from stable to labile, while it remains further labile. This result was formulated as follows by Nyquist: if the point —1 on the real axis lies outside the vector diagram, the system is stable, if the point lies within the vector diagram the system is labile.

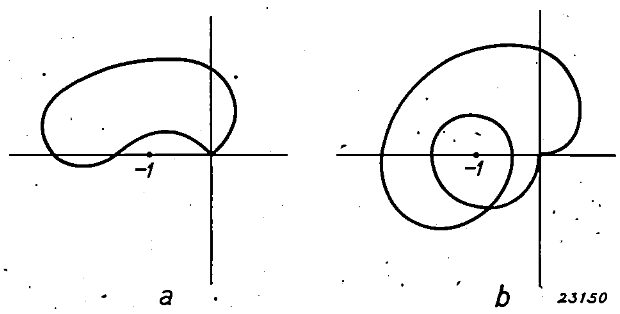


Fig. 5. *a*. Vector diagram of a stable system which upon decrease of the inverse feed-back becomes labile. *b*. Vector diagram of a labile system which remains labile upon increase of the inverse feed-back.
A system is stable when the point — 1 lies outside the closed curve of the vector diagram; labile when this point is within the curve.

Voltage and current inverse feed-back

Up to now we have not spoken of the loading resistance of the amplifier, and have considered this as a part of the amplifier. If we desire however to investigate the behaviour of the inverse feed-back amplifier when the loading resistance Z_u is variable, we must consider it separately from the amplifiers. We can then distinguish two cases. Either the ratio between the inverse feed-back voltage E_t and the output voltage E_u, which we have called β above, is independent of Z_u, or the ratio between E_t and the output current I_u, which we shall call γ, is independent of Z_u. Since $E_u = I_a\,Z_u$, the relation $\beta = \gamma/Z_u$ exists between β and γ. In the first case we speak of voltage inverse feed-back, in the second of current inverse feed-back. These two cases are represented in *figs. 6a* and *b*. If Z_u is made smaller, in case a the output voltage E_u (due to the internal resistance of the amplifier) will decrease, and the same holds

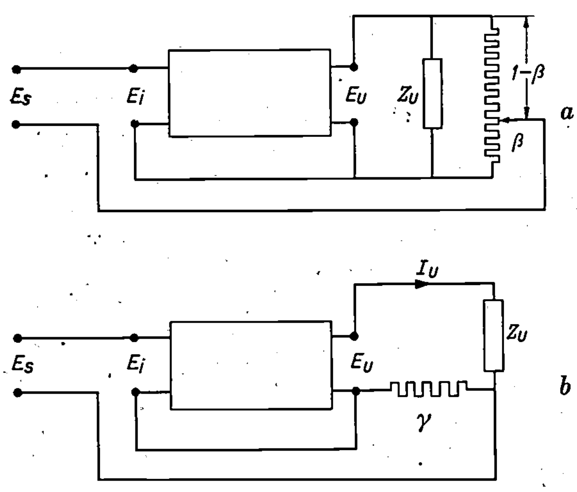


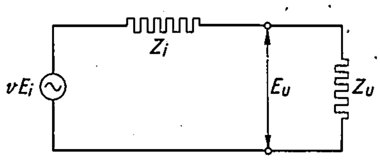
Fig. 6. *a*. Voltage inverse feed-back. The ratio between the inverse feed-back voltage and E_u is independent of Z_u. *b*. Current inverse feed-back. The ratio between the inverse feed-back current and I_u is independent of Z_u.

[6]) H. Nyquist, Regeneration Theory, Bell Syst. Techn. Journal, 11, 126 (1932).

for the inverse feed-back voltage βE_u. In case b on the other hand, upon decrease of Z_u the output current I_u will become larger, and consequently also the inverse feed-back voltage γI_u.

We saw above that with strong inverse feed-back the amplification approaches $1/\beta$, that it depends therefore only slightly upon Z_u with voltage inverse feed-back, and that with current increase feed-back it is about proportional to Z_u, or in other words that the internal resistance becomes small with voltage inverse feed-back, while the internal resistance becomes large with current inverse feed-back.

We should like to examine this change of internal resistance more closely. Every linear system which ends in a pair of terminals, as regards its external effects, may be replaced by its open-circuit voltage with its internal resistance connected in series. We shall apply this theorem to an amplifier for the calculation of the internal resistance. If an input voltage E_i acts on the amplifier, then the open circuit voltage at the output end may be represented by vE_i, and the internal resistance by Z_i. If the amplifier is loaded with a resistance Z_u we arrive at the arrangement



23151

Fig. 7. Equivalent circuit of an amplifier with input voltage E_i, no-load voltage $v\,E_i$, internal resistance Z_i and external resistance Z_u.

of *fig.* 7. The voltage on Z_u is thus,

$$E_u = vE_i \cdot \frac{Z_u}{Z_i + Z_u} \quad \cdots\cdots \quad (5)$$

and the amplification a is thus given by:

$$a = \frac{v\,Z_u}{Z_i + Z_u} \quad \cdots\cdots\cdots \quad (6)$$

When inverse feed-back is applied, we saw above (eq. (2)), that the relation between E_u and E^s is given by:

$$E_u = \frac{a}{1 + a\beta}\, E_s \,.$$

Substituting in this the value of a from (6):

$$E_u = \frac{v\,Z_u}{Z_i + Z_u + v\,\beta\,Z_u}\, E_s \,. \quad \cdot \quad (7)$$

If we are concerned with voltage inverse feed-back, β is independent of Z_u, and we write (7) in the form:

$$E_u = \frac{v\,E_s}{1 + v\,\beta} \cdot \frac{Z_u}{\dfrac{Z_i}{1 + v\,\beta} + Z_u} \quad \cdot \quad (8)$$

If we compare (8) with (5), we see that, due to voltage inverse feed-back, the internal resistance is changed to

$$\frac{Z_i}{1 + v\,\beta} \,.$$

At the same time we see that the open circuit voltage is changed to

$$\frac{v\,E_s}{1 + v\,\beta}$$

The stronger the voltage inverse feed-back, the smaller the internal resistance becomes. If Z_u, Z_i, a, β and v are all real and positive, it follows from (6) that a is smaller than v, so that the internal resistance diminishes more than the amplification due to the voltage inverse feed-back, since the latter decreases from a to $a/(1 + a\,\beta)$. If however the internal resistance without inverse feed-back is already small with respect to the external resistance, then a and v are about equal, and, with voltage inverse feed-back, amplification and internal resistance are about equally diminished.

If we are concerned with current inverse feed-back, we substitute γ/Z_u for β in (7), since γ is now independent of Z_u and then obtain:

$$E_u = v\,E_s \cdot \frac{Z_u}{Z_i + v\gamma + Z_u} \quad \cdot \cdot \quad (9)$$

Comparing (9) with (5) we see that, due to current inverse feed-back, the internal resistance is changed to

$$Z_i + v\,\gamma \,.$$

The stronger the current inverse feed-back, the larger the internal resistance becomes. If Z_u, Z_i, a, γ and v are all real and positive, it follows from (6) that a is smaller than $v\,Z_u/Z_i$. Now the internal resistance increases by the factor

$$(1 + \frac{v\,Z_u}{Z_i} \cdot \gamma/Z_u),$$

while the amplification decreases by the factor

$$(1 + a \cdot \gamma/Z_u),$$

as appears from equation (2) upon substitution of β by γ/Z_u. From this it thus follows that the internal resistance increases more strongly due to current inverse feed-back than the amplification decreases. If however the internal resistance without inverse feed-back is already large with respect to the external resistance, then a is about equal to $v\,Z_u/Z_i$, and with current inverse feed-back the internal resistance is increased by the same amount as the amplification is decreased.

By the application of a combination of current and voltage inverse feed-back, at a given degree of inverse feed-back, any desired internal resistance may be obtained.

We have seen from the above that by the application of inverse feed-back many characteristics of an amplifier may be influenced in a simple way, and other properties may be improved to an important degree. The one disadvantage is a loss in amplification. Due to the improvement of amplifier valves, however, it is now possible to get a high amplification with a small number of valves, so that this loss of amplification in many cases need no longer be considered a great disadvantage when compared with the advantages attained. Inverse feed-back has been used recently, not only in telephone amplifiers [7]), where, because of the large number of amplifiers connected in cascade arrangement, a great constancy of the amplification is necessary, while in carrier wave telephony, intermodulation is prevented at the same time, but also in radio receiving apparatus and amplifiers for sound reproduction [8]), in which the decrease in distortion and the simple way of modifying the frequency characteristic are of great advantage.

---

7) W. Six and H. Mulders, The use of amplifiers in telephone technique, Philips techn. Rev. 2, 209 (1937).

8) B. D. H. Tellegen en V. Cohen Henriquez. Inverse feed-back, its application to receivers and amplifiers. Wirel. Eng. 14, 409, 1937.

THE APPLICATION OF MAGNETIC OIL FILTERS TO LUBRICATING SYSTEMS

by L. H. DE LANGEN.

Summary. An apparatus is described which can remove small magnetic particles from an oil-stream. The decrease in wear on shafts and the like obtained with this apparatus is discussed.

After the lubrication of bearings and other parts of machinery had been the exclusive concern of the engineer for many years, this domain of technology was brought within the reach of science about the end of the last century. The problem was at first attacked chiefly from the mathematical point of view. In setting up the hydrodynamic lubrication theory [1]) Reynolds began with the proportionality between the sliding stress appearing in the lubricant and the fall in velocity. Sommerfeld, Michell, Gümbel and others later developed the theory further. Of the various assumptions which form the basis of the hydrodynamic lubrication theory the following two are of importance for our subject:

1) that the surfaces which slide over each other are quite smooth and
2) that the lubricant is a pure liquid.

In practice neither of these two conditions is entirely fulfilled. Recent investigations have shown that most surfaces, as obtained in the workshop by the usual means, even when they are observed under moderate magnification, resemble a mountain landscape. The heights of the mountains in this landscape are of about the same order of magnitude as the thickness of the lubricating oil layer. It is therefore not surprising that when these surfaces are allowed to slide over each other the theory of lubrication is scarcely applicable; the oil film is broken, and considerable wear occurs.

The writer has ascertained on various occasions that even with highly-finished bushes in which a shaft had run for several hours, the diameter had become several hundredths of a millimetre greater; the tops of the mountains had been worn away. Because of this wear the second assumption, namely that the lubrication is carried out with a pure liquid, is no longer justified. The particles of metal worn off circulate with the liquid and cause further wear. This experience leads to the statement of the following paradox: a bearing must be lubricated with as small an amount of oil as possible. When the oil is contaminated due to wear it is actually better to use little oil, so that only a few solid particles take part in the further wear and tear of the surfaces. Of two approximately equally loaded bearings of a certain construction, one of which received very much oil and the other, because of accidental circumstances, almost nothing from the same oil reservoir, the first showed by far the greater wear.

With the above in mind it is not difficult to indicate methods for improving lubrication and decreasing wear and tear. In the first place more ideal surfaces must be aimed at, and in the second place methods must be found of removing the metal particles carried by the oil before they cause further wear.

It is now possible, even in ordinary production, by making use of modern methods of treatment such as lapping and honing [2]) to give the surfaces of shafts and bearings such a finish that the above-mentioned first condition is fairly well satisfied.

This article will describe a method developed in this laboratory of continually removing the particles of iron from the oil, so that the second condition may also be fulfilled. This method is as follows: the circulating lubricating oil flows continuously through a strong magnetic field, so that the iron particles are attracted to one of the poles and separated from the oil.

The latest development of magnet steel has made it possible, with permanent magnets of moderate dimensions, to attain fields of such strength that even very small iron particles can be taken out in this way.

Fundamentally the magnetic oil filters consist of a cylindrical permanent magnet placed in a cylindrical space enclosed by iron walls. A simple

[1]) O. Reynolds. Phil. Trans. Roy. Soc. London 177, 157 (1886).

[2]) Lapping is a method of polishing in which a metal surface is polished with another softer metal surface, with the addition of an abrasive suspended in a liquid. By giving the correct form to the polishing surface it is possible in this way, not only to make the parts smooth, but also true in shape and dimensions. Honing is a similar process, in which however stone is used for the polishing surface instead of a fine abrasive embedded in a soft metal. The finished surface is somewhat less smooth.

construction is drawn in *fig. 1*. The permanent magnet A is placed in a steel tube B. Three brass wires C soldered to the magnet hold it in the middle

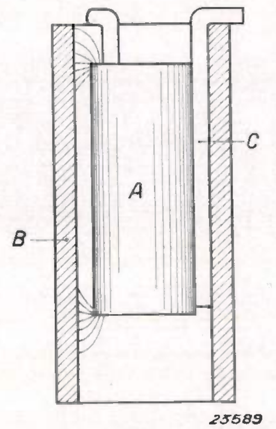


Fig. 1. Simple magnetic oil filter.

of the tube. The lines of force of the magnet are drawn in the left-hand half. If oil contaminated with particles of iron flows through this tube, the iron is deposited mainly on the upper side of the magnet at the points of highest concentration of magnetic flux, as may be seen in the photograph (*fig. 2*) of a contaminated magnet.

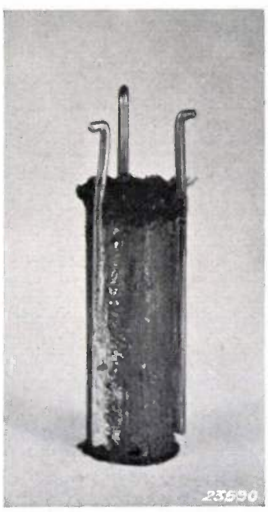


Fig. 2. Magnet with iron filings.

An oil filter of this construction is very satisfactory, but is somewhat limited in capacity; because the impurities are deposited as a narrow ring, the filter is quickly full.

An oil filter of more perfect construction is shown in *fig. 3*. The permanent magnet A is fastened to an iron cap B, which is screwed into the iron housing C. A brass cylinder D with holes E is placed around the magnet in order to direct the oil which enters at F and leaves at G. This construction has three important advantages. Firstly a good magnetic circuit is formed here with only one air gap so that with a given length of the permanent magnet

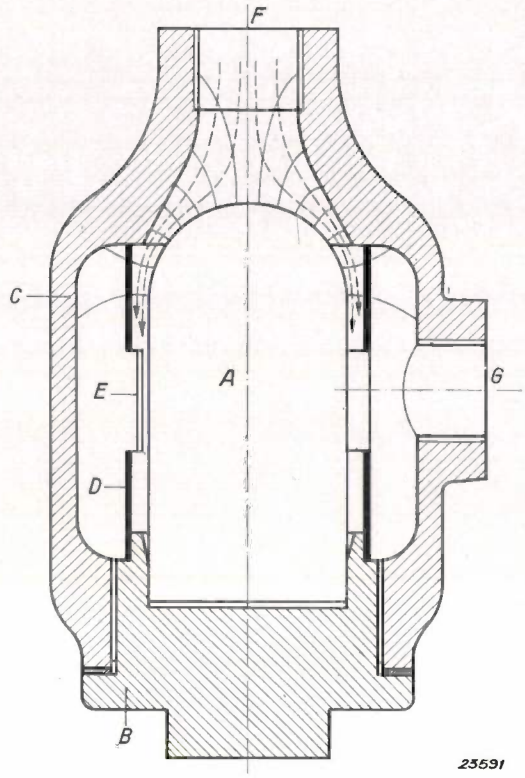


Fig. 3. Cross section of an oil filter. The magnetic lines of force are indicated by continuous lines and the stream lines are dotted.

the strongest possible field is generated. Secondly the convex pole of the magnet is so placed with respect to the housing that the particles of iron are deposited over the whole convex surface, and the filter can therefore collect a large amount of impurities. Thirdly the air gap is of such a form that the oil entering flows in about the same direction as that in which the iron particles are attracted. The lines of force and the stream lines cross each other at small angles; therefore the iron particles are subject to the attractive effect of the magnet over a long distance.

Various magnetic filters have been employed in the Philips organisation for some time. The first general application is to be found in the film projectors of the Cinema Department, which are now equipped with magnetic oil filters. Successful attempts have been made along two lines to limit the wear and tear on film projectors. The first step was that of making the surfaces of the parts truer and smoother. A great deal has been achieved in this respect. All shafts and bearing bushes are lapped or honed, the Maltese cross also is finished by a special lapping treatment to give great smoothness and high precision (tolerances 1 to 2 microns). The toothed wheels, however, have an ordinary finish, which is obtained by hobbing with a fine milling cutter. Grinding or lapping of these little toothed wheels is not yet possible. During working the toothed wheels produced iron dust which, acting as an abrasive, spoiled the smooth surfaces of the lapped shafts, the Maltese cross etc. In this case the use of simple magnetic filters effected a great improvement. Two of these filters are built into the housing of the Maltese cross. *Fig.* 4 is a cross section through one of these filters.

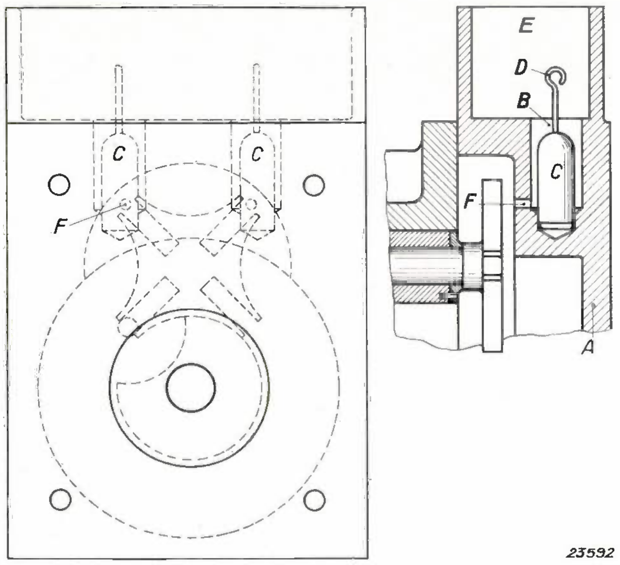


Fig. 4 The oil filters. *C* are here introduced into holes *B* in the housing *A* of the Maltese cross of a film projector. In the front view (left) they are dotted, and to the right is a cross section through a magnet.

A is a part of the housing, *B* is a hole in which the permanent magnet *C* is placed. *D* is a wire of non-magnetic material which is cast in. *E* is an oil reservoir which is continually refilled by the oil pump. The oil flows through the hole *F* into the housing. Of the oil passing through the pumps only that portion is filtered which is intended for the Maltese cross. Since, however, the circulation is rapid, the whole of the oil eventually finds its way through the filter to the Maltese cross. This method is very satisfactory; if a magnetic filter is introduced into a projector whose oil has become dark-coloured from the ground-off iron particles, the oil may be seen to become clear again after a little time. The length of the captured particles of iron seems to lie between 0.4 and 4 microns. The grinding off decreased rapidly after the initial working in, as may be seen from the graph of *fig*. 5.

Especially in the case of the Maltese cross is the limitation of the wear and tear extremely important; a Maltese cross in which there is play due to wear begins to knock harder and then wears off even more rapidly.

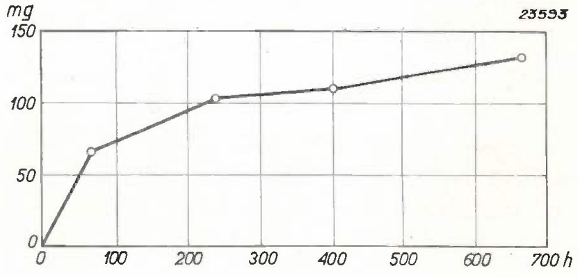


Fig. 5. Wear and tear on a projector: weight of the iron particles deposited by the magnetic oil filter as a function of the number of hours during which the projector had run.

We have also introduced a magnetic filter into an overhauled automobile motor, whose cylinders had been honed and whose pistons had been provided with new piston rings. A filter like the one in fig. 3 was placed in the pressure oil system in such a way that all the oil passed through it. After the car had run 500 km the filter contained 673 mg of iron particles whose size varied between 0.4 and 5 microns. *Fig. 6* is a photograph

Fig. 6. Oil filter magnet covered with iron filings.

of the magnet after it was taken out of the filter housing. Later on the wear became less, as may be seen from the graph of *fig*. 7. The graph representing the wear and tear has the same form as that found in the case of the cinema projectors.

Magnetic oil filters also have been used with success in various industrial machines. In the isolated case of an oil filter introduced into the lubricating system of a steam turbine, which had been running for some time, only a small amount of worn-off metal was found, as would be expected.

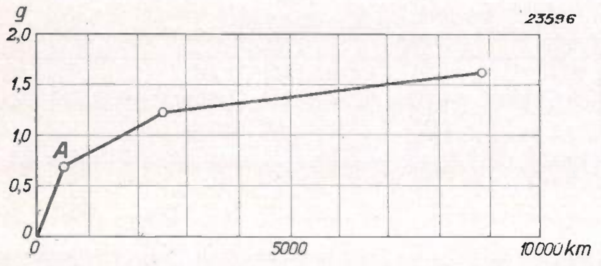


Fig. 7. Wear and tear on an overhauled automobile motor: weight of the iron particles deposited on the oil filter as a function of the distance the car had run. The magnet of fig. 6. was taken out of the filter housing at the moment represented in this figure by the point A.

In the employment of magnetic oil filters there are various methods to be followed. In the case of machines which are not too large it is best to introduce an oil filter into the main feedpipe of the pressure system. With larger machines having a large amount of circulating oil it may be sufficient to pass only a portion of the oil flow through the filter by installing it in a branch pipe of the supply. Another possibility is the introduction of a simple oil filter directly before each lubrication. In many cases it is not difficult to arrange the holes for the lubricant in such a way that a magnet can be introduced, for example in the way done in fig. 4.

With a magnetic oil filter of course only magnetic particles can be captured; particles of bronze or white metal thus remain in the oil. Fortunately the parts which are most subject to initial wear and tear are as a rule made of iron or steel, such as gear wheels, piston rings, cylinders etc.

In many cases where bronze is now often employed, cast iron may be used to advantage. Thus for example the use of cast iron bearings in the Philips projector was a success, and as an additional advantage any particles from the bearings were captured in the magnetic filter.

The wear due to metal particles in the lubricating oil has also been investigated by means of the following simple test. A silver steel shaft lapped as in normal production, 10 mm in diameter, ran for $1^1/_2$ hours in smooth bronze bearings. In order to be sure that the lubrication would not be unfavourably affected be too tight a fit the play was arranged to be large namely 0.06 mm (difference of the diameters). The shaft, which made about 1000 revolutions per minute, was driven by a belt of string; the tension of the string was the only load on the shaft. The bearings were lubricated with oil which had been artificially contaminated with iron particles taken from the magnetic oil filter of a projector. After the conclusion of this test the shaft showed ringshaped grooves visible to the naked eye. With a special Busch microscope for the examination of surfaces, a photograph was made of the worn part of the shaft and of a portion which lay outside the bearings. This photograph is shown in *fig. 8*. Two objects which are illuminated obliquily may be placed under the microscope for comparison of the surfaces. If the surfaces are completely smooth the light is reflected to the other side; the perpendicularly placed microscope receives no light. If there are grooves or inequalities in the surface, light is also reflected in a vertical direction. Looking through the microscope, light streaks or points may be observed. Part A of the photograph shows the undamaged portion of the shaft. The fine criss-cross scratches of the polishing treatment may be clearly seen. It is not difficult to produce a much more perfect surface which appears practically black in the microscope, but for ordinary purposes this extreme quality is not necessary.

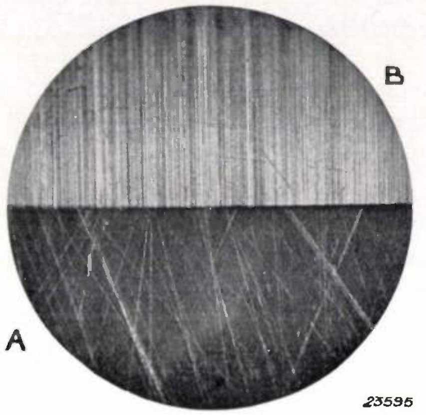


Fig. 8. Photomicrograph of the surface of a shaft with illumination from one side. A is the new and B the worn portion. Magnification 102 diameters.

Part B is a picture of the worn shaft. Nothing can be seen of the original polishing scratches; instead there are now the much coarser grooves due to wear. By means of this simple test which can easily be repeated it is shown convincingly that attention must be devoted to meaures for the purification of circulating lubricating oil.

Tests will be carried out on a still larger scale on the application of magnetic oil filters. We hope to announce the results of such tests in the future.

A DECIMETRE WAVE RADIO LINK BETWEEN EINDHOVEN AND NIMEGUEN

by C. G. A. VON LINDERN and G. DE VRIES.

Summary. A radio connection between Eindhoven and Nimeguen is described in which use is made of waves of about 25 cm length. Magnetron transmitters and superheterodyne receivers are employed. Directional aerials in the shape of paraboloids of revolution are used, having an amplification factor of about 20.

Introduction

Transmitting valves for waves of about 25 cm (1200 Mc/s) have recently been considerably improved, so that they are able to generate relatively high powers. This fact led us to make use of these waves in an experimental connection between Eindhoven and Nimeguen. The terminals of this link were about 50 km apart and set up on the 50 m tower of the lamp factory in Eindhoven and on a water tower situated on one of the highest points in the neighbourhood of Nimeguen. In a previous article the writers have already indicated the most important condition for satisfactory transmission: a free optical path. The free path is in this case of even greater importance owing to the shorter wave lengths used.

Fig. 1 shows how the path from Eindhoven—Nimeguen is situated with respect to sea-level. The line of direct sight is the line joining the transmitting aerial on the tower of the lamp factory with the aerial on the water tower. It may be seen from the figure that, just as in the case of the Eindhoven—Tilburg beam, the line of direct sight is always at least 10 m above the treetops and any buildings.

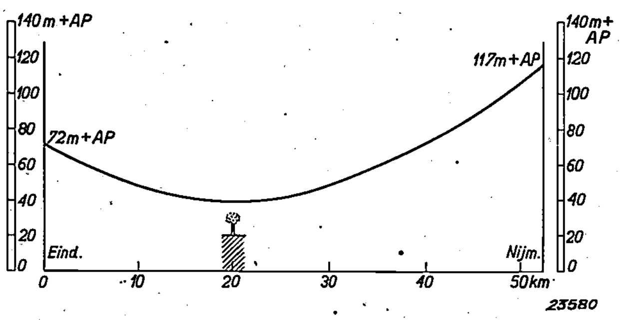


Fig. 1. The curved line represents the line of direct sight between transmitter and receiver in the link between Eindhoven and Nijmegen. The earth is drawn flat there as x-axis; this represents sea level in the Netherlands.

Transmitting tests in the direction of Nimeguen carried out with the aid of an experimental rotating Yagi aerial, which is shown in fig. 5 p. 302 of our previous article in this periodical, confirmed the fact that there was a free optical path, since the field strength received in Nimeguen corresponded with what would be expected, given a free path, to the same order of magnitude.

The aerial

In order to obtain some idea of the dependence of the received strength on the frequency used, one may in the first place consider the case in which transmitter and receiver are set up free from the earth. If transmitter and receiver are both free in space and far removed from the earth, the field strength at the receiving end is proportional to the square root of the energy W in the dipole of the transmitter and inversely proportional to the distance R between transmitter and receiver, thus equal to $k_1\sqrt{W}/R$.

While with 1 m waves we used Yagi directional aerials, with 25 cm waves the paraboloid of revolution must be considered. The directing action of these mirrors is appreciably greater; at a diameter of 3 m it is about six times as great as that of the previously described Yagi directional aerials whose amplification is about 3.5. The amplification of such a reflector is roughly equal to $\frac{\pi R'}{\lambda}$, where R' is the longest radius of the paraboloid and λ is the wave length, both expressed in the same unit [2]).

It is obvious that such an amplification is due to the small angle of divergence in the polar diagram of the paraboloid; this diagram is indeed much more satisfactory than that of the Yagi aerial. As may be seen from *fig. 2* the angle of divergence is about 11°. For secrecy, in the sense of desiring to communicate with only a definite region, this is an advantage.

If paraboloids are used for transmitting and receiving aerials with radii R_1' and R_2' respectively and amplification factors g_1 and g_2 respectively, the field strength becomes proportional to:

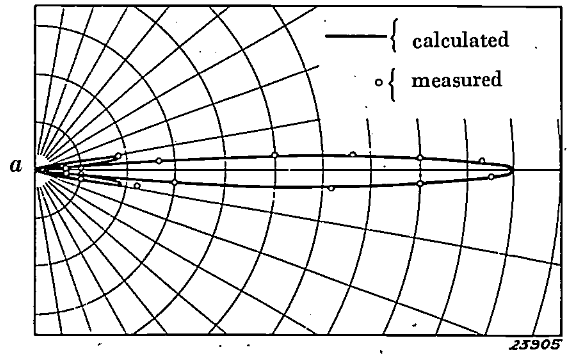
$$g_1\, g_2\, \frac{\sqrt{W}}{R} \text{ or } \frac{\pi R_1'}{\lambda} \cdot \frac{\pi R_2'}{\lambda} \cdot \frac{\sqrt{W}}{R}\,.$$

This field strength is however not important, it is

[1]) Philips techn. Rev. 1 (1937).

[2]) C. J. H. A. Staal, Transmitt. News 3, No. 3 (1936).

rather the induced electromotive force in the aerial which is important. The receiving aerial is half a wave length long, so that the included electromotive force is proportional to $g_1 g_2 \lambda \frac{\sqrt{W}}{R}$ or to $\frac{R_1' R_2'}{\lambda R} \sqrt{W}$. We therefore see that the signal which reaches the receiver increases proportionally with the frequency, when the diameters of the paraboloids



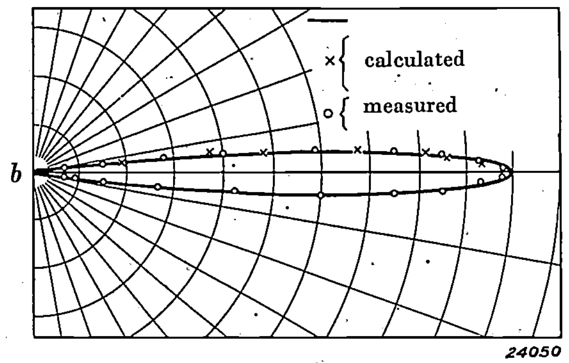


Fig. 2a) Horizontal, b) vertical polar diagrams measured and calculated for a paraboloid of revolution with a vertical dipole at the focus.

and the transmitting energy are kept constant. In addition the conditions for waves of 25 and 125 cm may be compared with each other with the aid of these expressions. If the diameter of the paraboloid at the transmitting end is 3 m and at the receiving end 1 m, and if the transmitting energy is 5 watts, then, at the wave length used, the electromotive force induced in the receiving aerial is larger than in the 125 cm wave connection Eindhoven—Tilburg, where two Yagi aerials are used, and the energy at the transmitting end is calculated to be 10 watts. The transmission on the shorter wave is possibly more favourable than on the longer wave, considering that the line of direct vision is a greater number of wave lengths from the earth. Moreover the final result depends not only upon the EMF induced in the aerial, but equally upon the sensitivity of the receiver used.

The magnetron transmitter

When a relatively large amount of energy must be produced at a frequency of 1200 Mc/s (wave length 25 cm) the most suitable energy source is at present a magnetron. In general a magnetron is a diode with a cylindrical anode in one or more sections, and an electron emitting filament which is placed approximately along the axis of the cylinder. A constant magnetic field is applied in the direction of the filament. The paths of the electrons are thus curved, since the strength of the electric field E acting on the electron and the centrifugal force $m v^2/r$ are directed toward the outside, while the Lorentz force $e v H$ due to the magnetic field is perpendicular to the path and toward the inside (*fig. 3*). The latter

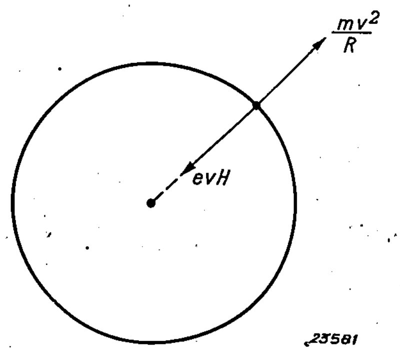


Fig. 3a. Forces acting on an electron in a homogeneous magnetic field when no electric field is present, but when the electron has a definite initial velocity.

component is proportional to the velocity of the electrons and the strength of the magnetic field. In the absence of an electric field the electrons move along a circular path whose radius depends upon the velocity of the electron. When a radial electric field is present, the path will be practically

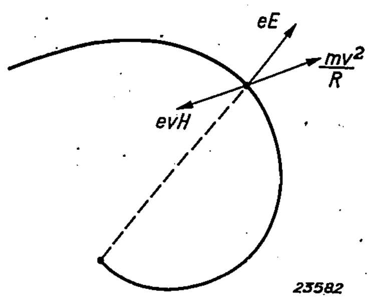


Fig. 3b. Forces acting on an electron in a magnetic field when a radial electric field is present.

heart-shaped. With a suitable anode voltage and magnetic field this electron motion leads to a negative resistance between the segments of the anode within a certain frequency range, so that it is possible to use the magnetron as an oscillator. If the negative resistance is greater than the ohmic resistance of the LC circuit of the tuned double wire feeder

(Lecher system) connected with the magnetron, any oscillation present increases in amplitude until an equilibrium state is reached at a greater amplitude due to the curvature of the characteristic. This holds of course for every oscillator circuit.

The problems which present themselves when a magnetron is to be employed for the purpose of communication include among others the question of modulation, and of the reliability of the wireless connection. Modulation of the anode voltage offers various difficulties. The frequency of a magnetron oscillator changes appreciably with the anode voltage, while in addition the characteristic which represents the high-frequency voltage as a function of the anode voltage is not only not straight, but exhibits discontinuities [3]). In order to have a steady signal strength, the energy sources must deliver a constant voltage. A constant magnetic field can easily be realized by the employment of a permanent magnet as shown in *fig. 4.* When

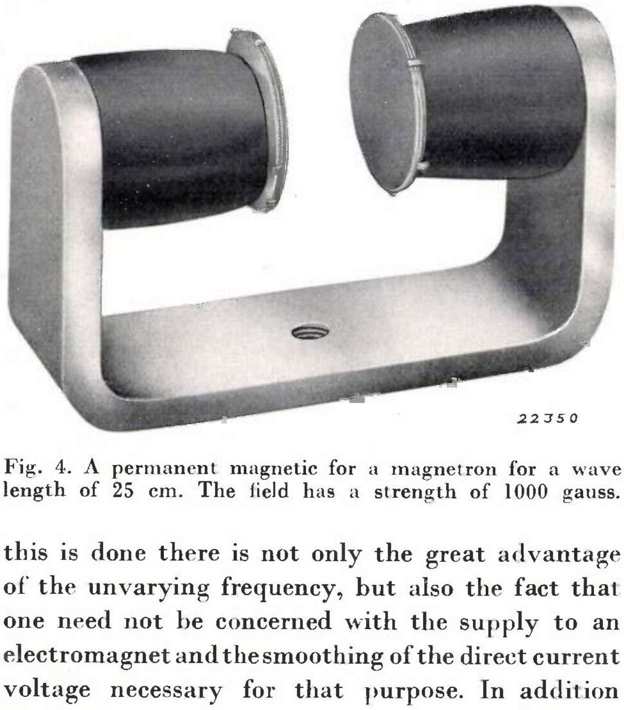


Fig. 4. A permanent magnetic for a magnetron for a wave length of 25 cm. The field has a strength of 1000 gauss.

this is done there is not only the great advantage of the unvarying frequency, but also the fact that one need not be concerned with the supply to an electromagnet and the smoothing of the direct current voltage necessary for that purpose. In addition to a saving in weight in the installation, this means an improvement in efficiency of the whole outfit. Another point is the question of the permissible variation in the anode voltage. This quantity depends upon the receivers used, as will be shown later, and may not be more than 0.1 volt. Thus the variation must be less than 0.1 per cent. This is achieved by means of a special voltage regulator based upon the following principle: a portion of the voltage which is to be kept constant is compared with a constant dry battery; any deviation of the voltage varies a resistance which is included between the source of the voltage and the transmitter. It is the principle of the inverse feed-back direct current amplifier. The filament current also must remain strictly constant. This can be achieved with the aid of a hydrogen-filled iron wire resistance lamp.

We have employed a special *system of modulation*, which was constructed for magnetron transmitters but is in general applicable for the modulation of oscillators whose frequency depends to a large degree upon the voltage to be modulated. The method of modulation comes to this, that *the amplitude* of the carrier wave is *not changed, but the time* during which the transmitter oscillates. When no low-frequency modulation voltages are applied to the transmitter, the oscillator is allowed to generate periodically in such a way that the period during wich it is generating is equal to the period during which it is not generating. The low-frequency modulation voltages cause a change in the ratio of these periods, in the sense that when the generating period is lengthened, the period in which no oscillations occur is equally shortened, and *vice versa.* The result is that the average high-frequency energy for each moment of the low-frequency period will correspond with the instantaneous amplitude of the low-frequency alternating current voltage when the interruption frequency is large with respect to the low frequency. This will in the end produce the same result in the receiver as if the aerial energy varied in the more usual manner. The attractive feature of the method is that neither the dependence of the anode voltage nor the disadvantage of the non-linear and discontinuous modulation characteristic plays the slightest part, although it must immediately be added that a disadvantage of the method lies in the fact that the range of frequencies transmitted is limited to about 20 000 c/s. for practical reasons which will be mentioned later.

The periodic oscillation and non-oscillation can be brought about in different ways. One of the methods which we used and which is employed in the installation described consists in the following. A variable resistance is introduced in parallel with the magnetron, and the magnetron (and therefore also the resistance in parallel with it) is fed through a constant resistance. If the variable resistance is infinite, the magnetron works normally; if the variable resistance is made smaller, it begins to take some current and the voltage over the common resistance will therefore change, and with this the voltage of the magnetron also. There are now two

[3]) C. G. A. von Lindern, Transmitt. News 2, No. 2 (195).

courses open: in the first place one may make the current consumption of the variable resistance so large that the magnetron ceases to generate, for which

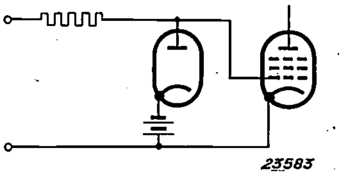


Fig. 5. The diode with series resistance to limit the voltages in the grid circuit of the pentode.

purpose the anode voltage need not be greatly changed. In the second place one may make the current consumption of the variable resistance smaller so that, while a large variation in frequency occurs, the oscillation is not interrupted. In the latter case therefore two waves are radiated so that the reliability of a connection is in general improved. It will depend further upon the receiving system used which method is chosen, since the two wave lengths lie relatively close to each other, and the result might be that with a less selective receiver one would hear nothing at all of the modulation. We have however chosen the latter method since a sufficiently selective superheterodyne receiver was being used, and moreover because a slight variation in current makes the transmission of the lower register easier, due to the fact that then the smoothing condenser of the high voltage feeder may be smaller. For the variable resistance we used a pentode, the first grid of which we made periodically highly negative "cut off" and then less negative. We achieved this by means of a saw-tooth alternating voltage whose amplitude was large with respect to the necessary variation in the grid voltage of the pentode. When this grid voltage is about zero (the saw-tooth voltage has the extreme values of $+A$ and $-A$) the change over takes place from infinite resistance to a smaller resistance. The way in which that resistance becomes constant after the change, and how at the same time the grid of the pentode is protected from overload will be explained later. At present the important fact is that the change takes place in a short interval of time; it must of course take place during a small part of the period of the saw-tooth voltage whose frequency is large with respect to that of the low-frequency voltages. It is therefore possible to cause this change to take place at any moment of a period of the saw-tooth voltage by introducing a direct voltage in series with the saw-tooth voltage. When this direct voltage is $+A$, the variable voltage will never be infinite in other words, the magnetron will no longer generate, while, when the direct voltage is $-A$, the generating of the magnetron will be uninterrupted.

If instead of a direct voltage we introduce in series with the saw-tooth voltage an alternating voltage whose frequency is small with respect to the auxiliary frequency, the magnetron will also generate when the alternating voltage is negative, and will not generate when the alternating voltage is positive. One may therefore modulate the magnetron transmitter in this way. When the amplitude of the low-frequency alternating voltage is A, that is,

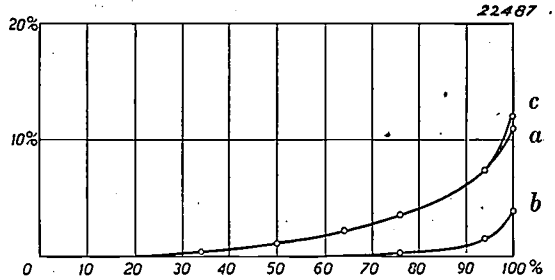


Fig. 6. The distortion as a function of the depth of modulation when a sinusoidal auxiliary voltage instead of a sawtooth voltage is employed.

equal to that of the saw-tooth voltage, the depth of modulation is 100 per cent. In any case of course when the low-frequency alternating voltage has the maximum amplitude A the depth of modulation is at a maximum.

In order to provide that immediately after the pentode has attained the correct resistance this state will be maintained during the rest of the saw-tooth period which is making the pentode grid more positive, a diode is introduced between the grid and the filament of the pentode, as shown in *fig. 5*. The diode has a very small internal resistance and the series resistance is of such dimensions that it is large with respect to the internal resistance of the diode. Thus when the anode of the diode is just slightly positive with respect to the cathode, the saw-tooth voltage will be taken up by the series resistance, and a practically constant voltage between the grid and the filament of the pentode remains. It is therefore clear that the grid of the pentode is protected against overloading, but also that the relaxation times in the grid circuit, which are determined by resistance and capacity, (and this holds also for the anode circuit) begin to play a part. This is the reason why the frequency band to be transmitted is limited to about 20 000 cycles per sec. A simplification of the installation is possible by the use of a sinusoidal auxiliary voltage instead of a saw-tooth voltage. *Fig. 6* shows the distortion caused thereby as a function of the depth of modulation. It may be seen that this distortion is not serious.

Figs. 7, 8 and *9* are photographs of the transmitter

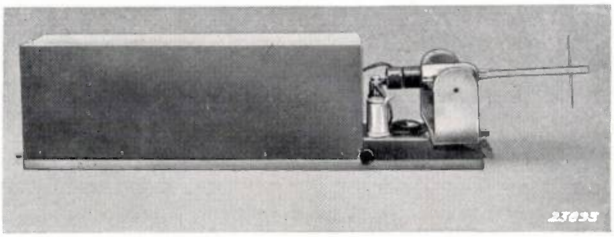


Fig. 7. Photograph of the transmitter.

installation used, of the magnetron and of the parabolic reflector.

The receivers

As in the case of the receivers used in the link with Tilburg which was maintained on a wave length of 125 cm, here also use is made of the superheterodyne principle.

In order to transmit with the same bandwidth as in the receivers for 1 m waves, the decision was made to try to keep the wave length of the transmitter and of the oscillator in the receiver sufficiently constant. One can then go to still shorter waves without encountering the practical impossibility of constructing a suitable intermediate frequency amplifier.

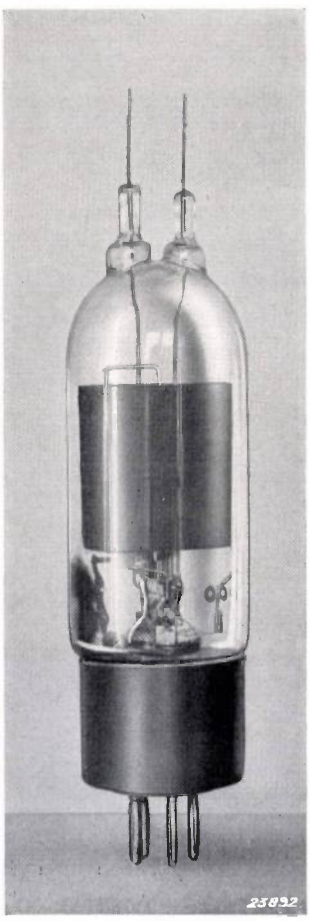


Fig. 8. Photograph of the magnetron used.

One may use an oscillator which generates either practically the same wave length as that which is to be received — namely, one which differs from it by 7.5 Mc/s, that is, the intermediate frequency which was also used in the receivers of the Eindhoven—Tilburg link — or one which generates a

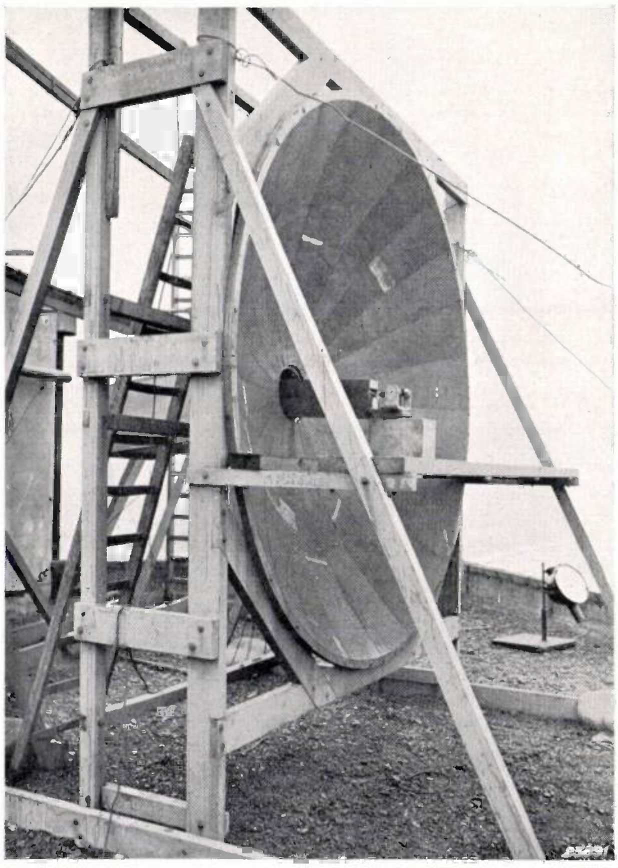


Fig. 9. Photograph of the parabolic reflector on the tower of the lamp factory in Eindhoven.

wave which is an integral number of times longer, and then make use of one of its harmonics. For the first method Barkhausen and possibly Pierret valves, as well as magnetrons, are available. For the second method Barkhausen or Pierret valves or triode generators may be used.

For generating Barkhausen or Pierret oscillations a triode with a tungsten filament, a cylindrical grid and a cylindrical plate is usually used. The grid is made positive with respect to the cathode, the "anode" slightly negative. A Lecher system, consisting of two parallel wires over which a plate, the bridge B, can be moved, is connected to grid and anode (*fig. 10*). With the proper distance

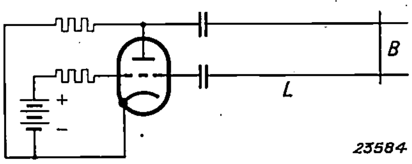


Fig. 10. Diagram of a Barkhausen valve with Lecher system.

between bridge and valve, oscillations are found to be generated whose wave length is dependent upon the anode voltage and also upon the position of the bridge.

In explanation it is usually assumed that the electrons carry out an oscillating motion around the grid which, being positive, attracts the electrons. A portion of them will shoot through the mesh of the grid instead of landing upon it, they will turn round in front of the negative "anode", shoot through the grid meshes again, turn round once more in front of the cathode, and so on. If all the electrons go through this process independently of each other, no alternating voltage will appear between the electrodes. Upon closer consideration however there seems to be a certain reciprocal action between the potentials of the electrodes and the motion of the electrons, and, as a matter of fact, between the electrons themselves. This reciprocal action causes the disordered motion to become ordered.

One may now conduct the oscillation picked up by the aerial to the Barkhausen valve. Because of non-linearity a certain degree of mixing will occur and the beat frequency between the Barkhausen oscillation and the oscillation picked up may be conducted to the intermediate-frequency amplifier.

The Pierret valve acts on the same principle as the Barkhausen valve, but there is no Lecher system between grid and anode for tuning. The grid itself (*fig. 11*), constructed in the form of a spiral without a supporting rod, is tuned in the Pierret valve.

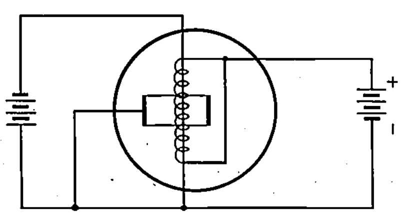


Fig. 11. Diagram of a Pierret valve.

It was found possible to construct sufficiently sensitive receivers in this way, but the control was less simple than that of triode generators, while there were no suitable indirectly heated valves available, and therefore alternating current supply could not be employed.

Further, there is the disadvantage that the frequency depends very much upon the grid voltage so that rigorous stabilization is necessary.

Later we chose the second method; the triode generator working on a wave several times longer than the wave to be received. The generator can be tuned to waves between 110 and 130 cm. The tuning is done by rotating a condenser. This oscillation of somewhat more than 1 m wave length is conducted to a diode together with the oscillation received. An acorn diode may be used here and the diode serves as a mixing valve. The circuit is shown in *fig. 12*. The coupling coil L is permanently tuned,

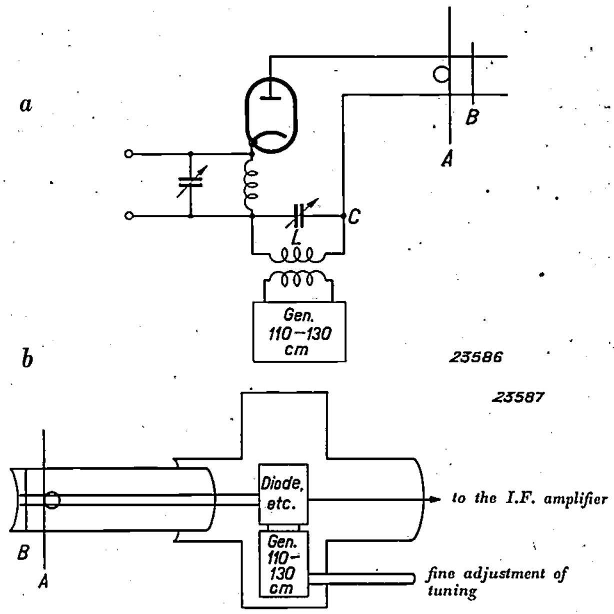


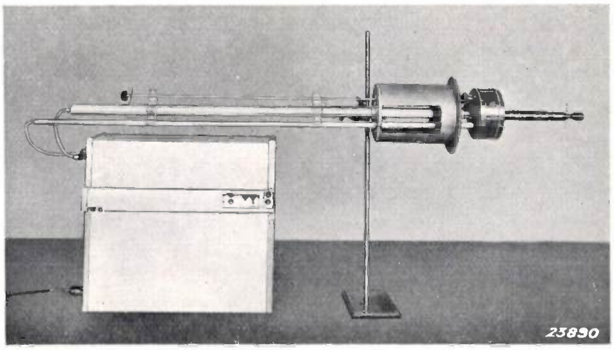
Fig. 12. *a*) Circuit, *b*) diagrammatic representation and construction scheme of the diode mixing stage.

for instance, to 120 cm. This tuning is so flat that no purpose is served by additional adjustment when the generator is turned. The wave to be received reaches the diode through the Lecher system BC which is tuned by shifting the bridge B. The aerial A is coupled inductively with the Lecher system. The position of the bridge must be correct within several millimctres for adequate sensitivity, fine adjustment being unnecessary. Since the Lecher system is connected directly to the intermediate frequency amplifier *via* coil L everything must be adequately screened. If the screening is inadequate, disturbing signals from transmitters working on a wave length of about 40 m penetrate to the receiver.

The middle of the loop of the aerial must also be connected to the screening arrangement.

In practice screening is obtained in the manner shown in *fig. 13*. The generator and the diode are mounted in a brass box provided with a tube 2 cm in diameter and 15 cm in length. A second tube slides inside the first. Within the second tube the insulated bridge is introduced; the Lecher wires run through the two tubes and pass through holes in the bridge. The aerial is attached at its middle point to the narrower tube and sticks out at both ends through large openings. The tube extends somewhat beyond the bridge in order to screen it and the remaining

a



b

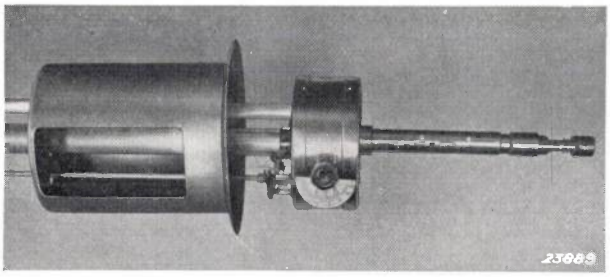


Fig. 13. *a*. represents the mixing stage, intermediate-frequency amplifier and low-frequency amplifier, *b*. the mixing stage separately on a larger scale.

part of the Lecher system. The intermediate frequency oscillation excited in the diode is conducted to the intermediate-frequency amplifier by a wire stretched in a tube fastened to the back of the box containing the generator and diode. This box is then fixed in the paraboloid or just behind it, depending upon the focal distance. The paraboloid is set up outdoors, outside a window of the water tower in our case, the intermediate-frequency amplifier, etc. are inside.

The fine adjustment of the tuning is carried out by turning a fine adjustment condenser, which is connected in parallel with the main condenser. For this purpose another rotating rod is introduced along the tube which leads in the intermediate-frequency voltage.

The voltage available at the input side of the intermediate-frequency amplifier is of the same order of magnitude as for the wave of 125 cm. Good telephone communication is possible with a

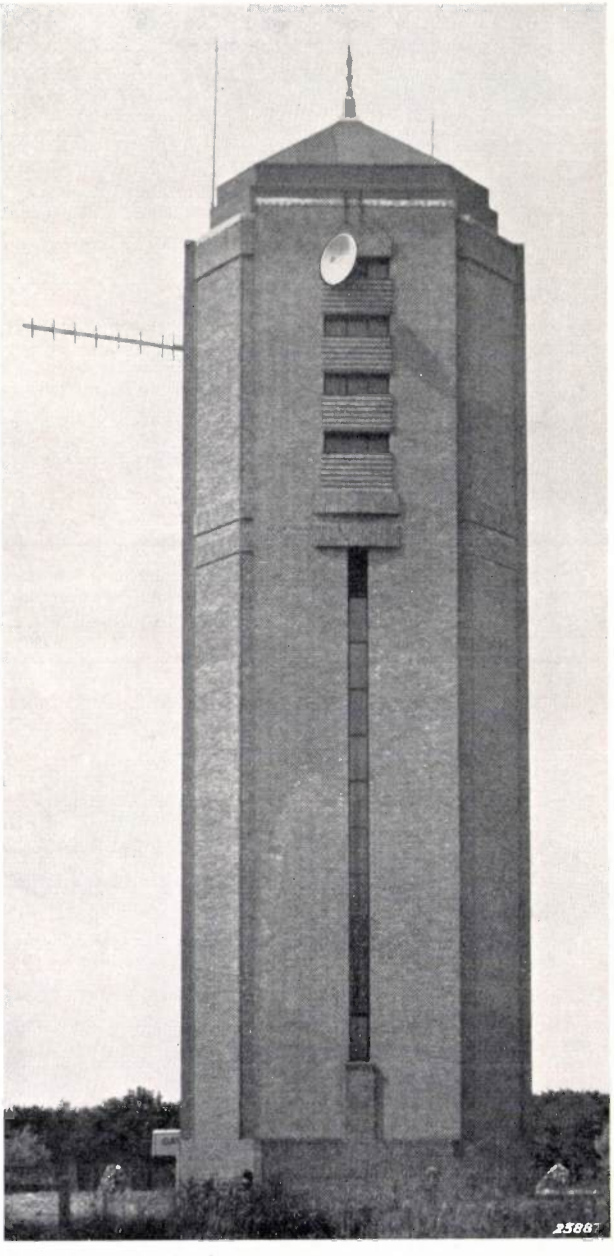


Fig. 14. Photograph of the tower in Nimeguen with the aerial system.

reflector at the receiving end having a diameter of about 1 m; for telegraphy a receiver without a reflector is sufficient. In both cases the reflector had a diameter of about 3 m at the transmitting end.

For use with large paraboloids with a focal distance of about 1 m, a long narrow receiver was developed, in which the intermediate-frequency amplifier with the second detector and the low-frequency amplifier were mounted end to end, and the whole combined with the mixing stage and the feeder to give a unit which was set up along the axis of the paraboloid, so that as little shadow as possible was caused.

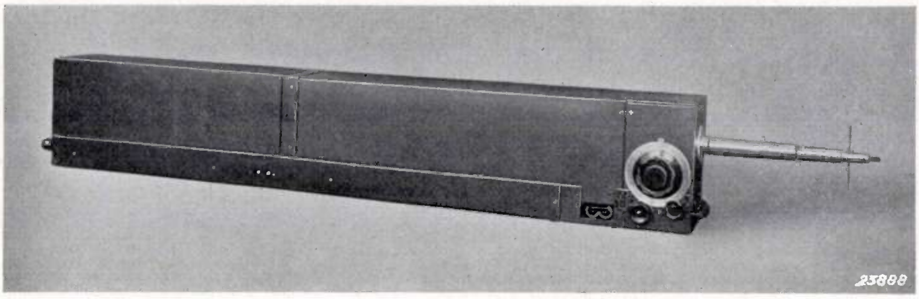


Fig. 15. Photograph of the long receiver in the paraboloid.

In *fig. 14* may be seen a picture of the long receiver and in *fig. 15* a photograph of aerial system used at the receiving end.

JOINTS BETWEEN METAL AND GLASS

by H. J. MEERKAMP VAN EMBDEN.

Summary. A survey is given of metal-glass joints and the demands made upon them in practice. The application of chrome iron to glass seals (particularly in gas discharge tubes) is discussed in some detail.

Introduction

Since Plücker and Geissler carried out their experiments with the forerunners of the modern neon tubes, gas discharge tubes have passed through a vast development, and are not only used for illumination but are finding a continually wider application in the field of radio and heavy current technique. Although originally the main object in using Geissler tubes was the study of interesting phenomena, in connection with technical applications it is now often important to be able to dissipate as great an electrical power as possible within a tube of given dimensions. This power, as in the case of practically all electrical machines, is limited by the heating permissible in relation to the construction and kind of material. Examples of heavily loaded discharge tubes have been described repeatedly in this periodical, some of these are the metal rectifier with mercury cathode [1]), the high pressure mercury lamp [2]), the water-cooled transmitter valves [3]), etc.

[1]) A new metal valve with a mercury cathode, Philips Techn. Rev. 1, 65, (1936).

[2]) The mercury lamp HP 300, Philips Techn. Rev. 1, 129 (1936).

[3]) The development and manufacture of the modern transmitter valves, Philips Techn. Rev. 2, 122, (1937).

The choice of material for heavily loaded tubes

The endeavour to attain maximum power in the tube leads to new aspects in the choice of material.

The wall of a discharge tube must be competely impervious and the current leads must pass through this wall without the possibility of gas leakage in either direction.

At the same time the wall must be of material suitable to insulate the electrodes from each other. But in addition to this the wall fulfills still another function: if the tube is used for illumination purposes, it must transmit the radiation, while it must give up to its surroundings by conduction and radiation the energy freed as heat in the tube. Because of this the use of glass-like substances for the walls of the tube is indicated.

In the technical development of discharge tubes, as is clear from the above, limitations are encountered especially with regard to the last point. Upon increasing the power the energy dissipated as heat increases also and the wall becomes steadily hotter.

When glass apparatus is used the danger of softening or cracking of the glass with great heat is considerable, and if an attempt is made to was developed, in which the intermediate-frequency amplifier with the second detector and the low-frequency amplifier were mounted end to end, and the whole combined with the mixing stage and the feeder to give a unit which was set up along the axis of the paraboloid, so that as little shadow as possible was caused.

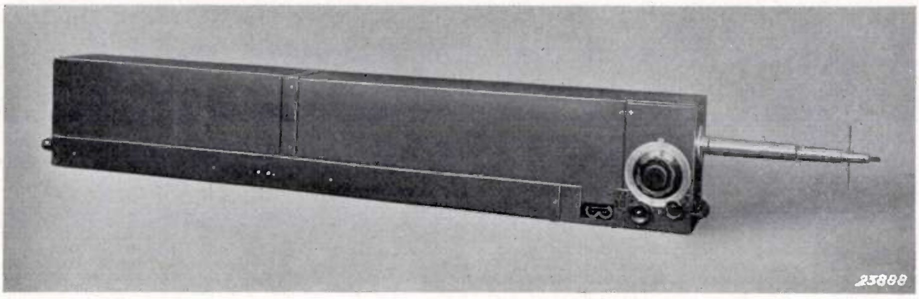


Fig. 15. Photograph of the long receiver in the paraboloid.

In *fig. 14* may be seen a picture of the long receiver and in *fig. 15* a photograph of aerial system used at the receiving end.

JOINTS BETWEEN METAL AND GLASS

by H. J. MEERKAMP VAN EMBDEN.

Summary. A survey is given of metal-glass joints and the demands made upon them in practice. The application of chrome iron to glass seals (particularly in gas discharge tubes) is discussed in some detail.

Introduction

Since Plücker and Geissler carried out their experiments with the forerunners of the modern neon tubes, gas discharge tubes have passed through a vast development, and are not only used for illumination but are finding a continually wider application in the field of radio and heavy current technique. Although originally the main object in using Geissler tubes was the study of interesting phenomena, in connection with technical applications it is now often important to be able to dissipate as great an electrical power as possible within a tube of given dimensions. This power, as in the case of practically all electrical machines, is limited by the heating permissible in relation to the construction and kind of material. Examples of heavily loaded discharge tubes have been described repeatedly in this periodical, some of these are the metal rectifier with mercury cathode [1]), the high pressure mercury lamp [2]), the water-cooled transmitter valves [3]), etc.

[1]) A new metal valve with a mercury cathode, Philips Techn. Rev. 1, 65, (1936).

[2]) The mercury lamp HP 300, Philips Techn. Rev. 1, 129 (1936).

[3]) The development and manufacture of the modern transmitter valves, Philips Techn. Rev. 2, 122, (1937).

The choice of material for heavily loaded tubes

The endeavour to attain maximum power in the tube leads to new aspects in the choice of material.

The wall of a discharge tube must be competely impervious and the current leads must pass through this wall without the possibility of gas leakage in either direction.

At the same time the wall must be of material suitable to insulate the electrodes from each other. But in addition to this the wall fulfills still another function: if the tube is used for illumination purposes, it must transmit the radiation, while it must give up to its surroundings by conduction and radiation the energy freed as heat in the tube. Because of this the use of glass-like substances for the walls of the tube is indicated.

In the technical development of discharge tubes, as is clear from the above, limitations are encountered especially with regard to the last point. Upon increasing the power the energy dissipated as heat increases also and the wall becomes steadily hotter.

When glass apparatus is used the danger of softening or cracking of the glass with great heat is considerable, and if an attempt is made to decrease the amount of energy given off per unit of surface by making the tubes larger, the limit of the technically possible is soon reached, since the implosion of large evacuated containers is very dangerous.

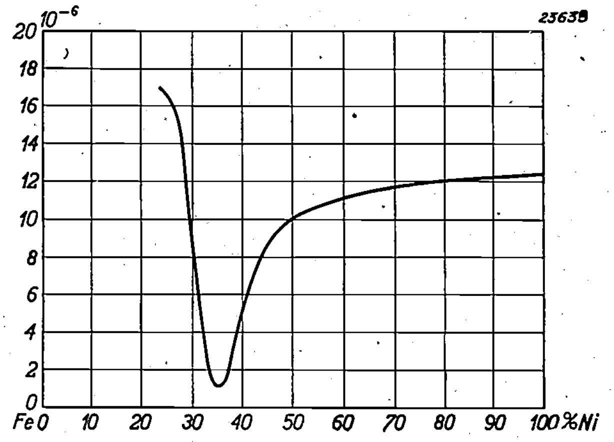


Fig. 1. Coefficient of expansion of nickel iron at 20° C as a function of the composition. With about 36 % Ni (invar) the coefficient of expansion has a sharp minimum. A coefficient of 100×10^{-7}/°C, thus equal to that of ordinary glass, is reached with about 50 % Ni. Int. Crit. Tables 2, 465, 1927.

It is obvious that a better material must be sought for heavily loaded tubes. For high pressure mercury lamps the use of quartz, for example, which softens only at higher temperatures than glass, is an important step. In general, however, glass seems to be an outstandingly suitable material, which should only be avoided at those points where the greatest heating effect occurs. If, for instance, one could substitute a metal for the glass at the points of greatest heat development, the danger of cracking and breaking would be avoided, while the removal of heat could be made much more rapid by the use of a water jacket. Moreover the

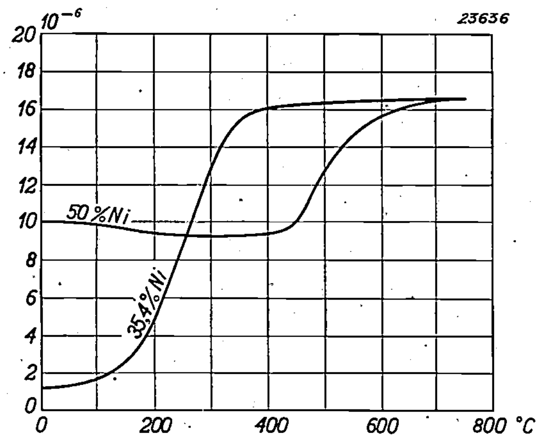


Fig. 2. Coefficient of expansion of nickel iron as a function of the temperature. a) 36 % Ni. The coefficient of expansion remains very low up to 300 °C. b) 50 % Ni. The coefficient of expansion is uniform up to 500 °C. Int. Crit. Tables 2, 465, 1927.

accuracy with which metal can be worked is much greater than can be obtained with glass, so that the construction and installation of a partly metal tube is much simpler. The sealing of glass to a metal is not so simple however. Due to the difference in coefficients of expansion the glass usually breaks upon cooling when it is fused to a metal.

In order to prevent this there are two possibilities:

1) make the metal so thin, that upon cooling it may be plastically deformed without causing the glass to break,
2) find a metal whose expansion corresponds with that of the glass [4]).

Both methods are employed; for the first copper (linear expansion coëfficient between 0 and 300 °C, average 168×10^{-7}/° C) is usually used as the metal. It must be very thin (about 0.1 mm) in order to be welded to glass. The procedure is, however, quite difficult and expensive, due to the tendency of the copper to be oxidized, while great strength cannot be expected of the thin material near the weld.

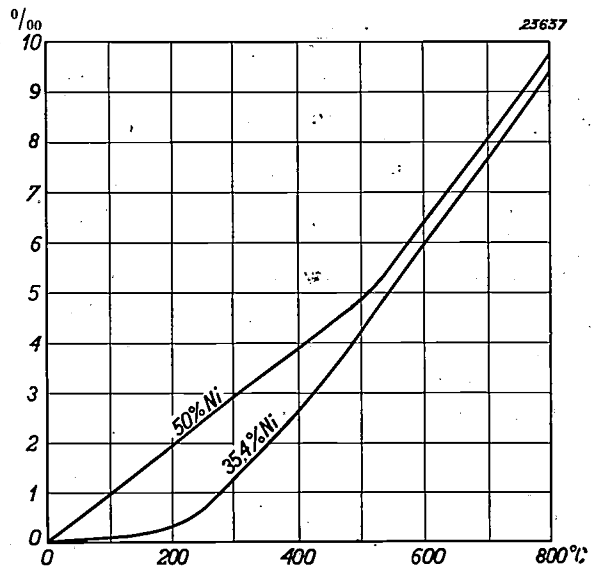


Fig. 3. Change in length as a function of the temperature. a) with 36 % Ni. b) with 50 % Ni. Int. Crit. Tables 2, 465, 1927.

For the second method, the use of a metal adapted to the expansion of glass, we shall first discuss the requirements which the metal must satisfy. They are the following:

1) Uniform expansion throughout a long temperature range, from room temperature to above the softening point of the glass.

---

[4]) This does not necessarily mean that the coefficients of expansion must be exactly the same. Conversely, with similar coefficients of expansion, differences in expansion may occur, because of the fact that the metal and the glass cool off at different rates. In this connection it is desirable that the metal does not have too great a heat conduction.

2) The metal must be sufficiently stable during the sealing-in.

3) The adhesion to the glass must be good.

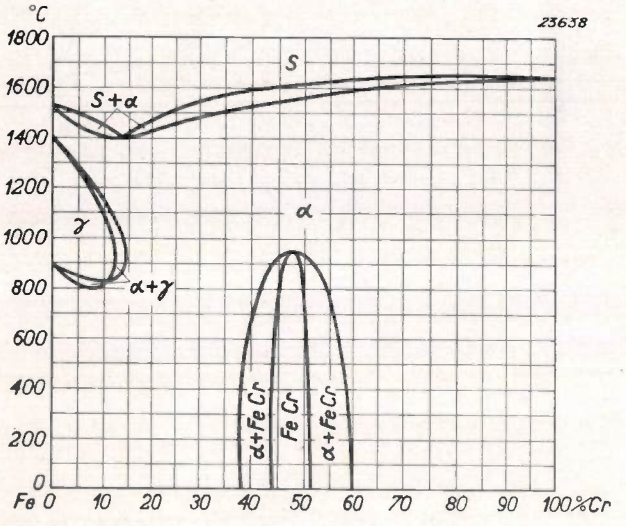


Fig. 4. The iron-chromium diagram. Taken from: Ed. Houdremont, Einführung in die Sonderstahlkunde, Springer 1935, p. 182.

4) The amount of gas given off by the metal must be small.

5) The price must be reasonable.

The significance of these requirements will be briefly explained in the following.

1) The expansion of glass is uniform from room temperature to the softening point, but varies considerably for different kinds of glass. Consequently a metal which is suitable for sealing in to normal calcium glass (coefficient of expansion about $100 \times 10^{-7}/°$ C), is not suitable for sealing

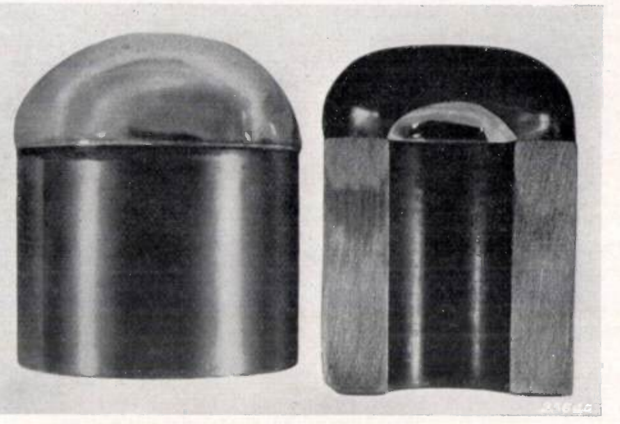


Fig. 5. Example illustrating the possibility of making thick chrome iron-glass joints.

into hard glass with a coefficient of expansion of 40 to $50 \times 10^{-7}/°$ C. The metal to be sealed in must possess uniformity of expansion over a wide range, from cold to the softening point.

If allotropic transitions occur in a metal upon passing certain temperatures, for example a transition from α to β-structure, the transitions are usually accompanied by volume changes. Stresses in the metal occur which may cause cracking of the glass. It is therefore necessary to choose a

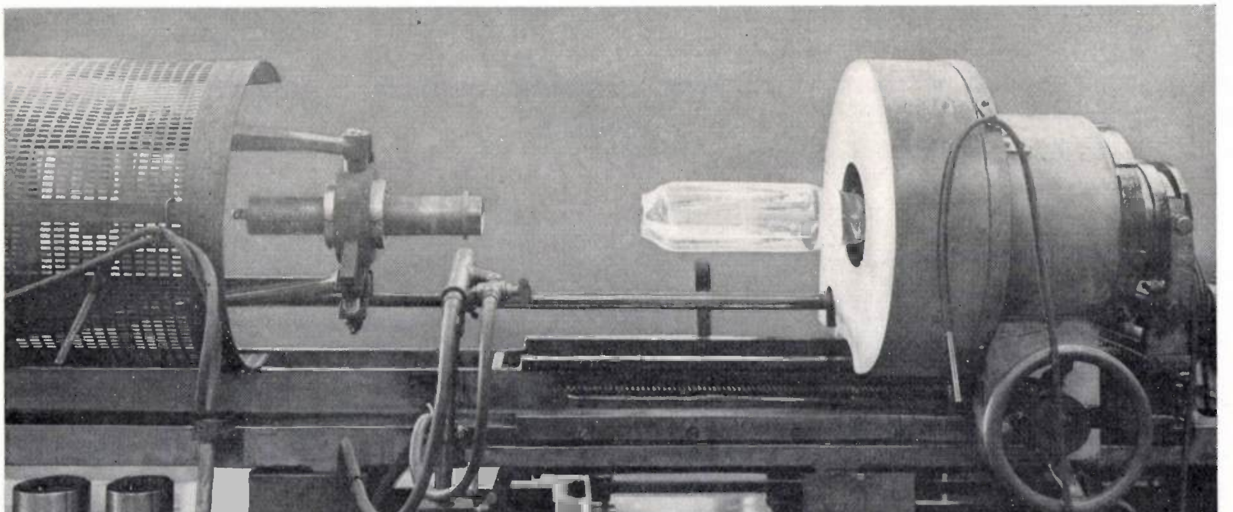


Fig. 6. Chrome iron tube with glass bulb (for transmitting valve) on the sealing-in machine.

metal or an alloy which exhibits no allotropic transitions in the temperature range in which the glass is hard.

2) The sealing-in to the glass takes place at high temperatures at which glass flows; at these temperatures the metal must not yet be soft and must not burn.

3) The seal may be sufficient in some cases due to adhesion alone. The classic example of this is the glass-platinum seal. The metal remains quite bright in welding and provides an excellent seal. In the case of most metals and alloys, however,

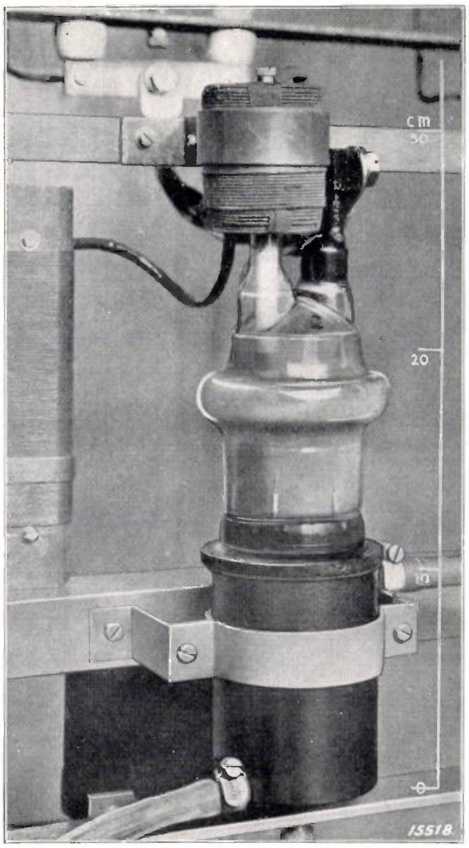


Fig. 7. A water-cooled metal valve with a mercury cathode, described in detail in Philips Techn. Rev. 1, 65 (1936). Average current strength 75 A, peak current 500 A. The chrome iron cylinder is water-cooled. It contains a pool of mercury and serves as cathode. The upper half of glass is necessary for insulation of the auxiliary anode (upper left) and the main anode (upper right) which is introduced by means of a chrome iron cap.

this is not the case and the adhesion between metal and glass is not sufficient, while at the same time the metal becomes covered with an oxide layer upon heating. This film may in some cases serve to promote a good seal when it functions as an intermediate layer between glass and metal. On one side the film is firmly bound to the metal, on the other it is to some extent dissolved in the glass forming a silicate and produces a gradual transition at this point. If, however, the oxide layer is thick and porous or has the tendency to form loose flakes, it

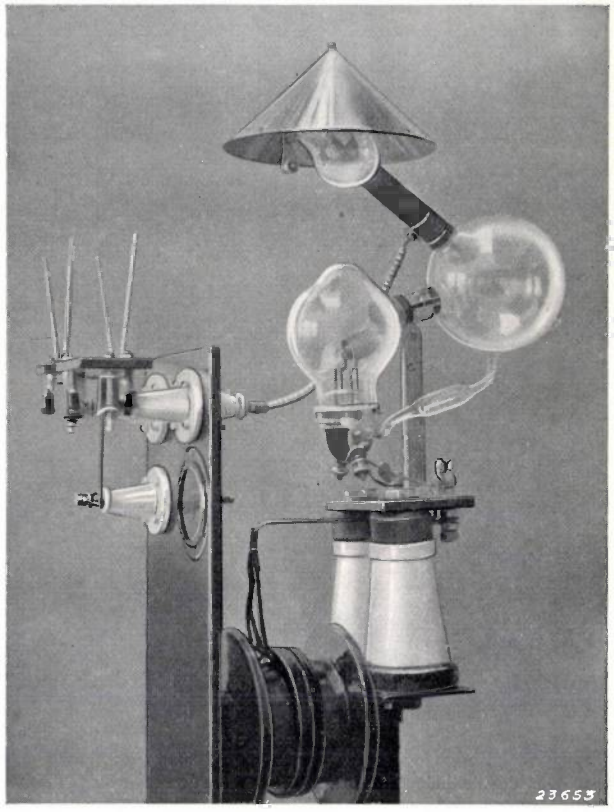


Fig. 8. The rectifier valve DCG 10/15, example of a more complicated construction of glass and chrome iron. Lower bulb: cathode space, upper bulb: anode space. The third bulb serves as a condensation chamber for the mercury evaporating in the lower bulb. The mercury pressure near the anode is thus kept low (little back ignition) with a high mercury pressure in the neighbourhood of the cathode (low cathode drop). The chrome iron tubes serve as auxiliary electrodes. They must be at a positive potential to ensure ignition. At the same time there is therefore the possibility of using the rectifier as a relay valve.

is obvious that there can be no question of a gas-tight seal.

4) During the sealing, and afterwards too, the metal should contain no gasses in the adsorbed

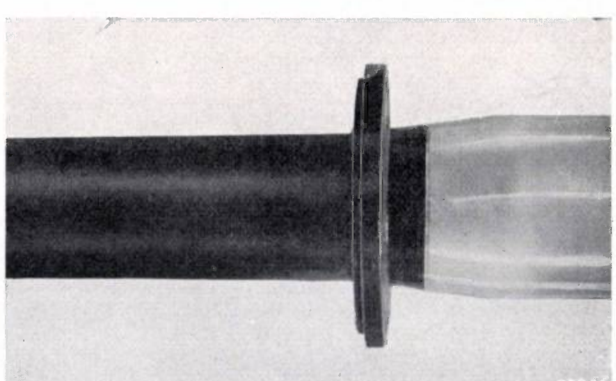


Fig. 9. Anode of a metal transmitting valve with chrome iron-glass seal.

or dissolved state, which could be freed during working. This would prevent a perfect seal. At the interface between glass and metal gas bubbles would occur which could cause leakage.

Also after the seal is finished no more gas may escape from the metal, since it would spoil the vacuum or the gas atmosphere in the tube.

It has long been known of platinum, with a linear coefficient of expansion of 90×10^{-7}, that it gives vacuum-tight seals, and indeed the first electric lamps had platinum leads. The price is however so high that attempts were immediately made to find cheaper alloys, and at present the platinum lead is of interest only in the laboratory because

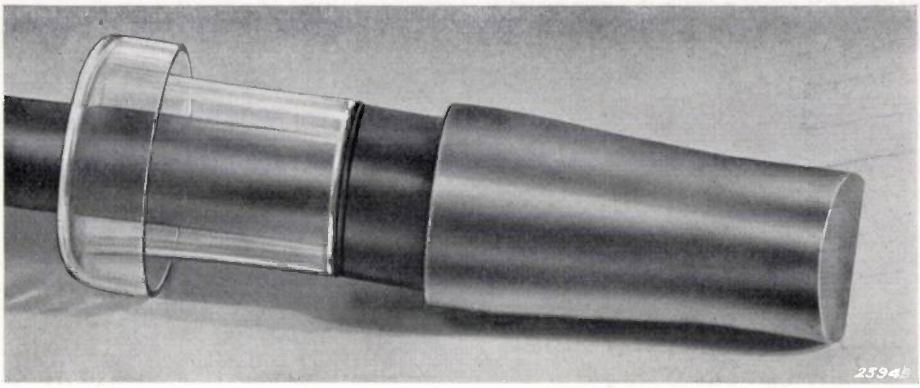


Fig. 10. The heavy anode of a "Metalix" X-Ray tube is sealed to the glass with a chrome iron ring.

5) The high costs in many cases prevent the use of a noble metal such as platinum.

Let us now list the metals which may be considered suitable for sealing to glass. For ordinary glass with a coefficient of expansion of 85 to $100 \times 10^{-7}/°$ C we find the following possibilities:

platinum
nickel iron
chrome iron

For hard glass with a lower coefficient of expansion there are other suitable metals and alloys which correspond in expansion with hard glass. A discussion of these would lead us too far afield. Use is usually made of ternary iron-nickel-cobalt alloys and also pure tungsten and molybdenum.

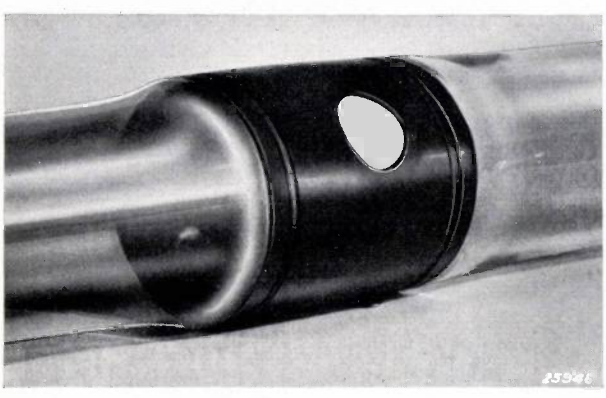


Fig. 11. The middle section of a "Metalix" X-Ray tube is made of chrome iron for screening- off undesired radiation. The X-rays are emitted exclusively through the glass window.

of the fact that the sealing-in of such leads is so simple, the ends of the wires are always bright and ductile and the metal is so extraordinarily stable chemically.

Of the nickel-iron alloys invar, with about 36 per cent of nickel, is known for its very slight coefficient of expansion. *Figs. 1* and *2* give the coefficients of expansion of Fe-Ni alloys as a function of the composition and of the temperature. If we heat such alloys we see that the coefficient of expansion between room temperature and 150° C is practically equal to zero, but that above that temperature (*fig. 3*) we obtain a horizontal line up to 200 ° which turns sharply upward at 200 °.

If we now add more nickel to the alloy we find that the horizontal line begins to slope upward and the point where the slope suddenly changes is displaced to higher temperatures. An alloy with about 50 per cent of nickel between room temperature and 500 ° has a coefficient of expansion equal to that of glass, above the latter temperature it is somewhat higher. This is however not dangerous, since at those higher temperatures the glass is soft and therefore will not crack.

With regard to expansion therefore the latter alloy is a suitable one, and is in fact very often used as sealing-in wire for lamps but it has however several disadvantages particularly with regard to adhesion. Upon sealing it in, the wire is strongly oxidized and the oxide has a great tendency to flake off, while in addition it is very difficult to make this material gasfree. Nickel absorbs all kinds of gasses very easily (its use as a catalyst is an example of the application of this property), and these gases are easily given off again. That is the reason why the sealing-in wires are often covered with a thin layer of copper

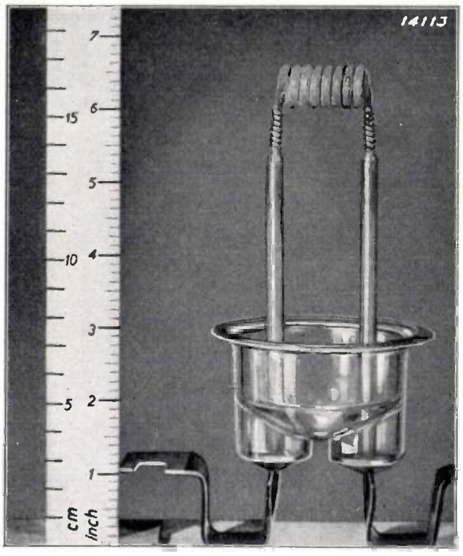


Fig. 12. Simple lead-in in a rectifier, consisting of a chrome iron plate with terminals soldered on to both sides.

(borated copper), since the adhesion of glass to copper by means of the film of copper oxide is very firm. Since copper has a much higher coefficient of expansion, the copper coating must be very thin so that it may be plastically deformed without cracking the glass. For larger seals nickel iron is out of the question.

On the other hand for large seals chrome iron has found extended application. Before we consider its application we shall first study the iron-chromium system. The diagram of the iron-chromium system is given in *fig. 4*. When pure iron is heated, it passes over into another modification, γ-iron, at 906 °C. γ-iron has a different atomic arrangement than the α-iron [5]) stable below this temperature.

Upon further heating γ-iron passes over at 1401° C into α-iron again which is identical with the α-iron stable below 906 °C. α-iron melts at 1528 °C. Upon cooling the transitions occur at the same points, solid at 1528°, α-iron to 1401°, γ-iron to 906° and α-iron lower. If we add chromium as alloying element we find that the two transition points approach each other with increasing chromium content; with about 15 per cent of chromium they coincide. An alloy with more than 15 per cent chromium and the rest iron consists therefore only of α-crystals between room temperature and the melting point.

That means that the expansion due to change of temperature proceeds quite uniformly without the transitions $\alpha \rightarrow \gamma$ which are accompanied by large changes in volume. Moreover the coefficient of expansion is 100×10^{-7} and corresponds therefore very well with that of glass.

Technical procedure in sealing

The alloy is melted in an electric furnace for which the electric arc furnace as well as the more modern high frequency furnace may be used. It is then poured into casting moulds and forms ingots. The ingots are now rolled into rods or sheets and the pieces intended for sealing in to glass are made from these by turning, drawing, etc. Thin sheet can very well be drawn, care must be taken however in rolling that the material is not given an anisotropic texture by cold working, since this would give bad results in the drawing [6]).

Chrome iron is fairly resistant to air. This stability is ascribed to the formation of an extremely thin film of oxide which entirely covers the surface and protects it from further attack by the atmosphere. Various chromium alloys are on the market as stainless steel. In our case the oxide film forms a

Fig. 13. Three electrodes of a rectifier valve are led in by means of sector-shaped chrome iron plates. The three plates are first welded together with a glass edge to give a circular plate and then sealed on to the circular opening of the rectifier bulb.

[5]) The atomic arrangement for α and γ-structure is shown in Philips Techn. Rev. 2, 253, (1937).

[6]) See Philips Techn. Rev. 2, 156, (1937). for the drawing of chrome iron cups (Example **23** for X-Ray testing of materials).

connecting substance between glass and metal, and it is obtained in the proper thickness for good adhesion by heating the chrome iron with the flame directly before the sealing into the glass.

Fig. 5 shows that even thick pieces of glass and chrome iron may adhere very well to each other. The fact that such a sealing-in may be carried out with chrome iron is due in part to the fact that the heat conduction of this alloy is very small.

For sealing large pieces a special lathe has been made (*fig. 7*) with two chucks opposite each other. In one chuck the chrome iron in tube form is held, in the other the glass part. While both are turning the edge of the chrome iron is first heated and oxidized, after which a thin layer of glass is deposited by hand upon the hot edge by applying a thin glass rod to it. After that the edge of the piece of glass to be welded is heated and the two chucks are brought together, still rotating. The glass seal is then obtained as an ordinary glass butt weld by fusing. The edge of the chrome iron tube in the above described process need not be sharp or even thin, a fact which is favourable to the strength of the glass-metal seal. In this respect chrome iron is distinct from all other ordinary sealing-in alloys which can only be sealed into glass in the form of thin, sharply beveled edges. *Figs. 7* and *8* show two important examples of the glass-chrome iron butt weld in the construction of large rectifiers. In the case of the very large water-cooled transmitting valves the sealing ring on which the glass is fused is made of chrome iron (*fig. 9*). This ring is soldered to a copper tube which is water-cooled. The heat conduction of copper is much better than that of chrome iron, so that more heat can be conducted away in this manner. The same principle is used in the case of the anodes of X-ray tubes (*fig. 10*), where a very high heat conduction is required. The anode therefore consists of a block of copper which is fused to one edge of a chrome iron ring, the other end of which is then joined to the glass.

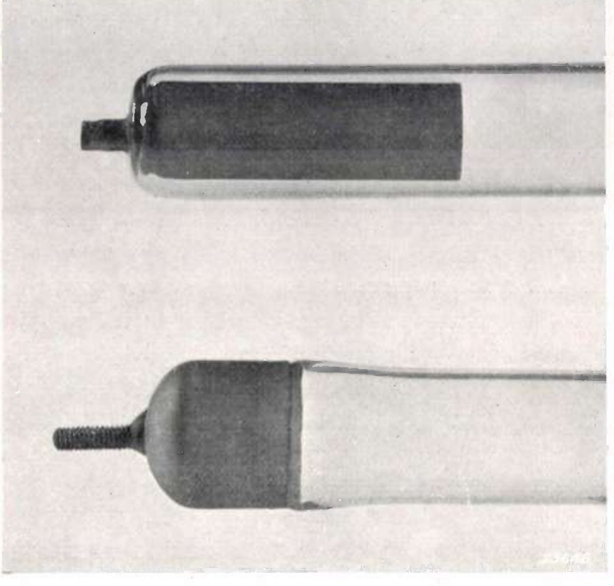


Fig. 14. The use of chrome iron anodes in neon tubes leads to a very simple construction.

The middle sections of the Philips "Metalix" X-ray tubes are also made of chrome iron (*fig. 11*) and a window for transmitting the beam of rays is sealed in. This would not be so simple with metals other than chrome iron. The tube is mounted in the apparatus by this metal middle section.

If it is desired to make current lead wires, for very heavy currents one may take a chrome iron cap with a turned edge on which a glass tube can be sealed in the manner described. The cap is then provided with a lead wire in the middle of the inner side which is soldered to the cap, and with a second wire on the outer side. Afterwards the whole can be sealed into the glass tube.

For smaller current strengths, for example in the rectifier given in *fig. 12*, a chrome iron plate is sufficient, which is provided on both sides with glass. It is then sealed into the glass tube. Several such plates may be "joined to a foot", which is sealed into the rectifier valve as a whole after assembly. *Fig. 13* gives an example of this.

Another example may be seen in the electrodes of a neon tube (*fig. 14*) in which a light cup of chrome iron which serves as electrode is sealed at its edge directly to the end of the glass tube so that the external current supply can take place directly through the base.

For additional technical applications reference is made to the possibility of introducing glass windows into pressure or vacuum boilers, and to the use of glass connecting pieces on metal reaction chambers, while the special application for telescope mirrors [7]) has already been mentioned.

[7]) Telescope mirrors, Philips Tech. Rev. 1, 358 (1936).

Illumination of the Eiffel-tower at the Paris World Exhibition (1937). The greater part of the electrical plant has been equipped by, or with material of, S. A. Philips - Paris.

THE EXAMINATION OF THE MACRO-STRUCTURE OF RAW MATERIALS AND PRODUCTS WITH THE HELP OF X-RAYS I

by J. E. DE GRAAF.

Summary. This article explains the essentials of the testing of raw materials and products by means of X-rays according to the absorption method. In a number of short articles the practical applications and the results obtained by this method will be studied in some detail.

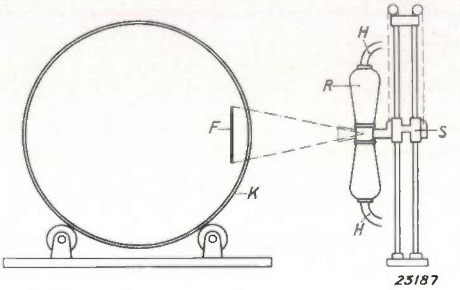
Introduction

In the application of X-rays to the examination of technical materials two large fields must be distinguished:

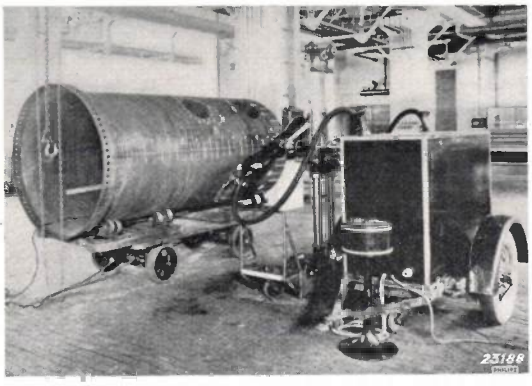
1) The investigation of the structure and texture of crystals (micro-structure) by means of the interference of the X-rays on the surface of the crystals [1]). The X-rays which produce the photograph are not the direct rays from the X-ray tube, but deflected rays. X-rays of rather long wave lengths are used (1 to 2 Å; 1 Ångström = 10^{-8} cm), which are unable to penetrate deeply into the object being examined. The micro-structure at the surface of the object is therefore examined.
2) The detection of cavities, cracks etc., in short, of faults which could be observed with the naked eye if the article were cut through in the proper place (macro-structure). The direct rays, which in this case are of very short wave lengths (for instance 0.1 Å) first pass through the article to be examined (*fig. 1*), in which they are absorbed to a greater or smaller degree, then they blacken a film to a correspondingly smaller or greater degree.

For homogeneous material the absorption of X-rays is the same throughout the mass; because of "inhomogeneities", such, for example, as enclosures of slag, the absorption may vary locally. On the film, which is chosen instead of a plate because of its durability, and which is provided with a sensitive emulsion on both sides, a picture of the object appears in which macro-faults may be seen more or less clearly. In this manner therefore the entire volume of the object is investigated, and not exclusively the surface layer.

Considering the fact that the blackening of the X-ray film depends upon the absorption of the X-rays in the film, and that this absorption is small, the time of exposure necessary to obtain a photograph with sufficient blackening becomes very long. The photographic action is therefore almost always reinforced by placing an intensifying



a



b

Fig. 1. *a*) Diagram of the X-ray testing of a boiler by the absorption method. *F* film, *K* boiler, *H* high voltage cable, *R* X-ray tube and *S* stand. *b*) Photograph of this test.

screen directly against the film. This screen fluoresces under the influence of the X-rays. It is usually covered with a layer of calcium tungstate. The fluorescent light of this layer contributes appreciably to the blackening of the film; the intensity of the light is proportional to the amount of incident radiation. The reinforcement obtained with such a screen may amount to a factor of thirty.

Influence of the thickness and nature of the material

If we first study the influence of the absorbing thickness of the material with otherwise similar conditions as regards time of exposure, distance

[1]) This was discussed in the August number of this periodical in the completed series entitled "Practical applications of the testing of materials by means of X-rays" by W. G. Burgers.

of the film from the X-ray tube, etc., we observe an absorption which increases rapidly with the thickness (*fig. 2*). It is found that in the first ap-

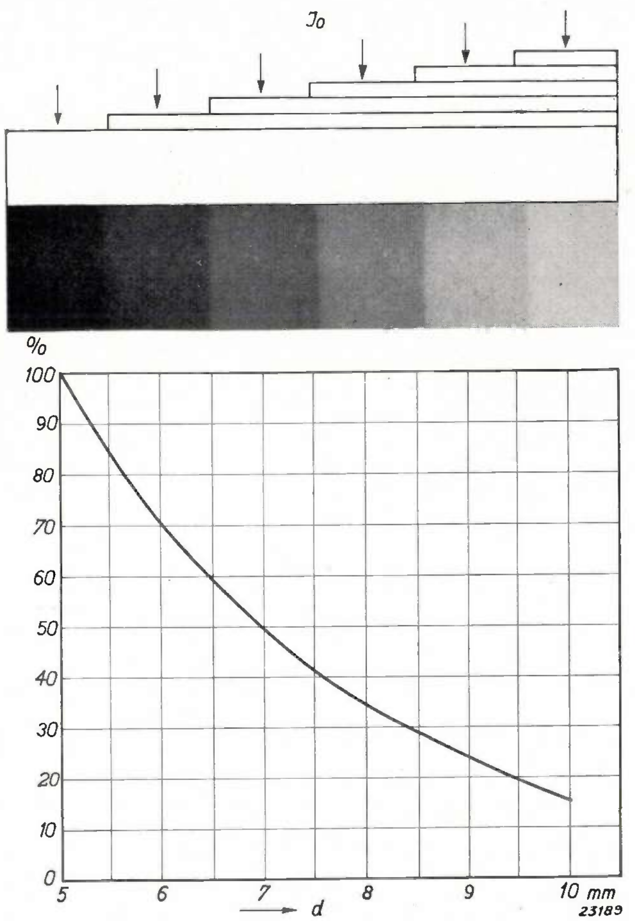


Fig. 2. Absorption of a beam of X-rays by different thicknesses of the same material. For this purpose an iron plate 5 mm thick was taken, upon which additional 1 mm layers of iron were laid until the final thickness was 10 mm.

proximation a certain thickness of material added to another layer of material always removes the same percentage of the radiation passing through the latter layer, no matter what the thickness of this layer (provided it is not too small). In reality this exponential increase of the absorption holds exactly only for monochromatic radiation, that is radiation of a single wave length. For the continuous spectrum emitted by an X-ray tube it holds however to a better approximation the thicker the layer of material through which the radiation passes, since the long waves are absorbed much more strongly than the short waves, and the radiation therefore becomes more homogeneous.

With equal thicknesses of material the nature of the material of the object examined still makes a great difference. Of the simultaneously photographed (*fig. 3*) equally thick plates of lead, tin, iron and aluminium, the aluminium plate is invisible because it has not absorbed appreciably. Behind the lead plate the film has remained practically clear which means that the lead has

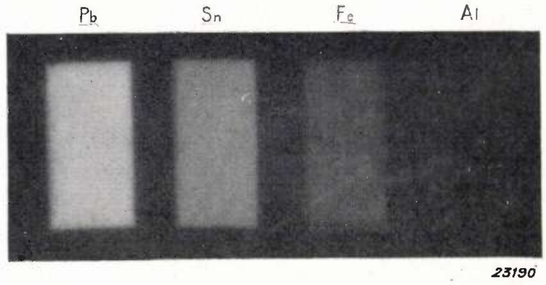


Fig. 3. Absorption of a beam of X-rays by the same thickness (0.2 mm) of different metals: lead (atomic weight 207), tin (119), iron (56) and aluminium (27) on an iron plate 10 mm thick. Exposure time 3 min., tube voltage 100 kV, current 4 mA. The aluminium because of the slightness of the contrast, is practically invisible.

absorbed nearly all the radiation. Roughly these results may be summarized in the following way: with increasing atomic weight substances absorb more strongly. This means therefore that in order to obtain the same absorption an appreciably thicker layer of aluminium is necessary than of iron. Or conversely: one can irradiate and examine appreciably greater thicknesses of aluminium than of iron under the same conditions.

Gases do not absorb appreciably even compared with aluminium. A hole filled with gas will be shown on the film in just the same way as when it was quite empty. Even when a cavity is filled with slag, this will not always be observable from the difference of blackening caused by this cavity. The greater the absorbtion of the metal itself, and the more gas bubbles the slag contains, the slighter the difference observable between a cavity filled with slag and one filled with gas. In such cases one must employ other peculiarities such as shape and position in order to be able to make any statement about the cause of the cavity.

On the other hand, greater differences in absorptive capacity, such as that between the metal connecting wires and the "Philite" in the base of the radio lamp in *fig. 4*, are clearly shown. For checking the work of assembly this fact has of course opened many possibilities of application. The photograph shows at the same time very clearly how sharply fine parts (thin wires) can be photographed with the help of X-rays.

Influence of the tube voltage

A third important factor in the detection of defects in materials, in addition to the thickness and nature of the materials, is the influence of the voltage between the anode and cathode of the

X-ray tube. The higher the voltage the shorter the waves occurring in the spectrum and the less the absorption of the radiation. Since in addition the output of the X-ray tube increases with increasing

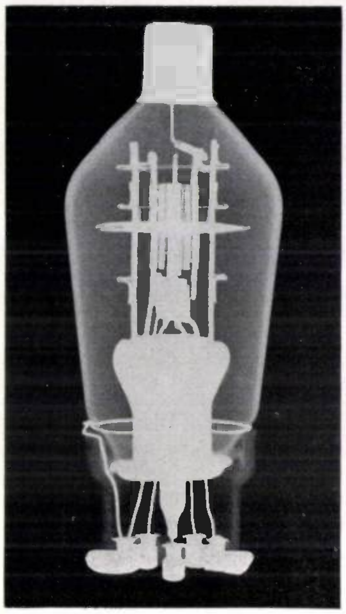


Fig. 4. X-ray photograph of a radio check of the assembly. Inside the visually non-transparent "Philite" socket may be seen the electrode leads. Inside the lamp in the stem and the pinch these wires are no longer visible. The glass of these parts contains a high proportion of lead (30 %) and the contrast effect through the fairly thick wires disappears. The assembly, however may be easily inspected.

voltage, the radiation passing through a certain thickness of material and also the blackening of the photographic plate caused thereby will increase rapidly with increasing voltage (*fig. 5*). This means again that one obtains a sufficiently blackened photograph of a definite thickness of material in a

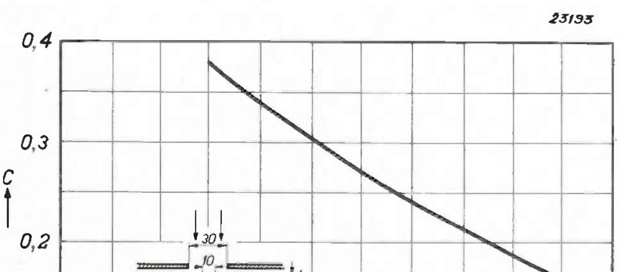


Fig. 6. Contrast C caused by a hole in a block of aluminium whose dimensions are given in mm on X-ray photographs taken at different voltages. A diaphragm of lead lies upon the aluminium in order to limit the field.

shorter time, the higher the tube voltage. It is however incorrect to conclude from this that one must therefore always use as high a voltage as possible. *Fig. 6* gives the difference in degree of blackening (contrast) caused by a hole in a block of aluminium when the block is photographed at different voltages in such a way that the main blackening remains the same. Between the tube voltages of 75 and 200 k.V. the contrast is found to decrease by about a factor of two, due to the decrease of absorption with increasing tube voltage which results in the fact that a given difference in thick-

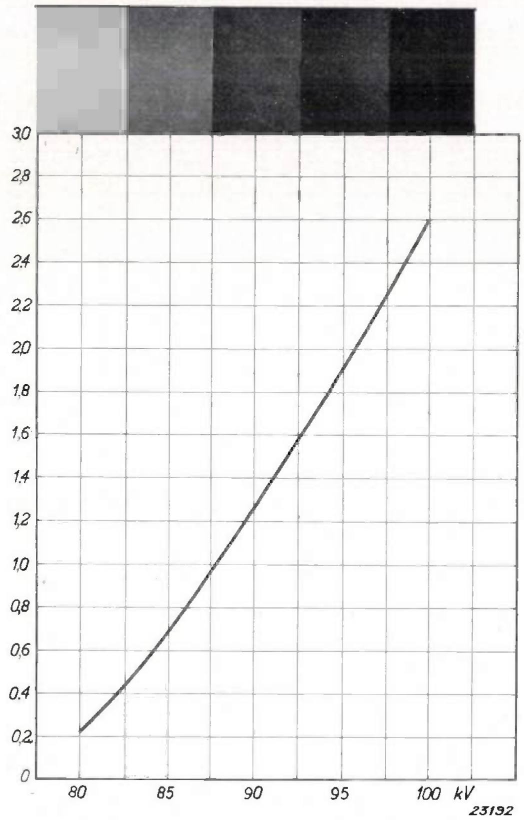


Fig. 5. Blackening Z due to X-ray beams with different tube voltages V with a current strength of 4 mA and an exposure time of 75 sec. behind 5 mm of iron. In the graph Z is shown its relation to V.

ness gives less contrast at higher voltage. This decrease in contrast may seem small, but it will be found that there is always a rather strong influence by the voltage on the visibility of fine details, since it is just in the appreciation of fine details which lie at the limit of visibility that a slight change of contrast makes a great difference. This dependence makes it important to establish the limit of visibility on each photograph by indication, for example, of the smallest difference in thickness distinguishable. For this purpose one may employ a frame with wires of different thicknesses made of the material of the object being examined. This frame is photographed at the same time as the object being placed upon the latter. The thickness of the thinnest wire just discernable must satisfy certain requirements: with objects up to 50 mm in thickness it must be at the most $1^1/_2$ per cent of this thickness. At the same time this affords a check on the dark room technique, since, when insufficient care is taken in developing, as a matter of fact well-taken photographs may be so badly treated that they give pictures with little contrast.

From the foregoing it will be obvious that every X-ray photograph for the testing of material is a compromise between the desire for a short time of exposure and the desire for good detail. As a rule exposure times of some minutes are chosen (of such length that the exposure times are short with respect to the time necessary for changing the films and adjusting the X-ray tube): the voltage necessary with a certain thickness of material for a photograph with an exposure time so chosen is taken from exposure graphs, such as that given in *fig.* 7 for iron.

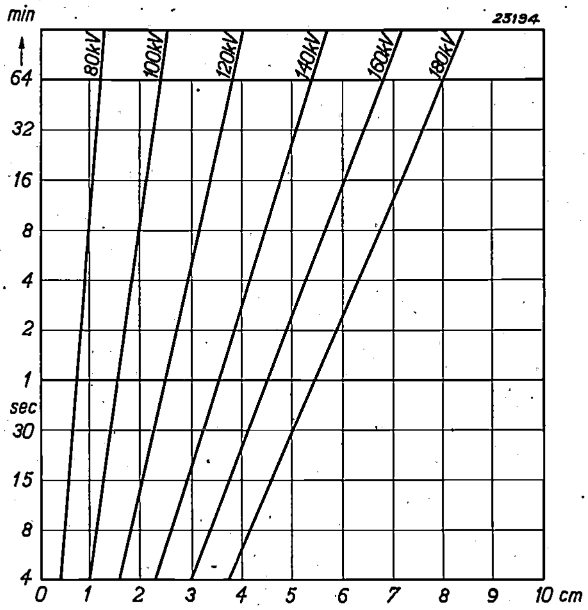


Fig. 7. Exposure graph for iron with intensifying screen and Agfa Spezialfilm. The thickness of the iron is plotted horizontally, the time of exposure vertically. The curves give the relation between the two at a constant blackening $Z = 0.5$ with the tube voltage as parameter, current 4 mA; distance between focus and film 40 cm, field 4×4 cm².

Visual observation

Besides the photographic method it is possible to apply direct visual observation with the aid of fluorescent screens. If such a screen is placed where the film would otherwise be, the fluorescent image can be observed on the screen with the eye. The radiation passed through the object must however not be too weak, that is to say, the objects must not be too heavy or too thick. In the case of iron the thickness for visual investigation must not be greater than about 20 mm. Since the contrasts are slighter in the fluorescence test than in the photographic method, and since there is no record made, the fluorescence method, although cheaper and quicker, is seldom employed for important tests. For the testing of technically less important parts visual observation is more desirable because the parts are, as a rule, much cheaper than the important ones, and thus require a cheaper method of testing.

Importance of the absorption method

When all the differences of density and thickness of the object to be examined are more or less clearly delineated on an X-ray photograph, the photograph can be used to detect undesired differences, that is, defects. The value of this method of detection by X-rays is twofold: firstly the photograph (in contrast to a cross section at one place) immediately gives a detailed view of the faults and of their relation to each other, and secondly the object is not sacrificed to the examination, which means that all objects whose importance makes it desirable may be tested. The first point is of significance not only in the development of new methods of manufacture in foundries and the like, where it is of immediate importance to be able to diagnose the cause of the fault, but also in testing production since from the cause of the fault, when once known, qualitative conclusions may very often be drawn as regards a decrease in strength due to factors not immediately observable such as changed structures. Thus the diagnosis of welding faults is of primary importance in the testing of welded boilers and other structures.

In a series of short articles we intend to discuss in more detail various examples of the application of irradiation by X-rays for the purpose of detecting defects in materials.

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

No. 1199*: R. Houwink and K. H. Klaassens: Die Viskositäts-Konzentrationsabhängigkeit in konzentrierter Lösung und ihre energetische Deutung, III (Koll. Zeit., 79, 138 - 148, May, 1937).

The article by the above authors summarised in Abstract No. 1130 led to a closer collaboration with H. L. Bredée and J. de Booys in investigating the nature of the relationship between the viscosity of colloidal solutions, their concentration and the size of the particles concerned.

In conjunction with two articles by Bredée and de Booys (Koll. Zeit., 79, 31 and 43, 1937), an empirical formula is given which satisfactorily expresses the viscosity of very different substances over a wide range of concentrations and which at low concentrations passes over into the Einstein formula.

The increase in the viscosity due to dispersed particles may be resolved into two components which are determined by the shape of the particles and by the space occupied by them respectively. It is found that elongated particles cause a relatively smaller increase in the viscosity of concentrated solutions than shorter, and particularly spherical, ones. This can be shown to be due to the differences in the change in energy transmission with the concentration between elongated and round particles

No. 1200: J. A. M. van Liempt and J. A. de Vriend: Über das Verbrennungslicht von Al-Zn- und Al-Cd-Legierungen (Rec. Trav. chim. Pays-Bas, 56, 594 - 598, May, 1937).

If zinc or cadmium is added to aluminium, its combustibility in oxygen is considerably increased. The rate of combustion and the actinic output are considerably augmented, without the colour of the light being altered to any appreciable extent.

No. 1201: J. A. M. van Liempt and J. H. M. Uden: Der photographische Nachweis von Thoriumoxyd in Glühdrähten (Rec. Trav. chim. Pays-Bas, 56, 607 - 612, May, 1937).

The presence of thorium oxide in tungsten incandescent filaments can be detected by means of the blackening produced on a photographic plate by the radio-active radiation from the thorium. Schumann plates are the most suitable for this purpose, next in suitability are the most sensitive ordinary plates. The sensitivity of this method increases with the thickness of the filament, the time of exposure and the atmospheric humidity. Drawn filaments give a deeper blackening than crystalline or single-crystal wires. Pure thorium wire gives a very deep blackening.

No. 1202: F. A. Heyn: De techniek der kerntransformaties (Ingenieur, 52, E51 - 56, May, 1937).

This paper gives a general survey of the various methods available for initiating nuclear transformations. The discussion shows how these transformations can be produced by means of alpha particles and the neutrons formed simultaneously with them, and how charged particles can be sufficiently accelerated by high voltages or by repeated passage through the same comparatively small potential difference so as to result in these transformations.

No. 1203: W. de Groot: Interpretatie van moleculaire verschijnselen door middel van potentiaal-krommen. I. Potentiaalkrommen van het tweeatomige molecuul (Ned. T. Natuurk., 4, 109 - 119, May, 1937).

Some properties of potential curves of the type introduced by Franck (1925) are discussed in this and several subsequent articles with special reference to problems in molecular physics, including predissociation and photo-ionisation.

No. 1204: W. Elenbaas: Die Gesamtstrahlung der Quecksilberhochdruckentladung als Funktion der Leistung, des Durchmessers und des Druckes (Physica, 4, 413 - 417, June, 1937).

For different pressures (0.1 to 25 atmos.) and with different loading L per cm of tube length (8 to 75 watts per cm), the total radiation of the high-pressure mercury discharge was measures for tube diameters d of 3.3 mm, 9.2 mm and 27 mm.

*) There is not a sufficient number of reprints available for distribution of the publications marked *). Reprints of the other publications will on request gladly be supplied by the Administratie van het Natuurkundig Laboratorium, Kastanjelaan, Eindhoven.

For power values above 20 watts per cm and pressures above 10/d atmos, the total radiation is equal to 0.72 (L - 10) watts per cm.

No. 1205: J. F. H. Custers: The ultraviolet absorption spectrum of potassium perrhenate (Physica, 4, 426 - 429, June, 1937).

The absorption spectra of an aqueous solution of potassium perrhenate were measured for various centrations in the region between 3134 and 2210 Å. At approximately 3150 Å the absorption becomes notice- able; at 2290 Å there is a strong absorption maximum and at 2390 Å a scarcely noticeable maximum. The absorption spectra of the permanganate and the perrhenate ions are compared with each other.

No. 1206: A. A. Kruithof and F. M. Penning: Determination of the Townsend ionisation coefficient for mixtures of neon and argon (Physica, 4, 430 - 449, June, 1937).

The ionisation coefficient α was plotted as a function of the quotient of the field intensity E in volts per cm and the pressure p_0 in mm Hg (reduced to deg. C) for pure neon and eight different mixtures of neon and argon. The results obtained are in agreement with previous estimates of the ionisation from the starting voltage of luminescent discharges. The most marked ionisation, corresponding to a doubling of the number of electrons on passing through a voltage of 18.7 volts, was found at $E/p = 3$ volts per cm and per mm Hg with neon containing 0.1 per cent argon. If the quotient E/p is made smaller, the ionisation coefficient is considerably reduced owing to the energy losses of the electrons on elastic collision with the atoms of the gas.

Since, according to earlier papers, the photoelectric current at constant field intensity increases more or less in stages, special attention was given to the exact determination of the current as a function of the distance between the electrodes.

No. 1207: A. A. Kruithof and M. J. Druyvesteyn: The Townsend ionisation coefficient α and some elementary processes in neon with small admixtures of argon (Physica, 4, 450 - 463, June, 1937).

From the Townsend ionisation coefficient, the probability χ of an excited neon atom ionising an argon atom was calculated by two different methods, viz.: a) assuming the probability that a neon atom is excited by impact of an electron, and b) taking as a basis the measurements described in No. **1026** in which the mean ionisation was measured as a function of the distance from the cathode. From the connection between χ and the percentage proportion of argon, it follows that nearly 90 per cent of the neon excitations give a metastable atom if the quotient of the field intensity in volts per cm and the pressure in mm is lower than 3.5.

No. 1208: M. J. Druyvesteyn; The mobility of electrons in neon (Physica, 4, 464 - 466, June, 1936).

The drift velocity of electrons was calculated from the velocity distribution obtained in measurements described in Abstract No. **1207**. The calculated curves for pure neon and neon containing a small proportion of impurities are in satisfactory agreement with the values found experimentally.

No. 1209: H. Bruining and J. H. de Boer: Secondary electron emission of metals with a low work function (Physica, 4, 473 - 477, June, 1937).

It is concluded from measurements on the secondary emission of pure metals and the effect of impurities that the more or less free electrons contribute only slightly to secondary emission - under bombardment by primary electrons having energies of several hunderd volts. The facility for secondary emission may be increased by impurities, which partly bind the electrons. The relationship between secondary emission and the thickness of the deposited layer also was investigated. Emission commences with a value corresponding to that of the base metal, reaches a maximum which may be a multiple of the initial value, and then diminishes again with increasing thickness of the deposited layer to a value equivalent to that of the coating metal.

VOL. 2 No. 11 NOVEMBER 1937

Philips Technical Review

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

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

THE INFLUENCE OF X-RAY TREATMENT ON FLOWER BULBS [1]) (HYACINTHS AND TULIPS)

by W. E. DE MOL.

Summary. A description is given of tests [2]) on genetic phenomena and improvement of flower bulbs, and particularly on the temporary and permanent changes brought about by X-ray treatment.

Cell division

For the assistance of the reader we shall begin with an explanation of several technical botanical terms and ideas relating to the multiplication of plant cells by division. We must distinguish between ordinary cells, those from which leaves, perianth leaves, stalks and roots are built up (the so-called vegetative or somatic cells), and cells which serve for reproduction, namely the male (pollen grains) and the female (embryo cells or germ cells) collectively called generative cells, sexual cells or gametes.

The somatic cells normally contain a constant number of longitudinal bodies, chromosomes, in their nuclei. The division (mitosis) takes place at a certain moment when all the chromosomes of a somatic nucleus split in two lengthwise (separation) and then the two products of the division (*fig. 1a*) move away from each other toward two opposite points (poles) in the cell (disjunction) to form two new cell nuclei in this way, each of which has the same number of chromosomes as the original cell before the division. This manner of division of the somatic cells is called typic mitosis or equivalent division because it gives exactly equivalent division products in the longitudinal halves of the chromosomes.

The division of sexual cells takes place in a different way. A cell which is destined to form generative cells (mother cell) undergoes a spontaneous process before the division in which the chromosomes group themselves in pairs. From each of these twin

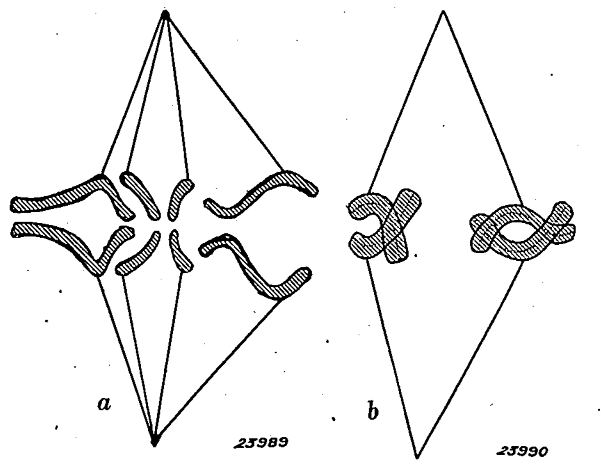


Fig. 1*a*. Ordinary division (typic division) of a nucleus with 4 chromosomes. All 4 chromosomes are split lengthwise.
Fig. 1*b*. Heterotypic division. The chromosomes are divided into 2 equivalent groups of two each. After twin chromosomes have been formed, they separate again, 2 chromosomes toward one pole and 2 toward the other.

chromosomes one moves toward the one pole and the other toward the other pole. (*fig. 1b*). The result is two new cell nuclei, each of which has half the number of chromosomes of the original cell. In this case one speaks of reduction division or of allotypic mitosis. The reduction division proper (first meiotic or heterotypic division) is followed as a rule by an ordinary division (second meiotic or homotypic division),

[1]) *Note by the editor:* In this article a new application of X-ray treatment is described. Dr. W. E. de Mol, who has carried out investigations on this subject, has consented to explain some of his results to our readers.

so that from one mother cell four cells are formed, each of which has half the normal number of chromosomes. In ordinary cases all of these cells develop into pollen grains in the anther. In the ovulum however from each four only one germ cell is formed. Further it must be kept in mind that in fertilization a male and a female nucleus are fused together to give a germ nucleus, which then again contains the original number of chromosomes. With the surrounding plasma the germ cell is thus built up, and is the first somatic cell of the newly formed individual.

Structure and development of the bulb

In order to understand phenomena occurring in a bulb it must be remembered that a bulb may be compared with a shoot. The leaves of the shoot correspond with the scales of the bulb. At the end of the shoot or in the axil of a "leaf" the future plant develops with its leaves and flowers. This development, during which the plant is still completely hidden inside the bulb, takes place in the plants under consideration (hyacinth and tulip), between June and November. Those parts which will be situated lowest on the plant are the first to be developed.

Just as in the axils of the leaves of deciduous trees buds are formed which will grow in the following year to new shoots, so bud formation also takes place in the bulb. These buds (bulbils) which are formed in the axils of the thick scales (*fig. 2*) are the bulbs for the following year or later. One bulb can in this way produce many buds, and so many new bulbs (asexual reproduction, vegetative propagation). By vegetative reproduction therefore, an extensive "family" can be formed from one bulb. When all the members of the "family" are alike, which is normally the case, one speaks of a clone. In addition to vegetative reproduction with bulbs, there is also reproduction by means of seeds, after previous pollination, self-pollination or crossing (sexual reproduction).

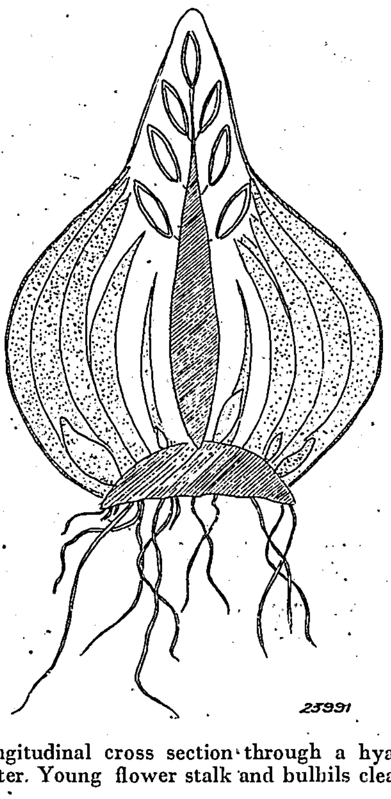


Fig. 2. Longitudinal cross section through a hyacinth bulb in the winter. Young flower stalk and bulbils clearly visible.

Preliminary tests

From 1908 up to the present time tests have been carried out by the writer on genetic phenomena and improvement, especially in the case of flower bulbs.

In the early years (1908 - 1914) attention was devoted exclusively to changes in the cells of leaves and perianth leaves. It appeared that not only modifications in the shape of the leaves and perianth leaves occurred, but also changes appeared in the colour of these organs. Many of these anomalies were found to be hereditary, so that we were concerned with somatic mutations. Up to 1914 the genetic anomalies were multiplied exclusively vegetatively. After that year the fact was also established by crossing that the deviations observed (of shape and colour) were hereditary.

From 1918 onwards the changes in the cells were studied with the help of the microscope (cytologic research). The first fact established was that in cultivation [2]) the number of chromosomes changes (increases) in many varieties. The wild hyacinth (Hyacinthus orientalis), the tulip (Tulipa Gesneriana) and the narcissus (Narcissus pseudonarcissus and poeticus) have respectively 16, 24 and 14 chromosomes in the somatic nuclei. In the cultivated form the following numbers usually appear: 24 chromosomes ($8+2\times 8$) in the hyacinth, 36 chromosomes ($12+2\times 12$) in the tulip and 21 chromosomes ($7+2\times 7$) or 28 chromosomes ($2\times 7 + 2\times 7$) in the narcissus. Other numbers of chromosomes are also found.

The probable explanation is that sometimes no reduction division takes place in the formation of the gametes, so that generative nuclei occur with the ordinary somatic number of chromosomes; thus with the hyacinth: $2\times 8 = 16$, with the tulip: $2\times 12 = 24$, and with the narcissus: $2\times 7 = 14$.

It was found that such a "doubling of the

[2]) By culture is meant the whole treatment applied by the grower for the multiplication of bulbs.

generative nuclei" can be produced by exposure to a certain temperature (cold, heat).

From pollination tests with the pollen of the hyacinth, among others, which contained, in addition to ordinary pollen grains, such "double" pollen grains, and from the cytologic examination of the germ plants produced with this pollen it was found that in addition to germ plants with the normal number of 16 chromosomes, plants with 24 chromosomes also occurred. It was concluded from this test that pollination with double pollen grains often leads to an increase in the number of chromosomes.

When plants occur with the same number of cells — and that is usually the case — those having 24 chromosomes are bigger than those with 16 chromosomes; often indeed very much bigger. In this respect the deliberate generation of double pollen grains, which the writer has been able to bring about with various kinds of plants, may be considered of great importance in practice, since in this way specimens may occur which are more robust than the original wild or cultivated variety.

Influence of X-ray treatment on the reduction division

With the ordinary garden tulip, *Tulipa Gesneriana*, as the object of experiment, it was found that "double" and even "quadruple" pollen grains may easily occur as a result of treatment with X-rays. While normally one parent pollen cell produces four pollen grains with 12 chromosomes, under abnormal conditions it produces two pollen grains with nuclei containing 24 chromosomes. In addition numerous other aberrations appear. Thus one or more chromosomes may be "retarded" in their passage toward the poles. They may also be eliminated. Such phenomena promote the occurrence of nuclei with differing, abnormal numbers of chromosomes. The cells with such nuclei, however, often die.

The technique of the irradiation (*fig. 3*) is as follows, according to the observations made with a variety of tulip which was studied exhaustively. The irradiation is carried out between June and November, with doses varying from 100 to 1200 *r* [3]). In June the formation of the new flower begins in the newly formed bulb. In November (or earlier) this is complete including the generative cells. The first phase of the process of growth, cell division, has then taken place. The second phase, cell enlargement, has yet to begin.

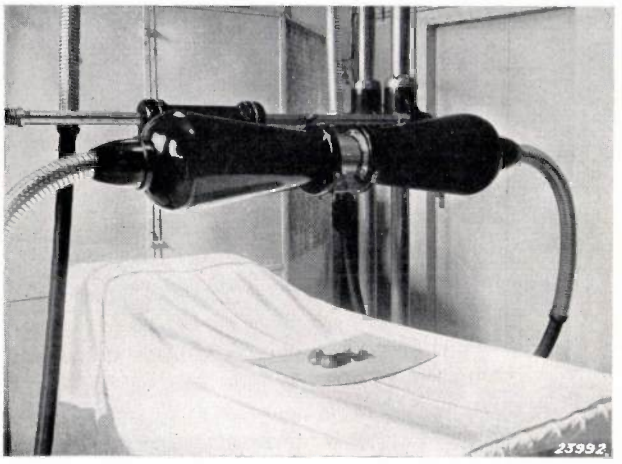


Fig. 3. Arrangement used in the X-ray treatment of flower bulbs.

Influence of X-rays treatment on somatic division

Not only allotypic mitoses but also the typical mitoses are sensitive to X-ray irradiation. It was found that with the hyacinth it was possible

Fig. 4. Bulb of a hyacinth on a "twin glass". The roots which have developed in the right-hand glass have not been irradiated; those in the left-hand glass have been irradiated (thus the shorter roots are the irradiated ones).

---

[3]) An irradiation represents a dose of 1 *r* when in 1 cm^3 of air at the spot where the object is placed 0.108 erg would be absorbed as a result of the same irradiation. This corresponds to 100 ergs in 1 cm^3 of water.

deliberately to cause the formation of somatic cells with a modified nuclear structure, which are able to divide further. For this purpose the bulb was placed on a so-called twin glass filled with water (*fig. 4*). It was possible in this way to divide the roots into two groups. One group was irradiated (400 r), the other was not irradiated. The irradiated and the unirradiated root tips are studied and compared with each other.

In the former case various anomalies had appeared. Two or more adjacent cells were often found with nuclei twice as large as normal. This proves in the first place that there were somatic cells in which separation of the chromosomes but not disjunction had taken place, and in the second place that such somatic cells with double the ordinary number of chromosomes afterwards underwent normal mitosis. In the irradiated root therefore there now exist groups of double cells next to normal cells (mixo-ploidy).

Excitation of somatic mutations in shape and colour

Considering the fact that one can exert influence on the cells during their division, it is obvious that the earlier one irradiates during the process of cell division of the flower formation, the more one can influence the organs occurring lower on the stalk which are formed; thus first the leaves, then the perianth, etc.

In the first place unusual forms of perianth sometimes occur with the tulip, (*fig. 5*) as a

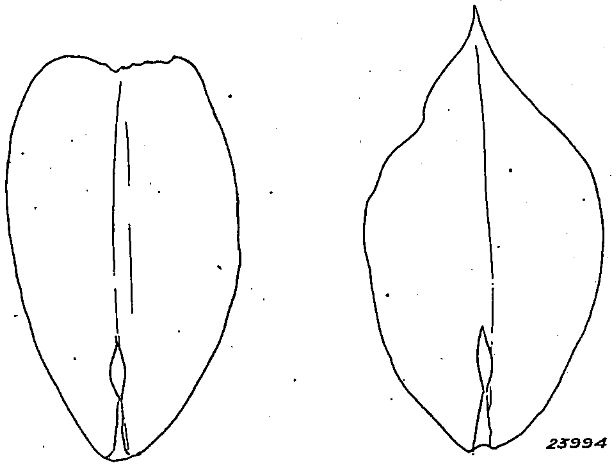


Fig. 5. Perianth leaves of the tulip Generaal de Wet. The left-hand leaf shows the original shape, the right-hand one, the shape as altered by irradiation.

result of irradiation. The perianth leaf becomes pointed (right), whereas it was originally rounded at the top (left). The perianth leaf (*fig. 6*) exhibits indentations on its edge and is curled to some extent (right), while these distinguishing marks were originally lacking (left). These changes have occurred because a group of cells produced from a given cell have died off, or have been enlarged so that they no longer fit into the normal cell tissue and often cause stresses in it.

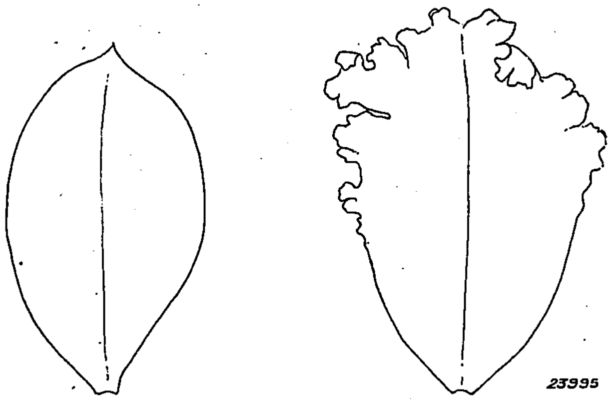


Fig. 6. Perianth leaves of the tulip Van der Neer. Left, the leaf with the original shape; right, the leaf as altered by irradiation.

At the same time, however, mutative phenomena, relating to coloration, occur in the soma. On the first occasion of their appearance they manifest themselves in a sector of a whole plant, in a sector of a whole flower or in a sector of one or more perianth leaves. The last mentioned occurs by far most frequently.

In the hyacinth a somatic mutation has remained constant (variety Grand Maître), in which three perianth leaves of the innermost corolla have changed into organs which more or less resemble anthers (*fig. 7*). On the edge, to the left and right, numerous pollen grains are developed (*fig. 8*).

Fig. 7. Two flowers of the hyacinth Grand Maître. The shape of the 3 perianth leaves of the inner whorl is altered by irradiation, and which have become stamen-like.

Moreover with the hyacinth since 1930 the rose-violet colour has been maintained after vegetative multiplication; this colour has taken the place of the blue-violet colour (also variety Grand Maître). While the blue-violet anthocyanin, as in the wild hyacinth, occurs in the lowest layer of cells of the skin of the perianth (subepidermis), the rose-violet anthocyanin is encountered in the top layer of the skin (epidermis).

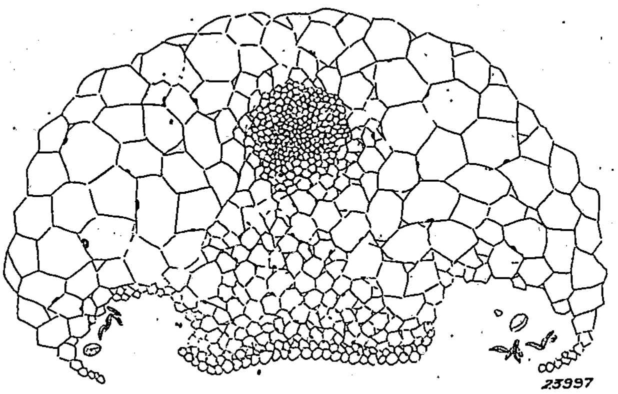


Fig. 8. Cross section of a mutated perianth leaf of the hyacinth of fig. 7. The drawing was made from a microtome preparation. It may be seen clearly that two loculi with pollen grains have developed as in the anther (lower left and right).

The following is a short description of the results of irradiation obtained in the spring of 1937, with respect to the excitation of somatic mutation in tulips.

Numerous somatic mutations in colour have occurred particularly in tulips. This year a large number of plants irradiated in 1928 bloomed in the experimental garden in Lisse. In addition the results could be seen of experiments carried out, in the majority of cases in 1933 and 1934.

We give below a schematic survey of the varieties of which one or more specimens bloomed this spring, and where it could be definitely ascertained that the flower has taken on a different colour or a different form. In this survey no statements are included about colour or shape change which up to now has been manifested only in larger of smaller sectors, or about cases in which further cultivation is necessary to determine definitely whether we are concerned with mutation (permanent change) or with modification (temporary deviation).

When it is kept in mind that for the last-mentioned tests (1933, 1934) not more than 8 or 16 bulbs of each variety were used, it is the more striking how large the chance of mutation is; at least when the experiment is well planned and well carried out.

The genetic change is not limited to a change of colour or shape of the flower. Changes of a permanent nature are also encountered in the leaves (variegated leaves). With the varieties Generaal de Wet, Herodiade, Oranje Nassau, Peach Blossom, Prince of Austria and Triumphator this phenomenon has appeared in various ways: white, yellow or yellow-green on the edges or in the middle.

In the case of the flower bulbs therefore it is the somatic mutations which are stimulated in such a surprising way by irradiation. First there is a small sector of a different colour on one perianth leaf (*fig. 9*, left), then half of a perianth leaf is of a different colour (fig. 9 middle). It is obvious that in these cases usually the corresponding portion of the stalk and of the whole plant including buds is concerned in the process of mutation, so that the following year half the flower exhibits the new colour (fig. 9 right), and finally the whole plant (*fig. 10*). This is often the progress of the development of such a vegetative mutation. This is true of course when a young bud (bulbil) has developed entirely or partially in the mutated sector.

As soon as one plant has undergone complete mutation, it may be considered as the beginning

Irradiated in 1928.

Variety	Original Colour	New Colour	Number of specimens
1. Eleonora	purple black anthers	pale violet yellow anthers	several rows (1 row = 8 specimens)
2. Roi d'Islande	pink	light pink	several plants
,,		more violet	,, ,,
,,		darker pink	,, ,,
3. William Pitt	cochineal-red	beautiful pink	14 plants, 8 flowers
,,		dark red, butterfly heart [4])	7 ,, 4 ,,
,,		dark red, blue heart [4])	50 ,, 30 ,,

[4]) By "butterfly heart" is meant that the blue spot on the inner side of the perianth leaves is bordered by a yellowish white stripe, as on the wings of a certain kind of butterfly (Vanessa antropa). If this border is lacking one sees only a "blue heart".

Irradiated in 1933 or 1934.

Variety	Original colour	New colour, one or more specimens
4. Berlin, the large-flowered mutation of Pride of Haarlem	cochineal-scarlet.	a. lighter red b. still lighter red
5. Clara Butt	salmon pink	a. darker pink b. red
6. Dido	orange pink	a. lighter orange b. darker orange
7. Flamingo, Darwin	pink	a. lighter b. darker and pointed c. beautifully striped
8. Generaal de Wet	orange	a. yellow b. redder and pointed
9. General French	cherry red	light pink, pure white heart instead of yellow
10. Herodiade	pink	lighter pink
11. Ibis	underneath small portion white, further red	a. narrower red border b. white
12. Valentin	purple violet	beautifully regularly striped, no "broken colour" [5])
13. Van der Neer	purple	perianth leaves bigger and indented.

of a culture of genetically changed specimens. After several years this clone, forming a new "X-ray variety" may have increased considerably in numbers.

We have mentioned above no less than 12 different varieties of which at present one or more specimens exist with changed colour of flower, and 6 varieties with changed colour of leaf. It is clear that such somatic mutations represent a material value, sometimes a very large one, when the new colour is more beautiful than the one already existing. Quite apart from this is the scientific value of these experiments.

The hyacinths under consideration were treated with a dose of 300 r. The doses used for the mutation of tulips varied widely, namely from 70 to 1600 r.

The parrot phenomenon in tulips

Extensive crossing experiments have up to now shown that the so-called parrot tulips (large perianth leaves, indented and irregular in shape and colour), cannot be multiplied by reproductive methods (by seed). In general this deviation can however be multiplied vegetatively. Considering that the genetic cells have their origin in a layer of cells situated under the outer skin (epidermis) and not in the epidermis itself, it seemed to the writer that the parrot character consists in a

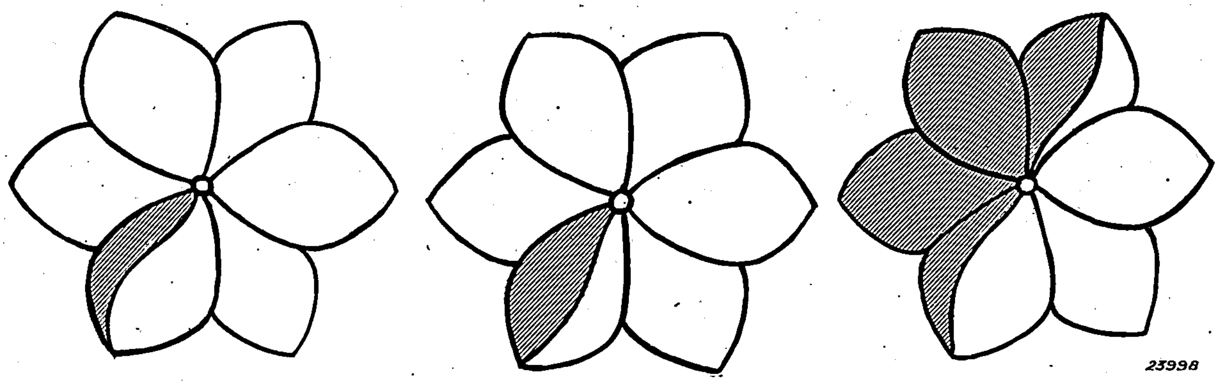


Fig. 9. Three tulip flowers drawn from below. Left: on one perianth leaf a small sector with changed colour has appeared. Middle: one leaf is half of a different colour. Right: half of the whole flower has changed its colour.

[5]) When one speaks of "broken colour", one means that the original colour is broken by stripes and spots in which no anthocyanin has developed. It is a mosaic effect, perhaps caused by a virus. In the variety Valentin the new colour pattern is not due to this cause, but is a consequence of the irradiation.

genetic change exclusively in the cells of the epidermis and is therefore based upon a permanent change in the outermost cell layers of the plant. Several times "parrot sectors" have been produced by X-ray treatment of a variety of tulip, but on the outside only of several perianth leaves and not on the inside. In this case this character is very

Fig. 10. From one bulb two flowerstalks have sprung. One of the flowers has entirely changed its colour, the other has retained completely the original colour. Such cases are extremely rare. In general two bulbs are formed with different coloured flowers.

probably excitated only in the epidermal cells limiting the outside of the perianth leaf, and not in those cells which form the inner side of that leaf. Consequently the parrot character was not completely developed. The perianth leaf had for instance retained its normal colour on the inner side, but had taken on the parrot coloration on the outside, etc. It is obvious that this character could not be retained in this case upon further vegetative multiplication, since at least one bud must be formed from this tissue if the genetic changes are to be permanent.

This fact furnishes a new basis for the abovementioned opinion that the genetic anomalies appear during the first phase of growth, the period of cell division, and are afterwards passed on from cell to cell.

The division hypothesis

In the opinion of the writer all genetic anomalies observed, that is cells observed with an abnormal number of chromosomes, or of abnormal colour or size, can be ascribed to the action of external influences acting during the extremely subtle process of nuclear division. These influences may be the more ordinary temperature influences as well as very unusual ones, such as treatment with X-rays. These often result in the process of division, i.e. the motion of the chromosomes toward the poles being entirely prevented, retarded or accelerated. The obvious conclusion is that in many cases this intervention acts only on certain genes, that is to say on the assumed material vehicles of the hereditary character (hereditary factors), either incorporated in the chromosomes, and therefore within the cell nucleus, or lying outside it. This hypothesis, proposed by the writer in 1928 and called in short "division hypothesis" will be able to guide investigators in the interpretation of the effects obtained, or in the search for new phenomena in this subject.

Literature

(Preliminary communications and short reports of lectures relating to irradiation of flower bulbs in the X-ray laboratory of the University of Amsterdam).

No. 1 (1929). Verandering der chromosomengarnituur door toediening van X-stralen en door blootstelling aan bepaalde temperaturen. (Retardatie en diversiteit). Handel. 22e Nederl. Nat. Geneesk. Congr., 134 - 139.

No. 2 (1930). Änderung der Chromosomengarnitur durch Röntgenbestrahlung und Temperaturwirkungen. (Retardation und Diversität). Zeitschr. ind. Abst. Vererb. **54**, 363 - 367.

No. 3 (1931a). Vertraging en versnelling in de processen der celdeeling en der celstrekking bij tulpen, veroorzaakt door X-stralen, en de gevolgen daarvan. Nederl. T. Geneesk. **75**, 1086 - 1088.

No. 4 (1931b). Wijziging van de bloemkleur der hyacint door röntgenbestraling enz. en een verklaring daarvoor. (Retardatie en acceleratie bij het deelingsproces der genen). Handel. 23e Nederl. Nat. Geneesk. Congr., 159 - 161.

No. 5 (1931c). Somatische Variation der Blumenfarbe der Hyazinthe durch Röntgenbestrahlung und andere äußere Umstände. (Teilungsretardation und -acceleration als Diversitätsursache). Zeitschr. ind. Abst. Vererb. 59, 280 - 283.

No. 6 (1932). Erfelijke en niet-erfelijke afwijkingen bij hyacinten en tulpen, door röntgenbestraling ontstaande, bezien in het licht der „Deelingshypothese". Natuurw. T. 14, 70 - 72.

No. 7 (1933). Mutation sowohl als Modification durch Röntgenbestrahlung und die „Teilungshypothese", Cellule 42, 149 - 162.

No. 8 (1934). Röntgenbestraling en veredeling der tulpen en hyacinten. Vlaamsche Wet. Congr., Beknopt Versl. 4e Congr. Land- en Tuinb., Agricultura 37, 176.

No. 9 (1935a). Practisch voordeel door röntgenbestraling ter verkrijging van knopmutaties („verloopingen"). Landb. T. 47, 4 - 17.

No. 10 (1935b). Verloopingen door röntgenbestraling (somatische mutaties), na bestraling ontstaan bij de Darwintulpen Roi d'Islande en William Pitt, Kweekersblad 38, 183 - 184.

No. 11 (1935c). Somatische mutatie door röntgenbestraling en de „Deelingshypothese". Handel. 25e Nederl. Nat. Geneesk. Congr., 270 - 272.

No. 12 (1936). Röntgenbestrahlung von Tulpen, Hyazinthen usw., Strahlentherapie 56, 215 - 227.

No. 13 (1937). De beteekenis van eenige röntgenbestralingsproeven met planten voor de geneeskundige röntgenologie. Handel. 26e Nederl. Nat. Geneesk. Congr., 265 - 267.

Further see Herba Topiaria (Study of ornamental plants) organ of the Netherlands Society for the promotion of scientific improvement of ornamental plants (Nederlandsche Vereeniging tot bevorderiug der wetenschappelijke veredeling van siergewassen.)

These tests were enabled to be carried out in the X-ray laboratory of the University of Amsterdam, thanks to the kind collaboration of Prof. Dr. J. van Ebbenhorst Tengbergen.

The microscopical examinations took place in the laboratory for physiological chemistry of the University of Amsterdam. The writer's thanks are due to the director Prof. Dr. B. C. P. Jansen.

In the department belonging to Prof. Combes (Sorbonne) at the World Exhibition in Paris (experimental biology) extensive material of the writer was exhibited, consisting of photographs, drawings and watercolours, relating to the subject treated in this article.

NOISE IN AMPLIFIERS CONTRIBUTED BY THE VALVES.

by M. ZIEGLER.

Summary. In this article the contribution of the amplifier valves to the noise in amplifiers and receiving apparatus is discussed in some detail. We have introduced the concepts "noise factor", "noise voltage" and "noise resistance" as suitable quantities for judging amplifier valves in this respect. A special study is made of the decrease in noise due to the influence of the space charge, and of the increase in noise in screen grid valves, due to fluctuations resulting from the division of the current. Finally the principle is explained of a "noiseless" screen grid valve developed in this laboratory and the advantages of this valve over existing valves are discussed.

Introduction

In a previous article in this periodical [1]) we became acquainted with the fundamental causes of noise in amplifiers: the thermal fluctuations of the electricity in conductors and the shot effect of currents of electrons in amplifier valves. We have seen that the irregular alternating voltage, which occurs between two points in an electric network as a result of the thermal fluctuations of electricity, has been established theoretically and practically. The contribution of this phenomenon to the total noise in amplifiers and radio receiving sets can always be calculated in a simple way from the formulae derived. In this article we shall not study further the thermal fluctuations, but confine ourselves to the noise contribution of the valves.

Noise factor

As a consequence of the corpuscular nature of electricity the anode current of every amplifier valve exhibits fluctuations about the mean value, even when the potentials of the electrodes are constant. In the very simple case of the saturated diode, that is a diode in which the whole emission passes over to the anode, one is concerned with the fluctuations in time of the number of independent electron emissions; the electrons emitted and therefore collected by the anode have irregular distribution in time; this is the case of pure shot effect. In the above mentioned article we derived the following formula for the value of these fluctuations:

$$\Delta\, \overline{(I-\overline{I})^2} = 2\, e\, \overline{I}\, \Delta\, \nu, \quad \cdots \quad (1)$$

in which the left-hand member of the equation represents the contribution to the mean square of the fluctuations of the current, which is due to the components included in the frequency interval between ν and $\nu + \Delta\, \nu$.

If we are not considering a saturated diode but an amplifier valve, the situation is not so simple, and the fluctuations of the anode current have a different value from the one given by (1). This can be expressed by introducing into the formula for the fluctuations of a current of electrons a factor usually indicated by F^2.

$$\Delta\, \overline{(I-\overline{I})^2} = F^2\, 2\, e\, \overline{I}\, \Delta\, \nu. \quad \cdots \quad (2)$$

We shall call this factor F^2 the noise factor.

By the noise factor of a current, therefore, we mean the ratio of the mean square of the fluctuations of the current to the mean square of the fluctuations of another current which is on the average equally large and exhibits pure shot effect.

Noise voltage and noise resistance

In the course of this article we shall be concerned with the factors which determine the magnitude of F^2. The first problem is the determination of the degree to which the fluctuations of the anode current of a valve act as disturbances in the reception of a signal.

In order to obtain some idea of this, these irregular fluctuations of the anode current of an amplifier valve may be compared with the variations of the anode current which are caused by a small voltage introduced between the cathode and the control grid of the valve. It is clear that the smaller the alternating grid voltage which causes current fluctuations whose mean square is equal to that of the fluctuations which are naturally present, the less disturbing these latter undesired fluctuations will be. The alternating grid voltage, which is called the equivalent noise voltage of the amplifier valve, is therefore an important factor.

---

1) The Causes of Noise in Amplifiers, Philips Techn. Rev. **2**, 136 (1937).

The noise voltage is found by dividing the value of the fluctuations of the anode current by the slope of the valve.

$$\Delta\, \overline{V^2} = \frac{\overline{\Delta\, (I - \overline{I})^2}}{S^2} = \frac{F^2\, 2\, e\, \overline{I}}{S^2}\, \Delta\, \nu \quad \cdots \quad (3)$$

One may calculate further the magnitude of the resistance which exhibits thermal fluctuations of voltage between its extremities equal in size to those of (3). In the article mentioned we found that the thermal fluctuations in voltage of a resistance R at the absolute temperature T are given by

$$\Delta\, \overline{V^2} = 4\, k\, T\, R\, \Delta\, \nu, \quad \cdots \quad (4)$$

in which k is the Boltzmann constant.

When (3) and (4) are combined one obtains

$$R = F^2 \frac{2\, e\, \overline{I}}{4\, S^2\, k\, T} \quad \cdots \quad (5)$$

With $e = 1.60 \times 10^{-19}$ coulomb, $k = 1.37 \times 10^{-23}$ erg/degree and $T = 290°$ K, R is found to equal 20 000 $F^2\overline{I}/S^2$ ohms, when $\overline{I}$ is in mA and S in mA/Volt.

This value of R, which is called the equivalent noise resistance of the amplifier valve is a useful measure of the quality of a valve with respect to noise. According to the definition of R the mean square of the fluctuations of the anode current is exactly equal to the thermal fluctuations which occur when the impedance between grid and cathode at the frequency under consideration is equal to R. When R is known, one immediately knows for any arbitrary circuit which source of noise furnishes the largest contribution; the amplifier valve or the electric circuits between grid and cathode.

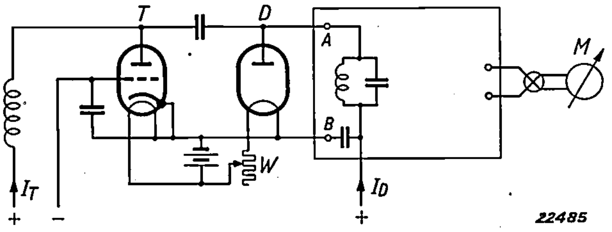
The determination of the noise resistance of an amplifier valve involves measurement of the noise factor F^2 with which we shall now deal.

Measurement of the noise factor

In order to determine directly the noise factor of a current with fluctuations of unknown intensity, one might use the measuring arrangement which was described in the article already cited on the causes of noise, and calculate the factor F^2 from the result of the measurement with the aid of equation (2).

A simpler method is the following. The well defined fluctuations of the current of a saturated diode are used as a standard to measure the unknown fluctuations. The arrangement used for this purpose is given diagrammatically in *fig. 1*.

T is the amplifier valve for which the fluctuations of the current to one or more electrodes must be measured. These fluctuations are amplified with a selective and linear amplifier which passes a certain band of the frequency spectrum so narrow, that it may be assumed that when the



Arrangement for the determination of the noise factor of the anode current of an amplifier valve T. By varying W, the anode current from the saturated diode D is adjusted so that the noise contribution of D is equal to that of T. The noise factor F^2 is then equal to I_D/I_T.

factor F^2 is dependent on frequency it may be regarded as practically constant in the frequency range transmitted. The amplified fluctuations are measured in an instrument M which indicates the mean square of the value of the output current.

In parallel with T, *via* a condenser of large capacity, is connected a diode D having a tungsten filament, whose temperature can be adjusted to any value by means of the filament control resistance W. The anode voltage of the diode is so high that all the electrons emitted are collected, and the fluctuations of this current therefore correspond with the pure shot effect. Furthermore, the internal resistance of this diode is always very high, so that the impedance between points A and B and thus the amplification of the fluctuations of T, is independent of the current adjustment of the diode.

If the amplification is adjusted for a certain value of the current I_T through T and $I_D = 0$ so that the meter M has a definite deflection, one may then, with the aid of W, regulate the current I_D so that the indications of M are doubled. Then the fluctuations of I_D are equally as large as those of I_T, since they are entirely independent of each other and therefore their mean squares may be added.

According to the definition given in the previous section F^2 is then exactly I_D/I_T.

When the fluctuations of the currents of various kinds of amplifier valves are measured in this way, very different values are found.

In general values will be found which are smaller than 1, but values larger than 1 will be found. The former values indicate that a certain regularity has appeared in the current of electrons which passes over to the electrode under consideration, and we shall see in the following that the space charge in an amplifier tube decreases the disorder in the transport of electrons to the anode, and therefore also decreases the factor F^2. The second case, $F^2>1$, is an indication that there must be another source of fluctuations besides the complete disorder with which the electrons are emitted. With complete disorder F^2 has the value 1. Only a quite different source of fluctuations such as a kind of abnormal modulation of the emission. due to temporary changes in the state of the cathode, the presence of other charged particles beside the electrons, or secondary emission, can provide an explanation of this.

The influence of secondary emission may be conceived as follows. Primary electrons which strike an electrode with sufficient speed may at the same time free more than one, for instance n secondary electrons: one speaks then of multiple secondary electrons which may be conceived as a charged particle with a charge equal to n times the charge on an electron. A current consisting of particles with the charge $n\,e$ must, according to the formula for the shot effect, exhibit fluctuations which are n times larger than the fluctuations of single electrons [2]).

Fluctuations in the state of the cathode itself may be caused by the presence of foreign particles which diffuse to the surface and cause local and temporary changes in the emission [3]). This phenomenon has nothing in common with the shot effect. The most striking difference is the dependence on frequency: these fluctuations are relatively slow and decrease sharply in intensity at high frequencies, in contrast to the shot effect fluctuations which, as we have seen in the article mentioned, are distributed uniformly over the whole frequency spectrum. At low frequencies the fluctuations due to the state of the cathode may exceed the shot effect in intensity by a large factor.

The presence of ions from the ionization of gas residues, or the emission of ions from the cathode may also lead to abnormal fluctuations [4]).

These phenomena however are not inherent in the principle of the amplifier valve, and need not necessarily appear in practice. We shall therefore further confine ourselves to the simple case in which the cathode (and only the cathode) emits electrons (and nothing else but electrons) and shows no changes with time, and the electron current is influenced by nothing else but the presence of the electrons themselves together with the constant voltage on the electrodes.

Influence of the space charge on the noise factor

If in our arrangement for the measurement of the noise factor, T is a saturated diode which satisfies all the conditions mentioned above (the standard diode D of course has been rigorously tested in this respect), we always find $I_D = I_T$, i.e. $F^2 = 1$. In a triode however in which the anode current is only a fraction of the emission, as a consequence of the presence of a cloud of electrons (space charge) around the cathode, a potential minimum occurs between anode and cathode, so that only the speediest electrons reach the anode. It is found that a change in the emission has an effect on the depth of the minimum, such that the number of electrons transported per second remains practically the same.

Fluctuations in the emission will therefore in this case be reproduced to a much smaller degree in the electron current to the anode, than when this is the saturation current. There appears therefore a decrease in the fluctuations due to the space charge; $F^2<1$. This decrease is a function of the saturation current I_s, the anode current I_a, and the geometry of the valve. For a given valve at a certain value of I_s one may determine the noise factor as a function of I_a/I_s. It is found that with decreasing value of I_a/I_s the noise factor is at first reduced. The decrease of F^2 with I_a/I_s may be very appreciable, but it is not unlimited. It is clear that, if by decrease of the anode voltage the anode current is made continually smaller, a state is finally reached at which a potential minimum no longer exists, but the potential of the anode is the lowest. In that case the space charge between cathode and anode has no influence on the current strength: all the electrons which have sufficient speed of emission to move against the negative anode voltage will reach the anode, and the current strength will fluctuate just as sharply as the number of these electrons. The emission of electrons with a speed higher than a definite value may, like the total emission, be considered as a series of independent events. The anode current thus again exhibits chance fluctuations and $F^2 = 1$. Thus F^2 as a function of I_a/I_s has a minimum. In *fig. 2* the curve for the noise factor is represented for a diode adjusted at an emission of 4.5 mA.

As an illustration of the decrease of fluctuations due to the space charge, we should like to mention measurements carried out on pentode EF 5, which was connected as a triode in this experiment by joining screen grid and plate and considering them as one anode. At the following adjustment:

$$V_{a+g} = 100\ \text{V},\ \ V_{g_1} = -2.5\ \text{V},\ \ I_{a+g} = 10\ \text{mA}$$

---

[2]) M. Ziegler, Shot Effect of Secondary Emission, Physica, 3, 1 and 306 (1936).

[3]) W. Schottky, Small Shot Effect and Flicker Effect, Phys. Rev. 28, 75 (1926).

[4]) S. Ballantine, Fluctuation Noise due to Collision Ionisation in Electronic Amplifier Tubes, Physics, 4, 294 (1933).

with the measuring arrangement described above we found:

$$I_D = 0.5 \text{ mA, thus } F^2 = I_D/I_T = 0.05$$

Since with this adjustment the slope of the triode is 2.2 mA/V, we calculate for this case a noise resistance of about 2000 ohms.

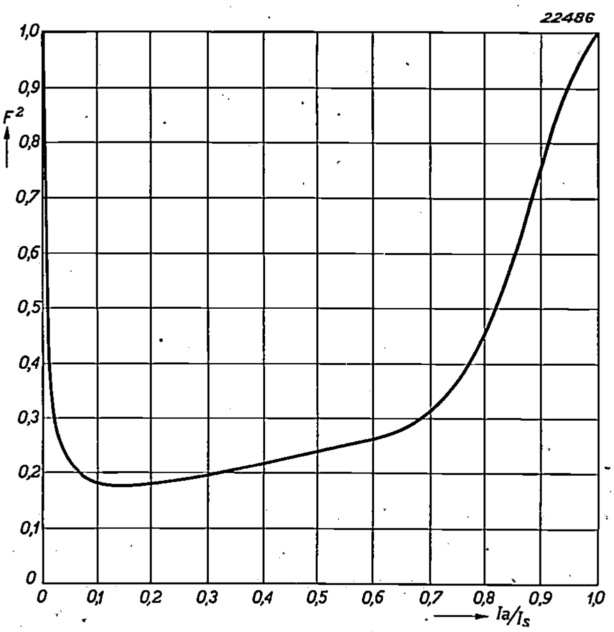


Fig. 2. Curve of the noise factor as a function of the current at constant emission from a diode with a tungsten filament.

As we have already seen this means, that in the presence of a resistance of 2000 ohms between grid and cathode, the thermal voltage fluctuations in this resistance cause anode current fluctuations which are as large as the fluctuation of the anode current itself. Since in general the resistance, i.e. the real portion of the impedance present between grid and cathode, in the frequency range under consideration, is much greater than 2000 ohms, the contribution thereof to the total fluctuations will be large compared with the contribution due to the amplifier valve mentioned.

Noise due to division of the current

It is known that because of the capacity between screen grid and anode a triode is not suitable for the amplification of voltages of high frequency [5]). Therefore a screen grid kept at a constant potential has been introduced between grid and anode, which makes the capacity very small. This however involves the fact that the total cathode current is divided into anode current and screen grid current. The quality of the valve is determined by the noise factor of the anode current: only the fluctuations of the anode current are superposed on the desired anode current variations which are caused by alternating voltage of the control grid to be amplified. If now the noise factor of the pentode EF 5 connected in a normal way is measured at the adjustment $V_{g_1} = -2.5$ V, $I_a = 7.5$ mA, $I_a + I_g = 10$ mA, which corresponds exactly with the adjustment given above for the same valve as triode, one finds $F_a^2 = 0.28$.

This means that, relatively as well as absolutely, the fluctuations of the anode current are much greater than the fluctuations of the total current. This may be ascribed to the appearance of the so-called division fluctuations: the total current fluctuates but little, each of its parts, however, fluctuates more strongly.

This phenomenon, like the shot effect, is due to the fact that the current consists of individual electrons.

The division of the total and practically non-fluctuating current into anode current and screen grid current depends at every moment on the chance position and direction of the electrons which are moving toward the positive electrodes: for a short moment more electrons will fall on the screen grid than correspond to the average, then again the anode will receive more electrons than the average number. How large are these fluctuations?

It is easy to calculate this for a special case. Assume that the electrons coming from the space charge follow each other at equal intervals of time so that no fluctuations appear in the total cathode current and the noise factor of the cathode current is therefore zero. Assume further that the screen grid is very widely spaced, so that by far the majority of the electrons reach the anode, or in other words that I_g is small compared with I_a. Now and again the direction and velocity of an electron will by chance be such that it must fall on the screen grid: the transitions of the electrons which go to the screen grid thus form a series of independent events, and the fluctuations of their number with the time are given by the formula for the shot effect. The fluctuations of the green grid are thus:

$$\Delta\, \overline{(I_g - \overline{I}_g)^2} = 2\, e\, I_g\, \Delta\, \nu.$$

The fluctuations of the anode current are of course equal and opposite, since the total current is strictly constant. The noise factor of the anode current is by definition therefore:

$$F_a^{\;2} = \frac{2\, e\, I_g\, \Delta\, \nu}{2\, e\, I_a\, \Delta\, \nu} = \frac{I_g}{I_a}.$$

[5]) See for example the article "Five-electrode Transmitting Valves (Pentodes)" Philips Techn. Rev. 2, 257 (1937).

For any arbitrary ratio between I_a and I_g for the case when the total current does not fluctuate and the division of the current obeys only the laws of chance, one finds the following:

$$F_a^2 = \frac{I_g}{I_a + I_g} \quad \cdots\cdots \quad (6)$$

Upon applying equation (6) to noise caused by the division of the current we must still take into account two deviations from the above assumption:

1) In reality the noise factor F_k^2 for the total current is not zero, so that F_a^2 is larger than the value given by (6). The difference $F_a^2 - F_k^2$ however is smaller than $I_g\ (I_g + I_a)$, since as the total current is not entirely free from fluctuations, the influence of chance in the division of the electrons is less noticeable. The less F_k^2 differs from 1, i.e. the more complete disorder is approached in the total current, the less is the disorderly division able to bring about new disorder; F_a^2 cannot be greater than 1 (total disorder).

2) In reality the division of the electrons between the screen grid and the anode is not always purely by chance, but may obey a certain geometrical regularity. It may happen that the electrons from certain points of departure on the cathode always land on the anode, and those from other points always on the screen grid.

This would occur if the screen grid did not consist of many thin wires but of a number of platelike strips. The portion between each strip and the cathode must then be considered as an individual triode, wherein therefore no division fluctuations occur. This phenomenon may be expressed mathematically by writing instead of equation (6) the following:

$$F_a^2 = a\,\frac{I_g}{I_a + I_g},$$

in which the chance factor $a < 1$.

In the case of the amplifier valve EF 5 described, where $I_g/(I_a + I_g) = 0.25$, one is concerned with both phenomena, as may be seen clearly from the values given for F_k^2 and F_a^2.

Principle of the "noiseless" valve

We have seen that the mean square of the noise voltage and the noise resistance of an amplifier valve are proportional to the noise factor F^2, the anode current I_a and inversely proportional to the square of the slope S. In order to attain a small noise resistance therefore I_a/S^2 and F^2 must be as small as possible. I_a and S are quantities which are more or less fixed for a given type of valve by the dimensions of the cathode and the requirements placed on the shape of the I_a-V_g-characteristic. A steep slope with a small anode current and the retention of all the desired properties is a goal which has always been aimed at for other reasons. Good valves are at present at the limit of the attainable in those respects. The noise resistance of the existing high-frequency amplifier valve could therefore only be made smaller by reducing the noise factor.

We have seen that the fluctuations of the anode current of a high-frequency amplifier valve may be ascribed chiefly to division fluctuations; one must therefore try to keep these division fluctuations as small as possible. Formula (6) shows us that they are about proportional to the screen grid current.

In a high-frequency amplifier valve developed in the Philips laboratory this current has been made very low. This was achieved by introducing between control grid and screen grid an extra grid which is wound with a pitch equal to the pitch of the screen grid, and is mounted so that the wires of the screen grid seen from the cathode are just in the shadow of those of the extra grid. If this extra grid is connected to the cathode, the distribution of the field is such that the electrons are, as it were, forced to move through the meshes of the extra grid and screen grid; only a small portion can reach the screen grid. Entirely in agreement with theory, it is then found that the increase of the noise factor due to division fluctuations is very small. The following are examples of values found for such an amplifier valve:

for $I_a = 8$ mA $\quad I_g = 0.2$ mA
$\quad F_k^2 = 0.06 \quad F_a^2 = 0.08.$

The slope is about 2 mA/V, so that the noise resistance of this valve according to (5) is

$$R = 20000 \times 0.08 \times 8/4 = 3200 \text{ ohms.}$$

For the normal EF5 with $I_a = 7.5$ mA, $S = 1.65$ mA/V and $F_a^2 = 0.28$, $R = 15\,500$ ohms. We see therefore that by reduction of the screen grid current an considerable improvement is achieved. The use of this special valve will thus result in a noticeable decrease of noise in the case where the real portion of the impedance between control grid and cathode is of the order of several thousand ohms.

THE ADJUSTMENT OF SYNCHRONIZERS FOR THE "PHOTOFLUX" PHOTO FLASH BULB

by J. A. M. VAN LIEMPT and J. A. DE VRIEND.

Summary. A description is given of a simple method of adjusting synchronizers so that the shutter of the camera is opened just at the moment of maximum intensity of the flash-light.

A synchronizer is a small subsidiary apparatus used in combination with a flash-light lamp and a photographic camera to provide that the evolution of light by the flash-light lamp takes place at the moment when the shutter is open.

In America this apparatus is in general use and practically every press photographer is equipped with one; in Europe it has not yet come into general use. We do not doubt that the situation here will soon change since synchronizers have important advantages.

In the first place they permit the photographer to operate both the shutter and the lamp with a single movement, which is a considerable simplification.

Further they provide that the light emission of the flash-light takes place at the right moment, that is, when the shutter is open widest. Considering that the time of the flash of a flash-light lamp is usually of the order of 1/40 sec, this means that with a shutter speed of 1/40 sec the whole of the illumination of the flash is used by the camera and unwanted (for instance frontal) illumination is reduced to a minimum.

In cases where the nature of the moving object to be photographed demands a short exposure time, of 1/100 sec or less, the synchronizer (at least one of the better kinds) permits us to make use of a portion of the total light emitted, and thus makes possible the taking of good snapshots in cases where lack of sufficient daylight or artificial light would otherwise prevent this.

Since the "Photoflux" flash bulb is made with a practically constant peak time, i.e. the time which passes between switching on the lamp and the maximum light development, the synchronizer, which operates not only the switch of the flash-bulb but also the shutter, especially with such short exposure times, must provide that the movement of the shutter begins and ends at such a time that as much light as possible from the flash-light lamp is used.

We are confining this discussion to Compur shutters. With focal plane shutters, each case must be examined individually since many focal plane shutters on the market are unusable with flash-light bulbs because the time of traverse of the slit over the plate is so long that the flash would expose only a portion of the negative.

Synchronizers of various kinds may be obtained. As will appear in the following, an adjustable synchronizer is to be recommended in which the instant of shutter operation may be varied.

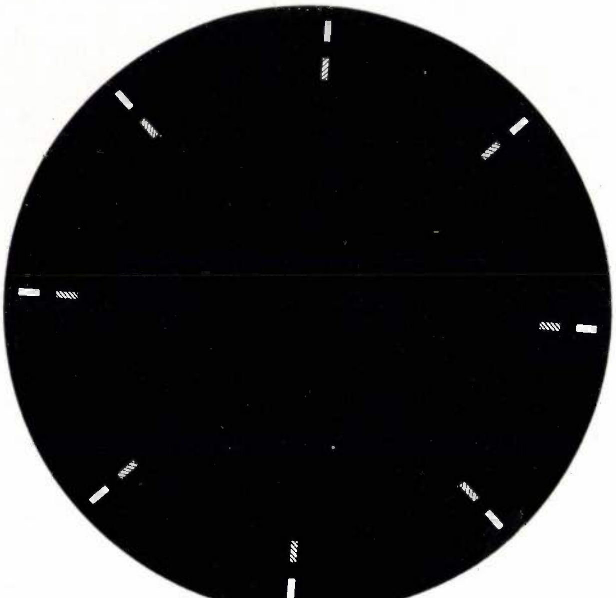
For the sake of clearness we will briefly describe one of the existing types. In this type the flash-light is ignited by switching on the current from a pocket battery. In parallel with this is an electromagnet, whose armature operates the shutter release. By variation of the distance between the armature and the core of the electromagnet, and by adjusting the movement of the armature to be quick or slow, it is possible to produce variation in the time between ignition of the flash and operation of the shutter.

When an adjustable synchronizer is to be employed it must be adjusted to suit the camera and the flash-bulb used. By means of a cathode ray tube it is possible to determine accurately not only the peak time of the flash-light bulb but also the time constant of the synchronizer at different adjustments [1]). This method, however, is too complicated for the ordinary photographer and therefore we are giving the following simple method of carrying out the desired adjustment, particularly for the case when the Philips "Photoflux" is used as flash-light.

A cardboard disc with a diameter of about 40 cm should be covered on one side with dull black paper. Round the edge of the blackened side of this cardboard, at equal distances, eight or more white, and the same number of dark grey strips of paper are glued, having the dimensions 3 × 15 mm (see *fig. 1* where the dark grey strips are shown

1) J. A. M. van Liempt and J. A. de Vriend, Physica 4, 703, 1937.

cross-hatched). This disc is fastened to a gramophone disc, a loose bicycle wheel, a motor shaft or the like. When a gramophone is used, which has the advantage of running with a constant speed



23627

Fig. 1. The cardboard disc with eight white and eight grey strips.

(78 revolutions per minute), it is preferable to set it on its edge.

A portion of the circumference of the disc is then photographed in a partially darkened room with the aid of the "Photoflux" and the synchronizer, at as nearly as possible full size (that is, with the camera close to the disc) and with an *effective exposure time of 1/100 sec and a stop of F . 11 or smaller.* A fixed black paper arrow set up in the direction of motion of the disc is photographed at the same time.

The picture of the strips will be quite different in appearance according as to whether the shutter was opened too early, too late, or at the correct moment. The case where the shutter was much too early or much too late, in which no picture at all is obtained, occurs very seldom. Usually in such a case the pocket battery will be found to have run down or the contacts in the circuits to have too much resistance. When these points are in order one should adjust the synchronizer until a picture is obtained.

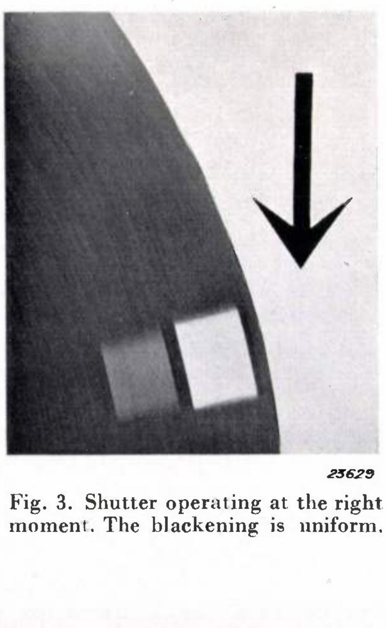
If the shutter opening is slightly too early, a picture like that in *fig. 2* is obtained. The distinguishing feature of this picture is that the intensity increases in the direction of the arrow. If the shutter is too late, one obtains a picture like that in *fig. 4;* this is distinguished by the fact that the intensity decreases in the direction of the arrow. If the shutter is open at exactly the right moment a picture like that in fig. 3 is obtained, with uniform intensity. By regulating the synchronizer it can always be set at the proper position in this way. In connection with the fact that most synchronizers and shutters and also the flash-light bulbs are subject to small variations, it is advisable to check the correctness of the adjustment several times in order to reach a good average.

The method here described is based upon the following considerations.

The light transmission of a Compur shutter with an effective opening time of about $^1/_{100}$ and $^1/_{200}$ sec is not the same at every moment, because of the fact that the movement of the sectors takes place with a finite velocity, but it may be represented by the ordinate of *fig. 5.* The continuous line in the figure holds for the case where the diaphragm is wide open (F . 3.5). When the opening is small (F . 11

23628

Fig. 2. Shutter too early. The blackening increases in the direction of the arrow.



23629

Fig. 3. Shutter operating at the right moment. The blackening is uniform.

23630

Fig. 4. Shutter too late. The blackening decreases in the direction of the arrow.

for instance) the maximum aperature is limited by the stop and is reached much sooner, as may be seen from the dotted line. The characteristic of the shutter is reduced therefore from approximately a triangle to a rectangle.

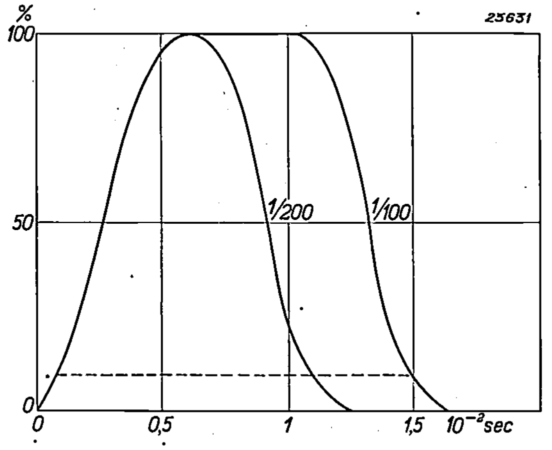


Fig. 5. The light transmission of a Compur shutter as a function of the time, for effective times of exposure of about $^1/_{100}$ and $^1/_{200}$ sec.

In *fig. 6* may be seen the flash curve of "Photoflux" Type II as it is recorded with a cathode ray tube [2]). The so-called "practical" flash time of this lamp is about $^1/_{40}$ sec and is thus shorter than the base of the curve in fig. 6, since the smaller intensities of light at the extremities give no negative blackening under the conditions under which the lamp is used, owing to the existence of a threshold value of the plate sensitivity.

The light transmission of the combined flash shutter can now easily be calculated from the

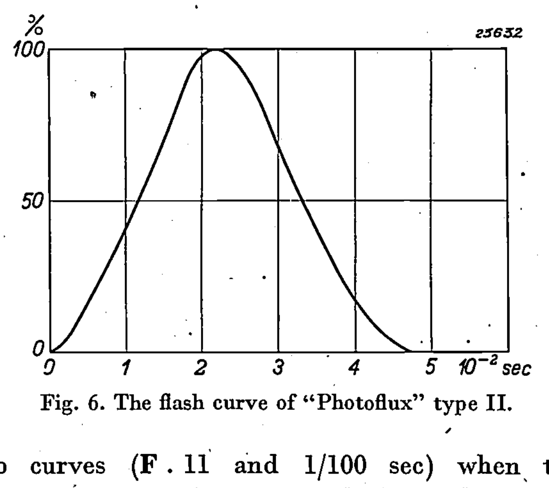


Fig. 6. The flash curve of "Photoflux" type II.

two curves (F . 11 and 1/100 sec) when the shutter opens 0.005 sec too early, or too late, or at the right moment. This is shown in *fig. 7*. From the figure it may be seen that if the shutter is too early or too late the light curve is asymmetrical with an obvious difference of intensity between the beginning and the end of the movement of the shutter, which is expressed on the negative as an easily observable difference in blackening. With greater deviations than 0.005 sec the difference is still

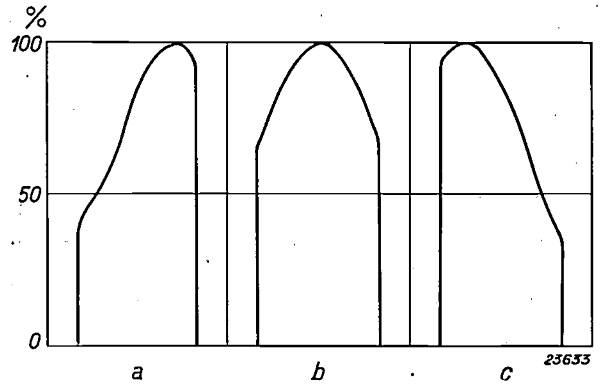


Fig. 7. The light transmission of the combination "Photoflux" (Fig. 6) with shutter $^1/_{100}$ sec and F . 11 (Fig. 5), when shutter is *a*, 0.005 sec too early, *b*, opened at the right moment and *c*, 0.005 sec too late.

more pronounced. In order to accentuate the difference in blackening the dark grey strip is glued to the disc and when the synchronizer is incorrectly adjusted, the grey strip fails in part to be shown in the picture because the brightness corresponding with its colour remains below the threshold value of the negative and thus simplifies the reading of the result. Reference is also made to fig. 2 in which the picture of the white strip after being interrupted appears again for a moment. This is caused by a common defect of the shutter consisting of the fact that after it has shut it opens slightly again and then shuts once more. This phenomenon is shown so plainly here because we were using a small opening of the diaphragm and because the light intensity in the case shown in fig. 2 was just at its maximum at the moment when the shutter reopened. The observation of this phenomenon is another proof of the fact that the shutter opened too soon.

We call attention to the fact that the "Photoflux" bulb [3]) is especially suitable for use with synchronizers because of its symmetrical somewhat flat-topped characteristic. We are concerned practically with small variations of the shutter movement caused by differences in humidity, temperature and friction, in addition there are variations in the flash-light bulb, the internal resistance of the pocket battery and also in the mechanism of the synchronizer itself.

When the synchronizer is once adjusted to an

[2]) See J. A. M. van Liempt and J. A. de Vriend, Physica 4, 353, 1937, and Philips techn. Rev. 1, 289, 1936.

[3]) For a detailed description see J. A. M. van Liempt and J. A. de Vriend, Philips techn. Rev. 1, 189, 1936.

average time of retardation, these variations will never lead to failures, since the shutter always remains open in that region of the light characteristic of the lamp where the light intensity is still adequate.

With the aid of figs. 6 and 4 the light transmission can easily be construed for the combination of shutter ($^1/_{100}$ sec) and "Photoflux" for the case where the half time value of the maximum opening of the shutter coincides with the peak value of the light emission of the lamp, and for the case where the deviation from coincidence is $\pm$ 0.005 sec. This has been done for a large lens opening (in this case **F** . 3.5) with which these lamps are generally used, and it is shown in *fig. 8*. As may be seen, the total amount of light which is allowed to pass scarcely differs, while the flash time is also unchanged. In addition it may be seen from this figure that the above described method of adjusting a synchronizer fails when a small opening of the diaphragm is not used. Only at a stop of **F** . 11 or smaller are the characteristic differences between too early and too late opening of the shutter manifested.

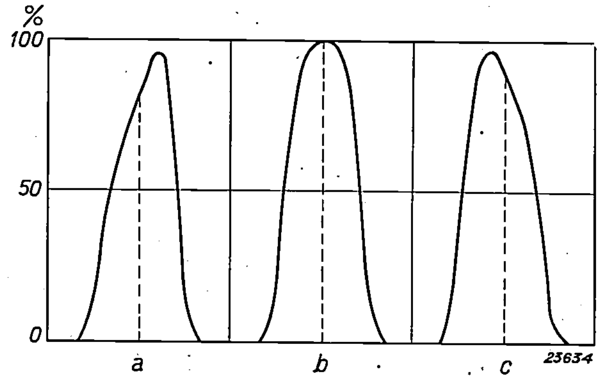


Fig. 8. The light transmission of the combination "Photoflux" (fig. 6) with shutter $^1/_{100}$ sec and F . 3.5 (fig. 5) when the half time value of the maximum opening of the shutter (given by vertical broken line) coincides with the peak value of the light emission of the lamp, and for the case when this falls 0.005 sec earlier (a) or later (c).

THE MOTION OF AN ELECTRON IN TWO-DIMENSIONAL ELECTROSTATIC FIELDS

by P. H. J. A. KLEYNEN.

Summary. The motion of a charged particle in a two-dimensional electrostatic field may be studied with the help of a model consisting of a stretched rubber membrane, over which a ball rolls.
The membrane is stretched in such a way that the tension of the surface is constant over the whole area, and the limiting conditions of the electrostatic field are applied to the model in the sense that the deviation of the membrane in the vertical direction is considered proportional to the electrostatic potential.
In several cases is studied the approximation of the path of a ball rolling over this rubber surface to the path of the electron in corresponding potential fields.

Introduction

The development of radio valve technique often demands a more or less accurate knowledge of the motion of the electrons in the electrostatic field of the valve electrodes. The mathematical investigation of this motion according to the ordinary laws of mechanics is impossible in most cases. The calculation of the potential field from the fairly arbitrary limiting conditions as they occur in practice is prevented as a rule by mathematical difficulties. This calculation is however only the first step in the right direction. If in some way or other the potential field has been successfully calculated, then the motion of the electron in this field must still be calculated with the aid of mechanics, which is in most cases no less difficult. It is obvious that other methods have been sought which lead more rapidly to a final result, although they may be less accurate than the mathematical analysis. In the first place one may try to determine the potential with the aid of probe measurements, or by making the lines of force visible. In addition one may look for other natural phenomena which may be described in the same way as the electric potential field by the differential equation of Laplace and which are often more amenable to measurement than the electrostatic potential. The most important method arising from these attempts is that of the electrolytic trough [1]). If the field is once known the equipotential lines may be drawn and the electron paths may be fairly easily constructed in this field of equipotential lines. Apparatus has also been developed recently which when coupled to an electrolytic trough, immediately draws the electron paths [2]). A disadvantage of all these methods is the unknown connection between the limiting conditions of the potential field and the final electron paths. If one desires to know the influence of the potential of one of the electrodes on the path, it is necessary to determine the field at various values of the potential, and then construct the path.

A method without these disadvantages is based upon the following principle [3]). If one imagines a membrane, in which only surface tensions can act, stretched horizontally in such a way that the surface tension is constant over the whole surface, and if the membrane is given a vertical deviation at several points, then the height of every point of the membrane when the deviations are small satisfies the differential equations of Laplace for two dimensions. If we give those points of the membrane which correspond to the electrodes of the valve a height which is proportional to the potential of those electrodes, then the height of every point is proportional to the potential in the electrostatic field. If we now allow a point mass to slide over the membrane without friction, with initial conditions similar to those of the electron in the potential field, then the horizontal projection of the motion of the point is similar in form to that of the electron, at least when the deviations are small. The deviation is considered small, when the slope of the membrane is small with respect to unity. Since

1) This method is based on the fact that the potential distribution is not disturbed when all the electrodes are immersed in an electrolyte. Because current flows between the electrodes, it is possible to measure the potential at any point with the aid of a probe electrode.

2) See for example D. Gabor, Nature **139**, 373 (1937) and D. B. Langmuir, Nature **139**, 1066 (1937).

3) See also W. R. Smythe, L. H. Rumbaugh and S. S. West, Phys. Rev. **45**, 724 (1934).

the membrane only approximately satisfies the two-dimensional Laplace differential equation, it is clear that all potential fields cannot be investigated by means of the above described model.

In the following we shall attempt to obtain an idea of the accuracy of this method by comparing several calculated electron paths with measurements taken from the model.

The differential equation of the membrane

A membrane in which only a surface tension acts will adjust itself upon deformation in such a way that the potential energy is at a minimum. Since this is proportional to the surface area, the membrane will assume a form which represents the smallest area under the given limiting conditions. Differential calculus gives the following equation for such a surface:

$$\left[1+\left(\frac{\delta h}{\delta y}\right)^2\right]\frac{\delta^2 h}{\delta x^2}-2\,\frac{\delta h}{\delta x}\,\frac{\delta h}{\delta y}\,\frac{\delta^2 h}{\delta x\,\delta y}+ \\ +\left[1+\left(\frac{\delta h}{\delta x}\right)^2\right]\frac{\delta^2 h}{\delta y^2}=0.$$

h is here the deviation of a point of the membrane from the horizontal, x and y are rectangular coordinates in the plane of the undeformed membrane. The first derivatives occur here in the second power, and if they are small we may neglect their squares with respect to unity.

The equation then becomes:

$$\frac{\delta^2 h}{\delta x^2}+\frac{\delta^2 h}{\delta y^2}=0$$

and thus becomes identical with the Laplace equation.

We can also derive this equation from the equilibrium conditions of an element $dx\,dy$ of the surface.

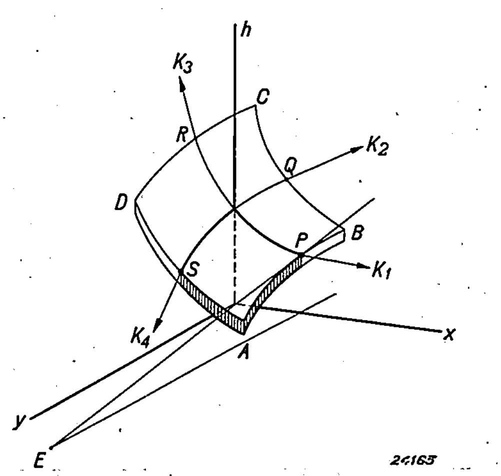


Fig. 1. Derivation of Laplace's differential equation from the equilibrium conditions of an element of the membrane.

In *fig. 1* $A\,B\,C\,D$ represents the surface element $dx\,dy$ which is formed by the intersection of the membrane by planes parallel with the coordinate planes. Forces act upon the intersecting planes, which touch the surface and are perpendicular to the curve of intersection. The element is in equilibrium, therefore the sum of these forces is zero:

$$K_1+K_2+K_3+K_4=0.$$

Of these forces we shall examine K_1. This force is perpendicular to the line of contact $P\,E$ on the intersection $A\,B$, and lies in the surface. If now the surface has only a slight slope from the xy plane, so that $P\,E$ makes only a small angle with the y-axis, then the forces K_1 and K_3 will lie practically parallel to the line $P\,R$ of the surface, and the forces K_2 and K_4 practically parallel to the line $A\,S$. In this case for the sum K_1+K_3 we find:

$$K_1+K_3=\sigma\,dy\left[\left(\frac{\delta h}{\delta x}\right)_P-\left(\frac{\delta h}{\delta x}\right)_R\right]=\sigma\,dy\,\frac{\delta^2 h}{\delta x^2}\,dx,$$

in which σ is the surface tension, and for K_2+K_4

$$K_2+K_4=\sigma\,dx\,\frac{\delta^2 h}{\delta y^2}\,dy\,.$$

From the condition that the sum of all these forces must be equal to zero, it follows that:

$$\frac{\delta^2 h}{\delta x^2}+\frac{\delta^2 h}{\delta y^2}=0\,.$$

Thus we see that the membrane satisfies Laplace's differential equation if the slope is sufficiently small at each point.

If we now apply to the membrane and the potential field similar limiting conditions, i.e., if we give the membrane at points corresponding with the electrodes a height proportional to the potential of these electrodes, then at every point of the membrane the height is proportional to the potential at the corresponding points of the electrostatic field.

The analogy between the model and the electrostatic field can be worked out somewhat farther. It is clear that lines of equal height correspond to the equipotential lines of the field. The maximum slope at one point of the membrane is proportional to the field strength. In electrostatics furthermore, the charge on an electrode is proportional to the integral of the field strength along the edge of the electrode, and we may therefore determine the charge with the help of the model by measuring the slope along the corresponding edge and integrating. Considering that the vertical component of the force acting on a linear element of the edge is proportional to the local slope of the membrane, this integral corresponds to the pressure of the membrane on the electrode.

The paths of the electron and the sliding point-mass

A consideration of the horizontally stretched membrane with the deviations applied, naturally gives rise to the question as to how far the horizontal projection of the motion of a point-mass, sliding over the surface without friction in the field of gravity, corresponds to that of the electron in the corresponding electrostatic case.

For the equation of motion of the electron we find

$$m\,\frac{\mathrm{d}^2x}{\mathrm{d}t^2} = -e\,\frac{\delta V}{\delta x}\,,\quad m\,\frac{\mathrm{d}^2y}{\mathrm{d}t^2} = -e\,\frac{\delta V}{\delta y}\quad\ldots\ (1)$$

We shall try to discover whether the differential equations for the projection of the motion of the point of mass on the horizontal plane have the same form.

Fig. 2 gives a vertical cross section through

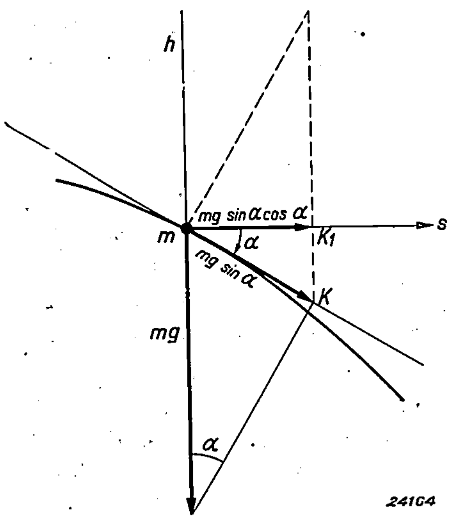


Fig. 2. Derivation of the forces acting on a point mass which slides on a sloping membrane without friction.

the membrane in the direction of greatest slope, with the forces which act on the point of mass m at a definite point of the membrane as a result of the force of gravity. It may be seen from the figure that the horizontal component K_1 of the resultant force K is equal to:

$$K_1 = m\,g\,\sin\alpha\cos\alpha,$$

where α is the angle of maximum slope.

Furthermore:

$$\tan\alpha = -\frac{\mathrm{d}h}{\mathrm{d}s}\,,$$

therefore:

$$K_1 = -mg\cdot\frac{\dfrac{\mathrm{d}h}{\mathrm{d}s}}{1+\left(\dfrac{\mathrm{d}h}{\mathrm{d}s}\right)^2}\,.$$

The components of the force in the x and y directions respectively are then:

$$\left.\begin{aligned} K_x &= m\,\frac{\mathrm{d}^2x}{\mathrm{d}t^2} = -mg\,\frac{\dfrac{\delta h}{\delta x}}{1+\left(\dfrac{\mathrm{d}h}{\mathrm{d}s}\right)^2}\,,\\ K_y &= m\,\frac{\mathrm{d}^2y}{\mathrm{d}t^2} = -mg\,\frac{\dfrac{\delta h}{\delta y}}{1+\left(\dfrac{\mathrm{d}h}{\mathrm{d}s}\right)^2}\,.\end{aligned}\right\}\quad(2)$$

If we now again assume that the slope of the membrane is small, we may neglect $\left(\frac{\mathrm{d}h}{\mathrm{d}s}\right)^2$ with respect to unity and equation (2) becomes:

$$m\,\frac{\mathrm{d}^2x}{\mathrm{d}t^2} = -mg\,\frac{\delta h}{\delta x}\,,\qquad m\,\frac{\mathrm{d}^2y}{\mathrm{d}t^2} = -mg\,\frac{\delta h}{\delta y}\quad\cdot\ (3)$$

These equations are indeed equivalent to (1) and with the proper choice of the boundary conditions the motions of the electron and of the point-mass will therefore also be equivalent.

Construction and testing of the membrane

Rubber is practically the only suitable material for the membrane. A sheet of rubber about 1 m in diameter, upon which rectangular and polar coordinates have been drawn before stretching, is stretched in a metal ring in such a way that the tension is constant over the whole area. If the coordinates drawn on the sheet are uniformly enlarged by the stretching, it may be assumed that the surface tension is everywhere constant. *Fig. 3* gives an idea of the arrangement. The preliminary check tests were carried out with a sheet 1 mm thick which was fastened with a stretch of 3 per cent. It was found that the irregularities in the tension at the edge which result from the method of fastening, extend no farther than about 5 cm toward the centre.

The sheet is set up horizontally on a table with a horizontal and plane surface. The plane projection of the electrodes of the electrostatic arrangement is now applied to the sheet on a suitable scale and at the positions of the electrodes the sheet is given deviations proportional to the potential of these electrodes by means of blocks or rings placed on or beneath the sheet. Since the charge of the electron is negative we plot the positive potential in a downward direction and the negative in an upward direction.

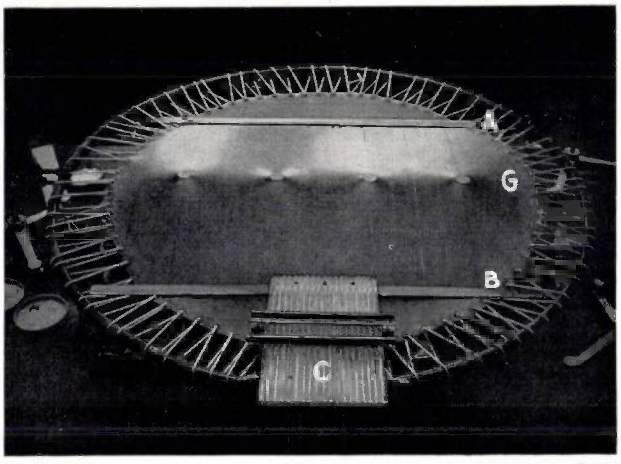


Fig. 3. Survey of the arrangement of the model. The rubber membrane is uniformly stretched in a horizontal iron ring. The height is fixed by two horizontal strips A and B. and by four cylinders G placed under the membrane.

From the form of the differential equations of the membrane and from the motion of the point-mass it follows that the path is independent of the scale on which the model is built. This holds not only for the linear dimensions but also for the ratio in which potential is transformed into height. It is known that in the electrical case also the shape of the path of the electrons does not change when the linear dimensions of the electrode system or the potential are increased proportionally.

Fig. 3 shows the potential field of a negative grid between two parallel positive plates. We shall return to discuss this in more detail later. The circles at G represent the intersections of the grid wires with a plane perpendicular to the grid wires. The straight lines at A and B are the lines where this plane cuts the parallel plates. The grid G consists in this case of four equally high cylinders which are placed on the table under the membrane. The field between the two middle grid supports is practically unchanged when the number of cylinders is increased to the left and right.

After the membrane is set up with the correct boundary conditions, the problem is to have a point-mass slide over the surface without friction. It is found, however, that with a membrane of rubber the coëfficient of sliding friction is much too large to produce a usable result. Therefore a rolling ball is substituted for the sliding point-mass, keeping in mind that the rotational energy of the ball and the rolling friction may be the causes of deviations in the path of the ball from that of a point mass sliding without friction. When we attempt to obtain an idea of this source of error in a mathematical way, we meet with great difficulties. We may introduce the moment of inertia and also the rolling friction into the differential equations of the motion in some form or other, but it is difficult to draw a conclusion from the form of these equations about the influence of these factors on the form of the path.

The only possibility then remaining is an attempt to show the errors in an experimental way. This can be done in different ways. In the first place with a given arrangement of the membrane we may change the friction and the moment of inertia of the ball, and then compare the paths of the ball. The friction for example may be changed by spreading a thin layer of powdered talcum on the membrane and the moment of inertia by taking balls with a heavy core and a light covering or the reverse. A simple calculation shows that a ball can be made with a platinum or a tungsten core and a aluminium covering, which has the same mass and the same diameter as a steel ball, but whose moment of inertia is one third as large. A second method consists in using an arrangement in which the motion of a sliding point mass is theoretically accurately known, and comparing this with the motion of the ball. A third method consists in the comparison of the path of the ball with the path of electrons, the shape of whose path is already completely known from mathematical or experimental investigation.

We have applied the second method to the motion of a ball on a flat sloping plane, and the third method to the motion of an electron in the field of a slit between two parallel plates.

The motion of a ball on a flat sloping plane

If the coördinates are placed as in *fig. 4*, the path of a point mass sliding without friction over

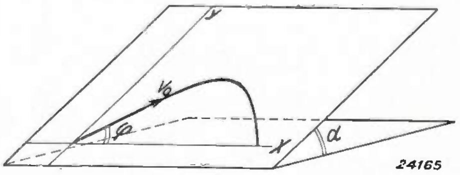


Fig. 4. Motion of a point mass on a sloping plane.

a flat sloping plane is a parabola with the equation:

$$y = x \tan \varphi - \frac{g \sin \alpha}{2\, v_0^2 \cos^2 \varphi}\, x^2, \quad . \quad . \quad (4)$$

where v_0 is the initial velocity and g the acceleration due to gravity.

The motion of the ball on the sloping plane may be recorded as follows. The highly polished ball is allowed to leave the source and is photographed during its motion while the arrangement is illuminated with a high pressure mercury lamp burning on single-phase alternating current. Since this lamp is periodically lighted and extinguished, and the period is equal to half that of the alternating current, the path of the ball is visible as a dotted line (see *fig. 5*). If the lamp is running on

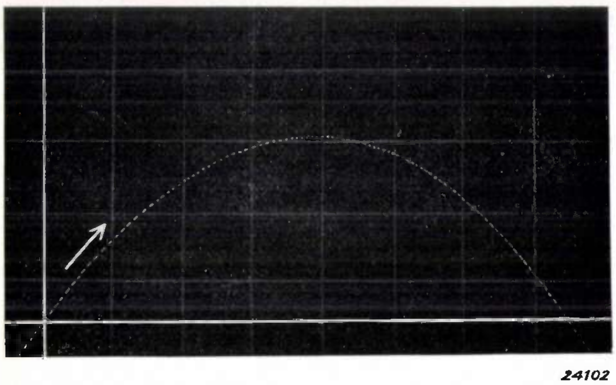


Fig. 5. Photograph of the path of a polished ball on flat sloping plane. The ball is illuminated by a lamp which is periodically extinguished, so that the track is visible as a dotted line.

50 period alternating current, the distance between two adjacent dots is covered in one-hundredth of a second, and the speed of the ball can thus be calculated directly from this distance. The angle φ at which the ball leaves the source and the slope of the membrane can be measured, and with the aid of (4) we can calculate the path which is followed by a point mass sliding without friction when it leaves the source under the same conditions. It must in addition be noted that the rotational energy of a truly rolling homogeneous sphere is $^2/_7$ of the total kinetic energy, so that we must multiply the velocity derived from fig. 5 by $\sqrt{7/5}$ in order to obtain the corresponding velocity of the sliding point mass.

If we carry out the calculation in this case and compare the parabola with the path of the ball of fig. 5 we find that the top of the parabola lies about 1 per cent higher than the top of the path of the ball, while the distance between the points at which the parabola cuts the x-axis is about 7 per cent greater than that between the corresponding points of the path of the ball. If the friction is increased slightly this distance decreases sharply, so that we have in this fact a good criterion for the influence of friction.

The electron in the field of a slit.

The foregoing case gives us an idea of the influence on the path of the friction alone. In the case now to be examined, the motion will be affected by sources of error which are connected with the fact that the slope of the membrane is not infinitesimally small, in addition to friction.

We imagine a plane plate S with a straight slit parallel to and situated at equal distances from two parallel plates A and K. *Fig. 6* shows a cross sec-

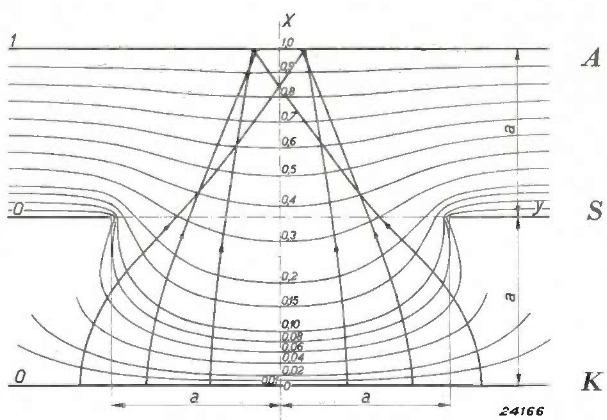


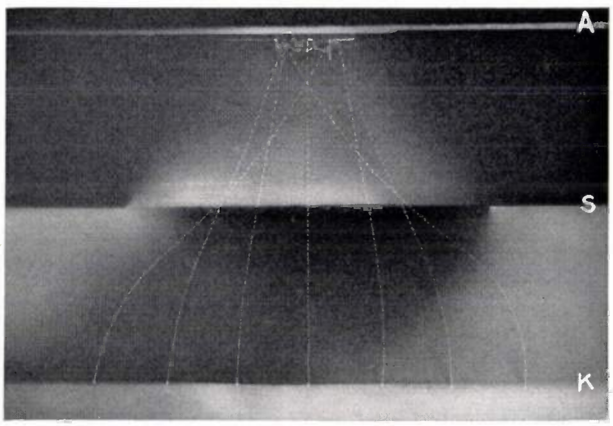
Fig. 6. The potential field of a slit between two parallel plates with the electron paths constructed in this field.

tion perpendicular to the plates and to the slit. With the aid of the theory of the function of a complex variable the potential field of such an electrode arrangement may be calculated. If we apply to plate A a potential equal to unity and to the slit and the other plate a potential zero, and if we choose the width of the slit and the distances between the plates as given in fig. 6, the potential V is given by the following:

$$V = \frac{x}{2a} + \frac{1}{2\pi} \arccos \left[\frac{\cosh^2 \frac{\pi y}{2a} - \sin^2 \frac{\pi x}{2a} - \sqrt{\left(6{,}295 + \cosh^2 \frac{\pi y}{2a} - \sin^2 \frac{\pi x}{2a}\right)^2 - 25{,}180 \cosh^2 \frac{\pi y}{2a} \cos^2 \frac{\pi x}{2a}}}{6{,}295} \right].$$

The calculation of V from this formula gives the equipotential lines shown in fig. 6, and in this field of equipotential lines we can construct the electron path in a simple manner. The construction is based on the same principle as the apparatus which derives the shape of the path from a potential and gradient measurement in the electrolytic trough (see for example D. Gabor and D. B. Langmuir, loc. cit.). The radius of curvature of the path can be calculated at every point from the potential and the gradient perpendicular to the path. If we now allow the electrons to leave with a velocity zero at the plate K, then the construction gives the paths reproduced in fig. 6.

In the rubber model the plates K and S are at equal heights, while A is somewhat lower. At K we allow 7 balls to roll away rapidly one after another with no initial velocity. We then obtain the paths as shown in *fig. 7*.



24103

Fig. 7. Photograph of the motion of balls on a rubber model which corresponds to the potential field represented in fig. 6.

If we compare figs. 6 and 7 we see that the agreement between the path of the electrons and that of the balls is quite good, at least when we do not require too great a degree of accuracy. Our experience has been that the rubber model may prove very valuable for design purposes in radio valve technology. In other cases, however, such as for example the determination of distortions of the image in electron optics, one must be careful, and only more elaborate tests of the above mentioned type will be able to give assurance that the accuracy attained is sufficient.

Limitations

As we have already mentioned, the application is limited to two-dimensional cases, or to three-dimensional potential fields in which the potential is independent of one of the coordinates. In general therefore rotation-symmetrical fields for example cannot be investigated by means of the model.

With greater electron densities we must moreover take into account the influence of the space charge on the potential field. The difficulty here is that the change of potential due to this space charge is dependent on the very electron path which we want to determine. If however the space charge is not too great, one can first determine the paths on the assumption that the space charge is zero, and with the help of these paths one may then make an estimation of the space charge. The lowering of the potential which results from this may then be applied to the model by increasing the height of the membrane locally proportional to this lowering of potential, and the path may then be determined once more. If this process is repeated the results usually converge quite rapidly to the correct solution.

Several other applications

The rubber model may also be used to measure capacities of two-dimensional electrode arrangements. If, for example, it is desired to determine the capacity of an electrode with respect to another set of electrodes, in the electrostatic case that electrode may be given a potential difference V with respect to the other electrodes which all have the same potential, and then the charge of the electrode under consideration may be determined. This charge is, per cm axial length, equal to $\frac{1}{4\pi}$ times the integral of the field strength along the edge of the electrode. The ratio between charge and potential difference then gives the capacity per cm. In the rubber model the potential difference V is equivalent to a difference in height h between the electrode named and the others, while instead of the field strength we integrate the slope of the membrane along the edge of the electrode. We then find the following for the capacity:

$$c_1 = \frac{1}{4\pi h}\int_S \tan\alpha \, ds = \frac{1}{4\pi}\cdot\frac{D}{\sigma h}, \text{ where}$$

c_1 = capacity per cm axial length
h = difference in height between the electrodes
α = slope of the membrane at the edge of the electrode
S = limit of the electrode
D = pressure on the electrode
σ = tension of the rubber membrane.

The dimensions of the electrode system may be applied to the model in any arbitrary enlargement,

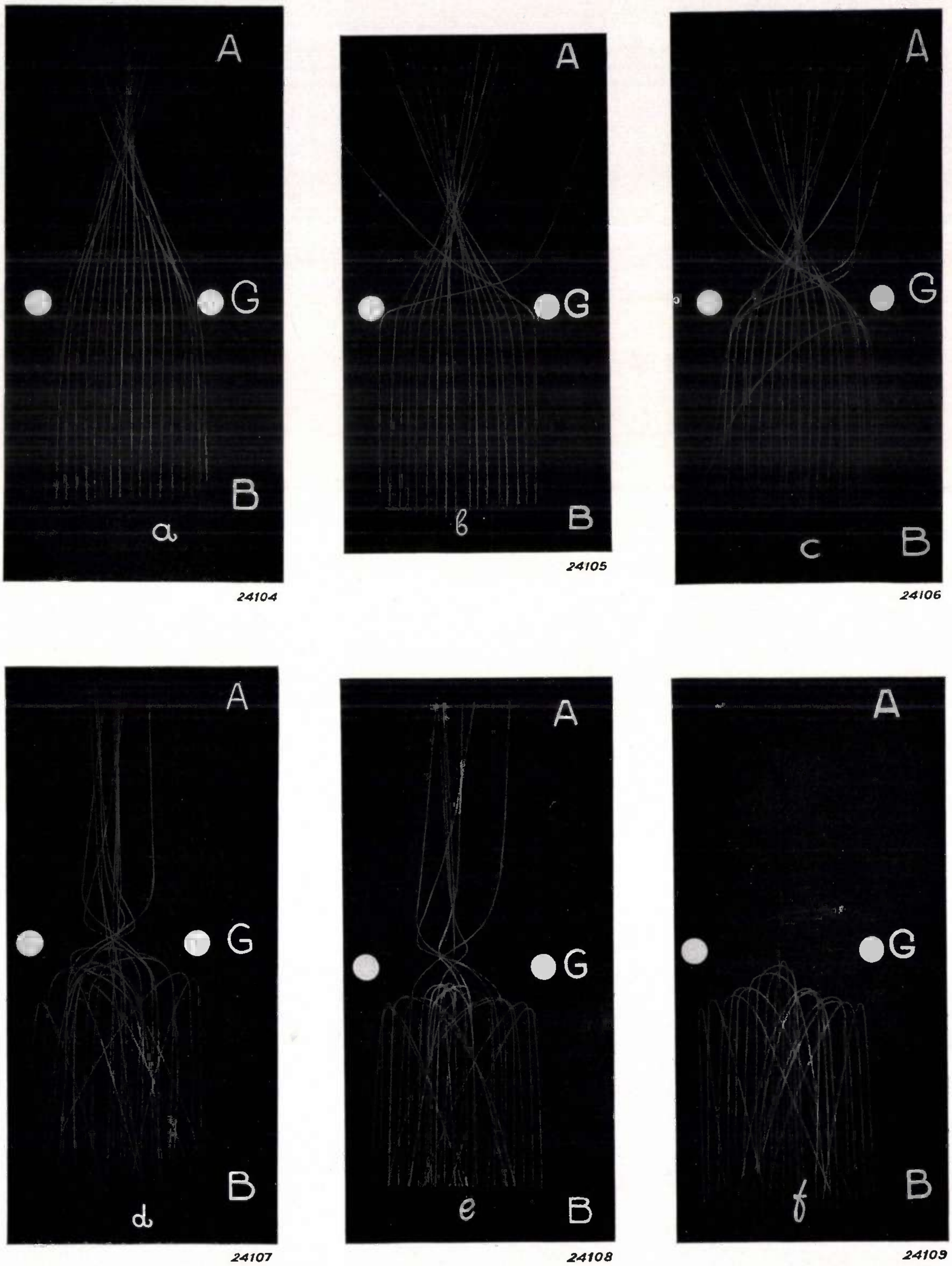


Fig. 8 a - f. Photographs of paths of balls, obtained by shooting a parallel "beam" of balls (in the arrangement of fig. 3) with a constant velocity perpendicular to the line joining the grid supports. The different figures are obtained by taking the height of G successively a little greater.

since the capacity of a two-dimensional arrangement per cm axial length is without dimensions. The slope may be measured by laying a small mirror on the membrane, and then with the aid of a telescope with a divided scale and a level recording the angle from the normal.

A photograph of the path of the ball further enables us to determine the time of transition of the electron in the corresponding electrostatic case. We have already seen that we may calculate the velocity of the ball at every point from the distance between two dots. This velocity is, however, dependent upon the dimensions of the model and also is much smaller than that of the electron. Since, however, the motions of the electron and of the point mass sliding without friction are similar in shape, we may consider it reasonable that there is a constant relation between the time elements of the motions of the electron and of the point. This relation may be determined from the differential equations (1) and (3), which for this purpose may be written in the following form:

$$\frac{d^2 x_e}{d t_e^2} = -\frac{e}{m}\frac{\delta V}{\delta x_e} \quad (1a), \qquad \frac{d^2 x_m}{d t_m^2} = -g\frac{\delta h}{\delta x_m} \quad (3a)$$

$$\frac{d^2 y_e}{d t_e^2} = -\frac{e}{m}\frac{\delta V}{\delta y_e} \quad (1b), \qquad \frac{d^2 y_m}{d t_m^2} = -g\frac{\delta h}{\delta y_m} \quad (3b)$$

The subscripts e and m indicate that the quantities concerned are taken from the electrostatic case and from the rubber model respectively. Since the ratio between the size elements (dx_e/dx_m and dy_e/dy_m) is known, and since we also know to what difference in height in the model the unit tension corresponds, we may determine from the equation the ratio of the time elements. We may for example divide equations (1a) and (3a) by each other, and then obtain from the resulting equation for the ratio of the time elements:

$$\frac{dt_m}{dt_e} = \frac{x_m}{x_e}\sqrt{\frac{eV}{gmh}}\,.$$

If we assume that the ball is homogeneous, and that it rolls without slipping, then, because of its rotational energy, its velocity at every point is at the most $\sqrt{5/7}$ times that of the sliding point mass under the same conditions, and in the case of the motion of a ball on a flat plane this factor is exactly $\sqrt{5/7}$. We shall be able to account for the moment of inertia to a large degree by choosing the time elements with the ball motion a factor $\sqrt{7/5}$ greater. We then find for the ratio:

$$\frac{dt_m}{dt_e} = \frac{x_m}{x_e}\sqrt{\frac{7eV}{5mgh}}\,.$$

Finally we shall give an example of the application to the subject of radio valves. If we imagine a positive anode and in front of it a grid at a lower potential, we may then ask how the electrons of a parallel beam, which are sent out with a definite velocity perpendicular to the anode, are deflected by the wires of the grid. The model then has the form shown in fig. 3. In this figure A and G represent the anode and the grid respectively. The balls are freed perpendicularly to B with a speed determined by the slope of a set of tubes C, in which the balls have attained their speed under the influence of gravity. The potential of plane B is determined at the same time by this speed, and with it those of A and G.

Fig. 8 gives a picture of the paths of the balls with various heights of G, that is with various potentials of the grid. In the first recording (fig. 8a) the potential of the grid was such that all the electrons were able to pass the grid, and we see how the electrons come together in a focus point between grid and anode. If we lower the potential of the grid, the focal distance becomes shorter and the electrons at the edge of the beam turn back. In fig. 8f the grid is so negative that all the electrons are reflected.

An important advantage of the above-described method is, that with very little trouble one may investigate the influence of a change of arrangement or potential of the electrodes. It is only necessary to shift the electrodes in the model slightly or to increase their height, and then to observe the paths of the balls.

A VIBRATOR FOR THE CONNECTION OF ALTERNATING CURRENT RECEIVING SETS TO THE DIRECT CURRENT MAINS

by J. W. ALEXANDER.

Summary. Receiving sets can be more advantageously constructed for alternating current than for direct current supply. In this article a vibrator is described which transforms direct voltage into alternating voltage. By introducing this instrument into the circuit, an alternating current receiver is rendered suitable for supply with direct current of the same mains voltage without further changes in its construction.

Introduction

In the case of radio sets which are fed from alternating current mains it is very simple to obtain the necessary heating voltage by means of a transformer for the valves connected in parallel. The necessary anode voltage is delivered *via* another winding on the same transformer and a rectifier valve. By giving the primary of the transformer several tappings the set may be made suitable for the ordinary voltages of 100-155 volts and 200-250 volts.

If, however, the apparatus is to be supplied from direct current mains, several difficulties arise which are all connected with the fact that no transformer can be used to adapt the apparatus to the voltage available. It is for instance necessary to connect the filaments of the different valves in series instead of in parallel, which introduces difficulties, since it is not easy with this method of connection to avoid the presence of hum from the mains. The provision of sufficient insulation to avoid the danger of accidental contact is much more difficult with direct current sets than with alternating current sets. Moreover it is practically impossible to insure that the set works equally well on all the mains voltages existing between 110 and 250 volts direct current.

The above-mentioned difficulties have led to an attempt to remove the disadvantage of direct current supply by transforming the direct voltage into alternating voltage by means of a vibrator. By connecting this instrument in circuit with the set, an alternating current receiver is rendered suitable for supply with direct current of the same range of mains voltages without the necessity of further alterations in its structure.

Principle and functioning of the vibrator

The circuit of the vibrator is reproduced diagrammatically in *fig. 1*. The vibrator may be considered as a double commutator, consisting of the springs A_1 and A_2, electrically separated, but mechanically connected, which can make contact alternately with K_{11} and K_{12}, and with K_{22} and K_{21} respectively. The springs are moved by an electromagnet M acting on an armature, which is mechanically

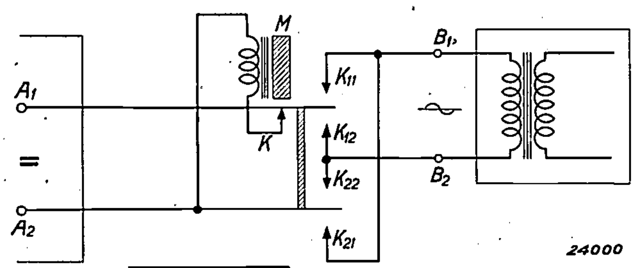


Fig. 1. Circuit diagram of the vibrator.

connected with the springs A_1 and A_2. If the vibrator is at rest and the terminals A_1 and A_2 are not yet connected to the mains then spring A_1 is connected to the coil of the magnet by means of the contact K. When the vibrator is switched on the electromagnet will attract the armature, thereby making contact between A_1 and B_1 (*via* K_{11}) and between A_2 and B_2 (*via* K_{22}). The contact K is broken. The springs bend farther to their maximum deviation, return and, owing to their inertia, pass the zero point to make contact between A_1 and B_2 *via* K_{12}, and between A_2 and B_1 *via* K_{21}. Upon passing the point of rest the magnet is again brought into circuit, etc. In this way the mechanism continues to function.

It is important to note, that an essential condition for the functioning of this mechanism is that growth of current in the magnet be retarded by its own self-induction.

When the contact K is closed and the armature is moving from the point of rest to the maximum deviation, the armature is retarded; when the armature is moving from the maximum deviation toward the position of rest, it is accelerated. If the self-induction of the electromagnet had no retarding effect on the current, the retarding force would be equal to the accelerating force, and the total energy applied to the armature would be zero. As a consequence of the self-induction an excess of applied energy occurs, and this is necessary in order to keep the mechanism working.

The result of making contact alternately at K_{11} and K_{22}, and K_{12} and K_{21} respectively is that the transformer is given a voltage of alternating sign, which, apart from the transformation ratio, induces the same voltage on the secondary side. If the transformer is shunted by a resistance, the secondary current has the form of *fig. 2b*, while a primary current flows as shown in *fig. 2a*, an interrupted direct current.

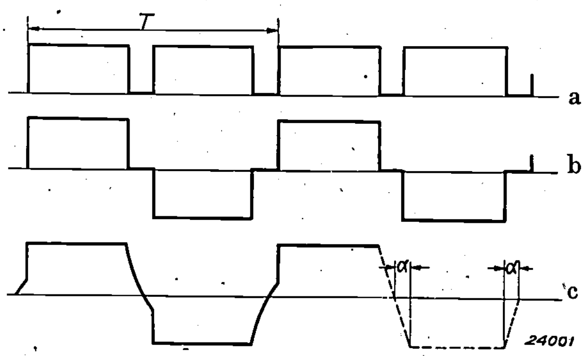


Fig. 2. a). Intermittent direct current taken from the input side of the vibrator.
b). Secondary voltage which would be delivered by a vibrator according to the circuit of fig. 1. T is the time of one period.
c). Alternating voltage delivered when the contacts are bridged with the help of condensers. The broken lines indicate how the curve may be idealized.

A vibrator constructed on this principle would have the disadvantage that at each half period upon commutation a high voltage would occur at the contacts where the circuit is broken, as a result of the self-induction of the transformer. This voltage surge can, however, be avoided by shunting the contacts with condensers. This is indicated in *fig. 3*. We assume that the vibrating spring is just on the point of leaving the contacts K_{11} and K_{22}.

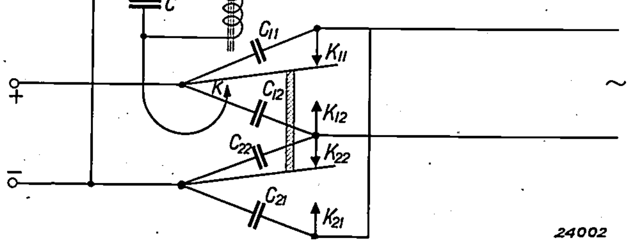


Fig. 3. By bridging the contacts with condensers high voltage commutation surges can be avoided every half period.

The current from the mains will then immediately be interrupted. The current through the transformer, however, will be able to continue and charge the condensers C_{11} and C_{22}, and will decay in a manner determined chiefly by the magnitude of the load impedance between B_1 and B_2, by the capacity of the condensers and the self-induction of the transformer. The slope di/dt of the current-time curve thus remains finite, and consequently also the output potential.

When the vibrating spring reaches the opposite contacts K_{12} and K_{21} the residual charge of the condensers C_{12} and C_{21} is short circuited. A strong short-circuit current surge is thereby caused, and the voltage at the output terminals immediately assumes the value of the mains voltage. In fig. 2c the diagram of the output potential at B_1, B_2 is given.

The disadvantage of the above-mentioned strong short-circuit current surge may be easily overcome by including in the circuit, in series with the condensers, a resistance which limits the short-circuit currents.

In addition to the parts sketched in fig. 1 the vibrator includes also two filters which protect the mains and the set respectively from high-frequency oscillations, which might be generated in the apparatus by the sudden current variations upon breaking the contacts. Reception entirely free of disturbances is by this means attained for frequencies of 150 kc/sec and higher, that is, over the whole broadcasting range.

Mechanical construction and mounting of the vibrator

Fig. 4 gives the mechanical construction. The vibrator consists of a frame R, which at the same time forms part of the electromagnet with the coil S. At a short distance from the core is the armature M fixed to and insulated from the two springs A_1 and A_2, to which the direct voltage is applied. These springs are insulated from the frame by means of mica plates P. Not far from the points where they are clamped the springs A_1 and A_2 have a u-shaped stiffened section, upon which the tungsten contacts K_{11}, K_{12}, K_{21} and K_{22} are fastened. The opposite contact plates are mounted a short distance away on side springs, which are insulated from the frame. Together with the u-shaped section a second spring is attached to spring A_2, with the contact K for connecting the current through the coil S of the magnet. When the contacts are open the middle springs A_1 and A_2 vibrate about the points where they are clamped;

during the greater part of their deviation, however, they are constrained by the side springs, so that the natural frequency of the middle springs weighted with the armature is considerably increased. It is about 100 c/s.

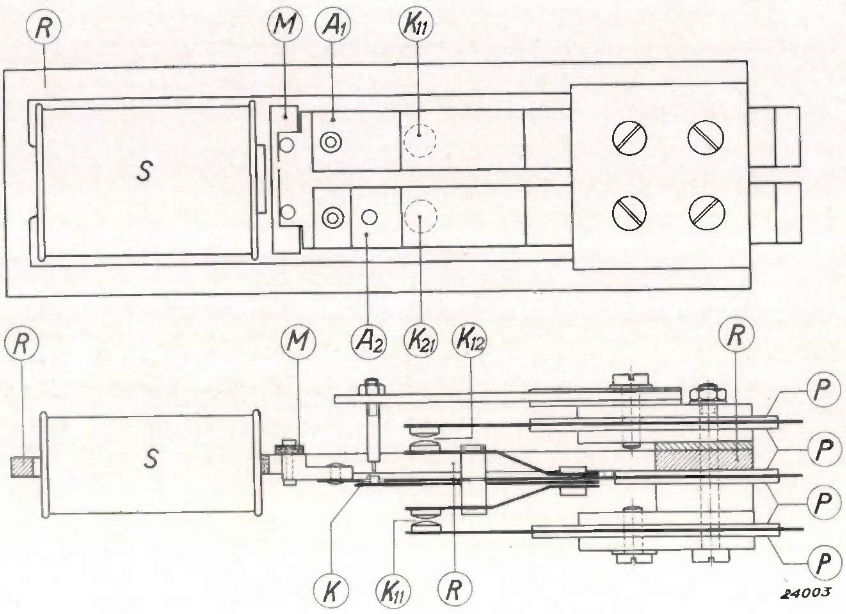


Fig. 4. Mechanical construction of the vibrator.

The condensers and coils of the interference filter are placed in a separate can, which is connected to the vibrator by means of six contact bushes. Two of the six pins serve to lead in the direct current, two to lead out the alternating current and two for the interference condenser C (fig. 3), which bridges the magnet coil and which also is placed in the separate can. *fig. 5* shows the different parts separately; *fig. 6*

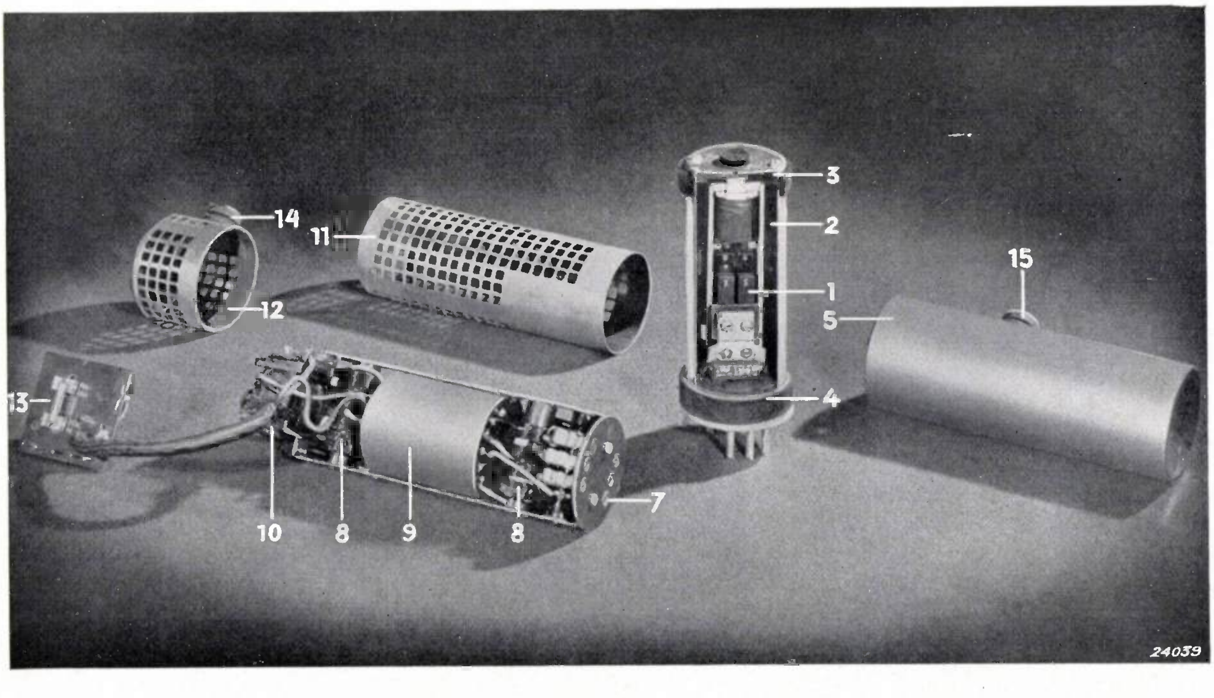


Fig. 5. Parts of the vibrator. The vibrating element *1* is hung on springs in the can *2* which is sealed by the rubber cover *4*, which also serves to close can *5* surrounding can *2*. The vibrator with its six pins is inserted into the contact bushes *7* of the interference filter component. *8* chokes and *9* condensers of the interference filter element; *10* coils with iron cores, *11* and *12* ventilation of the interference filter element. *13* fuse, *14* and *15* rubber studs for flexible and damped mounting in the receiving set.

shows the vibrator joined to the interference components.

In mounting the vibrator account must be taken of the fact that the alternating current generated is not sinusoidal, but contains higher harmonics, and chiefly odd harmonics of about 300, 500, 700 periods per second [1]). By careful layout of the receiver components, and, if necessary, by electric or magnetic shielding of the parts sensitive to those influences, the penetration of those harmonics into the receiving set can be avoided.

Another important point in mounting is the damping of the acoustic vibrations caused by the vibrator. These vibrations are conducted not only by the solid material, but also by the air. The solution of this problem has been found by making the can which contains the vibrator, air-tight, and surrounding it with a second can which is closed by a rubber seal through which the connecting wires are led. In this way chiefly vibrations with a higher frequency are damped. The vibrations with a lower frequency, which are propagated more through the solid material, are sufficiently arrested by mounting the vibrator itself in the inner can with the help of springs, so that it has some fredom of movement.

The value of the voltage delivered by the vibrator is closely related to the voltage of the direct current mains to which the vibrator is connected. If it is supplied with 220 volts, an alternating voltage of about 200 volts is obtained. This is clear when we remember that the maximum value of the alternating voltage delivered by the vibrator is equal to the direct voltage V_0. The effective value of the alternating voltage will therefore certainly be lower. If we represent the voltage schematically in the way used in fig. 2c, we find the following:

$$V_{\text{eff}} = V_0 \sqrt{1 - \frac{4\alpha}{3\pi}}\,.$$

The effective voltage is thus lower than V_0. The output voltage of the vibrator is in that case not connected to the 220 volts terminals of the radio set, but to a special tap which is introduced for this purpose.

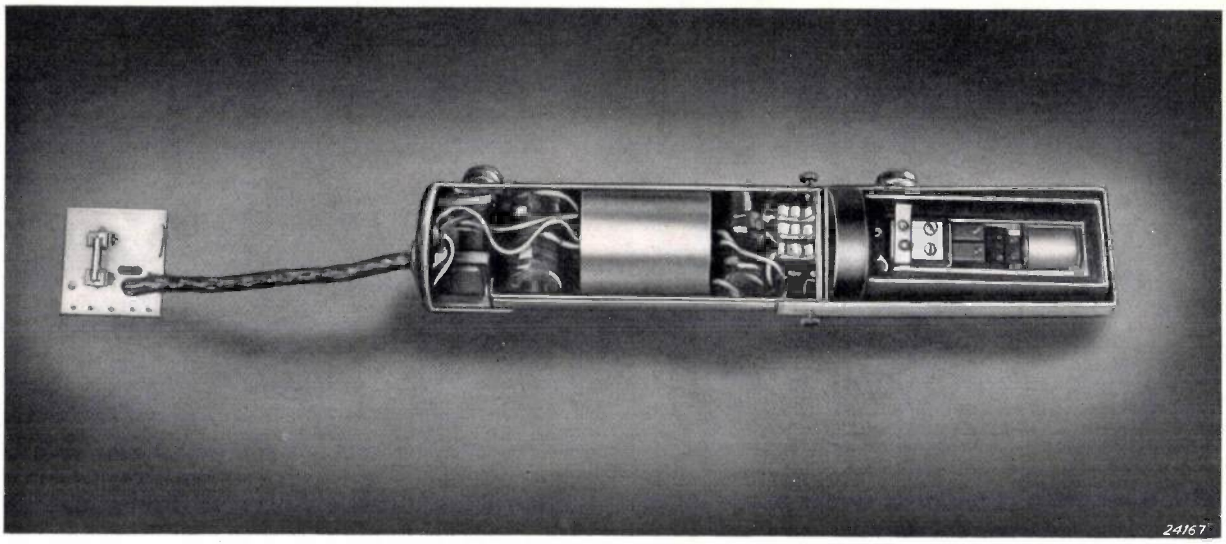


Fig. 6. Vibrator with non-interference components.

1) If we idealize the output voltage of the vibrator in the manner represented in fig. 2c we obtain the following Fourier series:

$$V = \frac{4}{\pi} V_0 \left[\frac{\sin\alpha}{\alpha} \sin\frac{2\pi t}{T} + \frac{\sin 3\alpha}{3\alpha} \sin 3\frac{2\pi t}{T} + \right.$$
$$\left. + \frac{\sin 5\alpha}{5\alpha} \sin 5\frac{2\pi t}{T} + \ldots\right].$$

Thus odd harmonics occur. At the input end of the vibrator an intermittent direct current is introduced, which, as may be seen from fig. 2, contains a main frequency $2/T$ and harmonics of this, which are therefore even harmonics of $1/T$. These alternating currents might superimpose alternating voltages on the mains, and thereby cause disturbances in receiving sets in the neighbourhood. For this reason the alternating currents are prevented from entering the main by the introduction of a choke coil.

THE INVESTIGATION OF THE MACRO-STRUCTURE OF RAW MATERIALS AND PRODUCTS BY MEANS OF X-RAYS II.

by J. E. DE GRAAF.

Two kinds of information may be obtained from X-ray examination: information about the strength as it is decreased by defects, and information about the cause of these defects.

The first type of information can seldom be obtained by a photographic method like the absorption method. The shape and dimensions of a flaw and thus the decrease in cross section may, it is true, be indicated, but this is only of value when the weakening due to the notch effect is not greater than that corresponding to the decrease in cross-section. Only with spherical flaws such as gas bubbles would one be able to determine the weakening. In all other cases the evaluation of the strength requires wide experience if it is to be at all reliable to any degree.

The question of the diagnosis of the flaw, although it requires much experience, may usually be answered conclusively. Examples of this will be given here and in subsequent articles.

The discovery of flaws in riveted joints

Although riveting as a method of making joints hast lost much ground to welding, and particularly

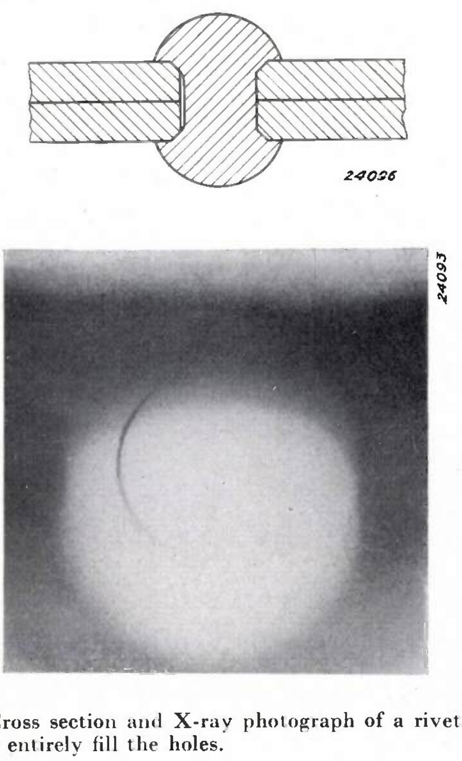


Fig. 1. Cross section and X-ray photograph of a rivet which does not entirely fill the holes.

to electric welding, it still remains an important working method. In certain cases it has retained a special position next to welding, for example, in the construction of parts of ships which are made of kinds of steel which easily crack upon non-uniform cooling. In such cases it is often undesirable to make the connections between parts welded in the shops by welding in the open air, while it is still quite possible to do this by riveting. As in welding, it is very important that the riveted joint be reliable, and not weakened by riveting flaws.

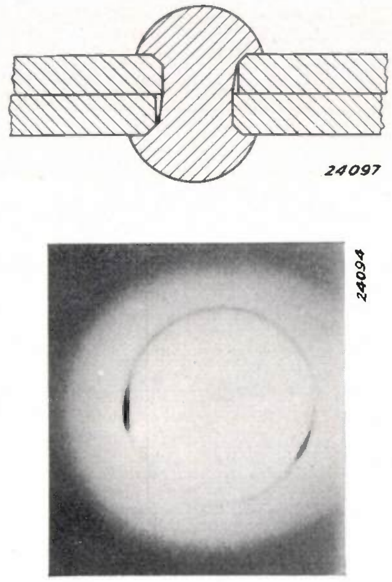


Fig. 2. The holes in the two plates of the riveted joint do not lie directly one above the other.

A riveting flaw which occurs fairly often is shown in *fig. 1:* the loose rivet. In this case, the rivet does not completely fill the hole, and the cylindrical fissure between the hole and the rivet is shown as a dark circle on the photographic negative. When the holes in the parts to be joined are not situated exactly one above the other (*fig. 2*), such circular arcs also will occur, but in this case, they do not coincide, but form parts of two non-concentric circles (for a true diagnosis it is of course necessary that the rays be incident perpendicularly to the plate). In the photographs of both of these flaws a sharply limited outer circle is seen which corresponds in diameter with that of the holes bored. By means of a circle drawn on transparent paper one may usually see clearly whether the flaw consists of arcs of one or of two circles. A third flaw which also gives rise to a circular image is shown in *fig. 3:* the head of the rivet was badly formed and later worked over to give the correct external shape. This flaw, however, in contrast to the two foregoing ones, shows an unsharp outer limit in the photograph, while the average diameter is greater than

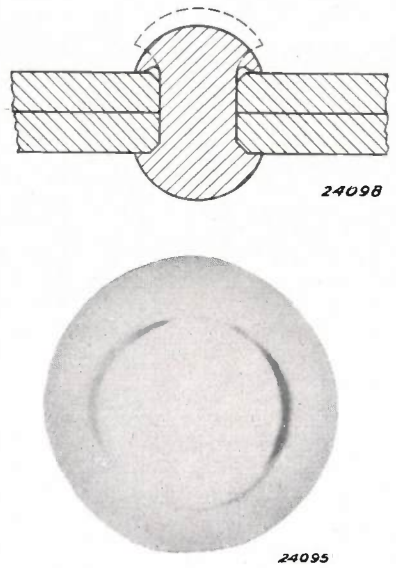


Fig. 3. The head of the rivet is badly formed and later worked over into the correct external shape.

the diameter of the rivet hole. This also can usually be detected with the circle drawn on transparent paper. Since the significance of these three flaws is not the same, the importance of their individual recognition will be clear.

The crooked head, which occurs quite often in hand riveting, is shown very clearly in an X-ray photograph as an irregular grey border, as may be seen in the upper left had corner of *fig. 4*.

In the boiler plate itself very dangerous cracks may occur, not only around the rivet holes (*fig. 4*), but between successive holes. Cracks are particularly important since they always have the tendency

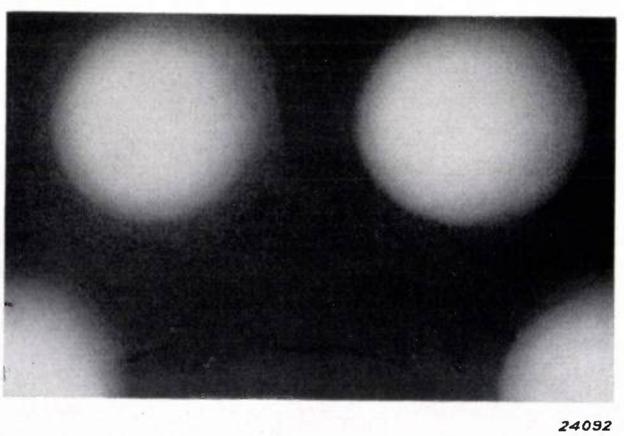


Fig. 4. X-ray photograph of cracks in a boiler plate in and around the rivet holes (cf. fig. 5a).

to become larger, which fact may be ascribed to the high concentrations of stress at their ends (notch effect). The detection of cracks is, however, not always simple. In the case of *fig. 5a*, in which the

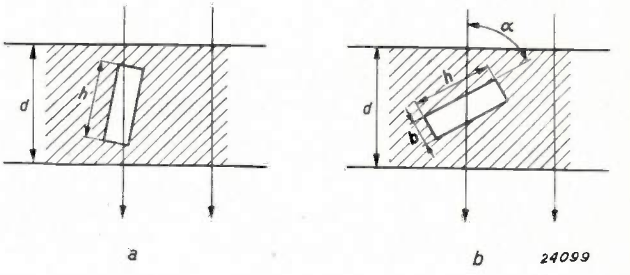


Fig. 5. The photography of cracks at different angles between crack and X-rays.

angle between the X-rays and the crack is small, the local decrease in thickness is equal to h, the whole height of the crack. In fig. 5*b* the decrease in thickness is only $b/\sin\alpha$. With increasing α therefore the difference in blackening caused by the crack will decrease rapidly. *Fig. 6* gives an

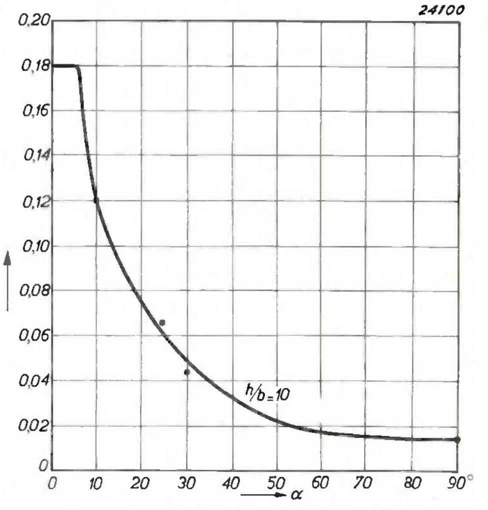


Fig. 6. Contrast between the picture of the crack and its surroundings at variable angles between the plane of the crack and the direction of the X-rays.

example of this for a crack with $h/b = 10$ in aluminium. For heavier metals the decrease in contrast is even greater. The angle α must not, therefore, by greater than 5 to 10° if the crack is to be clearly visible in the photograph. Suspicious vague lines must for this reason always be photographed again at somewhat different angles. The close dependence of the contrast on the direction of the rays offers an excellent means of distinguishing cracks from corrosion grooves which often occur with old boilers. These much less dangerous grooves also give vague lines on the photograph.

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

No. 1210: J. A. M. van Liempt and W. van Wijk: Die Löslichkeit von Krypton in verschiedenen Flüssigkeiten (Rec. Trav. chim. Pays Bas **56,** 632 - 634, June 1937).

The solubility of krypton in different liquids is determined. It is found to be much greater than that of argon. In pure glycerine krypton is only very slightly soluble: this liquid is therefore useful as a sealing liquid for krypton.

No. 1211: H. C. Hamaker: A system of colloid phenomena (Rec. Trav. chim. Pays Bas **56,** 727 - 747, June 1937).

Building further on the theoretical foundations given in No. 1154 and No. 1174 for the treatment of the lyophobe colloids, it is shown in this article that, as regards the attractive and repulsive forces between the particles, the possible properties of colloids may fall into two groups. These two groups of properties correspond to a large extent to the phenomena which are observed with lyohobe and lyophile colloids respectively. Theoretically however these two groups are not sharply divided, and the theory furnishes a clear description of the nature of the intermediate cases. Various results are in good agreement with experiment. Since the terms lyophobe and lyophile are in reality not applicable to the theoretical grouping, the writer proposes the distinguishing terms "reversible" and "irreversible". The classification into lyophobe and lyophile colloids can be retained at the same time. The two groupings are closely related but not identical. In this way a satisfactory classification of colloids and colloidal phenomena may be given.

No. 1212: J. H. Gisolf: Electron counters (Ned. T. Natuurk. 4, 129 - 149, June 1937).

In this lecture, given before the Netherlands Physical Society, a survey was given of the action and the construction of discharge tubes for counting electrons.

No. 1213: Balth. v. d. Pol: A new theorem on electrical networks (Physica 4, 585 - 589, July 1937).

With the aid of the modern form of Heaviside's symbolic calculus the following theorem is proved. If a constant unit EMF is applied to any given currentless network, the difference between the electrical and magnetic energy in the stationary final state is equal to half the derivative of the admittance to $j\omega$ for the limit $\omega = 0$. A special case of this is the well-known theorem that the efficiency with which a battery can charge a set of condensers is always 50 per cent.

No. 1214: J. Alfter and W. J. Oosterkamp: Ein Vorschlag für den Begriff Filmgüte bzw. Gütefaktor des Aufnahmematerials (Fortschr. Röntgenstr. **55,** 609 - 612, June 1937).

In order to obtain the same contrasts in the image of an object on different film material, in X-ray photographs of the lungs, the voltage of the tube must be chosen proportional to the gradation of the film at a blackening of 1. Only in relation to the gradation of the film does the tube voltage determine the character of the negative. As a standard of quality for X-ray film material a quantity is proposed which is proportional to the sensitivity and to the third power of the gradation of the film. This proposal refers to X-ray photographs of the lungs and is illustrated by photographs taken in practice.

No. 1215: Th. J. Weyers: Selectivity measurements of radio receiving sets (T. Ned. Rad.-Genoot. 7, 156 - 172 July 1937).

Measurements relating to the selectivity of radio receiving sets are carried out according to various methods. It is found that reliable selectivity curves are obtained only when in measuring the conditions of normal use are imitated as closely as possible by applying to the apparatus to be tested two signals, one a "desired" signal, to which the apparatus is tuned, and a "disturbing" signal which is modulated with music. With a given intensity of the "desired" signal it may be established by ear what is the permissible intensity of the "disturbing" signal, if no disturbance is to be heard during the pauses in the modulation of the desired signal. In this method of measurement the following features are taken fully into account: cross detection, cross modulatiou, combination tone of desired and disturbing carrier wave and the suppression of the modulation of a weak signal by a simultaneous strong signal.

VOL. 2 No. 12 DECEMBER 1937

Philips Technical Review

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

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

THE SODIUM LAMP

FROM LABORATORY EXPERIMENT TO STREET LIGHTING

by E. G. DORGELO and P. J. BOUMA.

Summary. A short survey is given of the technical problems encountered in the development of discharge lamps using sodium vapour. The development led a) to a low tension arc for direct current and b) to a positive column lamp for alternating current. The latter has meanwhile almost entirely superseded the direct current lamp. Following a technical description of the lamps, the properties of their light are discussed as well as the possibilities of application which follow from these properties.

Since street lighting is the most important application, the circuits and installation of sodium lamps for street lighting are gone into in some detail, while finally some results are given of investigations on the visibility on roads lighted with sodium light.

Introduction

When one considers the fact that the light from a discharge in sodium vapour consists chiefly of radiations of a single wave length (5890/5896 Å), and moreover that the sensitivity of the eye for this wavelength, is very great being 76.5 per cent of the maximum value [1]), it is understandable that attempts have been made for some time to apply such a discharge in the construction of a practical high efficiency source of light.

Innumerable difficulties and problems had to be solved before sodium lamps were developed to such a point that they could be installed in large numbers for street lighting; only when this stage had been reached was it possible to form a satisfactory opinion about the utility of sodium light, and the influence of its unusual composition on visual acuity.

Development of the lamp

Since the vapour pressure of sodium at room temperature is so low that no discharge can occur, the lamp is provided not only with sodium but also with a rare gas filling. This latter gas makes ignition possible, while in addition the rare gas discharge heats the glass wall, so that after some time the pressure of the sodium vapour becomes sufficiently high to cause the sodium light radiation to predominate. Two problems immediately arose, namely that of the heat insulation, which is necessary in order to raise the lamp to the correct temperature with the least possible energy, and which must ensue that this temperature is not too greatly affected by the temperature of the surroundings, and that of the kind of glass: the "ordinary" kinds of glass are strongly attacked by sodium vapour at these temperatures (250 - 280° C). They become brown or black in a short time and then absorb very much light.

The first problem was solved in this laboratory by surrounding the lamp with a removable double-walled evacuated glass container. The contents of such a container can easily be brought to a given temperature and kept at that temperature, since, of the three forms of dissipation of heat, radiation, conduction and convection, the last two are practically ineffective here. Another system which is also used, that of mounting the lamp permanently in an evacuated outer bulb, makes a somewhat simpler construction. A disadvantage of latter system is the fact that the lamp must always be replaced as a whole, while the removable vacuum glass can always be used again. Moreover, the circulation of air between the vacuum glass and the lamp promotes uniformity in the temperature distribution.

---

[1]) Characteristics of the eye with special reference to road lighting, Philips techn. Rev. 1, 102, 1936.
The representation of colour sensations in a colour space-diagram or colour triangle, Philips techn. Rev. 2, 39, 1937.

Since the reaction with the glass appeared to consist in the reduction of the silicates present in it, silicon-free glass or glass poor in silicon, borate glass among others, was used at first. Since, however, this glass is very difficult to work with, due to the short transition interval in softening, we changed over to an ordinary glass which was protected from attack by a thin layer of borate glass. This glass was easily workable and almost entirely unattacked.

The solution of these two main difficulties made possible the construction of an efficient sodium lamp.

Further development was carried out in two different directions: that of the low tension arc, where the distance between the electrodes is small and the working voltage low (10 - 30 volts), the lamp being usually bulb-shaped, and that of the positive column discharge with a greater distance between the electrodes, a higher voltage (100 volts or more) and straight or bent tube-shaped lamps. We shall not go into the fundamental differences between the two forms of discharge. The low tension arc is in many respects comparable with the discharge in a rectifying valve [2]); the positive column discharge with the discharge in neon advertising signs. The low tension sodium arc here described (*fig. 1*) has been able to hold its own only in the case of direct current installations mainly because its length of life with alternating current was not sufficiently long.

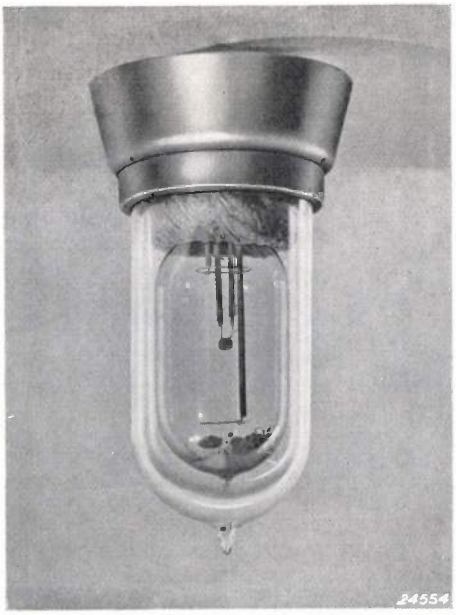


Fig. 1. Latest model of the low tension arc for direct current. Heat insulation by means of a removable vacuum glass.

In order to make use of direct current mains of the order of 220 volts a number of lamps were connected in series with a common series resistance.

The positive column lamp is now generally used with alternating current. In order to be able to include the connections for both electrodes in one cap the tube is bent in the form of a single or a double U (*fig. 2*). One of the problems encountered in the development of both types of lamps is that of the distribution of temperatures. In order to obtain a sufficiently high vapour pressure (about 10^{-5} atm) the temperature of the coldest part of the lamp must be about 250 - 280° C, while the rest of the walls of the tube may not be much hotter because of possible attack on the glass. A fairly uniform temperature of the walls is thus required. The method of sealing in, usual in the manufacture of electric light bulbs (see *fig. 3a*), (in which the lead-in wires are previously fastened into a glass foot which is fused into the bulb), is less suitable for sodium lamps since it is difficult to keep the place where the foot is fused into the neck of the bulb at the right temperature, while in addition the place where the connection wires are sealed in (the "pinch") becomes very hot, so that there is the danger that the glass along the wires will crack. Therefore, a special method of construction has been developed for sodium lamps (fig. 3*b*), in which the sealing-in wires are sealed in directly in the neck of the bulb (so-called reversed pinch). Especially in the case of the U-shaped positive column lamps has this method of sealing in proved effective. Both pinches are formed at

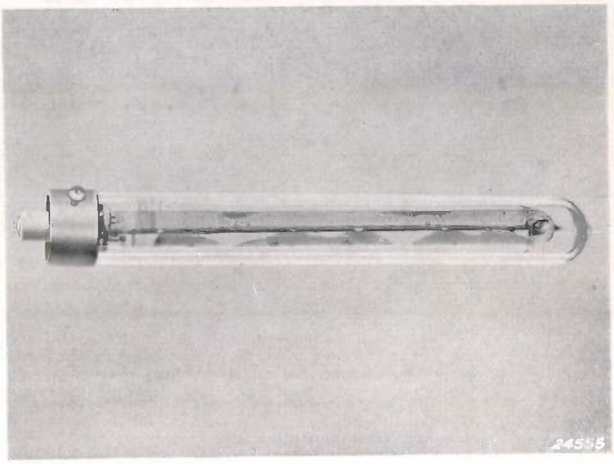


Fig. 2. Latest model of a positive column lamp for alternating current with vacuum glass. Tube bent in U shape. Running voltage 170 volts; current 0.6 A. Hot cathodes heated by the discharge.

2) Physical principles of gasfilled hot-cathode rectifiers, Philips techn. Rev. 2, 122, 1937.

the same time mechanically. For low tension arc lamps the construction shown in fig. 3*c* is also used, in which the space around the foot is isolated from the discharge by means of a chrome iron plate.

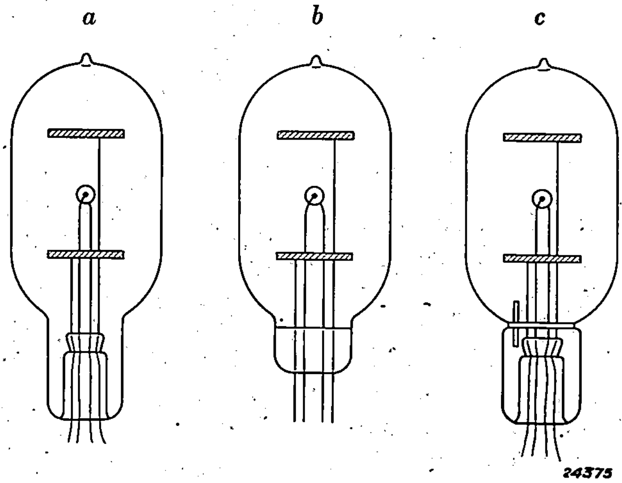


Fig. 3. a) Normal sealing in. The space around the foot is at too low a temperature so that when used as a sodium lamp the sodium condenses at this spot.
b) Leading-in wires attached directly to the neck of the bulb. The harmful cold space is here avoided.
c) Like a) but with a partition between the discharge space and the space around the foot. The condensation of Na vapour around the foot can also be avoided in this way.

In the low tension arc lamp the electrode system consists of a cathode in the centre of the discharge space, and two ring-shaped anodes situated about 2 cm above and below the cathode.

The cathode, which is wound in a spiral, is heated by current from a separate transformer. Its electron emission is large due to a film of alkaline earth oxides deposited upon it.

The positive column lamp, which is fed with alternating current, as already mentioned, has at each end of the tube one electrode which is wound in a spiral like the cathode of the low tension arc lamp. The function of this electrode is more complicated, however, since it must serve not only as cathode (in the negative phase) but also as anode (in the positive phase). Especially under the latter circumstance the heat to which the electrode is subjected due to the discharge is so great that a special heating current transformer is unnecessary.

A problem which must receive particular attention in the case of the positive column discharge is that of the distribution of the sodium. In order to obtain a uniform light from all parts of the lamp, it is necessary that the same concentration of sodium vapour be present in all parts of the bulb (or tube). The uniformity of distribution is threatened by different effects, such as one-directional movement of the ions due to a direct current component and differences in temperature. The influence of such effects is particularly great since the diffusion which seeks to promote uniformity is very weak in the case of sodium vapour in a gaseous atmosphere. With low tension arcs this fault is less serious due to the symmetrical construction of the lamp with the discharge in the centre. With positive column lamps the non-uniformity of the distribution of the sodium may cause some parts of the lamp to give less light than others; in extreme cases only the radiation of the rare gas remains in these parts. Little by little the causes of the non-uniformity have been successfully overcome, in the first place by providing for an even temperature distribution, and in the second place by keeping the transport of ions within definite limits (both the electrodes of the alternating current tube should be prepared in exactly the same way; since the conductivity is then the same in both directions, practically no direct current component occurs).

The problem of the d e p e n d e n c e o n t e m p e r a t u r e was found to be especially important in connection with the practical application. We have already seen that the temperature must be about 250 - 280° C. At a higher temperature and pressure the intensity of the light is found to increase only very slightly, and then to decrease again. Since more energy must be supplied for the higher temperature, the efficiency (lumen per watt) thus decreases. *Fig. 4* gives the light intensity as a

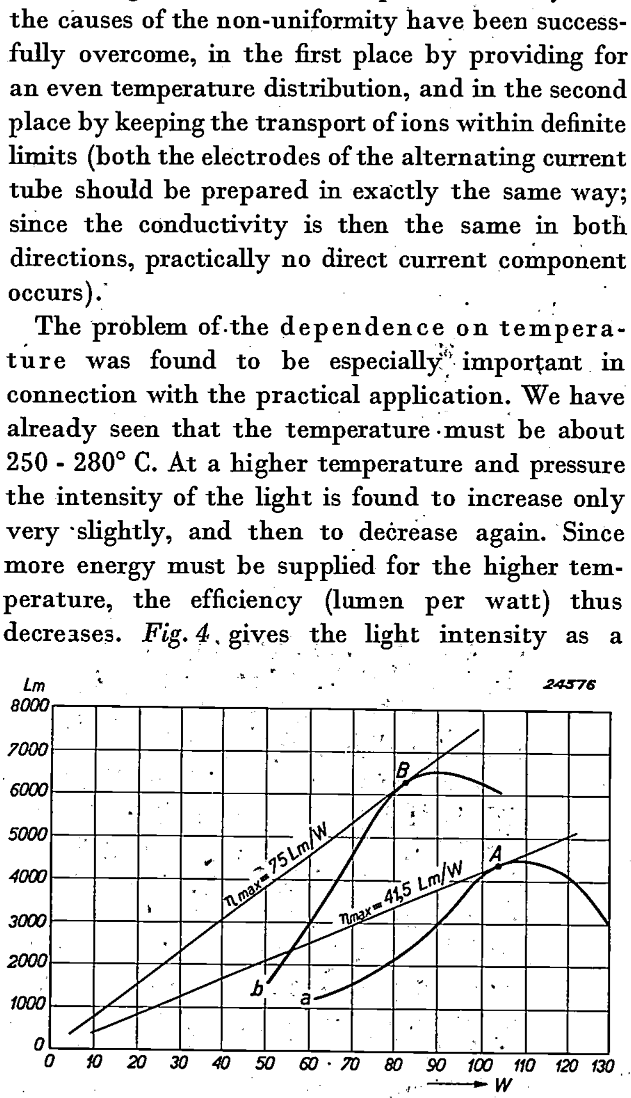


Fig. 4. Intensity of light (lumens) as a function of the power. a) Low tension arc on direct current. b) positive column lamp on alternating current. The curves do not represent average values but were measured on two individual lamps.

function of the energy input for two given lamps:

a) for a low tension arc lamp on direct current,
b) for a positive column lamp on alternating current.

The tangents to the curves from 0 give the points *A* and *B* where the lamps have their maximum efficiency. In practice it was found advantageous to let the lamps work in the horizontal portion of the characteristic: the variations in the energy supplied (mains voltage variations!) as well as variations in the external temperature then have only little influence on the intensity of the light.

The properties of sodium light

As the above-mentioned problems and difficulties were more and more already overcome, and the lamps reached the stage of practical application, the study of the properties of this new kind of light was also begun. Considering that results of several of these studies have already been described in this periodical, a brief outline will be sufficient here.

The monochromatic character and the favourable spectral position of the sodium lines in relation to the curve for the sensitivity of the eye have been pointed out in the introduction.

With the first experimental installations it was remarkable how sharply small objects could be seen by sodium light. It is therefore understandable that the first physiological-optical investigations were concerned with the acuity of vision [4]). In connection with the monochromatic nature of the light and the consequent lack of chromatic aberration, and also from earlier experiments by Ives an important difference from ordinary electric light was to be expected. The acuity of vision with sodium light was actually found to be considerably greater. It is remarkable that the magnitude of this difference was found to depend upon the nature of the test objects used.

In the case of speed of perception [3]) also, which is closely related to acuity of vision, important differences between sodium light and ordinary electric light were found.

Upon continuation of this study it became clearer and clearer that the improved acuity of vision, although one of the most striking phenomena, is certainly not the only factor which determines vision with different coloured light, and that especial attention must be devoted to still another side of the problem, namely that of contrasts [4]). In the beginning it proved very difficult to confirm by means of laboratory experiments the general impression of the better contrasts on a road lighted by sodium light. The differences in sensitivity to contrast were too slight to explain the striking phenomena; nor did the measurement of coefficients of reflection for the various kinds of light show any important differences. A critical study of the concept of brightness and the Purkinje effect led us by way of the concept of subjective brightness [5]) to that of richness in contrast [4]). Along these lines it could be shown how the greater contrasts with sodium light are to a large degree related to a shifting of the curve for the sensitivity of the eye which is strongest at just the brightness occurring on a well-lighted road.

Since all these characteristics of the eye can be influenced very strongly by glare [6]), it was interesting to make comparisons between the different kinds of light in this respect also. The result was that for simultaneous glare no important differences could be shown, but that sodium light presents appreciable advantages with respect to successive glare as well as with respect to disturbing effects. The low brightness of the sodium lamp (compared with that of the mercury lamp for instance) was found to be important in this connection.

Little is yet known about the psychological effects of this new kind of light; in general it may be said that on country roads the yellow colour makes a restful impression on the great majority of road users. Experiments carried out in workshops and offices showed that the fatigue phenomena were equally great for white light and sodium light.

The range of application

Simultaneously with the physiological-optical study the investigation of the possibility of application was carried out. Although this problem was often attacked by intuition and by the method of trial and error, we may now conclude directly from the results mentioned in the previous section as to the conditions under which sodium light may be used to advantage.

Because of its monochromatic character we can employ sodium light only where the reproduction of colour is of no importance [7]). Its use in living rooms, shops, streets in the centre of cities, etc. is therefore immediately eliminated.

Moreover, in cases where very low intensities are

3) Visual acuity and speed of vision in road lighting, Philips techn. Rev. 1, 215, 1936, other literature is cited in this article.

4) The perception of brightness contrasts in road lighting, Philips techn. Rev. 1, 166, 1936.

5) The definitions of brightness and apparent brightness and their importance in road lighting and photometry, Philips techn. Rev. 1, 142, 1936.

6) The problem of glare in highway lighting, Philips techn. Rev., 1 225, 1936.

7) The perception of colour, Philips techn. Rev. 1, 283, 1936. Colour reproduction in the use of different sources of "white" light, Philips techn. Rev. 2, 1, 1937.

usual, as for example in mines, sodium light cannot be used, since in these cases the Purkinje effect has an unfavourable influence on vision.

The use of sodium light will offer particular advantage in those cases in which acuity of vision is very important, such as the inspection of articles for fine cracks, or when quickness of perception plays a part (on tennis courts for example) [8]). When, as in the case of the lighting of country roads, the levels of brightness are of such a nature that in addition to these advantages that of the great richness in contrast becomes important, the sodium lamp will often be the most suitable source of illumination. Wherever glare must be avoided or kept at a minimum, the sodium lamp with its low brightness, its slight successive glare and its relatively pleasant colour offers great advantages.

Further there is a long series of applications in which special use is made of the pronounced yellow colour of the light. The advantages of this colour may be manifested in very different ways:

1. Aesthetic effects may be obtained when the light is used as flood lighting for buildings, as part of festival illuminations, illumination of groups of trees, etc.
2. Certain objects such as advertising signs, signals, etc., can be given a very striking appearence by means of sodium light; its use for boundary lights of aerodromes belongs to this category.
3. In various special cases small contrasts become much clearer in sodium light, because the differences in coefficient of reflection may be greater than with illumination by white light. This applies particularly to various cases of the testing of materials.

Sodium light is already being used in the photographic studio [9]).

Finally sodium light will often be used for purely economic reasons in cases where large areas must be lighted (shunting yards) or where for special reasons road lighting of a very high intensity must be used (traffic tunnels in the daytime). Sodium light in such cases presents entirely new possibilities, since illumination with ordinary electric lamps is often extremely expensive.

Circuit and electrical characteristics of sodium lamps [10])

In the low tension arc lamps used with direct current the two anodes are connected to each other so that the lamp has three connections: one positive terminal and the two ends of the heated cathode as negative terminals. *Fig. 5* gives a complete diagram of a number of lamps connected in series. For heating the cathode each lamp has a separate heating current transformer T which is fed with alternating current. In addition there is a cut-out cartridge S in parallel with each lamp, which breaks down when the lamp does not work, and thus keeps the circuit closed. Since in such a circuit the direct voltage applied is at first distributed unevenly over the various lamps so that the lamps may light one after another, the ignition voltage required per lamp may be considerably lower than that for one single lamp (this is about 30 volts). Consequently only about 10 per cent of the total voltage need be lost in a series resistance, which often takes the form of an iron wire in an atmosphere of hydrogen. This series resistance keeps the current constant in spite of variations in the mains voltage or changes in the

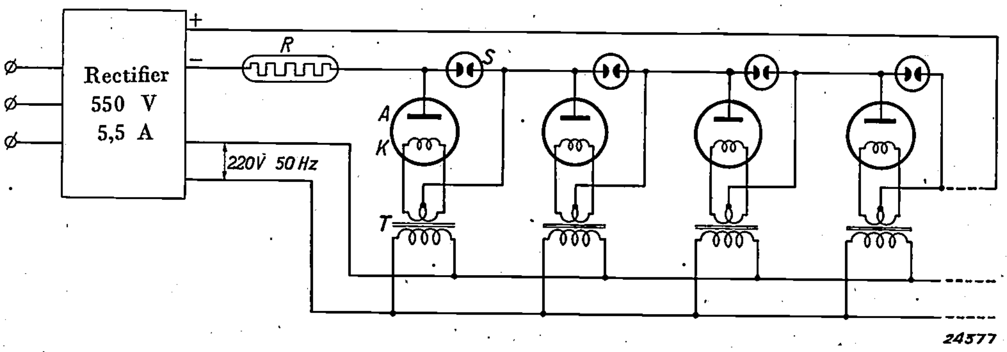


Fig. 5. Diagram of a series installation with low tension arc lamps. A anodes, K cathodes, T heating current transformers, R iron wire resistance, S short-circuiting cartridge.

[8]) Sodium lighting of tennis courts, Philips techn. Rev. 1, 252, 1936.

[9]) The "Philora" sodium lamp and its importance to photography, Philips techn. Rev. 2, 24, 1937.

[10]) Alternating-current circuits for discharge lamps, Philips techn. Rev. 2, 103, 1937.

load resulting from the falling out of lamps. Some of the objections connected with supply by direct current are the necessity of installing whole groups at the same time, the four-conductor system (see fig. 5), the relatively high sensitivity to variations in temperature and the occurrence of radio interferences which may, however, be suppressed by installing filters [11]).

We are giving herewith some data of a direct current installation (first street lighting with sodium lamps, June 1931, Beek - Geleen, Holland).

Total direct voltage: 500 volts.
Voltage across the series resistance (average): 50 volts.
Sum of the voltages of the individual lamps: 450 volts.
Number of lamps: 30
Current 5.5 A.
Consumption of lamp including heating current transformer: 102 W
Total consumption: 3.1 kW.
Luminous flux per lamp: 4000 lumens
Efficiency, gross: 38.5 lm/W.

The positive column lamp used with alternating current offered wider possibilities. In this lamp two hot cathodes are used for electrodes. These electrodes are raised and kept to the proper temperature by the discharge itself, and function as both cathode and anode alternately. A separate heating current transformer is therefore not provided. The current is limited by a transformer of special design. *Fig. 6* shows such a combination

Fig. 6. Leakage flux transformer.

which is called a leakage flux transformer. The self-inductance of this transformer causes a phase shift between current and mains voltage which may be eliminated by the introduction of a correcting condenser across the mains terminals.

The A.C. lamps may be installed quite independently of each other. The following table (*table I*) gives some data of the lamps in use at the present time:

Table I

Type	Power incl. losses in current limiting device	Current through lamp	Running voltage	Luminous flux	Efficiency
50 W	65 W	0.6 A	80 V	2 550 lm	39.3 lm/W
65 W	80 W	0.6 A	110 V	3 780 lm	47.3 lm/W
100 W	105 W	0.6 A	165 V	6 100 lm	58.1 lm/W
150 W	165 W	0.9 A	165 V	9 600 lm	58.2 lm/W

As may be seen the first three types take the same current. Because the current supplied by the leakage flux transformer due to its high impedance depends only slightly on the voltage of the lamp used, these types can be used with the same transformer.

Fig. 7 shows how the light flux increases as a function of the time elapsing after switching on.

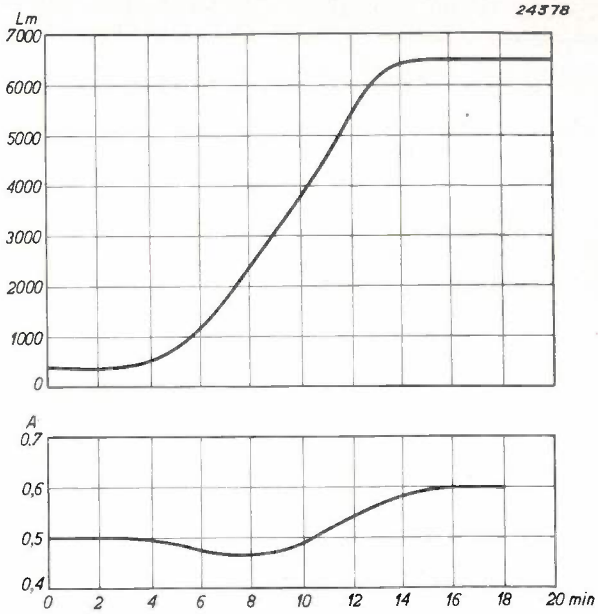


Fig. 7. Luminous flux (lumens) and lamp current (amperes) as a function of the time elapsed after switching on for a U-shaped lamp of 100 W on a transformer for 470 volts and 0.6 A.

The changes in light flux with variations in the mains voltage may be seen in *fig. 8*. For the sake of comparison the corresponding curve for an incandescent electric lamp is also shown.

We conclude this section with a few remarks about the life of sodium lamps, particularly of A.C. lamps.

While with ordinary electric lamps the length of life is chiefly determined by the rate at which the filament evaporates, and therefore has a value which may be foretold with fair accuracy, the length of life of sodium lamps depends upon a number of factors whose influence it is more difficult

[11]) High-frequency oscillations in sodium lamps, Philips techn. Rev. **1**, 87, 1936.

to discover. Evaporation of the electrodes and the material deposited upon them occurs here also, but the available reserve is so great that the length of life of the lamp is not generally influenced by this factor. After several thousand

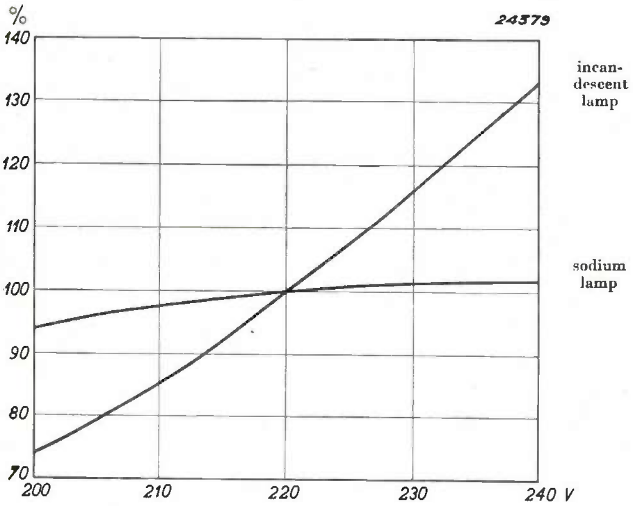


Fig. 8. Light flux of a sodium lamp and of an incandescent electric lamp for different mains voltages in per cent of the value reached at the nominal mains voltage of 220 volts. The sodium lamp is very insensitive to voltage variations.

hours, however, symptoms of age appear, which to a great extent may be ascribed to a less uniform distribution of the sodium. This may result in certain parts of the tube giving too little light. Local brown coloration of the glass may also result, while an abnormal accumulation of molten sodium in the neighbourhood of the electrodes may lead to cracking at the point of sealing in. On the average an A.C. sodium lamp lives 2 500 hours.

The installation of sodium lamps from the standpoint of lighting technology

We shall not go too deeply into this subject, and shall only discuss those points in which sodium illumination differs fundamentally from illumination by means of incandescent electric light.

These differences are due chiefly to the special form, dimensions and light distribution curve of the sodium lamp. Because of these differences special measures must often be taken in order to satisfy the general requirements of street lighting, such as the greatest possible uniformity, sufficiently high brightness of the road surface, little glare and good visibility.

If we consider the distribution of light from a sodium lamp (*fig. 9*), we see that the best position for the lamp is horizontal and perpendicular to the direction of the road. The greatest part of the light radiated downward then falls on the surface of the road and its immediate surroundings. In order to obtain illumination directly under the lamp the two arms of the U must lie in a horizontal plane.

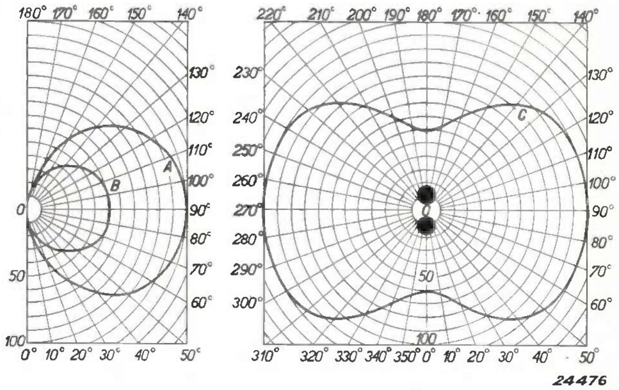


Fig. 9. Distribution of the light of a U-shaped sodium lamp in three different planes. *A*) plane through the axis of the lamp and perpendicular to that of the U-tube. *B*) plane through the axis of the lamp and coinciding with that of the U-tube. *C*) plane perpendicular to the axis of the lamp.

On the other hand for road lighting where it is desired to distribute the light more or less evenly over the whole surface of the road, it is better to place the two arms one above the other.

The part which serves chiefly to reflect the light radiated in an upward direction may be very simple in this case. In *fig. 10* may be seen a reflector, white-enamelled on the inner side and of such a shape that light rays which fall within an angle of 20° to the road surface are intercepted. Direct perception of the lamp at a greater distance

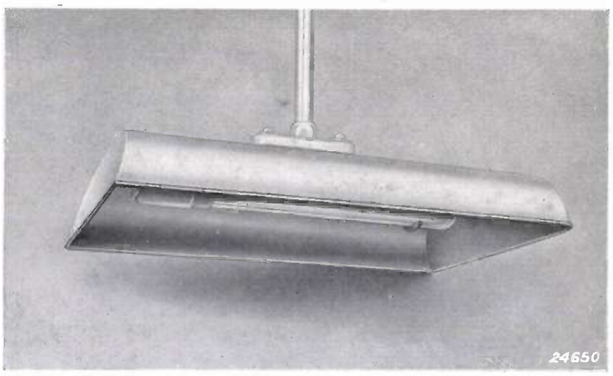


Fig. 10. Simple holder for sodium lamps.

is hereby impossible, so that glare is practically eliminated. Because of the fact that the vertical dimension of the light source is so small (about one inch) it is possible to cut it off sharply and yet retain a wide angle of radiation.

With existing installations the height of the light is usually 25 to 30 feet with a distance between the standards of 100 to 120 feet. Using lamps of about 100 W a good illumination is obtained with this arrangement.

In certain cases (workshops where machines, cranes, etc. are used) the fact that the intensity of the light pulsates 100 times a second may cause stroboscopic effects. If this must be avoided two lamps may be placed in one holder, and their transformers may be so connected that the lamp currents are 90° out of phase with each other. The total light flux is then never zero, and the difficulty mentioned is thus avoided.

Appraisal of the result achieved

When sodium lamps have been employed in a lighting installation, especially for the lighting of a country road, after taking into due account practical experience as well as results arrived at through theoretical considerations it is interesting afterwards to find out to what degree the result achieved may be considered satisfactory. The appraisal of the result may be carried out in very different ways.

One method which may give very misleading results when injudiciously applicated, is to stand on the road and look. In general more attention is paid to the light sources than to the road, more to the total amount of light entering the field of vision than to the correct distribution of brightness, more to the impressiveness of the installation than to the visibility on the road. It is much better to drive along the road and to notice how the lighting influences the ease of driving and the feeling of security. The objection to this method is that it is difficult to express the result in figures. As has been explained in another place in this periodical [12]), this goal can be reached with the help of a visibility meter, an instrument by means of which the weakest contrast still observable on the lighted road can be established in a simple way: the weaker this contrast the more successful the lighting system.

In *table II* are collected the average values of the weakest recognizable contrast for a number of different installations.

Table II.

No. of installations	Kind of light	Type	Power kW/mile	Weakest recognizable contrast (average)
14	Sodium	A	5.3	0.20
5	Mercury (and mixed light)	A	9.0	0.22
4	Mercury	V	15	0.395
5	Incandescent electric light	V	10	0.31

The letter A here indicates that a well-shielded light source was used, the letter V, that this was not the case. The great influence of glare due to insufficiently shielded sources of light may be seen: in spite of the high power installed the visibility is very low. With sodium lamp installations an excellent visibility is attained with a small power.

[12]) How can one judge the efficiency of road lighting?, Philips techn. Rev., 1, 349, 1936.

ILLUMINATION AT THE INTERNATIONAL EXHIBITION IN PARIS 1937.

by L. C. KALFF.

At every world exhibition the illumination forms an attraction which draws millions of visitors. Since the exhibitions at Barcelona and Seville (1929), it may even be said that the illumination has become one of the most important attractions. In the evening, after a day's work, when no particular desire is felt for studying the contents of the various pavilions, hundreds of thousands of visitors may still enjoy the outward aspect of the exhibition, and these visitors will judge the merits of the various countries and groups by the more or less successful illumination of their buildings.

In the series of exhibitions of the last 15 years there has been an interesting development in the technique of illumination.

After the example of Barcelona, which, under the gifted leadership of the engineer Carlos Buygas, and thanks to the unlimited financial support which he enjoyed, was an almost insurpassable technical and artistic success, an attempt has been made in succeeding exhibitions at Antwerp (1930), Vincennes (Exposition Coloniale 1931), Chicago (1933) and Brussels (1935) to continue the development of the new trend in illumination.

In Barcelona the illumination consisted mainly of floodlighting of the buildings combined with lighted columns of transparent glass along the avenues and gigantic illuminated fountains in changing colours. All these forms of illumination in all the possible colours were controlled from a small room in a tower. Buildings, columns and fountains could in this way be displayed in every colour combination from a central point.

Relatively few people in western Europe had the opportunity of admiring the illumination at Barcelona, so that for most of us the illumination at the exhibition in Antwerp was entirely new. The budget for this exhibition was much more modest, and the illumination was carried out on a much more modest scale, but if possible with even more effect.

In the first place the use of coloured light was given up, while as a new element of light decoration there was a large scale application of lighted columns of painted wood or plaster which give an indirect illumination. These columns were much cheaper than the glass ones of Barcelona, and had the additional advantage of making a pleasing effect by daylight, since they could be made of the same material as the surrounding buildings.

The chief impression which many will have carried away from Antwerp will be that of avenues flanked by rows of lighting elements, sometimes tall, in the form of stately columns, and sometimes very low, placed as lanterns along the narrower avenues among the shrubbery.

The colonial exhibition at Vincennes continued the development along the lines laid down at Antwerp. There, too, strongly illuminated surfaces were introduced as light carriers, often in fantastic forms such as palm leaves or roofs of pagodas to correspond with the different styles of architecture in the colonies. The architects Granet and Expert were chiefly responsible for presenting new attractions to the public in the shape of these lighting elements and particularly the magnificent fountains. The fountains, which were quite different from those at Barcelona with respect to form and illumination, enjoyed an enormous success. Illumination in colours was only rarely used in the exhibition. We recall particularly the splendid temple of Ankorvat which was metamorphosed by Jaccopozzi with golden yellow light into something from a fairy tale.

After Vincennes came the exhibition at Chicago, where from a technical point of view the illumination did not offer much that was new, but where it could be seen that more colourful effects were being sought than had been achieved so far. Very simple means were used for this purpose in Chicago.

The floodlights for the lighting of different buildings were provided with colour screens, and the surfaces of the buildings which were to be illuminated were simply painted in primary colours corresponding with the colours of the screens of the floodlights. No change of colour was thus possible and in daylight the buildings were also very brightly coloured.

In 1935 came the exhibition in Brussels where all the above-mentioned methods of illumination were applied on a large scale. In addition extensive use was made of neon tubes, often in combination with floodlighting. A new departure at this exhibition was the extensive use of metallic vapour lamps for illuminating the exteriors of buildings. By this method variety in the colour of the buildings was obtained after dark in an economical and natural way. Next to the white light of ordinary electric lights the blue-green of mercury lamps and the golden yellow of sodium lamps could be seen. The liberal use of these lamps in the magnificent park, where the century-old beeches were lighted up in a startling way by this colourful light, was especially new and proved a great success.

It may readily be understood that the organizers of the 1937 Paris exhibition, after having seen the exhibition at Brussels, while admitting that it was very fine, wished to have everything as different and as new as possible in Paris in 1937. Everyone concerned, the committee of the exhibition as well as the architects and the manufacturers of the lamps and fittings, have done their very best to realize this ideal, and we shall attempt in the following description to examine the degree of success achieved.

Each of three architectural bureaux received a commission to design the illumination of a part of the exhibition, and to act as advisers in full authority for the illumination of the pavilions lying within their territory. These advisers were: for the part on the "rive gauche" of the Seine the architect Granet, for that on the "rive droite" the architect Expert, while architects Beaudouin and Lods, after a prize competition, were appointed designers of the light festivals on the Seine.

In designing the illumination the decision was made to avoid all light columns and other elements constructed solely for illumination throughout the whole exhibition. The buildings themselves together with the lighted fountains and trees would have to supply the light necessary for the circulation of the public in the evening. This principle was very consistently maintained. Almost nowhere throughout the exhibition grounds are there visible light elements such as columns and the like.

Several interesting conclusions may be drawn from the result of this method.

It goes without saying that the collection of pavilions which forms the exhibition is extremely heterogeneous. Each pavilion tries to be conspicuous among all the others by shape, decoration or dimensions. Because of this the illumination of each building formed a new problem demanding a solution. The unity which can be obtained by placing rows of light columns along a boulevard, even though it be bordered by very different buildings, is here sacrificed, but in its place there is a very interesting diversity of silhouettes and systems of illumination. Moreover, the various buildings may be viewed separately to better advantage than would be possible if light columns situated in front of them interfered with the view.

The systems of illumination used for lighting the different buildings are numerous, but the colour of the light is limited in the main to a few tints. Besides the white light of ordinary electric lamps may be seen the golden yellow of sodium lamps, the blue-green of mercury lamps and the light of several other gas discharge tubes in different colours. A greater variety is shown only by the light fountains, which work with colour screens in all colours, and the huge decorative dome of gas discharge tubes in many different colours under the first story of the Eiffel Tower, which was designed by architect Granet.

We shall give a survey of the most striking illuminations of the various buildings.

In the first place there is the new Trocadéro (*fig. 1*). The low middle sections, executed in light stone, glow in the strong white of ordinary electric light which is projected upon them by several batteries of floodlights, while the two curved wings are bathed in golden yellow light by a combination of sodium and ordinary electric lamps which give a fine warm colour to the limestone.

Above the cornice the building is terminated by an Attic story which is lighted by hundreds of electric lights with built-in silver mirrors.

This mighty building forms a beautiful background and termination for the international part of the exhibition. The pavilions of the various countries lie at the foot of this building rather close together and partially hidden by all the trees.

At either end of the Pont de Jena beside the

Fig. 1. This picture shows the entrance of the great underground banquet hall and part of the right wing of the Trocadéro. From the terrace above the hall one has a magnificent view of the fountains and the Eiffel Tower which lie along the main axis of the exhibition grounds (see the accompanying photographs).

Seine the buildings of Germany and Russia dominate by their size. Russia has brought out its grey stone façade and the two gigantic metal figures which crown its pavilion by means of an abundance of white light which gives the monument a bluish-white, almost immaterial aspect. Opposite to this stands the huge square tower of the pavilion of Germany, which is rendered almost transparent in the evening by the indirect illumination of long rows of lamps hidden behind the pilasters which divide up the surface of the tower. The lighted surfaces consist of partially gilded terra cotta which reflects the flood of light in a remarkably beautiful way.

Opposite these groups of buildings stands the Eiffel Tower, which continually changes its aspect. In the first place there is the dome of neon tubes already mentioned, but when these are not working the whole Eiffel Tower appears in all its delicate tracery lighted by 750 large projectors which are introduced into the iron work. The effect is startling and its colour is changed each week by changing the coloured filters in front of all the projectors. In addition, there are large search lights which send out blue-green beams straight upwards. These search lights are provided with Philips super high pressure mercury lamps with water cooling [1]), which form one of the novelties of the exhibition. Besides these search lights which are set up at the corners of the three storeys of the Eiffel Tower, there are eight more of them on the first storey (see the accompanying photographs). These eight beams of light may be moved from the vertical position until they form a fan of light which gives the impression of an immense peacock's tail unfolding. This gives an especially splendid effect on misty evenings.

Of the other buildings we must not fail to mention the new permanent museum of modern art (*fig. 2*), where the shape has not so much been adapted to the purpose of illumination, but where an interesting "claire obscure" has been obtained by an ingenious arrangement of the light sources in such a way that the reliefs and their reflection in the pool in front of the building give particularly fine effects.

The illumination of the Hungarian pavilion is

[1]) Philips techn. Rev. 2, 165, 1937.

Fig. 2. The new permanent museum which forms a part of the exhibition and which is also striking in the evening due to its fantastic illumination.

also very well carried out (*fig. 3*). The tower with its glass top, lighted by mercury light, comes out especially well among the green of the trees among which reflectors with mercury lamps are also placed.

Along the Seine the pavilion of the "Stations Thermales" (*fig. 4*) makes a very striking picture. Huge frescos in semicircular niches are lighted from the side by hidden sources. The whole building is one great light monument and deserves particular attention.

Then there is the attractive "centre régional", where buildings in the style of the different provinces of France are grouped about a square (*figs. 5* and *6*). In the evening the buildings give many pleasing glimpses of the characteristic architecture of the provinces by means of abundant and carefully arranged lighting.

Interesting adaptations of sodium and mercury lamps may be seen in the pavilion of Luxembourg and the tower of the "Pavilion du froid". The lower part of the façade of the Luxembourg pavilion exhibits a large relief representing the picturesquely situated city of Luxembourg. Hidden sodium lamps bring out the details of this relief in a golden light. The lighting of the upper band of the façade with ordinary electric light forms a contrast.

In the "Pavilion du froid" the applications of artificial cooling are exhibited. The tower is covered with a thin layer of snow which is brightly lighted with the blue-green light of mercury lamps (*fig. 8*).

When one enters the exhibition through the entrance between Grand Palais and Petit Palais, one first passes under the 300 feet high mast which, with its decoration of gaily waving signal flags, gives a foretaste of the festive spirit to be encountered farther on in the Parc des Attractions. To reach this Parc one passes over the Pont Alexandre, one of the monumental remainders of a previous exhibition, which, however, due to the twelve aluminium pavilions which comprise the exhibit of the Philips factories, is totally changed in appearance (see accompanying coloured photographs). It goes without saying that light is here the important element in bringing about the desired effect.

Above the stands which, bathed in a sea of light, exhibit the different Philips products in photograph and in actuality, the pylons glow in a blue-white light from ordinary electric lights and mercury tubes. These pylons are built up of aluminium organ pipes and appear as transparent as glass under the fantastic illumination. Seen especially from the Pont de la Concorde, where the 12 pylons are

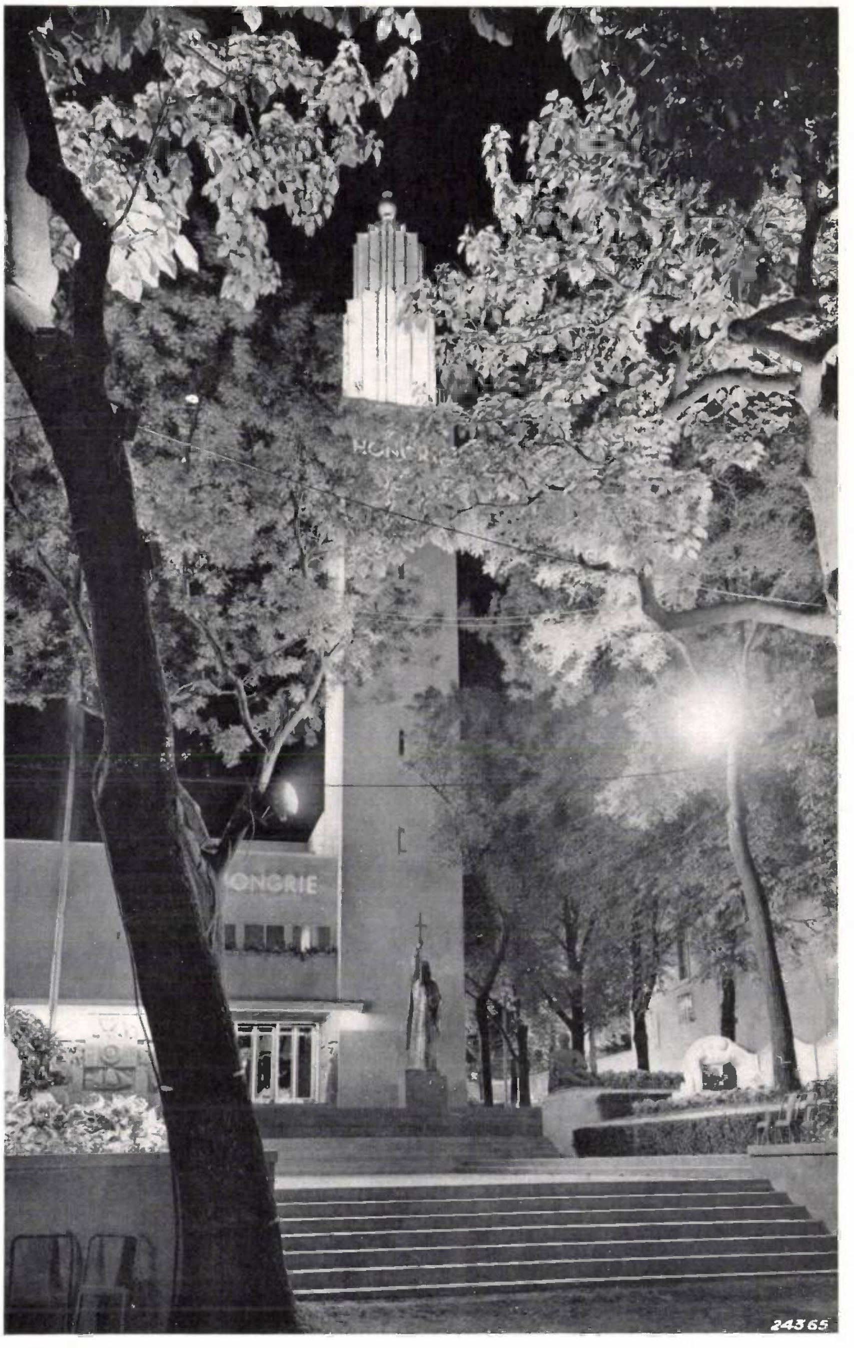


Fig. 3. Together with the trees surrounding it, the glass tower of the Hungarian pavilion glows in the blue-green light of mercury lamps.

reflected in the waters of the Seine, they form a magnificient termination to the exhibition grounds.

After this survey of the different notable buildings, we shall devote our attention to the illuminated fountains.

In the first place there are fountains in front of the Trocadéro (see enclosure). Between a double row of slender jets springs a mighty arc of water (*fig. 9*) which is lighted from below in different colours. The effect of this fountain is increased by the fact that a substance is dissolved in the water which fluoresces blue when irradiated with ultra violet light. On both sides of the fountain are long rows of Philips high pressure mercury lamps with bulbs of a glass which contains nickel oxide and which transmits only the ultra violet radiation. This invisible light causes the streams of water in the basin to shine with a blue glow. In addition the fountains are illuminated in other colours by light sources under the water, so that they are seen in a many coloured light. When the light from below is orange and it fades and changes slowly into blue, the effect is extraordinarily beautiful.

Somewhat farther in the direction of the Place de la Concorde, the fontaines de Jeumont (*fig. 10*) may be found on the bank of the Seine, opposite the fashionable restaurant "Roi George". These fountains play in an almost endless series of variations. The form and colour is continually changing. 450 lamps of 1000 watts with different coloured screens here provide the illumination. An especially ingenious invention is an instrument which has the appearance of an organ and which contains a series of switches for making the changes in form and colour of this fountain. The "organ" stands in the restaurant "Roi George" on the other side of the river, and the guests of the restaurant may play upon it.

It will be obvious that we have been able to discuss only the largest and most striking installations, and that many have had to pass without mention.

When we ask ourselves finally what is our opinion of the result of the principle followed of having the illumination consist only of the light reflected from the buildings, we are compelled to mention several objections. The great wide boulevards especially in the evening give somewhat the impression of darkness. This of course brings out the buildings

Fig. 4. The immense frescos on the curved surfaces of the façade are irradiated from the side by hidden light sources.

Beaudouin and Lods, which, with respect to lighting, is surely the most remarkable and the most successful to be seen at this elaborate exhibition. At certain times, in the evenings, particularly at the "soirée de gala", tens of thousands collect on the Pont de Jena and along both banks of the Seine. A display then begins such as surely never before was seen. To the acompaniment of music specially written by well-known composers, the light fountains on both sides of the bridge and in the middle of the Seine begin to play. These fountains change continually in height and colour to the rhythm and in harmony with the mood of the music. Among the fountains gigantic clouds of steam are developed now and again, which are also lighted by coloured lamps. At the climax of the music, enormous bunches of rockets mount into the sky and conclude the different themes with peals of thunder. All kinds of clever inventions increase the effect, such, as for example, thousands of small balloons which are freed and then pursued by two huge beams from large naval search lights to dizzy heights where the balloons are visible in the bright light as clouds

all the better, but strolling about is decidedly pleasanter and more sociable when there is a bit more light on the public. The execution of the various pavilions has not always been sufficiently under control to provide enough unity and the use of adequate light.

We may conclude from this that a general illumination of the façades may be used with success, provided that architecture and illumination are more completely under one control. We are reminded in this connection of the splendid exhibition in Stockholm, designed by architect Asplund in 1930. To give a truly gay impression the surface of the thoroughfares must certainly be lighted to a certain extent in order to form a connection between the lighted buildings.

We arrive finally at the work of the architects

Eiffel Tower.

The iron construction of this tower is lighted by 750 large coloured beam lights. Further there are on the first storey eight searchlights with super high pressure mercury lamps. The beams of these searchlights can be moved from the vertical to a fan-shaped position.

The new Trocadero.

The low buildings in the middle are lighted by ordinary incandescent lamps of a powerfull white colour. The two curved wings radiate in a golden yellow blended light of sodium and incandescent lamps. The fountain before the Trocadero is lighted in several colours and moreover irradiated with ultraviolet light. A substance dissolved in the water fluoresces blue under the influence of these rays.

Pont Alexandre.

Twelf pavilions on this bridge, crowned with aluminium columns and lighted by means of incandescent lamps and mercury lamps, have been installed on the bridge by Philips.

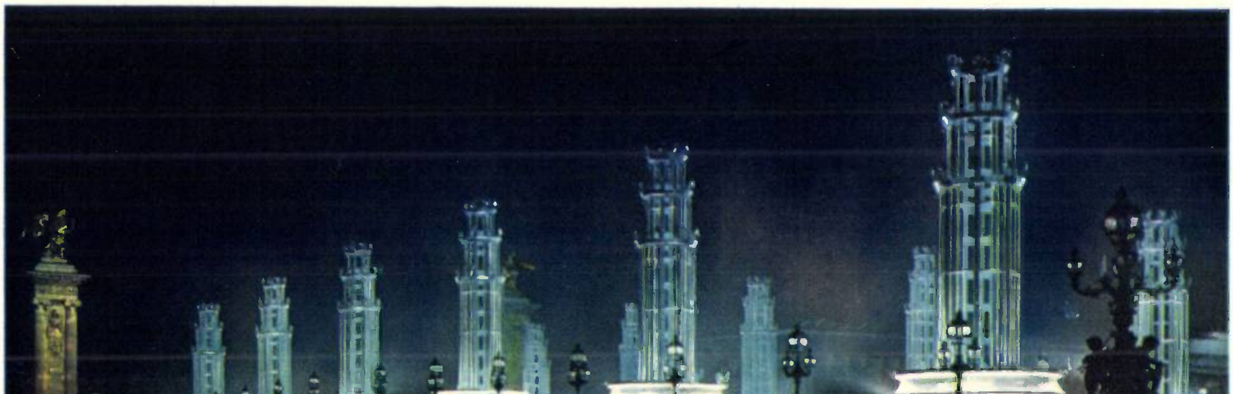


Fig. 5. In the „Centre régional".

of fine confetti against the night sky. Floating lights in different colours drift with the water of the Seine, under the bridge, and Bengal lights behind the clouds of steam and water vapour add to the glow from the projectors installed under water. Sometimes fireworks burst from the Eiffel Tower at the same time, so that the spectator is surrounded on all sides by this feast of colour, shape and sound.

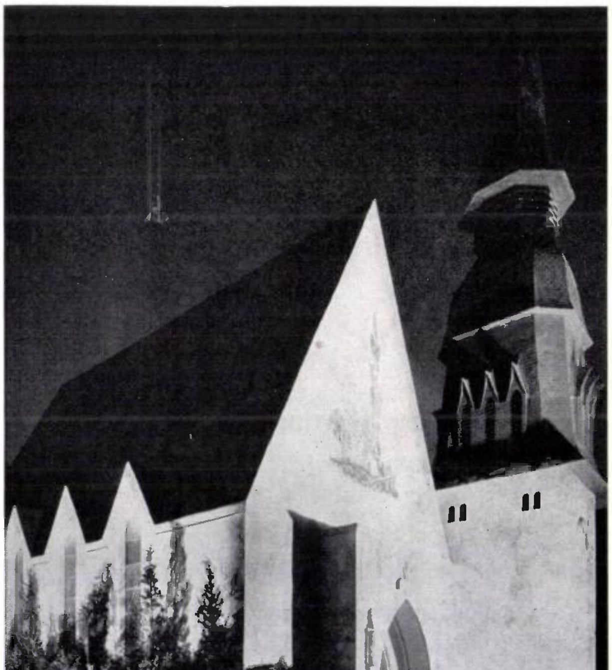


Fig. 6. Because of the carefully arranged lighting these buildings in the styles of the different French provinces offer many pleasing glimpses in the evening also.

The architects have achieved something quite unusual with this combination of music and light. It is obvious that the greatest ingenuity has been employed for this purpose. In the first place there was the problem of water-borne traffic. In the daytime this traffic could not be interfered with, so that all the fountains in the middle of the river had to disappear. Two means of doing this have been used. Part of the fountain is set up on floating buoys which are anchored on the river bottom. These buoys are connected to the bank by a cable so that the lamps can be lighted from the shore, while pumps working under water supply the water for the fountains. In the daytime the buoys are simply allowed to fill with water until they sink, while they can be brought to the surface again by being pumped full of air.

The larger fountains for water and steam are built on floating pontoons which can be towed away after the display.

The fireworks are also set up on floating pontoons and can be set off electrically, while the loud-speakers for the music are also floating and may be towed away.

Near the Belgian pavilion by the bank of the Seine lies a floating control station from which all these different instruments can be operated. An extensive switchboard with dozens of knobs and a small studio for the playing of gramophone records are housed in this station.

A complete score has been written for every display, and fountains, lights and fireworks are indicated on the different "staves". In this way the operator in the control room knows what he has to do from moment to moment during the whole display.

Such a combination worked out to the tiniest detail is of course unique up to the present, and it is particularly these displays which have again put Paris at the head of all cities which have organized world exhibitions.

In order to illustrate the enormous developments in the adaptation of artificial light in exhibitions which has been made in the last decades, we may give the following figures of the total power installed at various exhibitions:

1925: Arts Décoratifs, Paris	9 000 kW
1929: World Exhibition, Barcelona	7 500 kW
1930: World Exhibition, Antwerp	4 000 kW

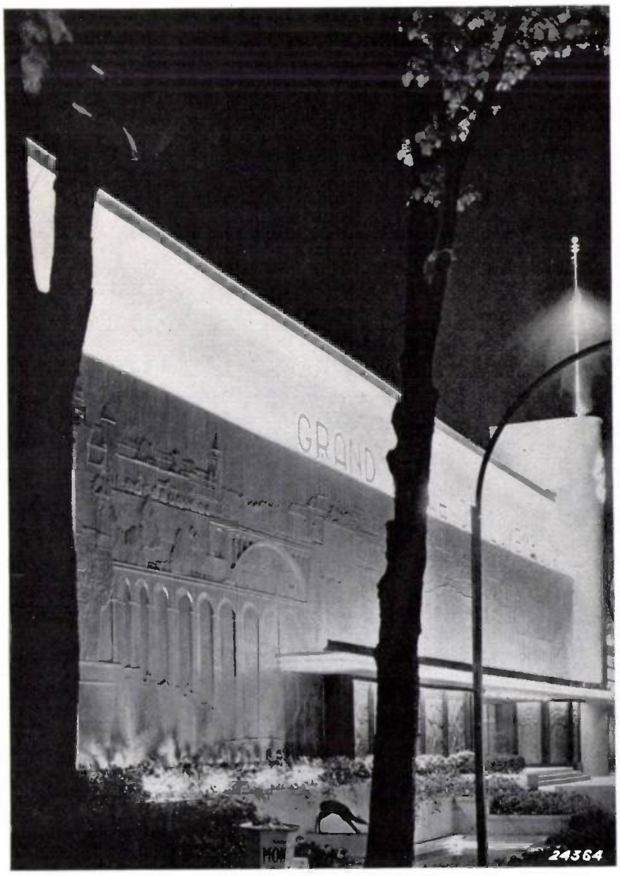


Fig. 7. Luxembourg's pavilion. The façade is lighted with different kinds of light. The huge relief of the city of Luxembourg is particularly striking in yellow sodium light.

Fig. 8. The snow-covered tower of this pavilion is lighted with blue-green mercury light.

1931: Colonial Exhibition, Vincennes near Paris	18 000 kW
1933: World Exhibition, Chicago	27 500 kW
1935: World Exhibition, Brussels	18 000 kW
1937: Exposition des Arts et Technique	62 500 kW

Apart from the range and extent of these different exhibitions, the increase by leaps and bounds in the power — the exhibition in Paris in 1931 showed double the power of that of 1925, and that of 1937 seven times that of 1931 — gives convincing proof that light plays a continually more important part in the organization of such exhibitions, but it also indicates how mankind is developing a need for more light. This need exerts great influence not only on social economy but also in our private lives.

It is interesting to be able to follow the development of lighting technique from one exhibition to another. Millions have been able to enjoy the latest novelties in illumination in Paris, but because of this very fact the organizers of the next

Fig. 9. Fontaines de Jeumont. One of the innumerable aspects which these fountains display to the public.

exhibitions find themselves faced with the task of going still farther. It is with some suspense that we await the results of the world exhibitions which are even now being planned. In the first place there is the exhibition in New York in 1939 and then that in Rome in 1942.

Doubts are sometimes felt about the benefits to be derived from world exhibitions, but for lighting technique they certainly form milestones along the road of progress, and each one is a stimulant toward further achievement. In this respect they are undoubtedly of the greatest importance.

RADIO LANDING BEACONS FOR AERODROMES

by P. ZIJLSTRA.

Summary. Following a short introduction dealing with the principles of radio landing beacons and a consideration of the advantages and disadvantages of long and short waves for this purpose, the radiation diagrams of modern landing beacons working on ultra short waves are discussed in detail. Finally a description is given of the Philips ultra short wave beacon transmitter B.R.A. 075/4.

Introduction

In an article on position finding and course plotting on board an aeroplane, in the June number of this periodical[1]), the functioning of the Philips long wave landing beacon B.R.A. 101 was explained. Different signals, are heard on either side of the course line which crosses the aerodrome. for example dots on the one side and dashes on the other. These signals complement each other in such a way that when both signals are equally strong an uninterrupted constant signal is observed. This is made possible by transmitting from a vertical aerial and a loop aerial, the ratio of whose currents can be regulated. By a combination of the circular radiation diagram of the vertical aerial with the figure of eight diagram of the loop aerial a heart shaped radiation diagram is obtained when the maximum field strengths are the same, as shown in *fig. 1*. The phase of the loop current is reversed in a dot-dash rhythm, so that the heart-shaped diagram is mirrored along the course line XY, and therefore the full line and the broken line diagram are sent out alternately in the dot-dash rhythm and together form the total radiation.

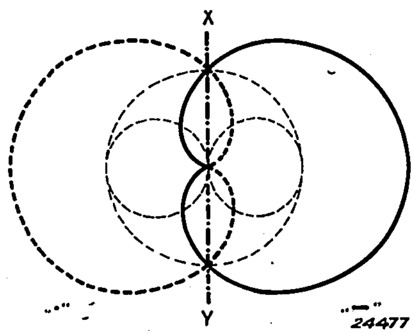


Fig. 1. Heart-shaped radiation diagram of a long wave beacon consisting of the circular diagram of a vertical aerial and the figure-of-eight diagram of a loop aerial. XY in this and the following figures is always the course line. To the right of the course line dashes are audible, to the left dots.

The nature of the signal heard when flying „off course" (either dots or dashes) determines the direction which must be steered to reach the course line. By means of the continuous signal on the course line it is possible to fly directly toward the aerodrome, and to estimate roughly the distance to the landing beacon from the field strength read from the corresponding meter. In addition two warning beacons are set up several kilometres apart along the course line, and close to the boundary of the aerodrome. These are weak transmitters which radiate a signal of their own audible only from directly above. The first indicates to the pilot that he may begin to descend and the second that he may land. Such a landing installation on long waves has already proved its usefulness many times.

It is desirable that one should hear an obvious difference in the intensity of the dots and dashes upon a slight deviation from the course line (cf. fig. 10). The smaller the angle at which the full line diagram and the broken line diagram cut each other on the course line, the sharper the landing beacon indicates the course line, since a correspondingly smaller deviation is observable.

By making the signal received from the vertical aerial weaker than that from the loop, the course line of the long wave beacon can be made sharper than is shown in fig. 1.

Choice of wave length.

One advantage of the use of long waves for landing beacons is that the beacon signal may be heard with the ordinary communication receiver of the aeroplane, and no special apparatus for blind landing by means of radio signals need be installed on board. One important disadvantage, however, is that in the long wave range only a very limited frequency band is available for beacon transmitters. With a multiplicity of beacon transmitters there is therefore the danger that they may interfere with each other. Atmospheric disturbances may also prove troublesome in the reception of long waves.

Landing beacons of the types B.R.A. 075/4 and B.R.A. 200/8 work on a wave length of 9 m which is internationally reserved for landing beacons, and

1) Philips techn. Rev. *2*, 184, (1937).

the corresponding warning beacons on 7.9 m. The combination of vertical and loop aerial usual for long wave landing beacons cannot be adopted without modification for short waves. The dimensions of the loop would be of the same order of magnitude as the wave length, and consequently the currents in the loop would no longer have the amplitudes and phases necessary for the formation of the figure-of-eight diagram. We must therefore choose a different aerial system for the beacon transmitter on short waves. It will be shown that this may be done in such a way that, in addition, the course line is more sharply demarcated than in the case of the long wave beacon whose radiation diagram was given in fig. 1.

For the sake of simplicity in the receiving apparatus on board the aeroplane, the beacon transmitter for ultra short waves is modulated with an audible frequency of 1150 cycles per second.

The symmetry of the radiation diagram

With short waves various different aerial systems may be employed for the beacon transmitter. For example the aerial may consist of two directional antennae at an acute angle to one another and which are energized in turn in a dot-dash rhythm so that together they form the radiation diagram which is given in an idealized form in *fig. 2*. Another

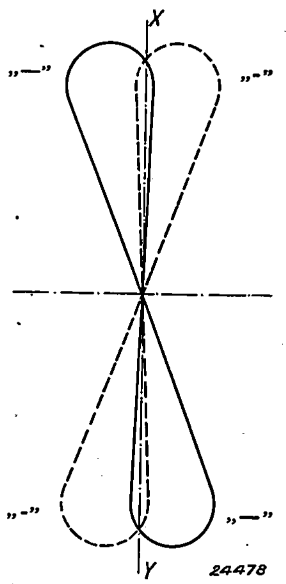


Fig. 2. Radiation of two directional aerials which make an acute angle with each other. In the first and third quadrants dots are heard, in the second and fourth dashes.

suitable aerial system consists of a vertical aerial A of a half wave length, flanked on either side by reflectors R. These reflectors are alternately rendered ineffective in a dot-dash rhythm by means of a relay so that the radiation diagram given in *fig. 3* is produced.

From these two simple examples we may now easily distinguish between the two possible methods of indicating the course line. The regions where dots of or dashes respectively may be heard alternate quadrant by quadrant in fig. 2, but in fig. 3 they

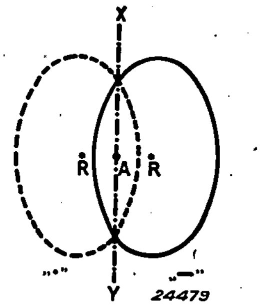


Fig. 3. Radiation diagram of a vertical aerial A with two reflectors R. To the right of the course line XY dashes are heard, to the left dots.

alternate only on either side of the course line, as was the case with the long wave beacon of fig. 1.

In the case of a beacon with quadrant zones as in fig. 2, one always hears dashes to the right of the beacon line and dots to the left, independently of the direction in which one approaches the aerodrome along this line. When dashes are heard, therefore, one must always turn to the left, and when dots are heard, to the right. In flying over the aerodrome another quadrant is entered and the nature of the signal changes. If the aerodrome has already been passed, dots are heard to the right and dashes to the left, as may be seen from fig. 2.

With a beacon having only two zones the same signal is always heard on one side of the beacon line, and this signal is not reversed when the aerodrome has been passed. In such a case the course must be followed by compass in order to know in which direction to turn to reach the beacon line when either dots or dashes are heard. Further, in the case of the two-zone beacon, there is the advantage that there can be no false course line perpendicular to the correct one, as may be the case with a beacon having quadrant zones when the field strength at the transition line from dots to dashes is not zero because of field distortions by large metal objects, such as for example hangars.

For economical operation of the landing beacon the radiated energy must be confined as much as possible to the landing line. In this repect the beacon with quadrant zones (*fig. 2*), which theoretically has no radiation perpendicular to the course line, is more satisfactory than the beacon with two zones (fig. 3) which sends out considerable radiation in that direction.

Aerial systems employed

As we have seen, a loop aerial cannot be used for the short wave beacon; in place of this two „half wave" vertical aerials were placed at a distance of one wave length from each other. Since the circuit is so arranged that the currents and the voltages in the two vertical aerials are always in opposite phases, we shall continue to call this combination the loop system (A and B in *fig. 4*). Let us now consider the field

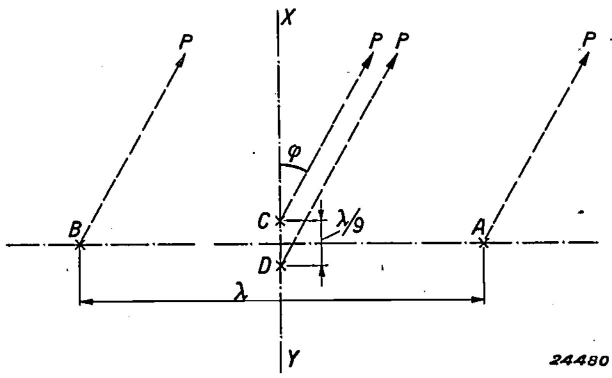


Fig. 4. Aerial system for an ultra short wave beacon. The two vertical aerials A and B are placed at a distance of one wave length λ from each other. A U-aerial, whose two arms C and D are separated by a distance of $\lambda/9$, is placed midway between A and B and with its plane perpendicular to AB. P is any given point in the horizontal plane which is so far away from the aerial that the various lines joining it to P may be considered parallel.

produced by A and B at point P at a great distance R from the middle of the line joining A and B ($R \gg \lambda$) and lying in the horizontal plane which makes an angle φ with the beacon line. At an angular frequency ω the two vertical aerials of opposite phase and the same amplitude F_m produce the following fields:

$$F_A = F_m \sin\omega\left(t - \frac{R - \frac{\lambda}{2}\sin\varphi}{c}\right) \text{ and}$$

$$F_B = F_m \sin\left\{\omega\left(t - \frac{R + \frac{\lambda}{2}\sin\varphi}{c}\right) + \pi\right\}, \quad \cdot \; (1)$$

since $\lambda/2 \sin\varphi$ is the difference in path for the waves from A or B reaching point P.

The mean value of the resulting field at point P is then:

$$F_P = 2\,F_m \sin\left\{\omega\left(t - \frac{R}{c}\right) + \frac{\pi}{2}\right\} \cos\left\{\omega\left(\frac{\lambda/2}{c}\sin\varphi\right) - \frac{\pi}{2}\right\}$$

$$= 2\,F_m \sin(\pi \sin\varphi)\cos\omega\left(t - \frac{R}{c}\right) \quad \cdot \cdot \; (2)$$

This field is therefore zero at the beacon line and perpendicular to it, while it is at a maximum at angles of 30° with the beacon line, and shifted in phase 90° with respect to the fields which each arm produces separately. For the mean value of the field strength we thus obtain the radiation diagram given in *fig. 5* consisting of four loops.

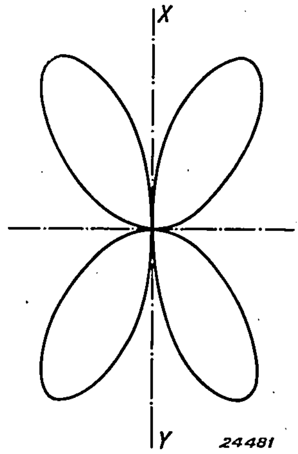


Fig. 5. Four-loop radiation diagram of the two vertical aerials A and B of fig. 4.

This four-loop diagram is still completely symmetrical with respect to the landing line, and we must add still another radiation in order to indicate the beacon line. A radiation diagram with quadrant zones may be obtained by introducing a U-aerial with its plane along the beacon line (C and D in fig. 4) between the aerials of the loop system. The U-aerial may be obtained by bending a tube of a half wave length $\lambda/2$, in a U-shape with a base of $\lambda/9$, so that the arms are somewhat less than $\lambda/4$ in length. The U-aerial is fed through its base, which is placed at the height of the middle of vertical aerials A and B, in such a way that the two arms C and D are in opposite phase, and each one is in phase with one of the aerials A and B. At point P the two arms of the U-aerial then produce the following fields:

$$F_C = F_u \sin\left\{\omega\left(t - \frac{R - \frac{\lambda}{2}\frac{\cos\varphi}{9}}{c}\right)\right\} \text{ and}$$

$$F_D = F_u \sin\left\{\omega\left(t - \frac{R + \frac{\lambda}{2}\frac{\cos\varphi}{9}}{c}\right) + \pi\right\} \quad \cdot \cdot \; (3)$$

The field produced by the U-aerial at P is thus:

$$F_P = 2\,F_u \sin\left(\frac{\pi}{9}\cos\varphi\right)\cos\omega\left(t - \frac{R}{c}\right). \quad \cdot \cdot \cdot \; (4)$$

This field is zero perpendicular to the beacon line and maximum on the beacon line, so that we obtain the radiation diagram for the U-aerial given in *fig. 6* and consisting of only two loops.

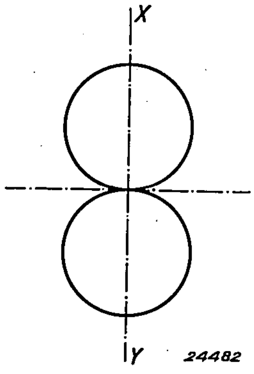


Fig. 6. Two-loop radiation diagram of the U-aerial (*CD* of fig. 4).

The total field produced at P by the whole aerial system is the sum of the two expressions (*2*) and (*4*):

$$F_P = 2\left\{F_u \sin\left(\frac{\pi}{9}\cos\varphi\right) + F_u \sin(\pi \sin\varphi)\right\}\cos\omega\left(t - \frac{R}{c}\right) \quad (5)$$

Since the phases of the fields produced by the loop and U-aerials (eq. (2) and (4)) are the same, we can obtain the total radiation diagram of *fig. 7*

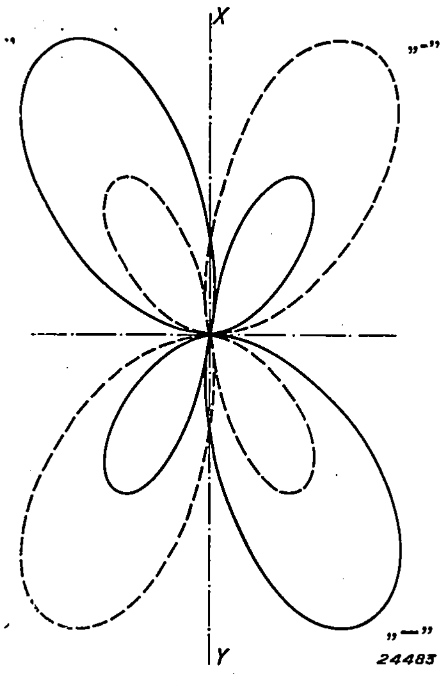


Fig. 7. Resultant radiation diagram for the aerial system shown in fig. 4.

by adding the two polar diagrams of figs. 5 and 6 with the correct sign. If the phase of the loop system is reversed, the total field at P is given by:

$$F_P = 2\left\{F_u \sin\left(\frac{\pi}{9}\cos\varphi\right) - F_m \sin(\pi \sin\varphi)\right\}\cos\omega\left(t - \frac{R}{c}\right) \quad (6)$$

In this way the full line and the broken line radiation diagrams of *fig. 7* are produced. If the loop current is reversed in a dot-dash rhythm the zones in which dashes are louder than dots and vice versa alternate in the successive quadrants.

If a beacon transmitter with half zones is desired, we may add an ordinary vertical aerial of a half wave length to the loop system having the radiation diagram given in fig. 5. This aerial is placed midway between the two wires A and B of the loop system, while the current is shifted 90° in phase with respect

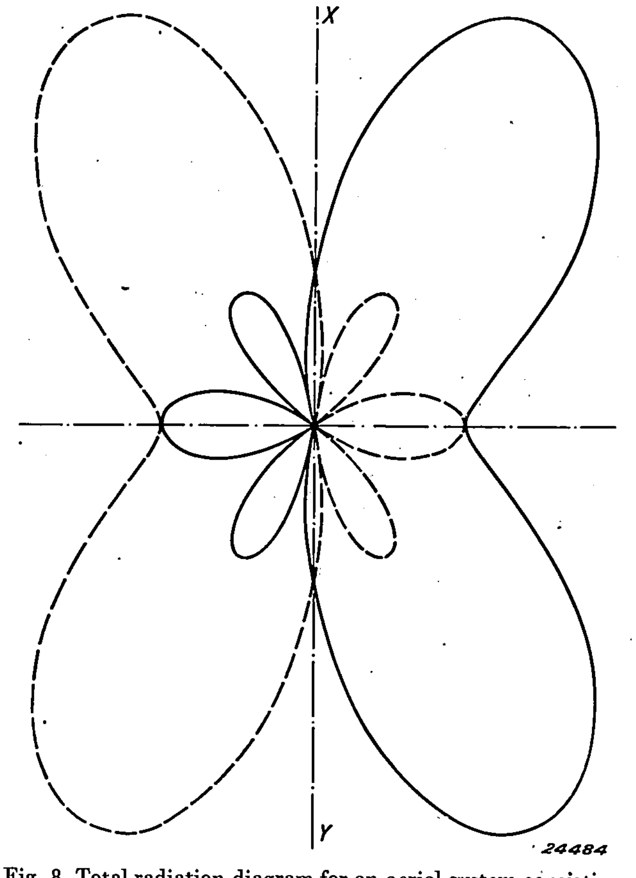


Fig. 8. Total radiation diagram for an aerial system consisting of two vertical aerials one wave length apart and working in opposite phase, with another vertical aerial half way between them whose current differs in phase by 90° from that in the first two.

to that of the loop system. The vertical aerial then radiates a field which is in phase with the field of the loop system, so that the polar diagram of the resultant average field is simply the sum of the circular radiation diagram of the vertical aerial and the four-loop diagram given in fig. 5 of the loop system, taken with the correct sign. The field at point P is:

$$F_P = \left\{F_a \pm 2\ F_m \sin(\pi \sin\varphi)\right\}\cos\omega\left(t - \frac{R}{c}\right), \quad (7).$$

when F_a represents the average value of the field of the added vertical aerial. In *fig. 8* is given the total radiation diagram for the case when F_m is taken equal to F_a; the full and the broken line figure refer again to the two directions of the loop current.

In the above-described manner we have obtained a beacon transmitter with half zones in which dots and dashes are heard respectively, while the energy radiated is fairly well confined to the direction of the landing line along which it is desirable to be able to hear the beacon signals clearly. One disadvantage however is that perpendicular to the beacon line dots and dashes are also heard equally clearly, so that a false landing course might be given in that direction. This can be prevented by a slight modification; we have to arrange the loop system whose phase is reversed in a dot-dash rhythm so that it fails to radiate exclusively on the beacon line. This may be done by placing the two vertical aerials at a distance of not exactly one wave length from each other. If we take a separation of $^5/_6\ \lambda$, for example, the resultant field of the beacon transmitter becomes:

$$F_P = \left\{ F_a \pm 2\, F_m \sin\left(\frac{5}{6}\pi \sin\varphi\right)\right\} \cos\omega\left(t - \frac{R}{c}\right), \quad (8)$$

the radiation diagram of which is given in *fig. 9*

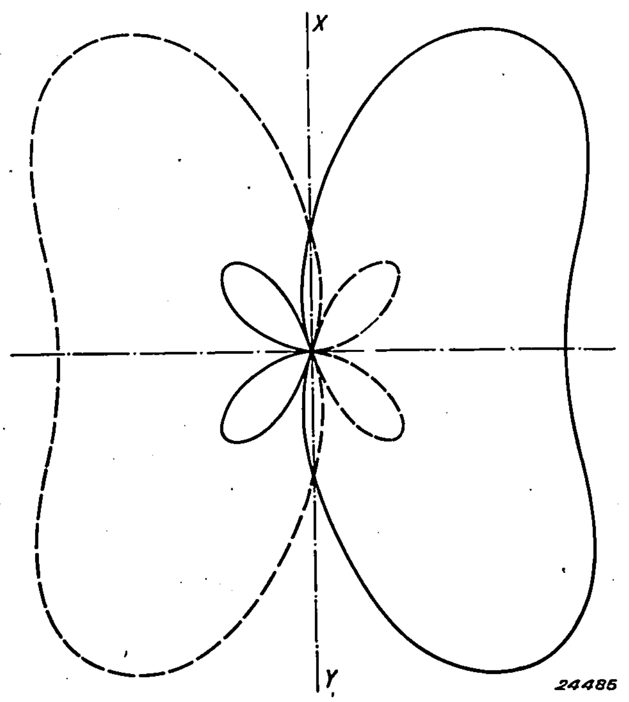


Fig. 9. Radiation diagram for the same system as in fig. 8, but in this case with the two vertical dipoles which form the loop aerial situated only $^5/_6\ \lambda$ apart.

for $F_m = F_a$. The avoidance of a false course has been made possible somewhat at the expense of the confinement of the radiation to the direction of the beacon line.

Sharpness and regulation of the course line

With a small deviation $\mathrm{d}\varphi$ from the course the difference in the field strengths for dots and dashes is $\mathrm{d}F$ as shown in *fig. 10.* For a sharp course the

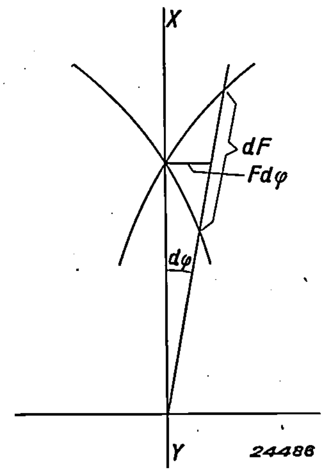


Fig. 10. Difference in intensity $\mathrm{d}F$ between dots and dashes with a slight deviation $\mathrm{d}\varphi$ from the course line.

difference $\mathrm{d}F$ in the field F must be audible at a small deviation $\mathrm{d}\varphi$. A good standard for the sharpness of the course line is therefore:

$$S = \lim_{\varphi \to 0} \frac{\mathrm{d}F}{F\,\mathrm{d}\varphi} \quad . \ . \ . \ . \ . \ . \ . \ . \quad (9)$$

For the combination of two rods with a U-aerial the sharpness of the course line thus becomes:

$$S = \frac{F_m\,\pi}{F_u \sin\dfrac{\pi}{9}} = 9.2\ \frac{F_m}{F_u} \quad . \ . \ . \quad (10)$$

For the last aerial system discussed in which the rods were placed $^5/_6\ \lambda$ apart, we obtain:

$$S = \frac{2\ F_m \cdot \dfrac{5}{6}\pi}{F_a} = 5.2\ \frac{F_m}{F_a} \quad . \ . \quad (11)$$

If we now keep in mind that the field F_m from one of the arms of the loop aerial is always several times as large as that of one of the arms of the U-aerial or of the vertical aerial F_a, we see that these aerial systems are always much sharper than for example the aerial system shown in fig. 3, whose sharpness S is only of the order of magnitude one.

For satisfactory observation by the pilot of the aeroplane it is desirable that the difference in intensity between dots and dashes be at least four decibels. This means that the ratio of the squares of the amplitudes F_1 and F_2 of dots and dashes respectively must be at least $10^{0.4}$:

$$\left(\frac{F_1}{F_2}\right)^2 = 10^{0.4} \text{ or } \frac{F_1}{F_2} = 1.7 \quad . \ . \ . \ . \quad (12)$$

In the combination of two rods with a U-aerial for which F_u is taken equal to F_a, the above relation is already reached at an angle of 1° 30′ with the course line; if F_m is made greater than F_u it is true for still smaller angles. This shows how sharply the course line is indicated by these beacons. Such a sharp beacon is also less sensitive to disturbances, because, due to the great difference in field strength between dots and dashes, the influence of a given interfering field is of course relatively less.

The beacons here described are not only inherently very sharp, they have in addition the advantage of allowing regulation. As may be seen from formulae (10) and (11), the sharpness of the course line may be regulated by changing the ratio between the fields from the different components of the aerial system. It is also possible to change the direction of the course line. This is done by allowing the two arms of the loop aerial to radiate in not quite exactly opposite phases. If the difference in phase is $\pi - \lambda$, then, with the combination of two rods and a U-aerial, the deviation from the course line becomes:

$$\varphi = \arcsin \frac{\gamma}{2\omega} \cdot$$

Thus for example if $\lambda = 15°$, $\varphi = 2° 23'$. Small deviations from the course line due to field distortion by metal hangars and the like may in this simple way be compensated.

Because of the symmetrical arrangement no total EMF is induced in the U-aerial by the current change in one arm of the loop aerial; there is said to be no radiation coupling. Upon changing the phase of the loop current we are therefore not concerned with a compensation for the changes thereby caused in the currents of the U-aerial, and this holds good for the aerial systems which produce the radiation diagrams of figs. 8 and 9.

Arrangement of the Philips ultra short wave beacon transmitter type B.R.A. 075/4.

Since the beacon transmitter must interfere as little as possible with other transmitters, its working radius must not be greater than 30 km. Because of the great sharpness of the ultra short wave beacon a high absolute field strength is not necessary. The final stage for the loop aerial may, for instance, consist of two T.B. 1/60 valves with a power of 60 watts, connected in push-pull. The final stage for the U-aerial may have even less power. The coupling of this final stage with the oscillator is variable, because with strong coupling the oscillator might easily be affected. For that reason the final stage for the U-aerial has been made similar to that for the loop aerial, and the coupling may then be looser.

The circuit of the beacon transmitter is shown in the block diagram of *fig. 11*. The frequency of the self-generating oscillator *1* is kept constant by placing the whole oscillator circuit in a thermostat and stabilizing the anode voltage. The heating voltage is supplied by a separate rectifier. The anode self-inductance is divided into two parallel portions which are coupled respectively with the final stage *2* for the U-aerial, and the phase-reversing stage *3*. The latter consists of two valves whose grids are in phase, and whose anode alternating voltages are in opposite phase. In a dot-dash rhythm the grids of these two valves are in turn made so negative that the valve concerned passes no anode current. The anodes of these two valves thus in turn supply the excitation voltage for the final stage of the loop aerial, whose phase is therefore reversed in a dot-dash-rhythm. The two valves are rendered inactive in turn in such a way that no intermediate state exists at which neither has a closed grid circuit. The occurrence of "click" phenomena in the receiver telephone is prevented in this way.

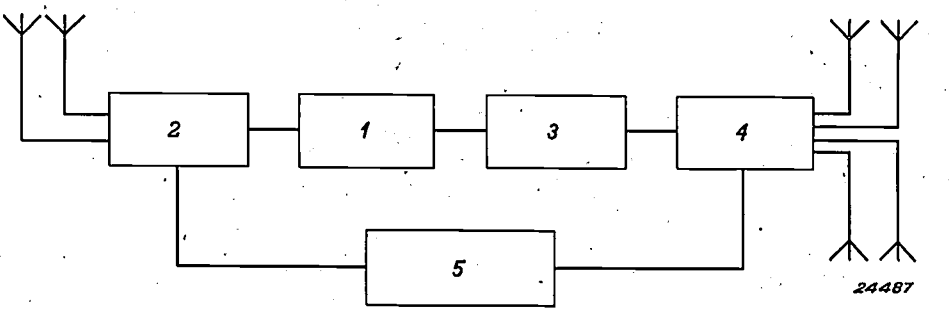


Fig. 11. Block diagrams of the ultra short wave beacon transmitter type B.R.A. 075/4. *1* oscillator stage; *2* final stage for U-aerial; *3* phase-reversing stage; *4* final stage for the loop aerial and *5* modulator.

The final stage *4* for the loop aerial also has an anode self-inductance divided into two parallel portions, each of which is coupled with the feeder of one of the two vertical rod aerials. The strength and the phase of these couplings can be altered for the purpose of regulating the sharpness and direction of the course.

The modulator *5* is a self-generating valve which feeds two parallel modulator stages each of which is coupled to one of the two final stages by means of a transformer. The various heating currents are alternating currents, except for that of the oscillator, and they are kept constant by regulator valves. The various negative grid voltages and the anode voltage for the modulator are supplied by separate rectifiers. The complete beacon transmitter is housed in a cabinet illustrated in *fig. 12.*

At present a new type of landing beacon type B.R.A. 200/8 is being built, and it will be installed at Schiphol. In addition to changes in the transmitter, the aerial system also differs somewhat from that described in this article. The radiation diagram corresponds in principle with that given in fig. 9.

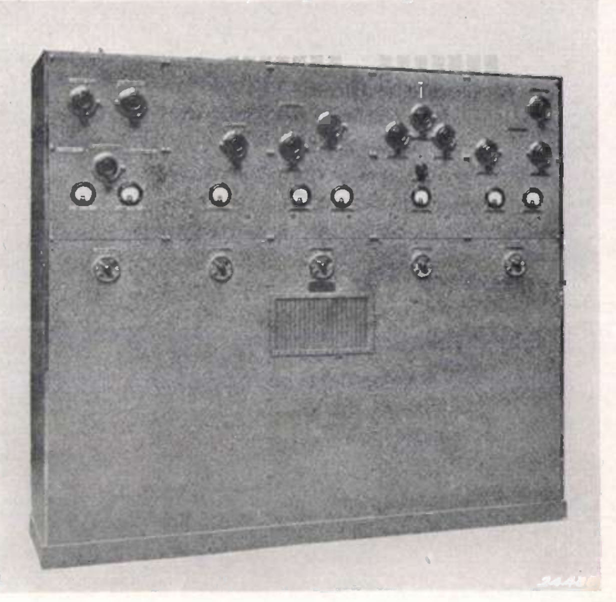


Fig. 12. Philips ultra short wave beacon transmitter type B.R.A. 075/4.

THE EXAMINATION OF THE MACRO-STRUCTURE OF MATERIALS AND PRODUCTS WITH THE HELP OF X-RAYS. III

by J. E. DE GRAAF.

The diagnosis of technical faults in casting.

We shall discuss several faults which may occur in castings without making any claims as to completeness either with respect to the faults or to their causes. It is also possible to examine the casting moulds with X-rays in order to establish the correct position of cores etc. [1]); this however will not be discussed in this article.

The faults in a casting may be divided into the following main categories: gas pockets, shrinkage faults and slag enclosures. In the case of the first two considerable difference in shape is possible according as the cavities occur in liquid metal, or in metal in a semi-solid state where solid crystals are already present in a still fluid mass.

Group I. Gas Pockets

24347

Fig. 1. Gas pockets occurring in liquid metal:

Characteristics: round spots, fairly large in size (up to several mm), with smooth, sharp outlines and heavy blackening compared with that due to flat faults; occurring sometimes alone and sometimes in groups.

Causes: gas bubbles arising out of the mould or from the melt itself into the liquid metal.

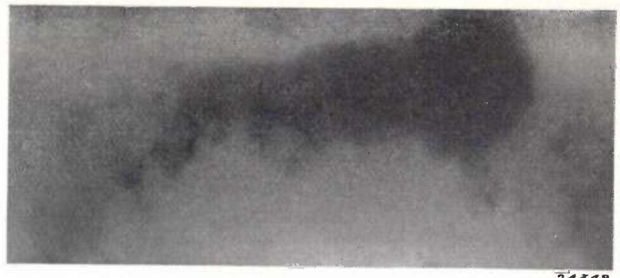
24348

Fig. 2. Gas pockets occurring in semi-solid metal.

Characteristics: small irregular spots with slight blackening and very vague outlines; they are distributed at random throughout the whole casting; with very fine cavities a cloudy image often appears.

Causes: development of gas in the melt itself during solidification.

Group II. Shrinkage Faults

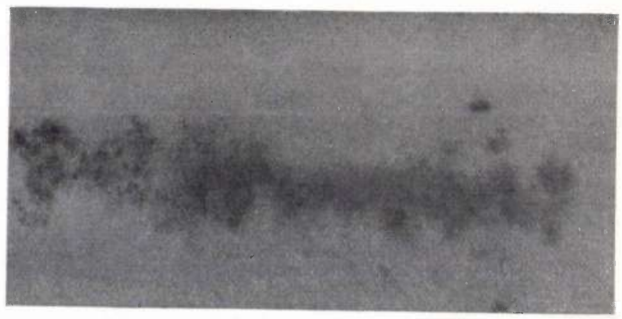


24349

Fig. 3. Shrinkage cavity which occurred in the presence of liquid metal.

Characteristics: large spots of quite random shape, often with rather sharp contours at the top and a more or less vague edge or offshoots below, heavy blackening.

Cause: a large, slowly solidifying volume of metal is cut off too early from the supply of material (for example because of the closing of a too restricted pourer channel due to freezing).



24350

Fig. 4. Shrinkage cavities which occur in semi-solid metal.

Characteristics: small irregular spots, with slight blackening and very vague outlines; situated at random in particular portions of the casting; difference from fig. 2: the gas cavities may occur throughout the whole casting, the shrinkage faults cannot occur in portions where shrinkage was

[1]) H. Reiniger: Die Giesserei 17, 40 (1930).

compensated by a steady supply of metal. Causes: the cutting off of the supply of material at a relatively late stage, so that shrinkage cavities appear only after many crystals have already been formed.

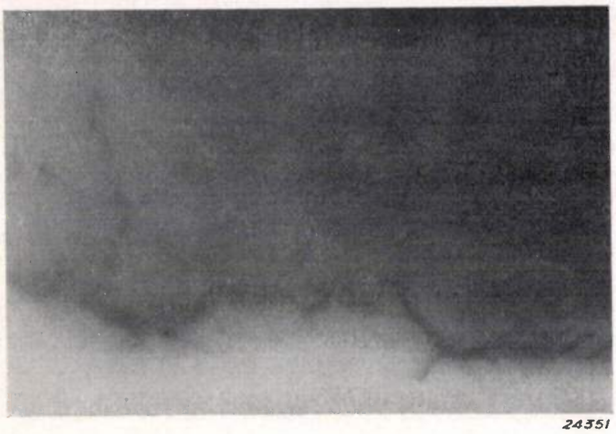


Fig. 5. Shrinkage cracks which occur in semi-solid metal.

Characteristics: the contrast depends very much upon the direction of the X-rays (cf. Philips techn. Rev. 2, 351 (1937)); relatively broad lines under favourable conditions, very black and with many branches.

Causes: due to a temporary state of stress a crack occurred rather than a cavity; the branches are due to the still only slight adhesion between the crystals already present in the mass; if the outer crust is already sufficiently strong the fault is often invisible from the outside.

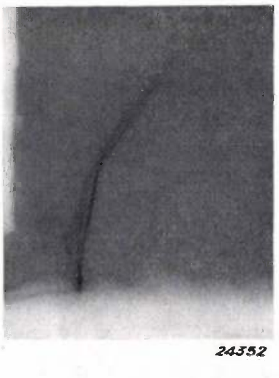


Fig. 6. Shrinkage cracks, which occur in solid metal.

Characteristics: the contrast is closely dependent on the direction of the X-rays, fine lines, sharper than in the previous case, very black and with few branches; the fault is visible on the surface.

Causes: excessive stresses in the solid state which have their highest values at the surface of the object.

Group III. Sand and slag enclosures

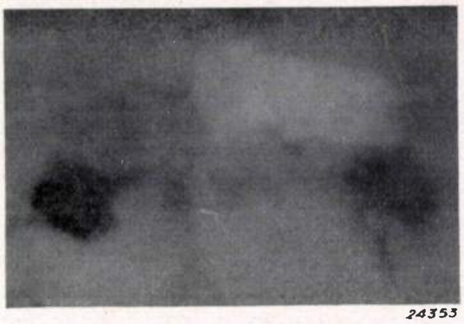


Fig. 7. Sand enclosures.

Characteristics: usually vague spots, irregular in shape and blackness, often with short branches and sometimes fine cracks; the fault is almost always directly below the surface, in contrast to shrinkage cavities.

Cause: a bit of sand has fallen out of the mould upon the rising iron (see the white spot in fig. 7: the casting is thicker at that point).

Group IV. Various Faults

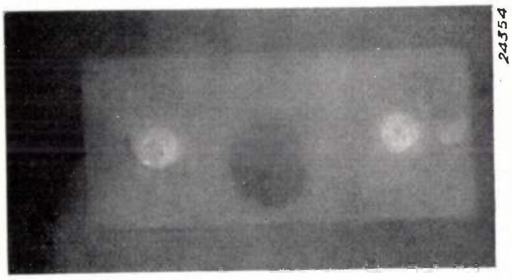


Fig. 8. Casting badly fused with the core support, (rectangular plate with two feet), immediately recognizable from the photograph. If black bands may be seen around parts cast in, these parts are practically loose (chance of leakage or fracture).

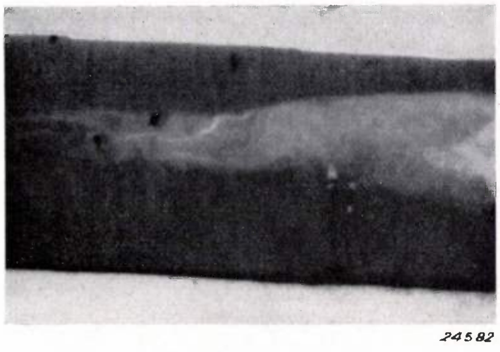


Fig. 9. Segregation (local increase of concentration) of a heavy metal from its alloy.

Characteristics: light spots which do not correspond with surface projections and must therefore be caused by heavier metal (cavities are darker).

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

No. 1216: K. F. Niessen: Eine Verschärfung des verbesserten Sommerfeldschen Fortpflanzungsformel für drahtlose Wellen zur Ausbreitung ihres Gültigkeitsgebietes nach kleineren Abständen (Ann. Physik **29**, 569 - 584, Aug. 1937).

Formulae are derived for the field at the earth's surface of a vertical, infinitesimally small dipole transmitter situated directly above the earth considered as flat, and particularly for the field at a distance of several wave lengths. The theory given by Sommerfeld in 1926 for the field at greater distances is made more rigorous with the help of an exact expression for the π function of Hertz derived by van der Pol and Niessen, in which account must be taken of terms which appear especially in this case because of the slightness of the distance. Different formulae are derived according as the electric or the magnetic field must serve for reception. These formulae coincide in the Sommerfeld region. The constants which describe the properties of the ground are chosen arbitrarily in this article. For several practical values the change of the intensities is shown as a function of the distance measured in wavelengths, in order to shown the deviation from the extrapolation to smaller distances from the formula given by Sommerfeld in 1926 which agrees with that of Weyl.

No. 1217: K. F. Niessen: Zur Entscheidung zwischen den beiden Sommerfeldschen Formeln für die Fortpflanzung von drahtlosen Wellen (Ann. Physik **29**, 585 - 596, Aug. 1937).

This article was written in response to the fact that up to recently use was made in the literature of the formula derived by Sommerfeld in 1909 for the field of a dipole transmitter, in spite of the fact that Sommerfeld himself in 1926 had given another in its place (in Riemann-Weber, for example) which agreed with a formula deduced in quite a different way by Weyl in 1919. The exact expression derived by van der Pol and Niessen in 1930 also leads to the same result as a first approximation. The purpose of the article was to discover where the error lay in Sommerfeld's first treatise.

No. 1218: E. M. H. Lips: An analysis of the structure of pearlitic cast iron (Gieterij **11**, 235 - 239, July-Aug. 1937).

In this lecture, given before the Netherlands Society of Foundry Technicians, the fact was discussed that normal pearlitic cast iron consists of a foundation mass of alloyed pearlitic steel which has a high tensile strength and slight wear, in which a larger number of graphite flakes are imbedded, with the result that the tensile strength is reduced to about one tbird of that of the foundation mass. The tensile strength of pearlitic cast iron can be raised by making the graphite inclosures smaller in number and more round in shape. In order to retain the low modulus of elasticity, the high damping capacity and the ease of working of cast iron, minimum content of carbon should be about 2.5 per cent, to which corresponds a maximum tensile strength of 35 to 40 kg/mm^2.

No. 1219*: R. Houwink: Elasticity, Plasticity and Structure of Matter (with a chapter on the plasticity of crystals by W. G. Burgers); 394 pages. 1937 (Cambridge Univ. Press).

In this book a survey is given of the modern views of elasticity and plasticity phenomena and of their relation to the structure of matter. Insight obtained from physical and chemical viewpoints as well as results attained from the technological side in the development of new materials are discussed. Various specialists have collaborated in the chapters on particular substances.

No. 1220*: J. H. de Boer: Elektronenemission und Adsorptionserscheinungen; 334 pages, 1937 (Joh. Ambr. Barth, Leipzig).

This book is the German edition of "Electron Emission and Adsorption Phenomena" by the same author, published by the Cambridge University Press in 1935. The author has made use of the opportunity to include the further developments in this subject. Further elucidations have also been introduced à propos of desires expressed in the discussions of the English edition. New conceptions

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

are given especially in regard to the ionization state of adsorbed atoms and the theory of the disturbed lattice positions in semi-conductors.

No. 1221: M. J. Druyvesteyn: Der Anodenfall in den Edelgasen He, Ne und Ar. (Physica 4, 669 - 682, Aug. 1937).

The degree to which the voltage on the tube depends upon the distance between the electrodes is determined for helium, neon and argon in a glow discharge as well as in an arc with a hot cathode. At a definite separation D the potential rises sharply since an anode drop occurs. The distance D is practically the same for a glow discharge and for an arc. The product of D and the gas pressure is found to be independent of the pressure at high gas pressures. For low pressures however it decreases with the pressure, since D can not be much greater than the diameter of the tube. By considering the Faraday dark space as a plasma, this behaviour can be explained. Measurements of the tube voltage as a function of the gas pressure at a constant distance between the electrodes are found to furnish results which agree with this.

No. 1222: J. A. M. van Liempt and J. A. de Vriend: Testing synchronizers (Physica 4, 703 - 714, Aug. 1937).

A description is given of the synchronizers used in photography with flash-light lamps, and their practical significance is explained. With the aid of a cathode ray tube the time constant and the fluctuation of a synchronizer is determined. It is explained that the flash time of flash-light lamps used with synchronizers must not be too short. The article ends with the description of a method for the correct adjustment of synchronizers, which has in the meantime also been given in Philips Technical Review 2, 334, 1937).

No. 1223: C. J. Bakker: On the number of neutrons emitted by a radium-beryllium source (Physica 4, 723 - 729, Aug. 1937).

The number of neutrons emitted by a radium-beryllium source per sec and per milli-curie of radium is measured according to a method given by Amaldi and Fermi, and is found to be $(2.1 \pm 0.2) \times 10^4$.

CONTENTS OF PHILIPS TRANSMITTING NEWS.

The editor proposes to announce regularly in future in the Philips Technical Review under "Abstracts of Recent Scientific Publications" the contents of "Philips Transmitting News" as soon as published. This journal, which is devoted to transmitter engineering, appears 4—6 times a year in German and English. For subscriptions kindly apply to Dept. Zendwezen of the N. V. Philips' Gloeilampenfabrieken, Eindhoven.

Of the fourth year's edition already appeared

in March 1937 Vo. IV, No. 1:

K. Posthumus and Tj. Douma: Some considerations of the frequency stability of ultra-short wave driving oscillators (continuation).
Short-wave broadcast transmitters type K.V.F.H. 10/12 for British India.
J. van der Mark: Television.

In May 1937. Vol. IV. No. 2:

F. M. Duinker: Filters for rectifiers.
A practical apparatus for radio-technical calculation.
The watercooled pentode type PA 12/15.
H. G. Boumeester and M. J. Druyvesteyn: Gas-filled triodes.

in July 1937. Vol. IV. No. 3:

K. Posthumus: The watercooled pentode type PA 12/15 an ultra short wavelength.
Aerodrome transmitters for Turkey.
H. G. Boumeester: Manufacture of modern transmitting valves.

in Sept. 1937. Vol. IV. No. 4:

J. P. Heyboer: The use of pentodes in radio-transmitters.
New Aircraft equipment for the Royal Dutch Airlines.
W. Albricht: A transmitter for television experiments.
Review of scientific publications.