PHILIPS TECHNICAL REVIEW

VOLUME 7

1942

PHILIPS RESEARCH LABORATORY – EINDHOVEN

CONTENTS

VOLUME VII – 1942

LIGHT AND ILLUMINATION

RADIO AND TELEVISION

SOUND RECORDING AND REPRODUCTION

TELEPHONY

CATHODE AND X-RAYS

ELECTROTECHNICAL APPARATUS

MATERIALS AND THEIR RESEARCH

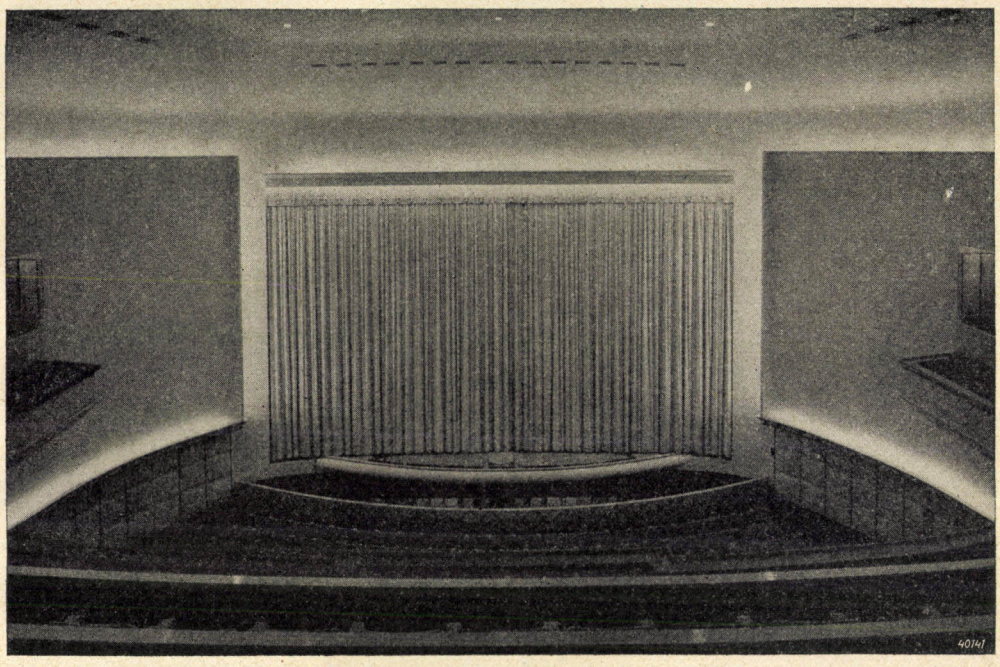
MISCELLANEOUS

VOL. 7 No. 1, p. 1-32 JANUARY 1942

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 ILLUMINATION OF THE NEW MUNICIPAL THEATRE IN UTRECHT

by L. C. KALFF.

628.972

In connection with the opening of the new Municipal Theatre in Utrecht, a discussion is here presented of the influence on the architecture of the interior exerted by the artificial illumination. The conclusion is reached that in this case illumination and architecture are so intimately related that the building represents a step forward in the progress of illumination architecture.

In visiting the new theatre in Utrecht it is interesting to compare the impression received with that made by older theatres, in order to see what signs can be found of a development in plan, architecture and technical installation. In this periodical we shall confine ourselves to a discussion of the interiors and expecially to the problems of illumination and acoustics[1]).

In this most recent creation of the architect D u d o k, the illumination plays such an important role in the architecture of the interior that it would be impossible to separate light and form from each other. This striking characteristic of the various parts of Utrecht's theatre points to an important evolution in illumination architecture.

Several episodes in the progress of the development of theatre illumination

Until 1920 all theatres were in fact illuminated in a way which had become traditional and which was derived from the candles of the oldest theatre halls. This tradition had grown up logically. All the halls were anything but fire-proof, for the curved surface of walls and ceilings necessitated the use of wood; the walls, chairs, floors, everything, in short, was made of wood and covered with cloth. The lighting of the hall had thus to be installed in such a position that

[1]) *Cf.* the following article by R. V e r m e u l e n.

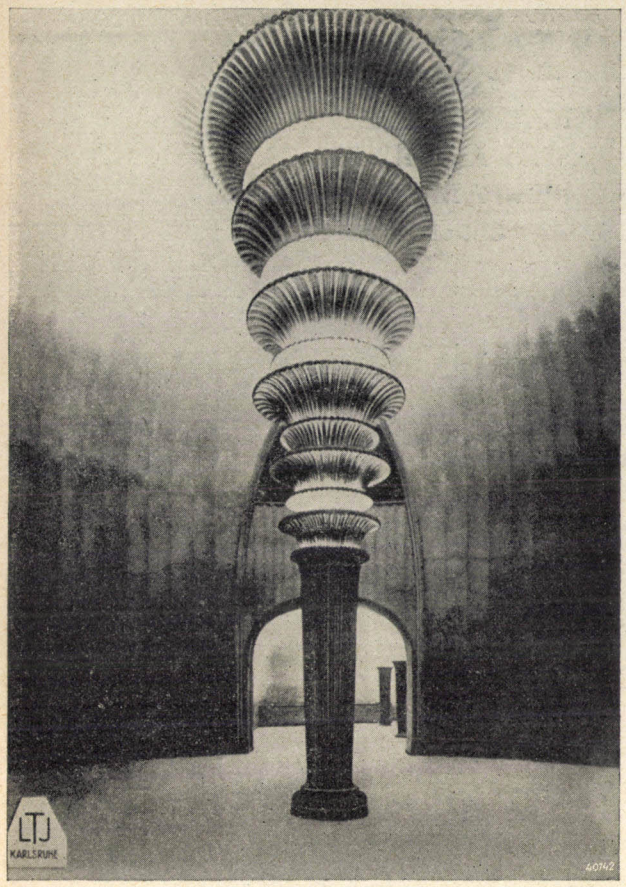


Fig. 1. One of the entrance halls in the „Große Schauspielhaus" in Berlin, built by architect P o e l z i g. (The auditorium of the theatre is not typical of such structures, since it is a rebuilt circus building). The picture shows one of the earliest examples of an intimate relation between light and architecture. Especially when it is kept in mind that this building dates from immediately after the world war, the merits of the architect can be rightly assessed.

the flames of candles and oil lamps, and later of gas, offered the least danger of fire. In most cases the position chosen was the centre of the hall, where a monumental chandelier bore the necessary number of light sources. Directly above the chandelier a grating was installed with a large ventilation shaft in the cupola above, in order the carry away the excess heat. As long as electric lighting was still unavailable the chandelier had to be lighted with tapers or candles on long poles, or, before the performance the whole chandelier was lowered with a winch and then lighted. Gas flames were then able to burn very low during the performance, but the oil lamps and candles remained lighted the whole evening.

Some theatres with boxes, like the famous Scala in Milan, had an eleborate system of lighting along the walls, on brackets between the boxes, so that they were easily accessible from the gallery.

Both of these solutions originated from necessity. But although electric light was introduced into most large theatres before 1900, no examples can be found before 1920, where the illumination was given any other than the traditional position.

This illumination had as yet only little influence on the form of the decoration of the hall; it may, however, be assumed that the frequent use of gilded mouldings and leaf-work, like the common use of lustrous textile materials such as silk and brocade, was a result of the dimness of the sources of light. By the use of these materials and also of crystal prisms in the

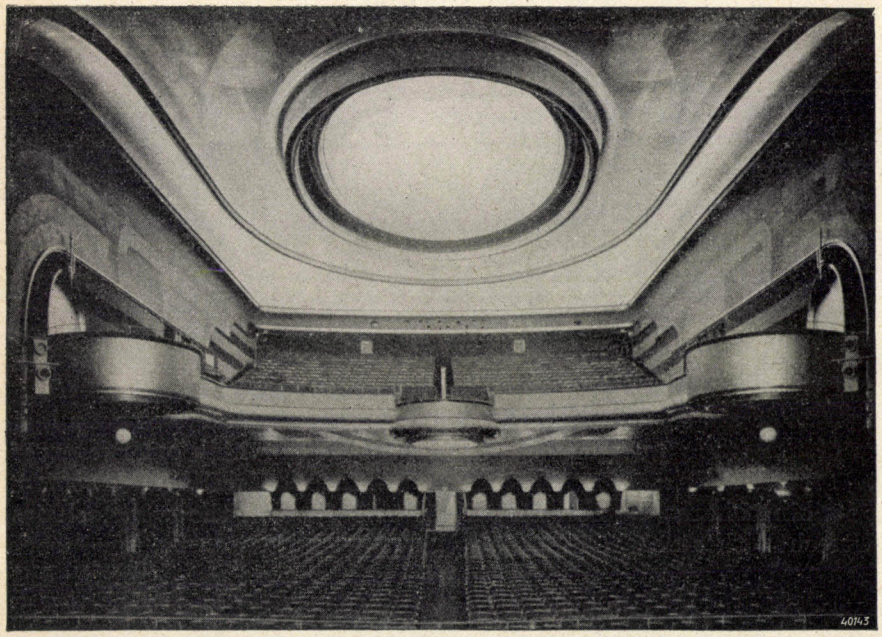


Fig. 2. The great auditorium of Titania Palast in Berlin by the architects S c h ö p f e r, Schoenbach and J a c o b i. The new possibilities of building with light have been employed here by the architects on a rather excessive scale, so that the lines of the hall are not organic and the impression now given is that the hall is already no longer modern.

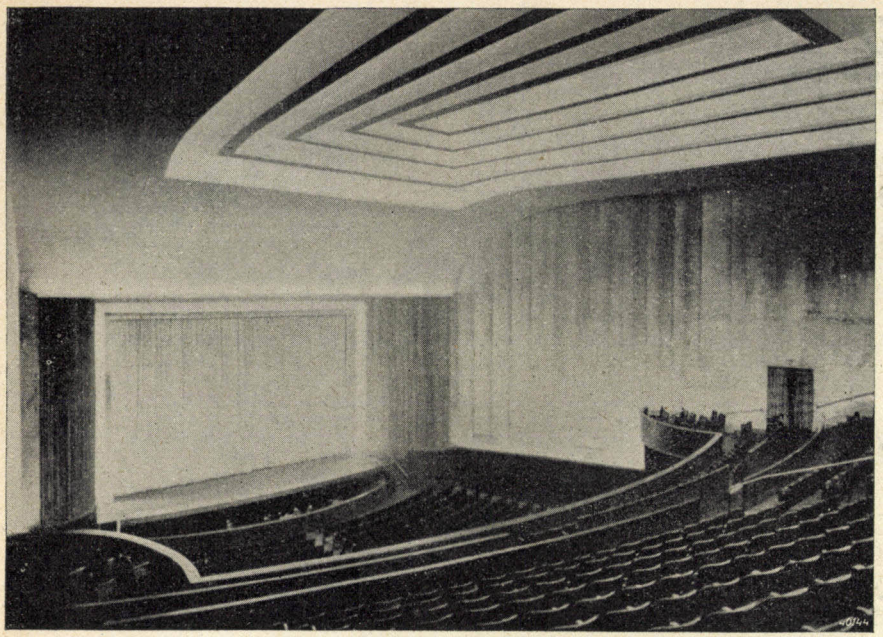


Fig. 3. Theater Lichtburg, Berlin, architect Rudolf Fränkel. The architecture here is much more restrained than in fig. 2. In a restful way the light has here been promoted to the position of the chief decorative motive. Although not all the possible consequences of the new light architecture have been carried through, this hall at the present time is still very pleasing and its lines are certainly not outdated.

chandeliers and of mirrors, the effect of the weak flames of candles and oil lamps was intensified.

While at first the introduction of the electric lamp had little effect on the decoration of the theatre auditorium, this was altered after the world war. People began to go out more and to seek in social entertainment and new impressions to drown their memories of those sad, dark years. In particular, however, it was the cinema which experienced an enormous development. It was in those years that the first huge cinema halls were built, such as Titania Palast in Berlin, Tuschinsky in Amsterdam and the Paramount Theatre in Paris. Originally, because films were still silent, and later because the sound can be amplified many times with loudspeakers, the acoustics of such halls presented less complicated problems to the architect than the acoustics of theatre auditoria. In cinema halls much greater distances between stage and audience can be tolerated than in theatres, and the architect feels less bound to the classic forms and materials of the playhouse.

In the field of illumination a large number of new designs were originated after 1920. The fact that the rapid growth of the film industry furnished the stimulus to this, resulted in an abnormally rapid development (with many excesses), which gradually also affected theatre construction. The architect Poelzig in Berlin, who rebuilt an old circus building into the ,,Große Schauspielhaus'' (see *fig. 1*), must certainly be mentioned as a pioneer. However, because of its quite exceptional form, this building has not had much influence on the development of the theatre auditorium.

The auditorium of Titania Palast in Berlin (*fig. 2*) is an example of the rebundant new wealth of line which artificial illumination puts at the disposal of the architect. It will now be generally agreed that this rebundancy failed to provide pleasing results. The lack of unity between the form and the purpose of the hall, and the unrestrained decorative forms provide evidence that it marked the beginning of a phase of development which still lacked the necessary clarification.

In contrast to this example we may consider the ,,Lichtburg'' (*fig. 3*) which has many more good architectural qualities. The architect has here confined himself to the use of a single lighting motive which is introduced as the only decoration in the ceiling, on the walls and around the stage opening. These light coves give a balanced and restful division into light and dark of walls and ceiling, while the unity in the decoration gives a pleasing effect.

We wish to point out several features in this hall which will enable us to make interesting comparisons in the discussion of the new Utrecht theatre.

The architect has perceived that a theatre auditorium must in the first place form a frame for the changing scene on the stage. He has therefore avoided the fussy decorations which could be varied by means of coloured lights, which were at that time the fashion. He probably considered rightly that in a hall which is repeatedly visited such expensive and temporary effects soon bore and become worthless. As a matter of fact we know of no hall possessing an installation for three or four different colours

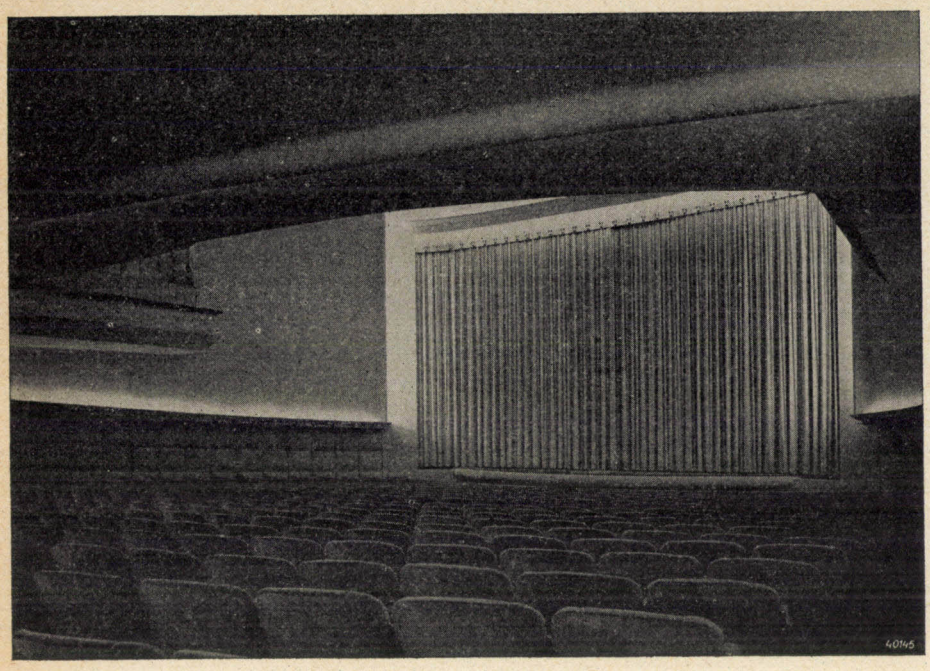


Fig. 4. View of the stage of the new theatre in Utrecht from the rear seats. The whole stage can be clearly seen. The ceiling of the balcony appears high and short due to its upward sloping form. The half-silvered „Cornalux" lamps in the light coves behind the side walls give strong relief to the very light curtain. A difficulty was found to exist in the difference in colour of lamps of different wattage. The light of 100 W „Cornalux" lamps is much whiter than that of show-window lamps „Philinea" lamps or ordinary lamps of 25 W. For this reason the side walls were painted in a slightly warmer tone and the inside of the light cove above the wainscot was made yellow. This produced an appreciable colour contrast between the light on the side walls and the light on the ceiling and the wall surrounding the stage.

of light in which any other than white light was used some time after the opening!

The position of the light coves is here indicated by the division of the hall, but no connection of the coves themselves and the light which issues from them with the shapes of the surfaces of wall and ceiling is yet noticeable. There is simply a band of light above the wainscot and around the stage opening, and the same element is repeated several times in the ceiling. In this ceiling, however, a different figure could also have been designed with the same construction, and the light from the frame around the stage opening does not fall upon surfaces especially intended to receive it. Here, therefore, there is not yet the intimate relation between form and lighting effect which might be denominated as consistently sustained illumination architecture.

The framing of the stage opening with light is typical. When this decoration is considered, it is found to be illogical. When the lights in the hall are burning the closed stage opening is not the most interesting part of the hall. At that time the audience will be finding their seats, conver-

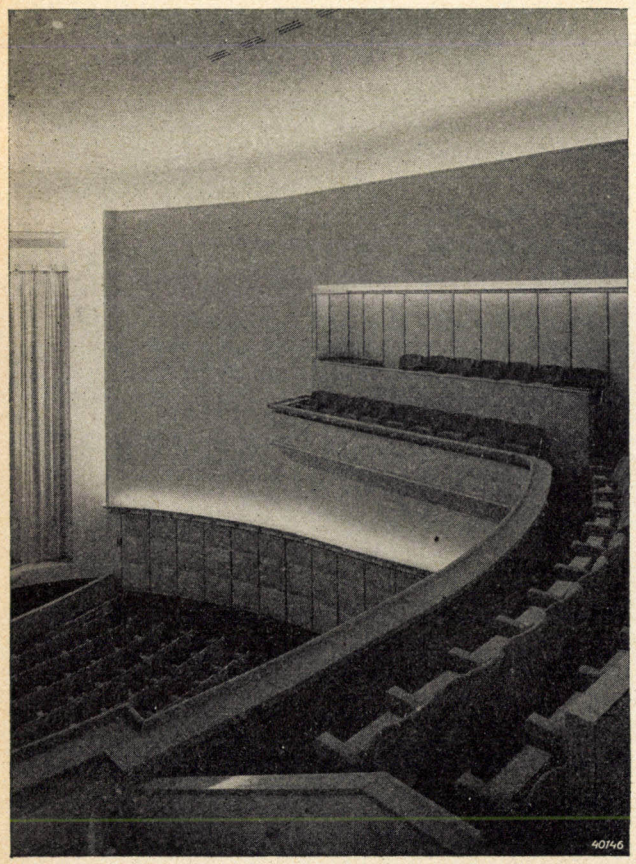


Fig. 5. View of the hall from the balcony. The different arrangements of the illumination are clearly visible here. Attention must be called to the slight unevenness of the ceiling illumination, which could if desired have been avoided by giving the ceiling a completely curved surface. This has, however, been avoided for acoustic reasons (*cf.* the following article). In practice the uniformity of the brightness of the ceiling was found to be very much affected by the structure of the surface of the plaster. The ceiling is rubbed with silver sand which produces a fine-grained surface. From the point of view of light technology this surface presents the difficulty that it exhibits great differences in brightness according as one observes it in the direction in which the light strikes it or in the opposite direction. This is caused by the innumerable tiny shadows cast by the grains of sand (*cf.* Philips techn. Rev. **5**, 125, 1940).

sing with each other or looking about the hall for acquaintances, etc. But if the attention is to be focussed on the stage, as it were, to increase the anticipation, the light from the above-mentioned frame ought to shine upon the curtain so that its beautiful material and rich drapery constitute an ornament to the wall. This has not been done, however, and the light is directed away from the curtain. And when the hall is darkened and the footlights or the projection apparatus are turned on, the light frame of the most interesting part of the hall must also be darkened !

Lighting of the Utrecht theatre auditorium

We shall now proceed to the discussion of the auditorium of the new Utrecht theatre, since this is certainly the most important part of the whole building, and since the auditorium discussed above furnishes points of comparison chiefly with this auditorium. We are aware of the fact that in doing this we are not doing full justice to the merits of the architect, because one of the good qualities of this building is the way in which the visitor, moving through the successive rooms and passages, each with its own proportions and atmosphere reaches a climax in the auditorium. We shall, however, speak of that later.

Above the wainscot of padded white artificial leather with gilt borders the hall is entirely in cream-white plaster. The colour scheme is extremely simple. The only colours present are those of the dark brown carpet and the striking brick-red velvet upholstery of the seats. The shape of the hall is such that every seat has an absolutely free view of the stage and that speech as well as music is excellently heard everywhere in the auditorium. In our opinion this is of more value than the intimacy lost by the arrangement, which could have been avoided by building two, less deep galleries, which, however, always contain a number of poor seats. The lighting here forms an integral part of the line and decoration of the hall, and in our opinion this is an extremely important advance in the art of interior decorating.

The side walls, curving gently toward the stage give the impression of being free standing shells from behind which light shines out. This light falls on the ceiling and thus takes away its heaviness, it also falls upon the gleaming curtain of the same colour as the wall and thus gives interest to that part of the wall even when the curtain is closed (*fig. 4*) .The side walls are relatively dark, but they therefore form as it were the fixed surface behind which the widening space recedes (see *fig. 5*).

The ceiling under the large balcony is treated in the same way. It gives the impression as if a free thin shell were suspended under the back of the balcony, from behind which the upward sloping front part of the ceiling of the balcony is lighted. This removes all the heaviness of this large surface, and there is no sense of depression in this lowceilinged part of the hall. The back part of the ceiling is pierced by five large flat domes, so that there also the impression is given that the lighted surface recedes far away above (see *fig. 6*).

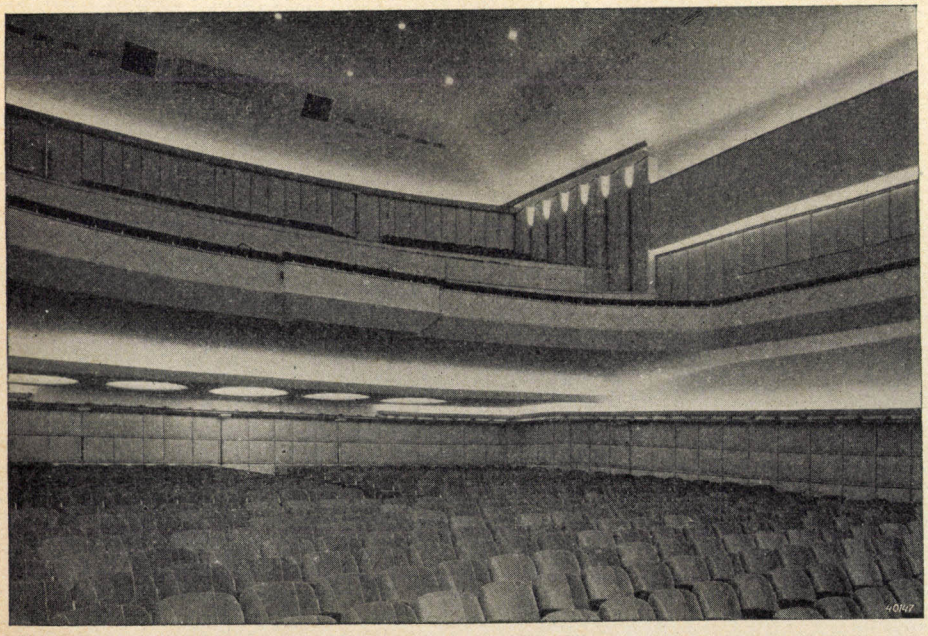


Fig. 6. View from the stage into the auditorium. Attention is called here to the form of the ceiling under the balcony which gives the appearance of a thin free-hanging shell in which circular holes are pierced and above which the upward sloping, brightly lighted ceiling of the balcony is suspended. In the ceiling of the auditorium may be seen the regularly spaced small holes through which direct light is thrown upon the audience.

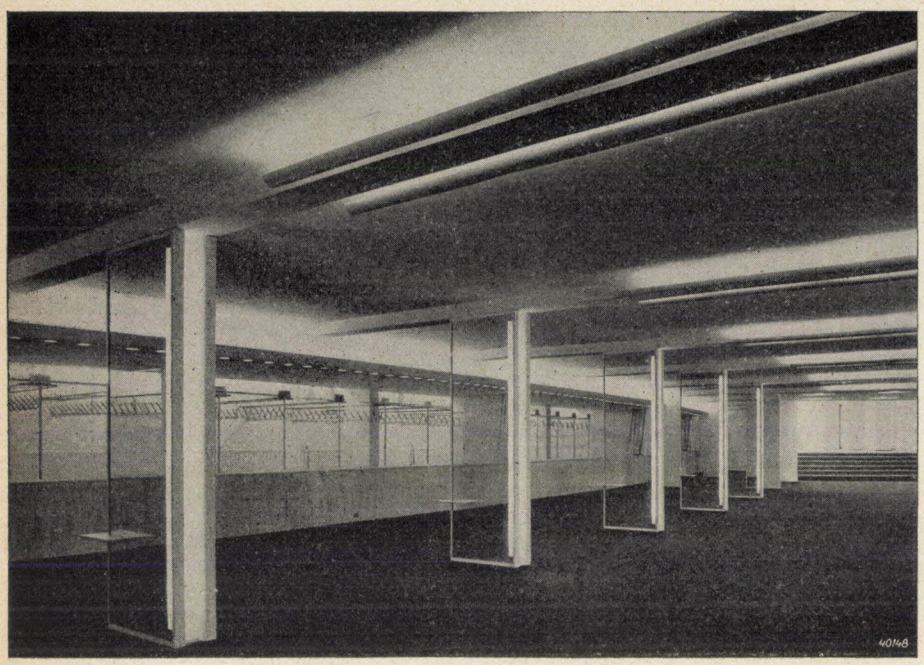


Fig. 7. The cloak-room is lighted entirely with indirect light from light coves above the counters and attached to the beams of the ceiling. The latter prevents an unnecessarily heavy impression being given by the beams. The illumination with the vertical mirrors, which also take away the heaviness of the columns which they hide, forms a pleasing accent.

It would lead us too far if we were to discuss many more of the details of this hall, but we hope with the help of the photographs to have shown that the work as a whole would be absolutely unthinkable without the light. Not only does it everywhere provide the necessary illumination. but it accentuates the distances between the different surfaces, provides an interesting variety of light and dark, it slides gently over the smooth curves of the padded wainscots, conjures up rich gleams in the folds of the stage curtain and falls unsuspected from the small openings in the ceiling on the gowns and jewels of the audience, so that also there is no lack of vivacity.

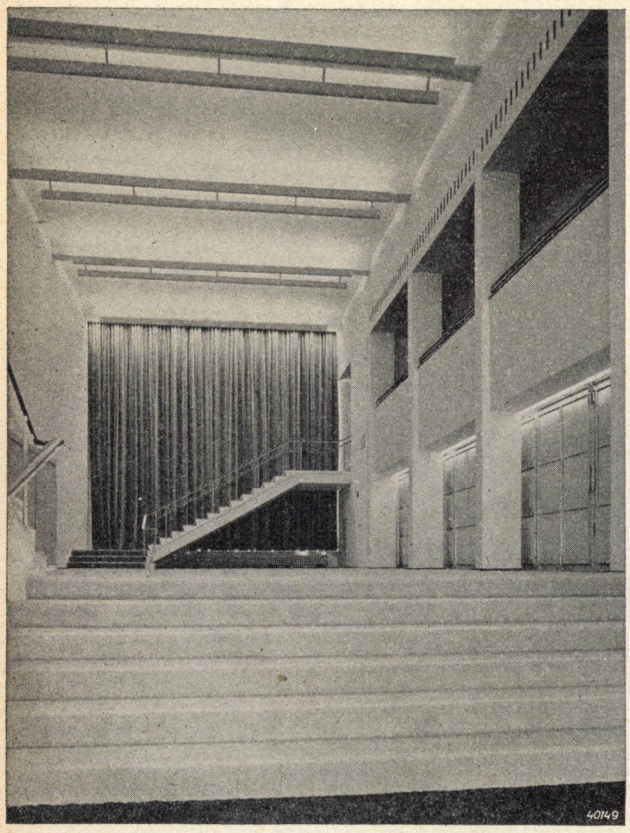


Fig. 8. Coming from the low-ceilinged cloak-room one enters the lofty hall which is lighted in the same simple manner as the cloak-room. Everywhere use is made of special lighting of the hangings to bring out the beautiful effects of the rich materials used.

The features indicated show the progress which has been made since the time of the examples previously mentioned. The builder uses light as a building material, no moulding, surface or colour is chosen before he has taken into account how the light will fall upon it and how therefore it will be seen.

We would by no means claim that this is the first example of which this may be said, on the contrary, we here see a growth and a continuation of the work of others. This piece of work is a link in a chain which will undoubtedly become longer in the future, and in which works of greater perfection will occur, In this connection it is instructive to consider the work of B ij v o e t in the auditorium of the theatre „Gooiland" in Hilversum, which exhibits the same principle in a less fully achieved form and which certainly must have constituted an example to our architect in many respects.

Illumination of the other parts of the theatre

When the other parts of the theatre are stu-

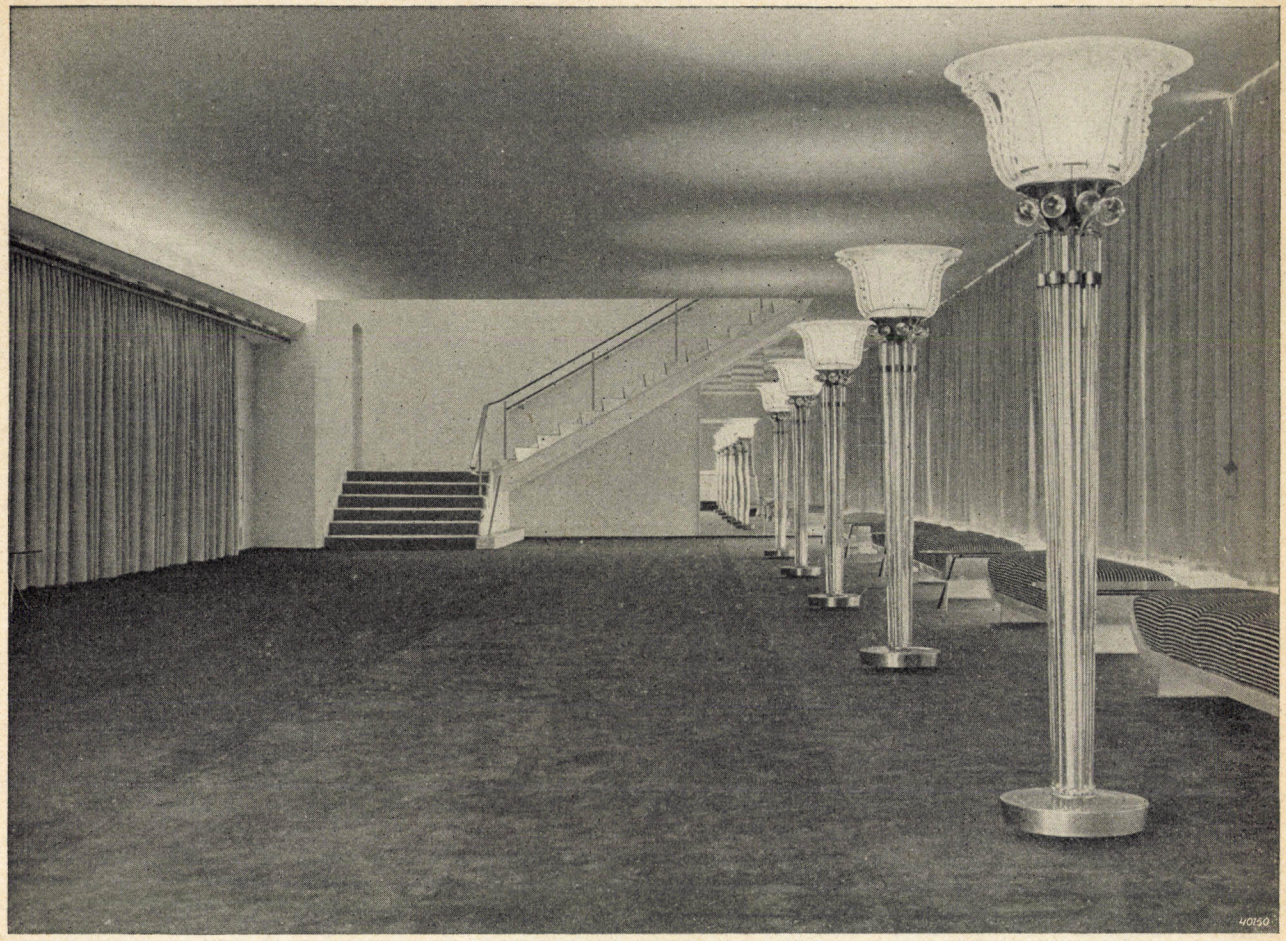


Fig. 9. Through wide doorways the foyer for non-smokers can be entered directly from the auditorium. Here there is a unilaterial indirect lighting installed along the side nearest the auditorium, while five monumental glass lamps by C o p i e r of Leerdam, in addition to the indirect lighting *via* the ceiling, give the necessary direct light which is indispensable for a festive atmosphere in such a hall.

died, it first strikes one that no space has been wasted in corridors, and that due to the fact that all the rooms open directly into one another, each room has a well considered relation to the succeeding one. The low-ceilinged cloak-room (*fig. 7*) has an indirect lighting which indeed makes the ceiling less impending but, nevertheless, emphasizes its horizontal chracter. A pleasant contrast is obtained by the vertical line of the lighting of the mirrors. The cloak-rooms proper can be closed with curtains which are lighted by a long row of holes in the under side of the light cove. This makes this part of the building a pleasant promenade during the intermission.

Passing through the cloak-room, up a short stairway one enters the lofty hall preceeding the auditorium (see *fig. 8*), which forms a surprising climax. The ampitheatre of the auditorium, the restaurant, and the foyer are connected with the hall by means of freely hanging staircases, so that this hall is as it were the traffic hub of the building. The simple indirect lighting enhances the appreciation of the more studied light decorations in the adjoining rooms.

In the first place we must mention the foyer for non-smokers (*fig. 9*) with the beautiful standard lamps by C o p i e r of Leerdam. Above this is the smokers' foyer and adjoining it the restaurant, which is also accessible directly from the street for the daily clientele (*fig. 10*). In this hall, besides the clever lighting with high glass ornaments in small domes, the beautiful lighting of the curtains is striking, which when it is dark outside and the lovely view of the Dom can no longer be enjoyed, forms a vital substitute for it.

The lighting of all these rooms separately formed only a part of the project. It will be clear that because there is such direct connection between them, the lighting of the different rooms, as far as system, colour and intensity are concerned, must be harmonious, while at the same time the necessary variety must also

be provided. Thus in the room where the public itself is the chief interest, as in the café and the foyers, there is a more generous use of direct light, while in the theatre the lighting serves primarily to accentuate the hall as back-ground and frame.

It will be understood that all these light effects necessitated years of collaboration between the architect and our lighting technologists. Many experiments were tried and many details designed to achieve the desired results. This collaboration has always given us the greatest satisfaction.

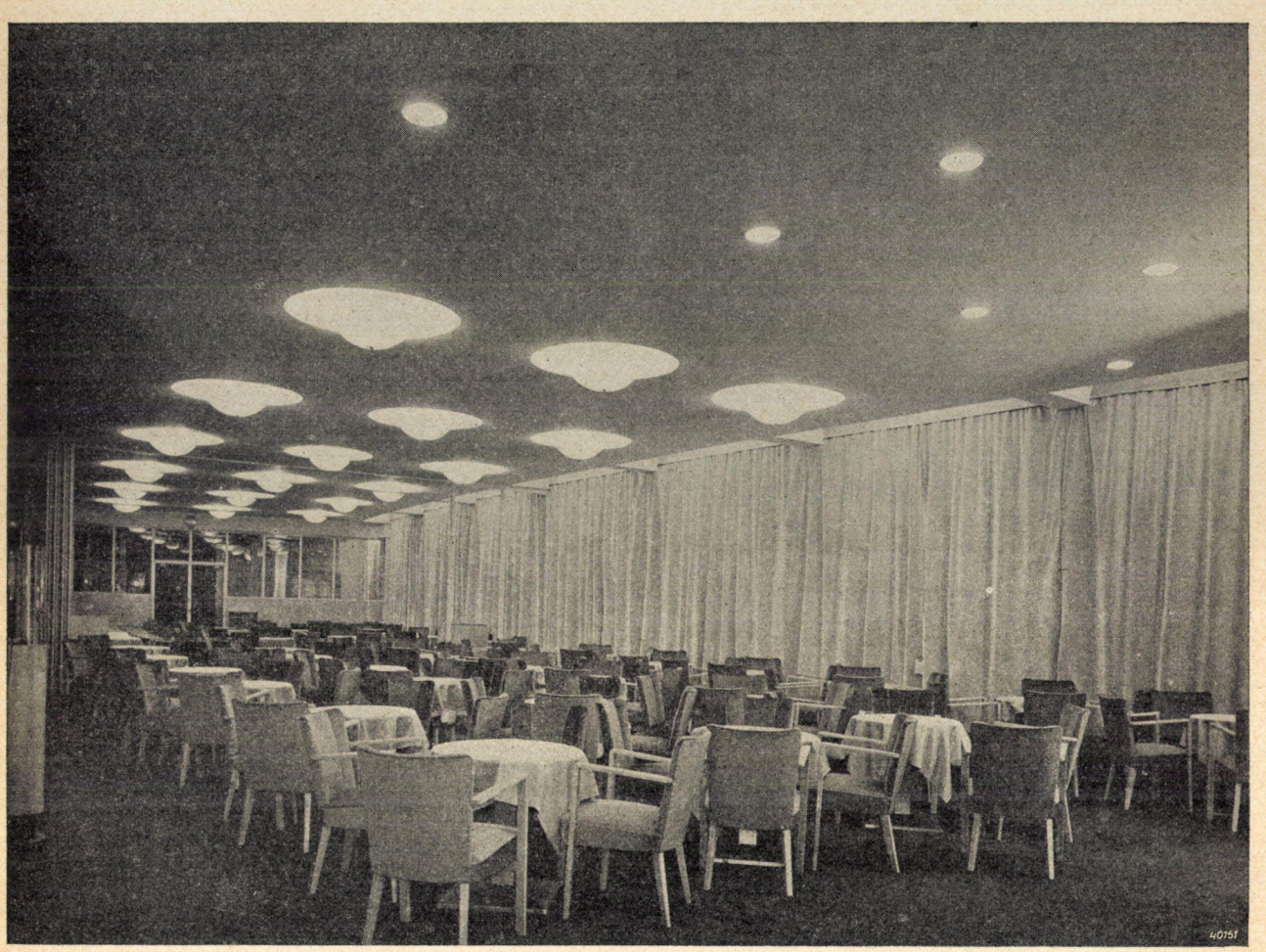


Fig. 10. In addition to the lighting by three rows of half indirect glass ornaments in white plaster domes the restaurant also possesses a very fine curtain illumination which gives to this not very large room a certain grandeur which is very pleasing.

In the text under the figures we have given various practical details and mentioned certain experience gained which seemed to us to be of importance to the technologist.

THE ACOUSTICS OF THE AUDITORIUM OF THE NEW THEATRE IN UTRECHT

by R. VERMEULEN.

534.84

In this article the influence is discussed of acoustic considerations in the design of the theatre auditorium in the Municipal Theatre which has recently been opened in Utrecht. At frequencies in the neighbourhood of 500 c/s a reverberation time of approximately 1.2 sec was measured, which agrees with what was expected theoretically.

In conjunction with the foregoing article on the illumination of the Utrecht Municipal Theatre, we shall here discuss the extent to which the requirement of good acoustics has affected the design of the auditorium, a subject about which we were permitted to exchange opinions with the architect W. M. Dudok at an early stage of the planning. These considerations constitute a constructive application of the general principles of the acoustics of auditoria about which a series of articles has already appeared in this periodical[1]).

The shape of the auditorium

The ideal situation would be that, in which every member of the audience receives the words directly from the lips of the speaker on the stage. In open-air gatherings where there are no reflecting walls experience has, however, shown that this is only possible over a very limited distance. This distance is found to become much greater when the speaker is placed on a platform high above the public. The sound waves experience great attenuation as they pass over the heads of the hearers, since the wave length for audible sound is of just the same order of magnitude as the dimensions of the heads. The sound waves are therefore refracted toward the spaces between the hearers and are there strongly absorbed. It is thus easy to understand that the sound is much more attenuated at some distance from the speaker, than would follow from the square law for the propagation of sound in space.

The solution of this difficulty for an open-air gathering, namely of placing the speaker high above his audience, cannot be applied in a theatre auditorium, since in that case the stage could not well be seen. It is, however, possible apparently to provide the speaker with this favourable position by causing his voice to be reflected at the ceiling. This reflecting surface should then be given such a form that the sound is distributed as fairly as possible over the whole audience.

If in the different designs for the shape of the ceiling the path of the sound waves is now drawn in, it is quickly noted that it is undesirable to have the ceiling slope upward from front to rear, which would be architecturally, the logical method of making room for the balcony. The sound in that case would be thrown mainly into the back of the hall, which of itself would be an advantage if it did not occur at the expense of the middle section. As in many cases, it was also found that in the design for the Municipal Theatre a horizontal ceiling furnished favourable results, since it provides the middle of the hall as well as the balcony with adequate sound.

In *fig. 1* the path of the sound rays is drawn in the main cross section of the auditorium, with the angles between successive rays being taken equal to 5°. The beams lying between these rays for directions of sound which do not deviate too greatly from the horizontal, will then all contain equal acoustic power and the intensity is thus inversely proportional to the surface upon which a given beam is incident.

It is obvious that the front rows receive adequate sound, not only because of the short distance, but also because of the fact that the position of the speaker is relatively high above them. For the sound which is reflected by the front part of the ceiling and which if the ceiling were horizontal would be thrown on the front rows of seats, a more useful employment can be found by giving this part of the ceiling a certain slope. The space under the balcony deserves first consideration in this matter since it receives nothing from the horizontal part of the ceiling. Since it is only the slope and not the position of the reflector which is under control, it was only possible to direct a small part of the sound into this space, the remainder is received by the audience on the balcony. In order, nevertheless, to capture as much sound as possible under the balcony the opening of this space was made as high as was compatible with other requirements, while the lower side of the balcony was given such a shape, that the incoming sound was conducted downwards upon the audience. It is scarcely permissible to speak of a reflection in the case of such an almost tangential incidence, since it is here actually the refraction of the sound which predominates.

The corner between rear wall and ceiling

[1]) Philips techn. Rev. 3, 65, 139 and 363, 1938.

always forms a dangerous point, since when the surfaces are mutually perpendicular the rays are there reflected back to the stage parallel to their path of incidence[2]). In general, it is true, it is desirable that the players receive sound back from the hall in order that they should not obtain the impression that the hall is difficult to play to. This reflection, however, was unsuitable for the purpose because the long distance (2 × 26 m) which the waves must cover would produce such a retardation that the sound would be heard as an echo. Therefore at the rear of the ceiling there is a section with a slope such that the sound falling upon it is reflected to the balcony.

In the case of the other rays also, care must be taken that no echo may be formed due to too great time differences. In the front of the hall the time difference is only 13/330, *i.e.* about $^1/_{20}$ sec, which is quite permissible.

While in the case of the longitudinal cross section the form of the ceiling was for the most part determined by the acoustic requirements, in the design of the transverse cross section the architect may allow other considerations, for example those connected with illuminating technology, to be the determining factors. Many variations are possible since the only condition to be fulfilled is that the sound shall be uniformly distributed over the audience and that it shall not be thrown against the side walls, since otherwise it covers longer distances and a longer reverberation results.

In the case of the form chosen by the architect a slight curvature of the ceiling concentrates

[2]) Compare also the function of a mirror, which has been described in Philips techn. Rev. 5, 335, 1940.

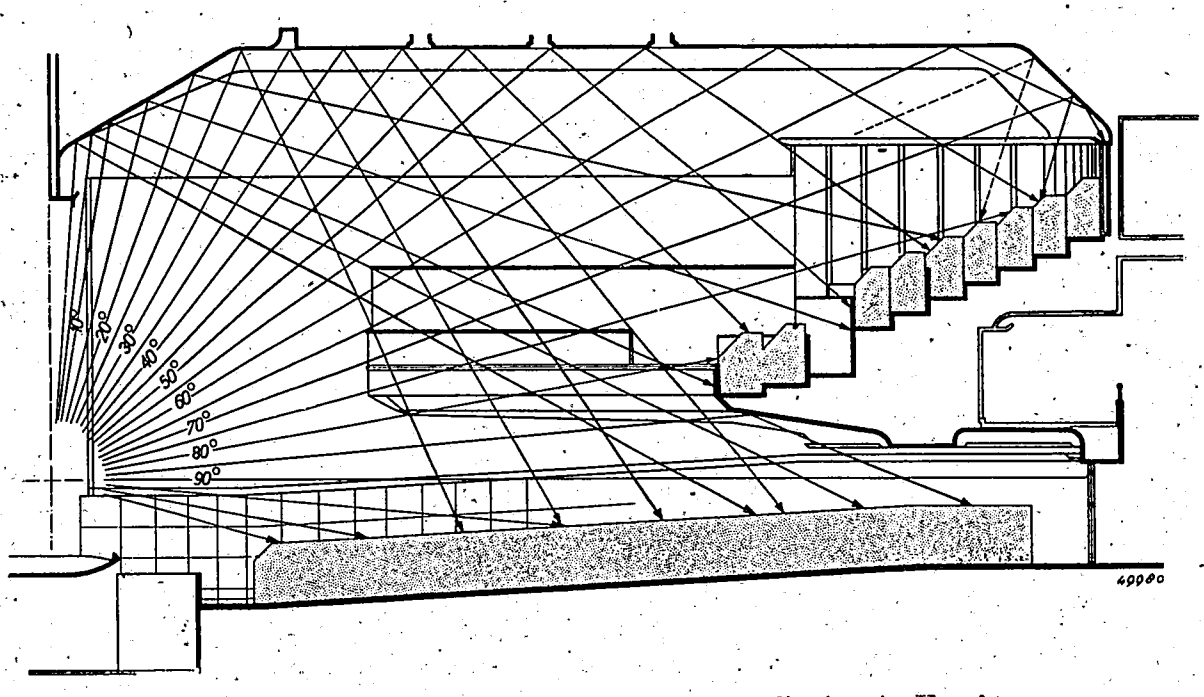


Fig. 1. Longitudinal cross section of the theatre auditorium in Utrecht.

The path of the rays of sound is drawn for sound from a source placed in the opening of the stage wall. From the surface covered by the beam an impression can be obtained of the intensity to be expected. The ray indicated by a broken line, which is only added to make the path of the rays clearer, must be left out of consideration in estimating the distribution of intensity. The space occupied by the audience is shaded. The beam between

105° and 100° falls upon the first 2 m of audience,
100° and 95° falls upon the following 4-5 m of audience,
95° and 90° falls upon the remaining 11 m of audience,
90° and 85° falls upon the underside of the balcony and thus reaches the back rows,
85° and 80° falls upon the front edge of the balcony and is fairly diffusely reflected,
80° and 75° as direct sound, reaches almost the entire balcony,
75° and 60° falls, partly *via* the ceiling, into the back of the balcony,
60° and 45° serves the front of the balcony *via* the ceiling,
40° and 25° falls upon 7-20 m in the middle of the auditorium,
25° and 10° is reflected to the balcony by the oblique part of the ceiling,
10° and 5° reaches the audience under the balcony.

the greater part of the sound on the audience, while the cutting off of the corners directs the sound which would otherwise reach the side walls to the outermost rows of seats (*fig. 2*).

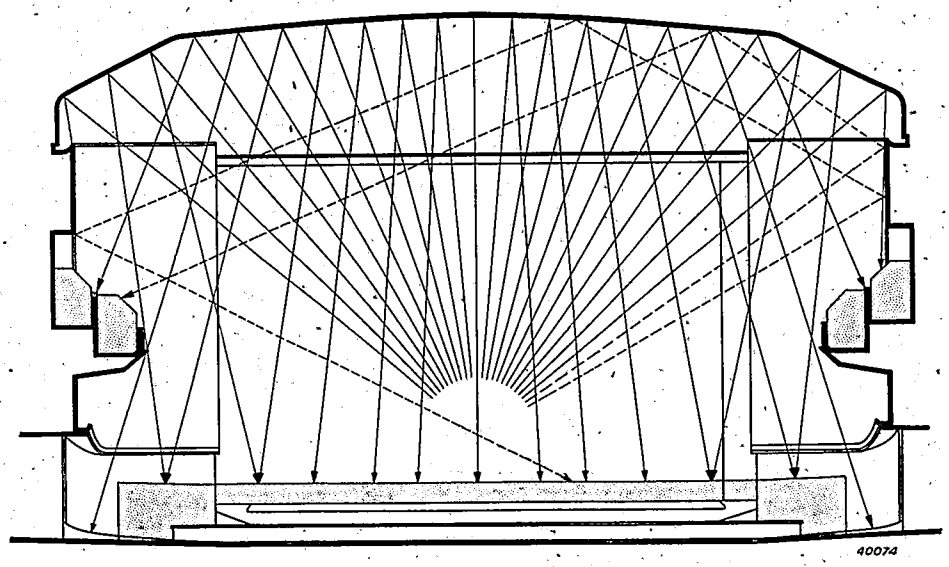


Fig. 2. Transverse cross section of the theatre auditorium in Utrecht.
The source of sound is considered to be in the same place as in fig. 1, and the figure shows the reflection of the sound rays by the ceiling. The vaulted part gives a slightly spreading beam upon the audience. The rays which would fall upon the walls after reflection are reflected into the auditorium by the sloping parts of the ceiling.
The path of a few rays which fall directly on the side walls is also shown by dotted lines; these furnish no effective contribution, but will go to make up the reverberation. For the rays drawn the intensity provided can be approximately estimated as in fig. 1. This is not the case for the rays not drawn which are directed immediately toward the audience. In fig. 1 the rays travel in the plane of the drawing, in fig. 2, however, they are projections; the source of sound lies far behind, and the receptive audience in front of the plane whereas the rays are reflected at the walls. Therefore beams of rays with equal intensity would have to form very different angles in the drawing.

These beams of sound will be of especial advantage when the speaker does not stand exactly in the centre of the stage and the sound after reflection at the middle section of the ceiling would just miss reaching the rows of seats on the same side as the speaker.

Reverberation time

In addition to the distribution of the sound which has been reflected once and which provides the chief contribution to intelligibility, the remaining sound, the reverberation, must also still be considered: on the one hand the reverberation time, especially in the low tones, must not be too long, since otherwise, as for instance in a church, the intelligibility decreases, and on the other hand it may not be too short since this produces an uncomfortable feeling and music suffers in quality. Although a good theatre auditorium will never be at the same time an ideal concert hall, the requirement must nevertheless be made that it may also be used for music. The immediate direction of the sound on the audience, necessary for a theatre auditorium, makes music sound dry and sharp, since due to their long wave length the lower tones a e much less easily directed. It is therefore desirable that the reverberation should here provide an amplification and should last longer for low tones than for the intermediate region of tones. Since the absorption of most materials increases with the frequency this requirement is automatically fulfilled. The increase in absorption with increasing frequency is even greater than would be expected, so that it still remains necessary to take measures for the sufficient absorption of the bass tones. The plan of the architect of covering the part of the walls within arm's reach from the floor with padded artificial leather seemed to be extraordinarily suitable for this purpose. This covering is not porous and will not therefore absorb high frequencies to any extent. For slower vibrations, however, with the layer of air behind it, it causes the occurrence of a resonance which is damped by the padding.

In general the average value of the reverbera-

tion time in a theatre auditorium gives little cause for anxiety if only the height of the hall is correctly chosen. The absorption of the sound takes place chiefly by the audience, whose absorption coefficient closely approaches 100 per cent. In the plan for the Municipal Theatre the surface area of the balcony was somewhat less than one half of that of the floor of the hall. If we represent this latter quantity by the letter S the height by H, then the volume of the hall is $V = H \cdot S$. For the absorbing surface of the audience we calculate with the value S in the hall and $0.4\ S$ on the balcony, and add another $0.2\ S$ for the absorption by the walls, etc. The desired reverberation time was estimated to be 1.2 sec on the basis of various data. Then from the wellknown formula of Sabine for the reverberation time:

$$T = 0.16\frac{V}{A},$$

where

V = volume of the hall in m³,

A = absorption of the hall in m² of equivalent open window surface area we find the following value for the height of the hall:

$$H = \frac{V}{S} = \frac{TA}{0.16\,S} = \frac{1.2}{0.16\,S}\,(S + 0.4\,S + 0.2\,S) = 12 \text{ m},$$

which was indeed chosen by the architect.

In a consideration of the different parts which contribute to the absorption it strikes one immediately that the audience furnished by far the greatest share. This has the disadvantage that an incompletely filled house would have a considerably longer reverberation time, and an empty house would even have eight times as long a reverberation time, *i.e.* 9.6 sec. !

Increasing the absorption of the walls would of course make the influence of the audience relatively smaller, but at the same time it would very much decrease the reverberation time unless the hall were made much larger, which would not only be an expensive solution but one which would be unsuitable from the point of view of intelligibility. A much better solution is to upholster the seats in such a manner that they can take over the function of the absent audience, but are relieved of their absorbent function by the audience which is present.

The measurement of the reverberation for the empty auditorium already gave the desired value of 1.2 sec, from which the conclusion may be drawn that the chairs actually do perform the part intended for them satisfactorily. The variation with frequency is also satisfactory (*fig. 3*): the increase for the low tones, desirable for music, is found by no means to give the hall an echoing or „hollow" character. It must still be noted that this frequency characteristic is only valid for the reverberation, which results from repeated reflection at the walls. The contributions of importance for intelligibility reach the audience after a single reflection at walls which absorb only a few per. cent. The variation of this with the frequency is therefore unable to influence the sound appreciably.

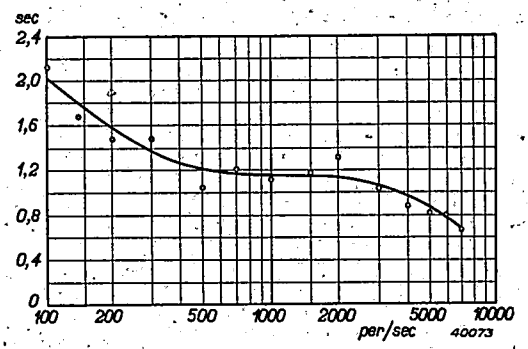


Fig. 3. Reverberation time of the theatre auditorium in Utrecht as a function of the frequency. The reverberation time was measured with the source on the stage, where a normal scene had been set up. The measurement was carried out for each frequency at three points in the auditorium and one on the balcony. The reverberation time at 200-2000 c/s amounts to 1.2 sec. and increases at low frequencies to 1.0 sec for 100 c/s; in the high tones the reverberation time decreases to 0.7 sec. for 7 000 c/s.

CONSIDERATIONS ON THE TEXTURES OF METALS

by J. F. H. CUSTERS.

62.918

When in a polycrystalline material the crystallographic axis directions of the crystal grains are not entirely at random the material is said to possess a texture. In many practical cases the texture of the material is of great importance. It is our intention to discuss in this periodical a succession of examples in which this is the case. As introduction it is explained in the following article how textures can be represented graphically (stereographic projections of the so-called reference sphere) and how the texture of a given specimen can be determined experimentally by means of X-ray photographs.

It is a well known characteristic of a crystal that its physical and chemical behaviour is in general different in different directions. When for example the tensile strength of a block of copper consisting of a single crystal is measured, it will be found that in the direction of a r i b of the cubic crystal lattice it is less than half that measured in the direction of one of the s p a c e d i a g o n a l s of the cubic lattice.

In technology such metals in the form of single crystals only play a subordinate part, because they possess properties which are usually undesired from the technical point of view. The p o l y c r y s t a l l i n e state is of much greater importance, where the metal is built up of a collection of larger or smaller single crystals or crystal grains.

Fig. 1 shows the surface of a polycrystalline metal as it appears after being etched with a suitable reagent. The various grains can here be distinguished clearly because, after having been attacked by the reagent, they reflect the light differently. Each of the grains is a single crystal. They are, however, usually not bounded by crystal planes as is often the case with single crystals encountered in nature, which commonly possess a symmetrical external shape; the boundaries in this case are rather determined by the more or less accidental meeting of grains growing from different points, which are later perhaps deformed, due, for instance to mechanical treatment.

Fig. 1. Photomicrograph of polished and etched *a*-brass (magnified 100 times). The polycrystalline structure is here clearly visible because of the fact that the different crystal grains reflect the light in different ways.

It is important to note that each of these grains has its own crystallographic axis direction and that these directions, due to the varied positions of the grains, are in general different from grain to grain as represented schematically in *fig. 2a*. It may, however, also occur that each crystal grain occupies a definite, more or less sharply distinguished, preferred position, as far as the position of its crystal axes is concerned, as represented in fig. 2*b*. In this case the polycrystalline material is said to exhibit a t e x t u r e. A material thus has a texture as soon as the grains no longer occupy absolutely irregular positions, as is assumed in fig. 2*a*.

The examples of figs. 2*a* and *b*, represent the two extreme cases of orientation: in the first case no texture, in the other case a texture of such a nature and completeness that the material may be said to be a pseudo single crystal. In addition to these limiting cases all types of textures can be found. As an example we may mention the texture of a drawn aluminium wire. When, beginning with a polycrystalline wire with non-orientated crystal positions, the wire is strongly stretched, the particles take up such positions that in each grain a space diagonal of the (cubic) crystal lattice lies in the direction of stretch, while for the rest the orientation still remains random (fig. 2*c*).

In general in a polycrystalline metal the above mentioned dependence on direction of the different properties of a crystal will not be noticed if the crystalline particles are non-oriented. Due to the random position of the particles the crystal properties in each direction of the polycrystalline metal constitute a sort of average over all

directions; the material thus behaves as an isotrope as long as the observations involve the collaboration of a sufficient number of particles. If on the contrary the material exhibits a texture, it may be expected that a dependence of the properties on direction will be observable. A typical example of this which has already been mentioned in this periodical[1]) is reproduced in *fig. 3* : upon drawing cups of chrome-iron sheet very different results were obtained according to whether or not the original material exhibited a texture.

In judging a material for a given use all kinds of other properties in addition to the texture naturally also play a part. For example, the dimensions of the crystal grains (fine or coarse) may also be of decisive importance. The presence or absence of stresses in the crystal particles may also be important. In general therefore a whole complex of properties should be studied in the case of a metal intended for a given purpose. We shall here, however, direct our attention especially to questions connected with the texture. For that purpose we shall in this article discuss the way in which textures are investigated and described, while it is our intention to discuss a series of examples in the following numbers of this periodical. The relation which is found between the texture and other properties of metal will there also be emphasized as far as possible.

[1]) W. G. Burgers, Philips techn. Rev. **2**, 156, 1937.

Description of a texture

Let us begin with the methods which are used for describing a texture. We consider a polycrystalline piece of metal in which there are hundreds of crystal grains. Let the crystal lattice be of the cubic type. In each single crystal the atoms which occupy the lattice points are arranged in a definite way so that planes can be distinguished which are regularly filled with lattice points. In *fig. 4* several of these planes are indicated. These are the lattice planes at which X-rays can be „reflected" (see below). If we confine our attention to the cube planes (fig. 4*b*), three mutually perpendicular sets may be distinguished. It is obvious that the position of a crystal grain is completely defined when the position of its cube planes is known, while of course the directions of the normals to these planes also fully determine the position of the crystal particles.

Returning to the piece of metal mentioned, a sphere is drawn about it, the so-called reference sphere whose radius is large compared with the

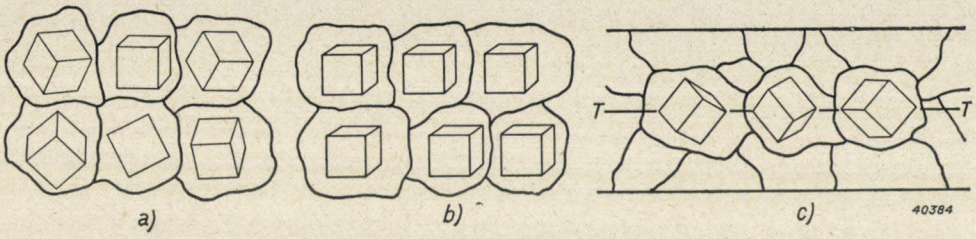


Fig. 2. Position of the crystal axes in the particles of a polycrystalline material with a cubic crystal lattice.
a) No texture: the particles lie with their axes entirely at random.
b) Sharp texture: the axes of all the particles are in the same position.
c) Texture of an aluminium wire drawn in the direction *TT* : in each particle a space diagonal of the cubic lattice is parallel to *TT*.

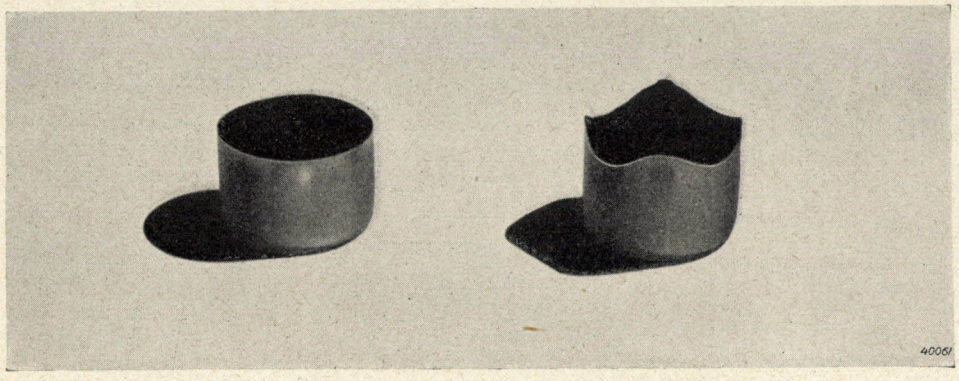


Fig. 3. Cups made of chrome-iron sheet, obtained by drawing. In the case of the left-hand cup the original material had no texture, in the case of the right-hand cup a texture was present.

dimensions of the piece of metal. The position of the three cube planes of each particle is investigated; the three normals to these planes are drawn through the centre of the sphere and extended until they pierce the sphere as shown in *fig. 5* for a single grain[2]). This is done for

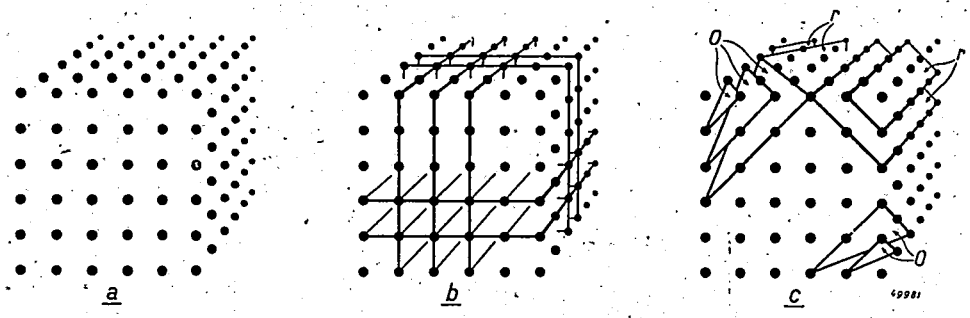


Fig. 4. *a*) Diagrammatic representation of a cubic lattice. *b*) The three sets of cube planes. *c*) Octahedron planes *o* and rhombododecahedron planes *r*. (After M. Polanyi).

every grain. In this way the surface of the sphere becomes filled with points. The density of the points on the sphere is the same at all points when the piece of metal exhibits no texture. In contrast to this, the density will be different from point to point when the crystal

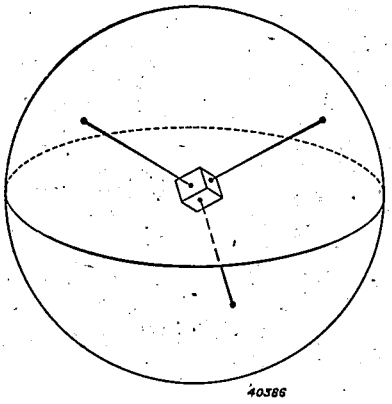


Fig. 5. Reference sphere drawn around a single crystal. The normals to the three cube faces passing through the centre of the sphere are indicated. These normals pierce the surface of the sphere at three points which are called the poles of the cube planes.

grains are more or less oriented. In *fig. 6b* this is shown for the texture of fig. 2*b*: at only three spots, which are more or less extended according to the sharpness of the texture, is the surface of the sphere „blackened". The texture of the abovementioned aluminium wire (fig. 2*c*) is represented in the same way in *fig. 7b*.

Use may be made not only of the cube planes but also of the so-called octahedron or rhombododecahedron planes[3]) (fig. 4*c*) to describe a texture. In the case of cubic crystals one usually confines oneself to the cube and octahedron planes.

In order to draw and measure a texture easily, it is desirable to represent the reference sphere just considered on a plane by means of a projection. The ordinary method is that of stereographic projection, which is explained in *fig. 8*. A plane is drawn through the centre of the reference sphere, the projection plane *Pr*. The normal to this plane through the centre *C* pierces the sphere at the „North" and „South" poles, *N* and *Z*, respectively[4]). Points lying on the surface of the upper hemisphere are now projected upon *Pr* by connecting them with the south pole and finding where the connecting lines cut the plane *Pr* (*S'* is the projection or image of *S*), while the points lying on the lower hemisphere are projected by connecting them with the north pole and marking the points of intersection of them with the north pole and marking the points of intersection of the connecting lines and the plane *Pr* (*T'* is the image of *T*). In this way all the projected points are made to lie within the so-called basic circle *G*. Points on the upper and lower hemispheres can be distinguished in the projection by marking them blue and red, respectively, for example. In practice this will usually be unnecessary because the texture of the material almost always possesses a plane of symmetry; this plane is then chosen as projection plane so that the projections in blue and red then have exactly the same appearance and one of them is sufficient.

[2]) Actually two normals are enough to determine the position of the crystal grain, but for the sake of the symmetry all three are always considered.

[3]) The planes are so called because upon reflection at the different planes of symmetry of the lattice, an octahedron or a rhombododecahedron, respectively, results from one such plane.

[4]) From this nomenclature it is evident that it is here a question of a procedure whose oldest and most important field of application is in geography in representations of the earth.

The pattern which is obtained when the points of intersection on the sphere of the normals corresponding to a given type of lattice plane are projected is called the p o l e f i g u r e for the corresponding plane. From two such pole figures, for example, that for a cube and that for an octahedron plane, a good idea of a texture can be obtained. In fig. 6*c* and 7*c* the pole figures are drawn for the cube plane of the two textures of fig. 2*b* and *c*.

How is the texture of a given material now determined? The method which is generally used for by this makes use of the diffraction of X-rays by crystals.

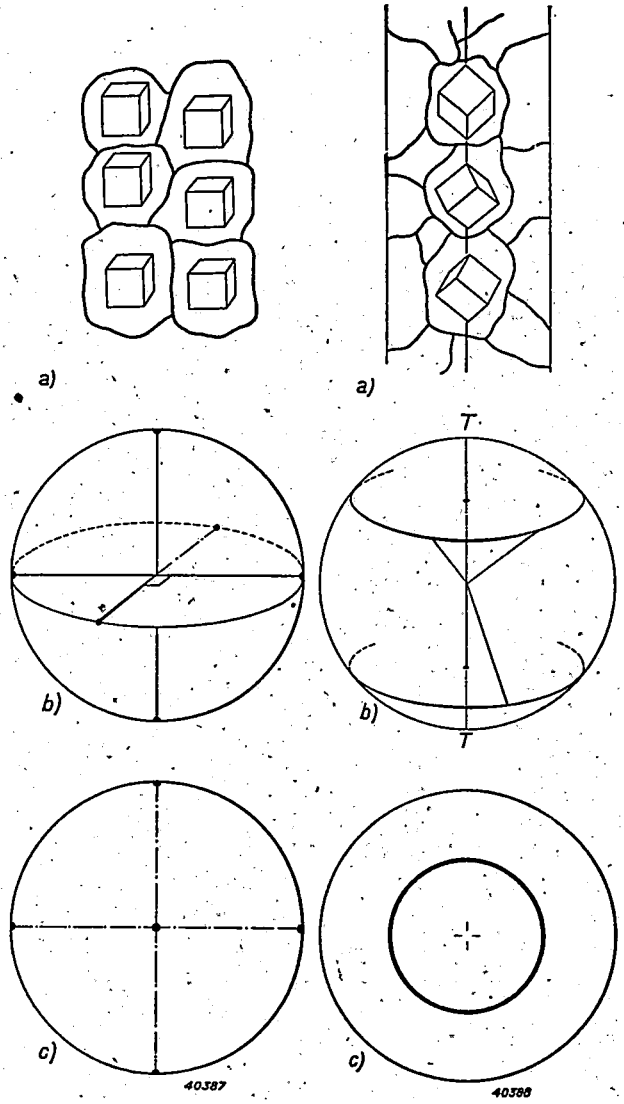


Fig. 6. Fig. 7.

Fig. 6. The texture of fig. 2*b* shown again in *a*) is represented on the reference sphere. The reference sphere is blackened at only three poles (and their three „counter poles"). In *c*) the projected pole figure is given (see below).

Fig. 7. Like fig. 6, for the texture drawn in fig. 2*c*. The reference sphere is blackened in two circles (parallel and counter parallel circle); they indicate the geometrical position of the points of intersection of the normals to the three cube planes in each particle.

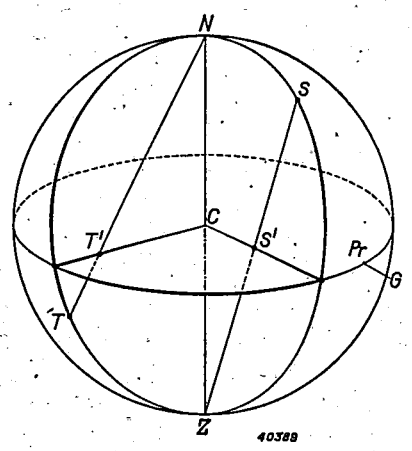


Fig. 8. Stereographic projection on the plane *Pr* of points lying on the reference sphere. *S* is the image of *S*, *T'* that of *T*.

X-ray determination of a texture

When X-rays of a given wave length λ (monochromatic radiation) fall upon a crystal, no diffraction will in general be observed, since the rays refracted at the lattice points cancel each other by interference. Only when B r a g g's condition:

$$n\,\lambda = 2\,d \sin \Theta,$$

is satisfied, does the interference of the rays deflected in certain directions lead, not to cancelation, but to mutual reinforcement. In this equation d is the distance between some set or other of parallel lattice planes. Θ the angle at which the beam of X-rays strikes the lattice plane (thus $90° - \Theta$ is the angle between the beam and the normal to the plane), and n is a whole number. The direction in which the deflected rays reinforce each other and which can therefore be determined experimentally, can now be described for a given lattice plane as if the X-rays were reflected against that lattice plane as against a mirror. With given values of d and λ this reflection, according to B r a g g's condition, can occur at various angles of incidence Θ which correspond to the values $n = 1, 2, 3, \ldots$ One then speaks of reflections of the 1st, 2nd, 3rd ... order. (The physical significance of n is then the number of wave lengths difference in path between X-rays which are reflected by successive lattice planes).

Example: with the help of an X-ray tube with copper anode, using suitable filters, a strong, practically monochromatic X-radiation can be obtained with the wave length $\lambda = 1.539$ Å (Cu Kα radiation). If this beam is allowed to fall upon a crystal of aluminium of which a first order reflection of the octahedron plane is desired (the lattice plane distance is $d = 2.333$ Å in this case), the beam of X-rays must strike that plane at an angle of 19°16'.

In *fig. 9* let *V* be the representative of some set of mutually parallel lattice planes in a

crystal grain. V is assumed to be in such a position relative to the incident X-ray beam that Bragg's condition is satisfied. Then the normal n to V includes an angle $90°-\Theta$ with S. It is clear that at every position of V in which n lies on the surface of the cone indicated in fig. 9 (apex O, axis S, semi-apex angle $90°-\Theta$), reflection of the X-ray beam will occur. We now draw a sphere around O as a centre which can be used as the above described reference sphere. The cone mentioned cuts the sphere along the so-called reflection circle R, and on this circle therefore the point of intersection (pole) of the normal to the surface V must always lie if that surface is to cause a reflection of a given order of the X-ray beam S. In *fig*

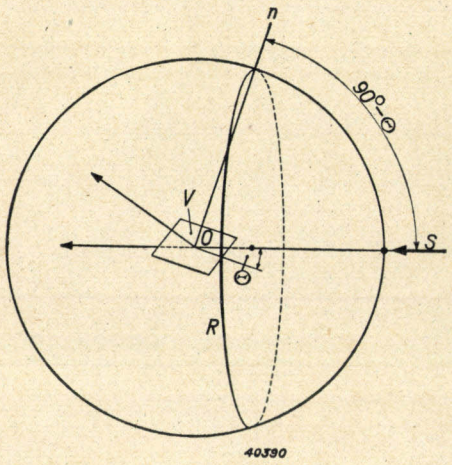


Fig. 9. A beam of X-rays S is reflected at a lattice plane V of a crystal only when the angle between S and the normal n to V is equal to $90°-\Theta$ (Bragg's condition). Therefore for each position of V in which n lies on the surface of a cone with the semi-apex angle $90°-\Theta$ will there be reflexion. The surface of this cone cuts the reference sphere drawn around the crystal in a circle, the so-called reflection circle R. In the case of reflection, therefore, the pole of the lattice plane in question must lie on this circle.

10 three positions of V are assumed which satisfy this condition (points of intersection N_1, N_2, N_3).

Let us now consider the reflected radiation. In the case of fig. 10 it will cause three spots (P_1, P_2, P_3) on a photographic film F placed behind the crystal; these spots may be considered as the „images" of the positions V_1, V_1, V_3 of the lattice plane. It is clear that the angles α_1, α_2, α_3, which the projections (on a plane perpendicular to S) of the normals n_1, n_2, n_3 inclose, are the same as the corresponding angles between P_1, P_2, P_3 on F; in other words in this projection the angles are retained.

At the point O we now place a fragment of polycrystalline metal with a cubic lattice, and we again consider a single type of lattice plane of the single crystal grains, for instance the cube planes. In the absence of a texture in the fragment the reference sphere will be uniformly covered with the poles of those planes, with the result that a complete circle of „images" is

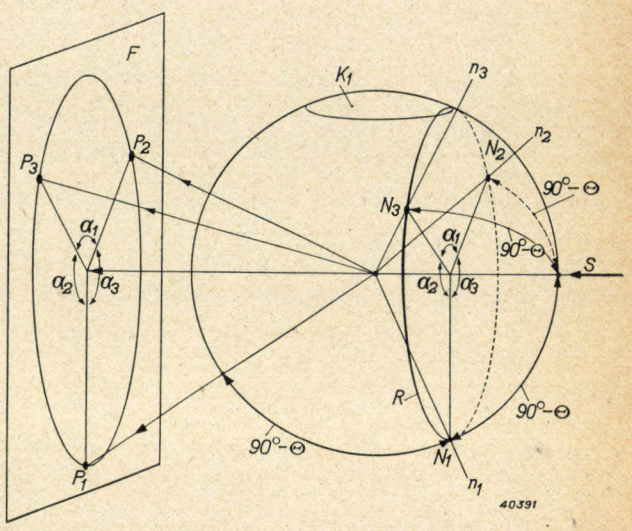


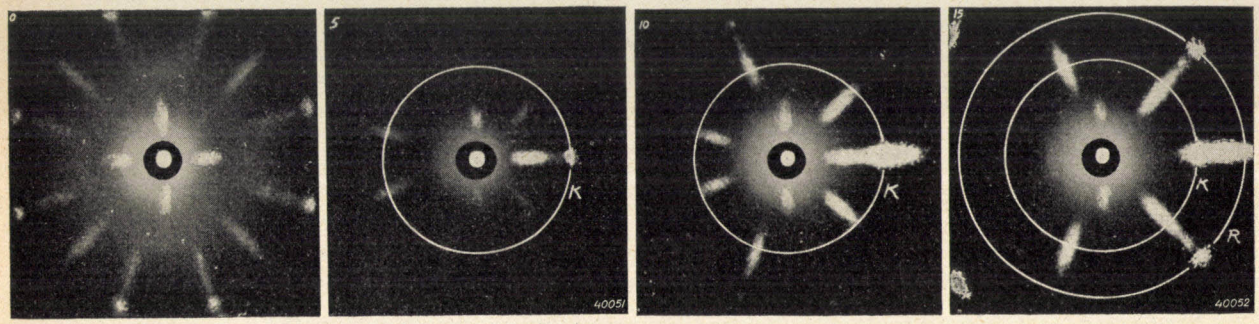
Fig. 10. The points P_1, P_2, P_3 on the photographic film F are the images of three positions of the lattice plane at which reflection takes place. The normals n_1, n_2, n_3 to the lattice plane in the three positions pierce the reference sphere at the poles N_1, N_2, N_3. From this it is easily deduced that the „blackened" parts of the reflection circle R are projected on F in such a way that the angles (α_1, α_2, α_3) are retained.

formed, a so-called Debye-Scherrer ring, whose radius is determined not only by the wave length λ and the distance between the lattice planes d but also by the distance at which the film is situated from the crystal fragment. In *fig. 11* an X-ray photograph may be seen with a number of such quite uniformly blackened circles. If on the contrary the crystal fragment does possess a texture, so that only certain parts of the reference sphere are „blackened", then in general the Debye-Scherrer

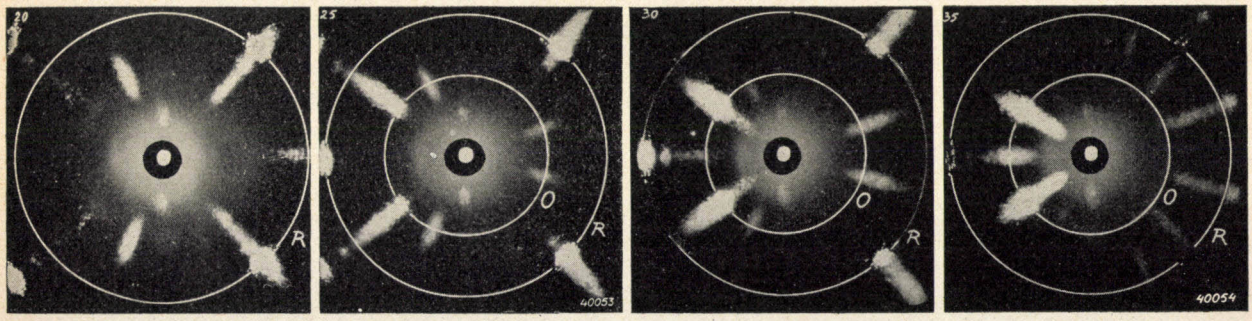
Fig. 11. X-ray photograph of a substance in powder form which naturally possesses no texture and therefor gives entirely uniformly blackened Debije-Scherrer rings.

ring on the film is not uniformly blackened, but only those segments of it which are the image of the parts of the reflection circle where this circle passes through ,,blackened'' regions on the reference sphere.

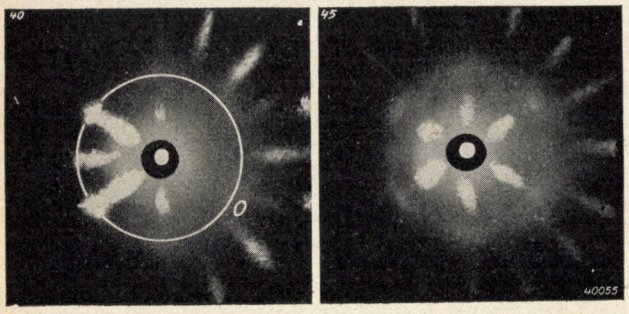
When one has thus obtained such an X-ray diffraction photograph of a piece of metal, then by measuring the angles within which a given Debye-Scherrer ring is blackened it can immediately be found along what part of the corresponding reflection circle R a blackening must be present on the reference sphere. The reflection circle, however, covers only a small part of the reference sphere, while for judging the texture we must know the blackening over the whole reference sphere (unless this requirement may in part be neglected because of symmetry properties of the texture). One of the means of ascertaining this would be to choose a succession of different wave lengths of the X-ray beam, so that the reflectin circle E moves parallel to itself over the reference sphere and the whole reference sphere is as it were scanned. In practice, however, this method encounters insuperable difficulties, the chief among which is the fact that one cannot have at one's disposal a set of X-ray tubes each of which furnishes radiation of a different wave length. The more obvious and commonly employed method of scanning the reference sphere is usually to give the crystal fragment in question a slightly different position each time. If the reference sphere is assumed to be rigidly connected with the crystal, while the positions



0° 5° 10° 15°



20° 25° 30° 35°



40° 45°

Fig. 12. Series of X-ray diffraction photographs of a polycrystalline substance with a definite texture. For the successive photographs the specimen has been rotated each time through an angle of 5°. The sphere of reference, blackened according to a definite pattern, thereby rotates as it were past the reflection circles for the different kinds of lattice plane. The circles on the photographs (Debij e-Scherrer rings), which are faithful images of the reflection circles, — K for the cubic planes, O for the octahedron planes, R for the rhombododecahedron planes —, are thereby blackened at certain points. From the whole series, therefore, the pattern of blackening on the reference sphere can be derived. (In the photographs there are also a number of other, in some cases quite intense, spots, which may not be used for the construction of the reference sphere, since they are due to the unavoidable continuous spectrum which the X-ray tube produced in addition to the monochromatic radiation).

of S, R and F are constant, different parts of the sphere turn with the crystal past the reflection circle R. The crystal fragment (for instance a piece of rolled strip or the like) is usually rotated through a certain angle, 5°, for example, from a given initial position about a vertical axis passing through the centre of the sphere, and in each position a photograph is taken. In *fig. 12* is shown a series of photographs taken in this way of a certain material. By this method of procedure the whole reference sphere is scanned with the exception of caps at the top and bottom, K_1, K_2 (in fig. 10 the cap K_1 is indicated). In order to include these caps also the fragment is rotated about a different axis, for instance about the line perpendicular to the plane of the drawing.

The construction of the pole figure

In the manner described the blackening pattern on the reference sphere of the fragment investigated can be constructed for the lattice plane considered, for example the cube plane. Actually, however, as explained above, we desire to obtain a plane pole figure by projecting the reference sphere on a plane passing through its centre. As plane of projection it is preferable to choose a plane of symmetry of the object in question (at least if the object possesses such a plane), since it may be expected that this will also be a plane of symmetry of the texture. Beginning with the X-ray photographs the pole figure can now also be drawn immediately on this plane, without first following the circuitous method *via* the reference sphere. For this purpose on the projection plane for each photograph the projection is drawn of the reflection circle which corresponds to the Debye-Scherrer ring considered, and only those segments of the projected circle are „blackened" which can be recognized as blackened segments of circles of the Debije-Scherrer ring on the film. It must be kept in mind that only in one position of the object (for instance, the initial position) is the plane of symmetry, which has been chosen as plane of projection, and which is bound to the object, parallel to the plane of the reflection circle which is bound to the incident X-rays (namely perpendicular to their direction). Thus in working out the diffraction photographs taken in rotated positions of the object, one must project the successive reflection circles provided with the appropriate blackenings in each case on the projection plane which has rotated farther with the object. This apparently difficult operation is carried out in practice easily and quickly by means of a so-called Wulff net. We shall not go unnecessarily deep into the details here, however.

In this way, for every desired lattice plane of the object, the pole figure can be constructed by considering the Debije-Scherrer ring corresponding to that lattice plane. In *fig. 13* and *14* we show the pole figures thus obtained, which are derived from the series of photographs of fig. 12 for the cube plane and for the octahedron plane, respectively.

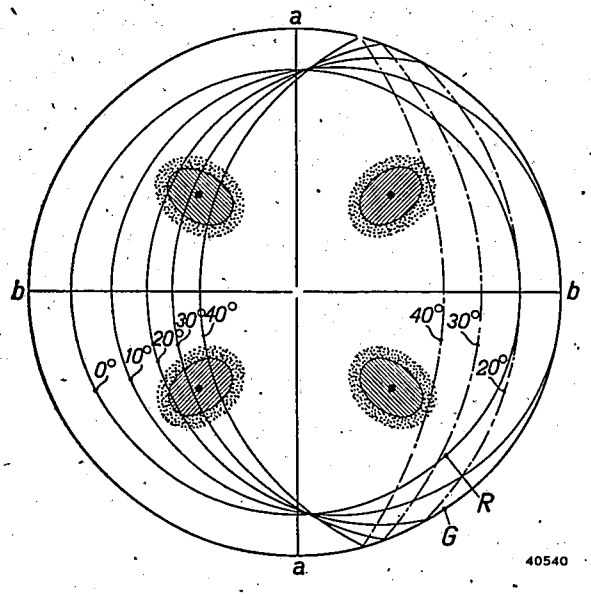


Fig. 13. Pole figure for the cube plane derived from the photographs in fig. 12. The projection of the reflection circle is indicated for a number of rotated positions of the piece of metal. It may be seen that at about 10°, it passes through the middle of the blackened region at the right-hand end of the axis *b-b*. In the photographs at 10° and 15° of fig. 12 the ring *K* exhibits an intense blackening at the corresponding point.

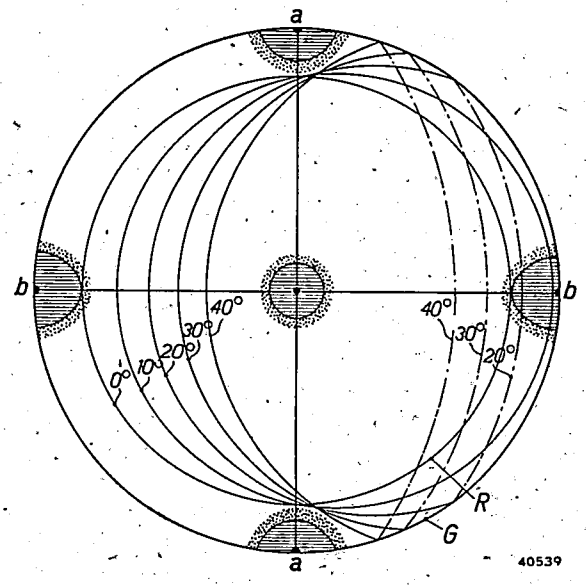


Fig. 14. Pole figure for the octahedron plane derived from the photographs of fig. 12. In the positions 30 to 40° the reflection circle passes through blackened regions which can be recognized in fig. 12 in the corresponding photographs as blackening of the corresponding segments of the ring *O*.

THE MEASUREMENT OF PEAK VOLTAGES IN A STUDIO INSTALLATION

by F. de FREMERY and J. W. G. WENKE. 621.317.726 : 621.396.712.3

A measuring apparatus is described which is used in a broadcasting studio for a continuous control of the alternating voltages which are sent to the transmitter. In connection with the purity of the broadcast these voltages must not exceed certain limits. Since there is a danger of their doing so, particularly at the peaks of the recorded sound, special provision has been made for convenient reading off of the peak voltages by means of a retardation arrangement.

The amplitudes of the alternating voltages which are generated in the electrical transmission of sound vibrations must lie between two limits: the voltages must on the one hand project far enough above the level of interferences (noise), and on the other hand the amplitude must not become so great that a disturbing non-linear deformation occurs in one of the links of the transmission.

An especially important part is played by these two conditions in the running of a broadcasting studio. In this case the aim is to secure the greatest possible freedom from interference and purity of reproduction, while the amplitude of the sound vibrations, at least when it is a question of music, exhibits great variations. It is in just these variations that the dynamic quality of the music lies. As an illustration a registrogram of the sound intensity in a concert hall during the performance of a piece of music is reproduced in *fig. 1*. In order to provide that the alternating currents which are sent from the studio to the transmitter are sufficiently free of distortion even at the p e a k s of sound, it will often be necessary to reduce the peaks artificially by decreasing in a suitable way and at the correct moment the amplification which is applied to the microphone voltages. On the other hand in the s o f t passages of the music it will be desirable to increase the amplification somewhat in order that the reception of those passages may suffer as little as possible from interferences. The regulation of the correct degrees of amplification for each part of a piece of music is the task of the technician who,seated at a control table, supervises and controls the whole broadcast.

In order to carry out this regulation the technician must know in advance the moments at which peaks may be expected as well as their intensity. In this he is aided by the score which lies before him and the experience gained at the rehearsals which precede the actual broadcast. In addition, however, the technician must have some method of checking whether the regulation he performs is having the desired effect. The desired check is obtained not only by means of a loud speaker which is connected with the line to the transmitter and with which it is possible to determine subjectively whether or not distortion occurs, but also by means of a measuring instrument which indicates continuously the amplitude af the alternating voltage on the line to the transmitter.

The principle of the measuring instrument which was used for this purpose in a studio installation previously described in this periodical[1]) is shown in *fig. 2*. It contains two amplifier stages in push-pull connection, followed by a rectifier and an output amplifier stage. The anode current of the final stage, which is a measure of the amplitude of the input A.C. voltage of the apparatus, is measured with a moving coil instrument with a light pointer

[1]) F. de Fremery and J. W. G. Wenke, Philips techn. Rev. **6**, 139, 1941.

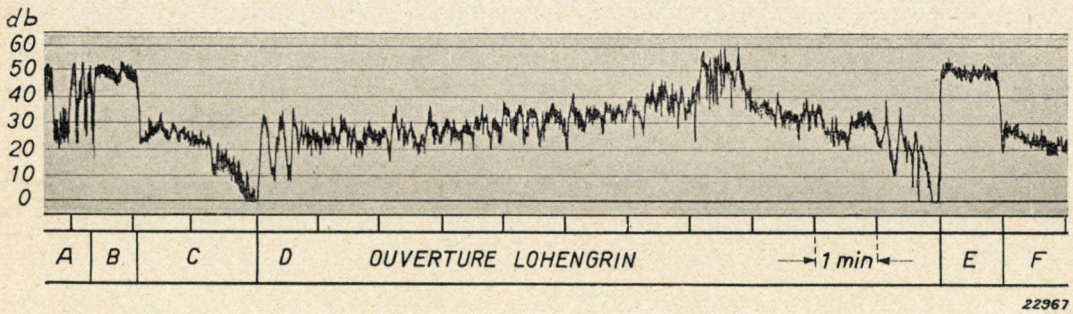


Fig. 1. Variation of the sound intensity in a concert hall during the performance of a piece of music: Overture to Lohengrin. (From: R. V e r m e u l e n, The relationship between fortissimo and pianissimo, Philips techn. Rev. **2**, 266, 1937).

which is built into the control table, see *fig. 3*. In series with this instrument a second meter is connected, which, together with the actual amplifier connection, is assembled on a panel of the amplifier racks of the studio installation, so that the broadcast can be controlled there also. The scale of the instrument is calibrated in decibels and runs from —45 to +45 dB. The value 0 dB thereby corresponds to an input A.C. voltage of 1.55 V R.M.S., which is considered as the normal, permissible zero level in broadcasting transmissions over lines with an impedance of 200 ohms (transmission power 12 mW). The part of the scale above 0 dB is coloured red as an indication that the A.C. voltages here exceed the prescribed limit and that there is therefore danger of distortion.

Upon rapid alternation of the amplitude of the input voltage it would be very fatiguing for the person at the control table to follow the light spot on the scale of the indicator as it continually jumps back and forth. For this reason a device has been employed which very much facilitates the reading off of voltage peaks in particular, with which we are chiefly concerned. The rectified signal voltage is not fed to the output amplifier valve directly, but *via* the time circuit indicated by T in fig. 2. This consists of a connection in parallel of a condenser C and a potentiometer, the latter being composed of a resistance R and a high resistance S, which we shall discuss later. If an A.C. voltage of constant amplitude is now fed to the input of the connections, the condenser C is charged to a certain voltage and a very small current flows through R and S. The grid of the output amplifier valve thereby assumes a certain negative voltage. If the amplitude of the input A.C. voltage increases, the condenser C is very quickly charged to a higher voltage and simultaneously the grid of the output valve becomes more negative, whereupon the indication of the anode current meter immediately adjusts itself to the new value. If, however, the amplitude of the input A.C. voltage now becomes smaller again, the condenser C can only adjust itself s l o w l y, by discharge over S and R, to the new, lower voltage value, so that the light pointer of the anode current meter moves back only slowly. In this way provision has been made that upon the occurrence of a voltage peak the indicator reaches practically its final value (maximum deviation 1 dB) within 0.005 sec, while it falls back at a rate of only about 20 dB per second.

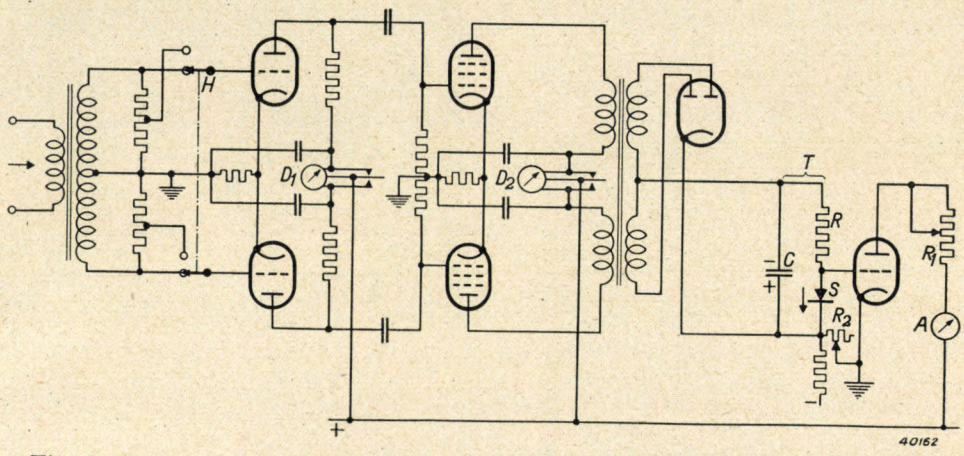


Fig. 2. Connections of the peak voltage meter. T time circuit consisting of a condenser C and two resistances R and S. For reasons explained below the leakage resistance of a blocking-layer rectifier is used for S. A ammeter calibrated directly in dB of the input signal. D_1 and D_2 differential ammeters.

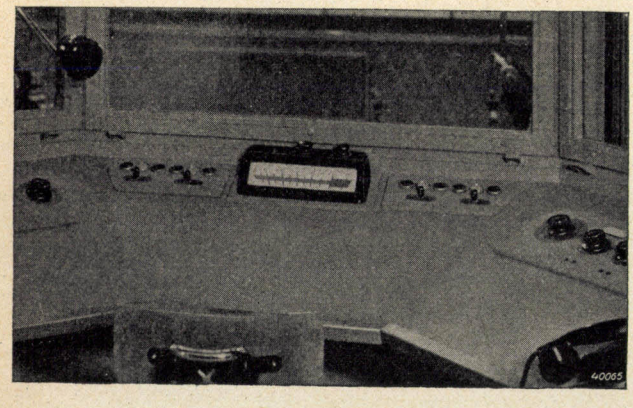


Fig. 3. Indicating instrument of the peak voltage meter built into the control table of the broadcasting studio.

The relation between the anode current of the output stage and the input A.C. voltage, *i.e.* the calibration of the scale of the indicator, depends in the first instance on the characteristics of the amplifier valves used, in particular on that of the output valve. Since with increa-

sing input A.C. voltage to the whole circuit the grid of the output valve becomes more strongly negative, with increasing input A.C. voltage the anode current will decrease. Due to the curvature of the i_a-v_g-characteristic the variation in anode current thereby becomes smaller and smaller just in the region of high input A.C. voltages in which we are especially interested, so that there as it were the scale is compressed. In order to avoid this undesired effect, a normal resistance is not used for the resistance S of the potentiometer in fig. 2, but one which is dependent on voltage, namely the

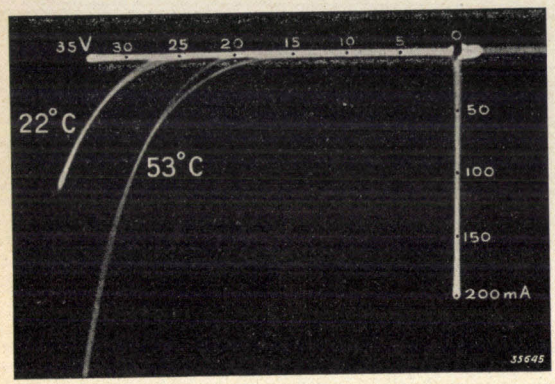


Fig. 4. Relation between leakage current and blocking voltage (blocking characteristic) of a selenium valve, at two different temperatures. (From: D. M. Duinker. The use of selenium valves in rectifiers, Philips techn. Rev. **5**, 199, 1940).

leakage resistance of a blocking-layer rectifier (selenium valve). In *fig. 4*, where the blocking characteristic of such a valve is given it may be seen that the very high resistance of the valve in the blocking direction (leakage resistance) decreases with increasing voltage and changes correspondingly the voltage division through the potentiometer R-S in such a way that with increasing input A.C. voltage the negative grid voltage of the output valve increases less than proportionally. By giving S and R suitable dimensions a fairly linear dB scale could be obtained for the indicating instrument.

Upon changing valves, etc. the calibration of the scale is checked in two steps: an A.C. voltage of 1.55 V R.M.S. is applied to the input terminals, and by adjusting the regulatory resistance R_2, which affects the grid bias voltage of the output valve, the pointer of the dB-meter on the amplifier rack is set exactly at 0 dB. In order to fix a second point on the scale, by reversing the switch H, the grid A.C. voltage of the first set of amplifier valves is then reduced by a factor which corresponds to an attenuation of the input signal by 30 dB. By adjustment of the resistance R_1 the indication of the anode current meter is then set at —30 dB. Since

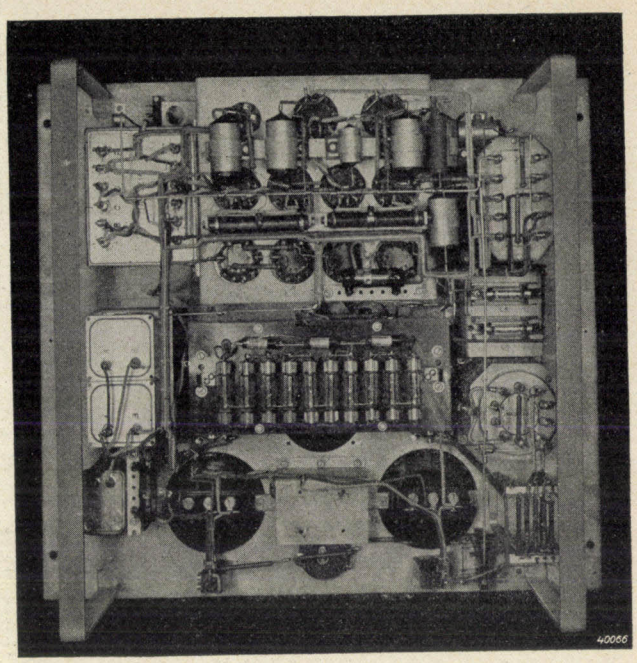


Fig. 5. The peak voltage meter opened. At the top centre is a panel for all the amplifier valves with their grid resistances, decoupling condensers, etc. Below it the potentiometer of the time circuit. The blocking-layer rectifier in this part is composed of three selenium valves connected in parallel. The potentiometer is covered with a double-walled metal cap.

only a small anode current flows at a high input A.C. voltage (0 dB), so that the adjustment of R_1 has little effect, the first calibration point remains practically unaffected in the fixing of the second.

Fig. 4 shows that the blocking characteristic of the selenium valve still depends very much upon the temperature. This dependence can be

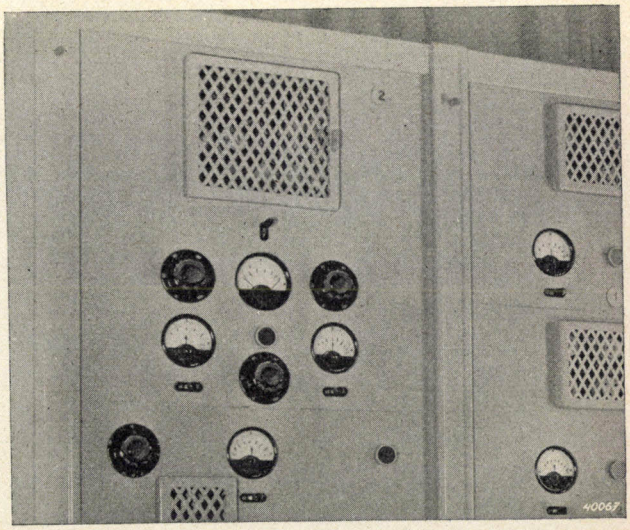


Fig. 6. Front view of the peak voltage meter mounted in an amplifier rack. The dB-meter is in the middle and below it the two differential meters with the corresponding keys. The amplifier valves are accessible through the door above.

expressed by a temperature coefficient of the leakage resistance, which coefficient is found to be approximately independent of the voltage in the voltage region in question. It is thus possible to make the calibration of the instrument independent of the temperature by making the resistance R of a material whose specific resistance has the same temperature coefficient as the leakage resistance of the valve. When that is done the voltage division through the potentiometer remains unchanged upon variations in the temperature. Care must now of course be taken that the resistance R and the valve are always at the same temperature, but this is easily ensured by housing them both in a double-walled heavy metal box, which because of its large heat capacity makes it possible for temperature fluctuations of the surroundings to influence only the temperature inside the box and not the temperature distribution. In *fig. 5*, in which the peak voltage meter is shown open, this box has been removed in order to show the parts of the potentiometer.

The influence of mains voltage fluctuations on the calibration is also eliminated by stabilizing with neon lamps the plate voltage which is provided by the supply apparatus. Moreover, the plate voltage of the first two amplifier stages can be checked with two meters. These are differential meters (see fig. 2) which in normal use indicate the difference in current between the two valves in push-pull connection of each stage. By reversing keys which are situated under the meters, see the photograph *fig. 6* of the complete apparatus, the plate current of each valve can, however, be measured separately.

A p_H-METER WITH A VERY HIGH INPUT RESISTANCE

by C. DORSMAN.

545.37 : 621.317.723

In very many chemical processes the degree of acidity is an important factor. This is expressed by the quantity p_H, defined as the negative Briggsian logarithm of the concentration of the hydrogen ion. In this article a discussion is given of several electrical methods, based upon Nernst's law of measuring the p_H. The p_H-meter manufactured by Philips is described; it contains an electrometer with compensator as well as complete chemical apparatus. The electrometer differs from existing instruments in that the D.C. voltage to be measured is converted into an A.C. voltage. This has the advantage that the input resistance is very high, since it is determined only by the resistance of two condensers with air as dielectric. Because of this high input resistance the instrument is particularly well adapted to measurements with the glass electrode. At the same time, an amplifier for A.C. voltage can be made stabler than one for D.C. voltage.

During the last twenty years the importance of an accurate knowledge of the degree of acidity of the most diverse liquids has come to be more and more commonly recognized. Not only in purely scientific research, but also in the chemical industry, in that of food products, in soil chemistry, in laundries and in all kinds of physiological work, the measurement of hydrogen ion concentrations now belongs to the standard methods daily applied in innumerable laboratories.

The concentration of the hydrogen ions varies between about 1 gram-ion per litre in strongly dissociated acids at a concentration of 1 normal, to 10^{-14} gram ion per litre for alkalies of 1 normal. These figures are inconvenient to use. For that reason Sörensen in 1909 proposed the introduction of the negative Briggsian logarithm of the hydrogen ion concentration, expressed in gram ions per litre, as a measure of the p_H. An advantage of the quantity p_H is that the potentials of the electrodes which are used to measure the degree of acidity of liquids are in a linear relation to it. This is expressed in Nernst's law which we shall now discuss.

Nernst's law

The relation between the difference ΔE in voltage between two solutions separated by a semi-permeable wall and the concentrations C_1 and C_2 in which a certain kind of ion which can pass through this wall is present in the two solutions (*fig. 1*) is given by the following formula:

$$\Delta E = \frac{RT}{F} \ln \frac{C_1}{C_2}, \quad \ldots \quad (1)$$

which among other things describes the equilibrium of the hydrogen ions at a glass electrode. In this equation R represents the gas constant, F the charge per gram molecule, T the absolute temperature.

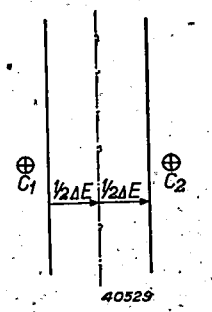


Fig. 1. A semipermeable wall separates two liquids in which an ion occurs with the concentrations C_1 and C_2, respectively. Only this ion can pass through the wall. There is a potential difference ΔE between the two liquids.

We shall give a brief derivation of formula (1). If the wall is permeable for a certain kind of ion and not for another, a number N_1 per second will pass from left to right, which number is determined by the concentration C_1. At the same time a number N_2, determined by C_2 will pass from right to left. Due to a potential difference ΔE between the two solutions the positive ions will show a preference for one direction. According to the well-known probability relation, the chance that a monovalent ion which strikes the wall will reach the middle plane of the wall from the left is:

$$e^{-q\frac{\varphi-\frac{1}{2}\Delta E}{kT}} \quad \text{and from the right:} \quad e^{-q\frac{\varphi+\frac{1}{2}\Delta E}{kT}}.$$

In these expressions e is the base of the natural logarithms, a a potential difference which determines the probability of reaching the middle when there is no external potential difference ΔE present. k is Boltzmann's constant and q is the charge of the electron. The total number of ions which passes from the left to the middle of the wall is proportional to this factor times the number of ions, which number is proportional to the concentration C, so that

$$N_1 \sim C_1 e^{-q\frac{\varphi-\frac{1}{2}\Delta E}{kT}} \quad \text{and} \quad N_3 \sim C_3 e^{-q\frac{\varphi+\frac{1}{2}\Delta E}{kT}}.$$

The difference in potential ΔE at which the same number of particles reach the middle of the wall from the left as from the right may be found from the condition $N_1 = N_2$:

$$C_1 e^{-q\frac{\varphi-\frac{1}{2}\Delta E}{kT}} = C_2 e^{-q\frac{\varphi+\frac{1}{2}\Delta E}{kT}},$$

so that:

$$\frac{C_1}{C_2} = e^{\frac{q\Delta E}{kT}}.$$

The difference in potential thus becomes

$$\Delta E = \frac{kT}{q} \ln \frac{C_1}{C_2}.$$

By multiplying numerator and denominator by the number of particles N in a gram molecule, formula (1) is obtained..

Entirely analogous to this equilibrium, a formula can be derived which describes the equilibrium state at an electrode in a liquid which contains the same kind of ion as the electrode can take up. This is the case for instance with the hydrogen electrode. Ions from the liquid are deposited on the electrode; ions of the same sort are freed from the electrode. The potential difference at which as many are deposited as are liberated is given for monovalent ions by Nernst's law:

$$\Delta E = \frac{RT}{F} \ln \frac{C_1}{K}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

in which K is a quantity with the dimensions of a concentration specific for the electrode. We shall make repeated use of this form of Nernst's equation. For n-valent ions it is only necessary to add a factor n to the denominator.

Methods of measuring the quantity pH

The colorimetric method was often used in the past. This is based upon the change in colour of certain, usually organic, indicator substances when the degree of acidity of the liquid changes between certain critical limits. As examples we may mention phenolphthalein which is colourless at $pH = 8$ and red at $pH = 10$, and methyl orange which is red at $pH = 3$ and yellow at $pH = 4$. In such colorimetric measurements it is possible to determine the pH accurately within one tenth of a unit.

In addition there exist a number of electrical methods of measurement which are based upon Nernst's law and which we shall now discuss briefly in turn.

Measurements with the hydrogen and the quinhydrone electrode

The so-called hydrogen electrode is an electrode whose surface is covered with atomic hydrogen. It can be prepared by allowing gaseous hydrogen to bubble along a platinum wire covered with platinum sponge. When the electrode is immersed in a fluid containing hydrogen ions, an exchange of ions between liquid and electrode will take place on the platinum sponge as a catalyst. No current flows if the following holds:

$$\Delta E = \frac{RT}{F} \ln \frac{C_H}{K}.$$

The value of the pH is now determined by measuring the difference of potential ΔE. Since the pH is the negative Briggsian logarithm of C_H the following is true:

$$\Delta E = -2{\cdot}30 \frac{RT}{F} pH - 2{\cdot}30 \frac{RT}{F} \lg K,$$

or $$\Delta E = E_0 - pH\,[58{\cdot}1 + 0{\cdot}2\,(t-20)] \quad . \quad . \quad (3)$$

Here $E_0 = -2.30 \frac{RT}{F} \log K$ and t is the temperature in centigrade degrees.

In the case of the so-called quinhydrone electrode an atmosphere of hydrogen around a platinum electrode is also used. For the sake of brevity we shall not go into it more deeply.

Measurements with the glass electrode

In measurements with the glass electrode a liquid of unknown pH is separated by a glass wall, which transmits hydrogen ions but no others, from a liquid with a known pH. An exceptionally suitable kind of glass for this purpose is the Corning glass 015, which consists of 72 per cent of SiO_2, 22 per cent of Na_2O and 6 per cent of CaO and contains extremely few impurities. Here again no current flows when the potential difference ΔE between the two liquids is equal to;

$$\Delta E = 2\,30 \frac{RT}{F} (pH_1 - pH_2) \text{ or}$$

$$\Delta E = (pH_1 - pH_2)\,[58{\cdot}1 + 0{\cdot}2\,(t-20)].$$

This formula holds for the kind of glass mentioned with very slight deviations between the values $pH = 0$ and $pH = 9$.

We assumed that the wall transmits only hydrogen ions and no others, so that we could immediately apply formula (1) to it. If, however, it also transmits other ions, an equilibrium must also be established for them in the same way. Fortunately, however, Corning glass 015 is impermeable to almost all other ions. Only when ions of alkali and alkaline earth metals are present does a so-called salt error occur, which has been quantitatively determined by Dole[1]) for Li, Na, K and Ba. In the case of strongly basic liquids whose pH is greater than 9, this must be taken into account. It is understandable that this error occurs exactly in the basic region, since there the number of hydrogen ions available for transport of electricity is very small, and therefore a number of other ions will be able to play a relatively more important part.

Comparison electrode

In the electrical determination of the degree of acidity the potential difference between the unknown liquid and platinum must be measured, or that between the unknown liquid and the liquid with the known pH. In order to do this

[1]) M. Dole, The theory of the glass electrode, J. Amer. Chem. Soc. **53**, 4260, 1931.

a second electrode, the so-called comparison electrode, must be immersed in the unknown liquid. This electrode must form a well defined, unvarying transition from the liquid to the electrical connection wire.

In the apparatus here described a saturated calomel electrode is used for that purpose (*fig. 2*). For the measurements with

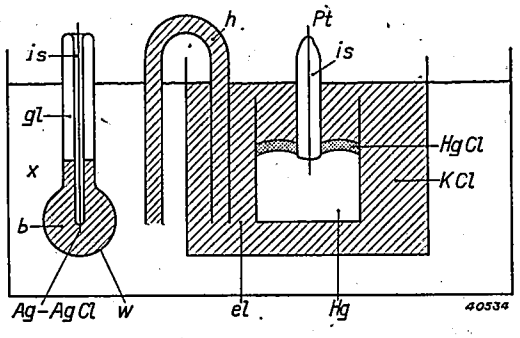


Fig. 2. In the liquid for which the quantity pH is to be measured a glass electrode (*gl*) and a comparison electrode (*el*) are immersed. The glass electrode, formed by a wall (*w*) which is permeable only for hydrogen ions, is filled with a buffer solution (*b*) with a very constant p_H value. This liquid makes contact with a silver wire covered with silver chloride (*Ag-AgCl*). The latter is insultated by a glass covering (*is*) and is led out at the top through the fused glass. The comparison electrode is filled with saturated KCl solution (*KCl*), which is connected with the liquid to be measured by a siphon (*h*). The KCl solution is in contact with the mercury (*Hg*) *via* a layer of calomel (*HgCl*) covering the mercury. The platinum wire (*Pt*), again insultated with glass (*is*) forms the second connection.

this electrode formula (3) is again valid, taking for E_0 the following values:

+159 upon use of the glass electrode
−249·5 „ „ „ „ hydrogen electrode and
+453·4 „ „ „ „ quinhydrone electrode

The method by which this potential difference E can be measured will now be discussed.

Methods of measuring the potential difference

The following requirements are made of the method of measurement by the chemist:

1) It must be possible to measure with a glass electrode. In practice it is important that this electrode should be sturdy and thus blown from glass several millimetres thick. Then, however, the electrical resistance for current conduction is very high, namely of the order of magnitude of 100 million ohms.

From *fig. 3* it may be seen that upon direct measurement of potential a measuring error will occur if the resistance R_i of the measuring instrument is comparable in magnitude to the glass resistance R_g. Therefore the requirement is made that the resistance R_i should be for instance one thousand times as high. It would thus have to be 100000 million ohms.

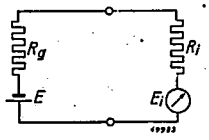


Fig. 3. An instrument with an internal resistance R_i measures the voltage E of an electrode circuit whose resistance, determined by the glass electrode is R_g. The voltage E_i on the instrument is given by $E_i = E \frac{R_i}{R_i + R_g}$. If R_i is one thousand times Rg, the instrument measures with a deviation of 0.1 per cent.

In a compensation measurement such as is represented in *fig. 4*, the potential difference between the input terminals of the instrument will be fully compensated in the equilibrium state. No current then flows through the input resistance and no error occurs. The sensitivity of the instrument does, however, decrease with increasing value of R_g.

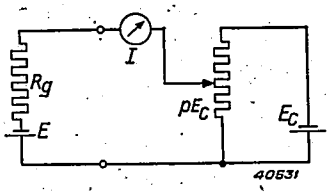


Fig. 4. An apparatus with compensation connections is so adjusted that the instrument I gives no deviation. pE_C is then equal to E. The resistance of the glass electrode results only in a decrease of the sensitivity of the connections. The current I is equal to $\frac{pE_C - E}{R_g}$. If $pE_C - E$ is equal to 1 millivolt and R_g amounts to a million ohms this current is 10^{-9} amperes, thus very small.

2) A difference of 0·01 in the pH value must be observable, while the absolute value must be able to be measured accurately to within several hundreds of a unit. This means that 0.5 millivolt must be observable.

3) The influence of temperature changes on the results of the measurement must be eliminated.

There are various possibilities for measuring the potential difference which we shall now deal with in turn.

Sensitive galvanometer

Good results can in general be obtained by the use of a compensation method with a very sensitive galvanometer. Such galvanometers are, however, so delicate that they are only seldom used in practical cases. Moreover, care must then

also be taken that the resistance in the glass electrode is not too high. Kinds of glass could be chosen for this purpose with very low specific resistance, but from a chemical point of view those glasses are much less resistant than the Corning glass 015 ordinarily used for glass electrodes.

Direct-voltage amplifier

The potential difference can also be amplified immediately or after compensation with a D.C. voltage amplifier, and then measured. In both cases a much sturdier instrument can be used (*fig. 5*). This method also, however, has two disadvantages. In the first place the current through the instrument varies when the anode current of the amplifier valve changes. The latter changes not only due to the change in the grid voltage of the valves, but also due to changes in the temperature of the cathode, of the structure of the cathode surface and of the contact potential of the grid.

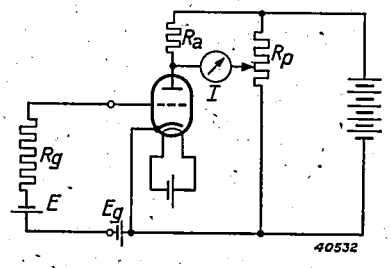


Fig. 5. The voltage E is fed to the grid of an amplifier valve. For correct adjustment a negative grid voltage is given with the battery E_g. The anode current of the valve flows through resistance R_a, so that a difference in voltage occurs over R_a. The potentiometer R_p is so adjusted that at a certain value of p_H no current flows through the instrument I. In the subsequent measurements the current through the instrument is proportional to the deviation of the value of the p_H.

In the second place the input resistance of the valve cannot be high enough compared with the resistance R_g of the glass electrode. If the valve contains gas residues, these may become ionized and give off charge to the grid. Electrical leakage along the insulation materials and photoelectric current from the grid due to the light from the hot cathode also decreases the input resistance. In a so-called electrometer triode[2]) these disadvantages are indeed avoided, but the result is the disadvantage of only slight sensitivity.

A variation on this method, which has been successfully employed, is the following (*fig. 6*). With the help of the voltage to be measured a condenser is charged; when the condenser is full, which in micro-analyses may, however, take about a minute, the current is zero. This D.C. voltage is then measured with a compensation connection with valve amplifier. The disadvantages which may be mentioned are that here also the variation of the anode current produces a deviation and that the contact of the switch must be closed again each time before a measurement can be carried out.

[2]) *Cf.* also : Philips techn. Rev. **5**, 54, 1940.

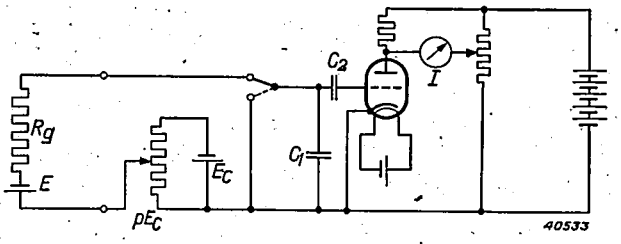


Fig. 6. The connections are largely similar to those of fig. 5. The electrode circuit with voltage E which is compensated with the voltage pE_C is reached. The condenser is then suddenly discharged by means of the reversing switch. The changes in voltage are applied to the grid with the condenser C_2. The current impulse is observed with the instrument I. As soon as the compensation is complete so that pE_C is equal to E. the current impulse will be zero. The voltage E is read off on the compensator.

Alternating voltage amplifier

As just explained a D.C. voltage amplifier has several disadvantages. In our apparatus we have attempted to avoid them by first converting the D.C. voltage into an A.C. voltage and then amplifying the latter.

Description of the apparatus

The p_H-meter GM 4491 contains an electrometer and a compensation part. The latter compensates the voltage between the saturated calomel electrode and the glass electrode. With the electrometer it is possible to ascertain when the compensation is accurate to within 0.5 millivolt. We shall now describe the two parts separately.

1) *The electrometer*

A condenser is connected to the electrode circuit by a resistance of a high value. When the voltage is V, the charge Q of the condenser C then amounts to:

$$Q = CV.$$

The size of the condenser is varied by allowing the plates to vibrate periodically with respect to each other, charge must then flow back and forth with the same periodicity and an A.C. voltage occurs on the high resistance. If, however, the D.C. voltage is zero no A.C. voltage occurs. The A.C. voltage is amplified and when present made visible with a cathode ray indicator. As an A.C. voltage amplifier this can be made very

stable, and since the A.C. voltage can be fed to the grid with a condenser, except for dielectric losses, the input resistance is infinitely large.

We shall now go somewhat more deeply into the action of the vibrating condenser.

The vibrating condenser

The electrometer contains a valve oscillator which generates an electrical A.C. voltage of 125 c/s. This electrical energy drives a mechanical vibration system. The latter consists of a shaft bearing a flat plate and a coil and is fastened flexibly with two menbranes. The coil is situated in a magnetic field (see *fig. 7*) and

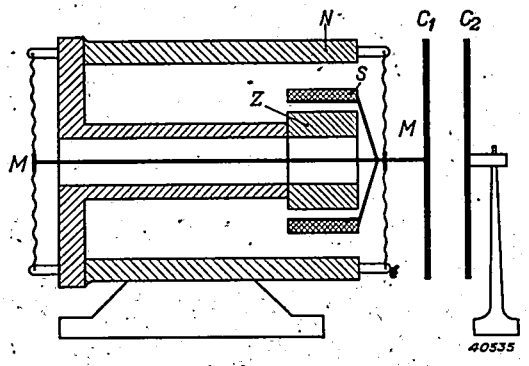


Fig. 7. The vibrating condenser. The condenser plate C_1 is fastened to the end of a shaft which is hung on two membrane springs M. Through the coil S fastened to the shaft flows an alternating current of 125 c/s. In the magnetic field between the north pole N and the south pole Z the coil experiences an alternating force. The coil with shaft and the condenser plate C_1 begin to vibrate with this frequency. The capacity of the condenser formed by the electrode C_1 and a second electrode C_2 therefore also varies at a frequency of 125 c/s.

begins to vibrate when an alternating current flows. Since the mechanical resonance frequency is equal to the frequency of the A.C. voltage a satisfactorily sinusoidal vibrational state occurs. The vibrating plate, together with a fixed one parallel to it and lying opposite it, forms a condenser whose separation has a sinusoidally varying component and whose capacity thus varies by one third of its value.

In order to obtain a good idea of the sensitivity of the mechanism it is instructive to calculate the electrical system.

At each moment the charge Q is equal to capacity C times voltage V:

$$Q = CV.$$

The current i is equal to the derivative of the charge with respect to time

$$i = \frac{dQ}{dt} = C\frac{dV}{dt} + V\frac{dC}{dt}.$$

Ohm's law gives the following for R:

$$\frac{V_0 - V}{R} = C\frac{dV}{dt} + V\frac{dC}{dt} \quad \ldots (4)$$

Let the variation of the distance be described by $(1 + a \cos \omega t)$ times an average value. The capacity then has the form

$$C = \frac{1 + a \cos \omega t}{C_0}.$$

In the apparatus here described a is about $^1/_3$ and ωRC_0 is equal to 30. This value of 30 is obtained by using a very high value of the resistance R. This could be realized with a column of liquid which has a resistance of 1000 million ohms.

For a sufficiently high value of ωRC_0 the approximate solution of the differential equation (4) for the variation in voltage is as follows:

$$V - V_0 = a\ V_0 \left(\cos \omega t - \frac{1}{\omega RC_0} \sin \omega t\right) \quad \ldots (5)$$

Since $^1/_3$ of the capacity oscillates and a is therefore equal to $^1/_3$, and since ωRC_0 is large enough, namely equal to 30, we may finally use the approximate expression:

$$V - V_0 = \frac{1}{3} V_0 \cos \omega t \quad \ldots (6)$$

for the voltage variations without hesitation. An A.C. voltage is thus obtained on the resistance R whose peak voltage is equal to $^1/_3$ of the D.C. voltage to be measured.

The effective value of the A.C. voltage obtained is then $1/\sqrt{2}$ times this part, *i.e.* 23 per cent of the D.C. voltage.

It is required that a deviation of 0.01 pH unit should be easily observed. According to the formula already given, this corresponds to a difference in D.C. voltage of 0.58 millivolt. The A.C. voltage amplifier must therefore make observable a voltage of 0.23 times 0.58 millivolt, or 0.1 millivolt.

Work functions

In addition to the externally applied potential differences there is an internal electromotive force in the apparatus.

The voltage that corresponds to the energy which is necessary to cause an electron to leave a metal surface is called the work function. Two metal objects with different work functions are brought close to each other. If any transport of electrons is possible through the intervening space, the metal with the higher work function will lose fewer electrons than that with the lower work function, until this nonstationary state is compensated with the help of charges on the surface.

The vibrating condenser which has just been described will in general therefore receive a charge. The difference in work function is given, the capacity varies in magnitude, the charge must therefore also vary. Even when the connection terminals of the instrument are short-circuited, therefore, an A.C. voltage will be generated. (This is of course by no means contrary to the main laws of thermodynamics, but means that mechanical energy of the vibrating system is converted into electrical energy). For any two plates of the same material the work function may differ as much as several tenths of a volt. It is found that the effect can be reduced to a few millivolts by using two plates of a very pure metal.

In the apparatus here described this small voltage must be compensated with an externally applied voltage before the apparatus is used. A potentiometer is introduced for this purpose.

The amplifier

The deviations from the equilibrium state in the case of incorrect compensation are made visible, as already stated, on the fluorescent screen of a cathode ray indicator. These tubes are already familiar as tuning indicators[3]) in modern radio receivers. The important point here is that a voltage of 1 V is easily observed. An amplification of 10000 times is thus necessary. This is possible with two pentode valves in cascade connection (*fig. 8*). Since an input voltage of 0.1 millivolt must be readily observable, small noise voltages and voltages from the supply mains may also become visible as interferences. By means of two resonance circuits introduced into the anode connections, only a very small frequency region Δf in the neighbourhood of the oscillator frequency is amplified (*fig. 9*). As already discussed in this periodical[4]) the noise voltage $V_R = \sqrt{4kTR\Delta f}$, or, after substitution of the numerical values, $V_R = 6\ \mu$ V. On the input terminals therefore this voltage may be neglected compared with the quantity to be measured. The resonance frequency of 125 c/s is so chosen that 50 c/s as well as its second and third harmonics, 100 and 150 c/s, fall well outside the resonance region.

In measuring entirely unknown quantities it is important that the sensitivity should be able to be reduced. This is possible since the cathode ray indicator is not equally sensitive in all sectors. At the same time there is also a switch with which a further reduction by a factor of 30 can be obtained.

In the performance of titrations this possibility of less sensitive measuring is also desirable, because in that case the approach to the turning point is clearly seen in advance. It is very important to simplify this observation, and a holder has been constructed for this purpose which makes it possible to fasten the removable cathode ray indicator beside the burette, so that attention can be concentrated simultaneously on these different points. The voltages are supplied *via* loose cables.

2) *The compensator*

The compensator offers the following possibilities.

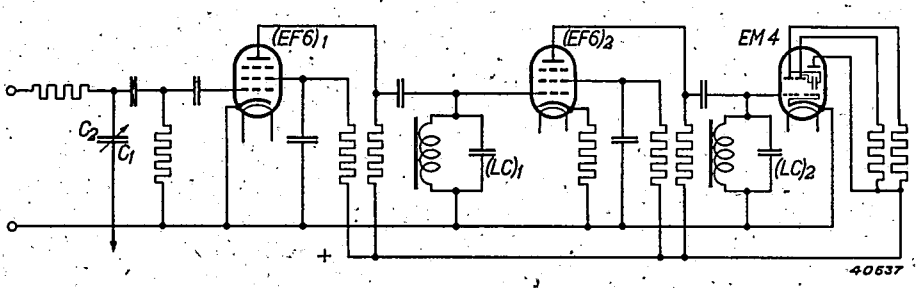


Fig. 8. Diagram of the electrometer. The D.C. voltage to be measured is fed to the vibrating condenser plate C_1; this D.C. voltage is compensated by the application of a compensation voltage to the other electrode C_2. If the compensation is not complete, an A.C. voltage is generated which is applied to the grid of the first pentode $(EF6)_1$. The voltage is selectively amplified with the resonance circuit $(LC)_1$. With $(EF6)_2$ and $(LC)_2$ this is repeated; the cathode ray indicator $(EM4)$ then makes the A.C. voltage visible.

1) The measurement of the quantity p_H with an electrode of Corning glass 105, with a hydrogen electrode, or with a quinhydrone electrode. As comparison electrode a saturated calomel electrode is used in each case. The previously derived formula (3):

$$\Delta E = E_0 - p_H\,[58{\cdot}1 + 0{\cdot}2\,(t-20)]. \quad (3)$$

always serves as the starting point for these measurements. The last term containing the

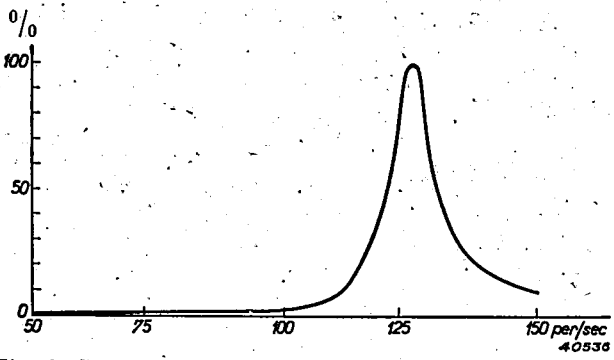


Fig. 9. Characteristic of the A.C. voltage amplifier. In order to eliminate the interference by the supply voltage of 50 c/s and its higher harmonics of 100 and 150 c/s, the amplification is made selective for the oscillator frequency of 125 c/s. The two resonance circuits (fig. 8) are tuned to this frequency. Thus at 50 c/s the amplification is 0.5%, at 100 c/s 2.5% and at 150 c/s 9% of that at 125 c/s. The influence of interference voltages is thus made so small that they fall entirely outside the limit of observation.

[3]) See for example: Philips techn. Rev. **2**, 270, 1937.
[4]) Philips techn. Rev. **6**, 129, 1941.

temperature t is taken into account with the help of the potentiometer R_3 between the limits 10 and 40 °C.

2) The performance of voltage measurements on

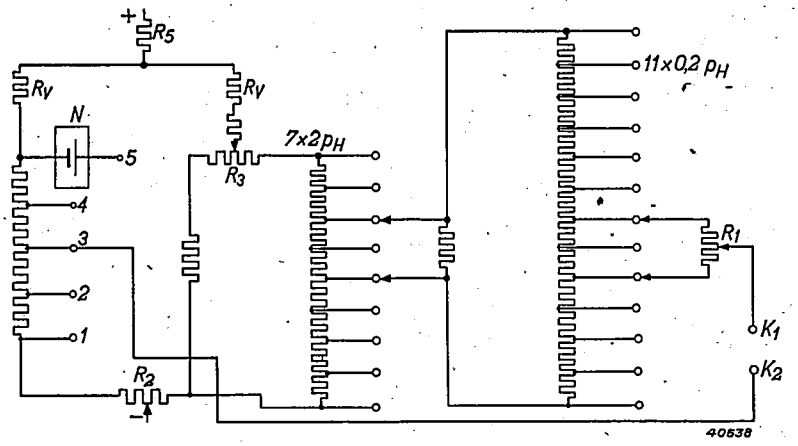


Fig. 10. The compensator. This figure represents the connections as arranged for measurements of the P_H. By commutation the electrical voltages can be measured in millivolts, or a control measurement with the standard element N can be carried out. The stabilized D.C. voltage of the supply is connected between + and —. The compensation voltage is connected at the terminals K_1 and K_2 with the plate C_1 of fig. 8. The voltage distributor is constructed in three stages and calibrated in steps of 2 and 0.2 P_H units, respectively, the potentiometer R_1 in 25 scale divisions of 0.01 p_H unit. The constant voltage E_0 in formula (3) is compensated by connecting the voltage distributor to the terminals *1*, *3* and *4*, respectively. Position *2* serves for the measurement in millivolts, position *5* for the control measurement, which is not dealt with in detail in this description. A small difference in voltage can be produced with the potentiometer R_2, by means of which the difference in work function of the electrometer condenser and the so-called asymmetry potential of the glass electrode may be compensated. The potentiometer R_0 is calibrated in degrees Centigrade. A shift of the contact results in a change of the voltage on the voltage distributor which is calibrated in p_H units. The variation of the term $[58.1 + 0.2 (t—20)]$ is realized in this way. The large resistances R_v provide that the variation of the potentiometer R_0 shall not affect the constant voltage of the terminals *1*, *3* and *4*. R_0 is a series resistance for the regulation of the input voltage.

a source of voltage with either the positive or the negative pole earthed. Between 0 and 1425 mV the voltage can be adjusted to within 1 millivolt per scale division with the switches and the potentiometer. The accuracy of this voltage measurement is determined by the precision of the resistances of the compensator connections and amounts to 1 mV plus 2 per cent of the value measured.

A few striking points in the construction of the compensator are given below.

a) All the important resistances are wound on a ceramic core of metal wire having a low temperature coefficient. After having been subjected to great temperature differences for a long time (aged) they are made accurate to within 0.05 per cent.

b) In order not to depend upon a continuously adjustable potentiometer which is by nature fairly inaccurate, the reading takes place in three steps. A switch with 7 stages of 2 p_H-units, a switch with 10 stages of 0.2 p_H-unit and the potentiometer R_1 which divides 0.25 p_H-unit. In this way an accuracy of 0.02 p_H-unit plus 2 per cent of the value measured is obtained over the whole region.

c) The compensator connections are fed from the A.C. mains. After rectification the voltage is stabilized with a gas-filled stabilization tube. This tube keeps the voltage accurately constant wi hin 1 per thousand for several hours. In the course of a year the voltage may vary by several per cent. The voltage of the stabilization tube is regularly checked with a saturated standard element according to W e s t o n and any deviation is eliminated with the series res stance R_1. The standard element produces no current; no battery or accumulator which has a limited life or must be charged is thus used in the apparatus.

d) The difference in work function of the condenser plates can be compensated with a potentiometer.

e) The whole apparatus is made as tight as possible against moisture with rubber and felt. Granular silica gel containing a cobalt salt as moisture indicator serves as drying agent. If it is observed through the window that it is moist the holder can be taken out and the silica gel dried at 140 °C.

The whole apparatus is shown in *fig. 11*.

The chemical part

Since an electrical instrument should be kept away from apparatus which makes use of chemical reagents, the chemical apparatus is housed in a separate box, whose cover can be used as a measuring box (see *fig. 12*). A rod with a clamp can be fastened into the cover and used as support. A short description of the mechanical construction of the glass and the calomel electrodes and of the principle of the buffer solution follows below.

1) T h e g l a s s e l e c t r o d e

Since the resistance may be high, the wall is several millimetres thick, giving a very sturdy unit (*fig. 13*). Care is taken that the potential differences due to differences in the internal

Fig. 11. The apparatus is built into a metal box which is made moisture tight. An opening covered by a magnifying lens for the cathode ray indicator may be seen. Above it is a second window giving access to a quantity of silica gel as drying agent. By using a strong baked lacquer, by chromium plating certain parts and by making it very tight against moisture, every attempt has been made to make the instrument suitable for use in chemical laboratories. Further protection is given by a heavy wooden case.

and external surfaces are smaller than ten millivolts. The electrode is filled with 0.1 normal HCl solution. This combination has a very low temperature coefficient for the internal voltage difference.

2) The saturated calomel electrode

In *fig. 14* it may be seen that the internal part is fused into a tube only 6 mm in thickness. By means of a piece of porous stone the space containing platinum wire, mercury and calomel is in contact with the surrounding reservoir containing saturated KCl solution. The advantage is that this part is very well protected mechanically and chemically, making the cleaning and filling of the KCl container very simple. It is still further simplified by the fact that the electrode can be taken apart (*fig. 15*).

When a small quantity of the liquid to be measured penetrates into the capillary there is a danger that the KCl solution will be contaminated. By means of the small stopcock at the end of the side tube a bubble of air is then admitted, a drop of KCl then escapes and the capillary is rinsed clean in this way.

3) Buffer solutions

It is desirable at certain intervals to carry out a *pH* measurement of a liquid having a known hydrogen ion concentration in order to check the correctness of the indications of the measuring electrodes. Such liquids are realized in the form of so-called buffer solutions. These are solutions of measured quantities of acid with a quantity of a salt, which possess the property that their *pH* value is very insensitive to slight contaminations by acids or bases which are either already present in the water or go into solution from the glass. Kolthoff has given a method of preparing such mixtures in tablet form. A tablet dissolved in 20 cm^3 of water gives a good buffer solution suitable for checking the electrode circuit, and if necessary correcting it. Three tubes with twenty tablets for the *pH* values 3, 6 and 8 are included in the apparatus.

LITERATURE

E. Madelung: Messung elektrostatischer Potentiale mit der Elektronenröhre, Verh. Deutsch. Phys. Ges. **6**, 14, 1935.

W. A. Zisman: A new method of measuring con-

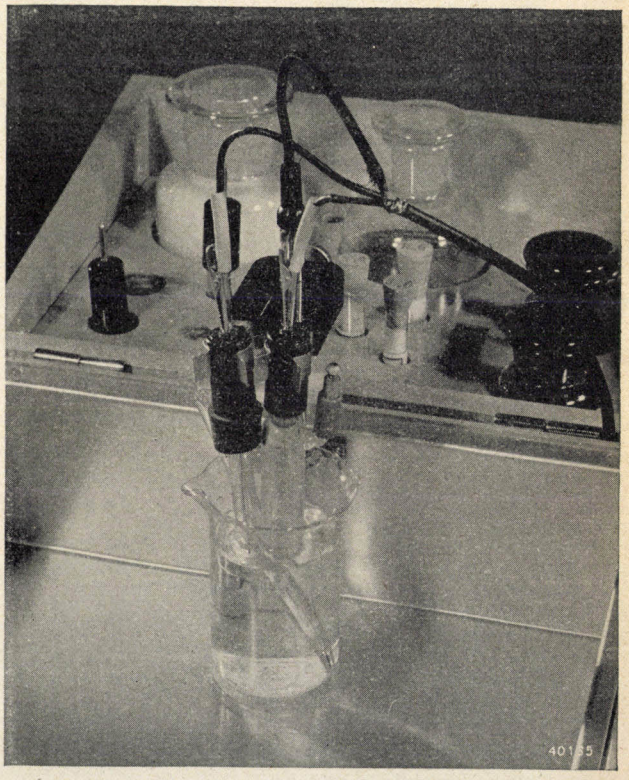
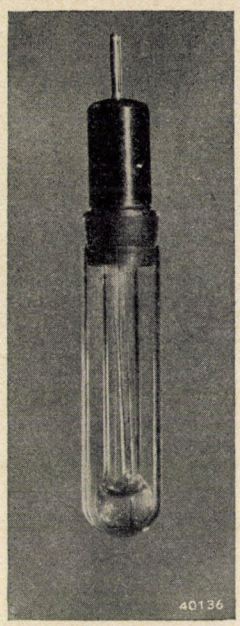
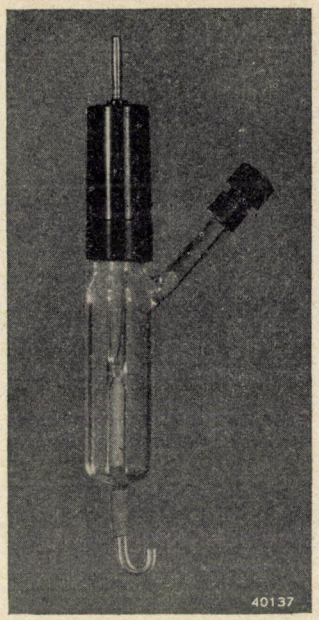


Fig. 12. The chemical part of the apparatus is housed in a separate wooden case. Two glass electrodes, a calomel electrode, platinum electrode and thermometer, bottles with buffer tablets for the *pH* values 3, 6 and 8, potassium chloride and quinhydrone and a set of beakers all have places in the case. In order to make measurements possible even outside the laboratory the cover of the case is constructed as a measuring container. In the photograph may be seen a measuring set-up with standard, clamp and two electrodes.





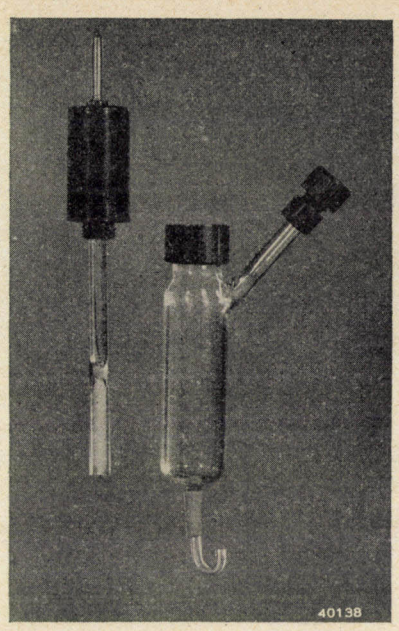


Fig. 13. The *pH*-meter GM 4491 makes it possible to carry out accurate measurements with glass electrodes of very high resistance. The glass electrode shown here has an electrical resistance of 100 million ohms. This value could be obtained with a strong, thick wall of Corning glass 015. The electrode is filled with a buffer solution in which is immersed a silver wire as electrode. The filling is so chosen that the jump in voltage at the silver and at the inside of the glass together have a temperature coefficient which is practically zero.

Fig. 14. A saturated calomel electrode is used as comparison electrode. A space filled with a saturated solution of KCl is connected through a capillary with the liquid to be measured. This solution is connected with an inner electrode through a piece of porous stone. The inner electrode contains a platinum wire immersed in mercury. On the mercury calomel (Hg_2Cl_2) is formed which makes very good contact with the solution of potassium cloride. Through the stopcock projecting to one side air bubbles can be admitted which push the solution through the capillary and rinse it clean in this way.

Fig. 15. The inner section of the calomel electrode can be taken out of the potassium chloride reservoir. The latter can then be easily cleaned and filled; in case of breakage this simple outer section can be replaced.

tact potential differences in metals, Rev. sci. Instr. **3**, 367, 1932.

H. R. Kruyt: Inleiding tot de physische chemie (Introduction to physical chemistry), Paris, Amsterdam 1936.

W. M. Clark: The determination of hydrogen ions, Baillière, Tindall and Cox, London 1928.

L. Michaelis: Die Wasserstoffionenkonzentration. Springer, Berlin 1922.

H. Jörgensen: Die Bestimmung der Wasserstoffionenkonzentration, Steinkopff, Dresden-Leipzig 1935.

W. Kordatzki: Taschenbuch der praktischen *pH*-Messung, Müller und Steinicke, München 1938.

M. Kolthoff: Die kolorimetrische und potentiometrische p_H-Bestimmung, Springer, Berlin 1932.

VOL. 7 No. 2, pp. 33-64 FEBRUARY 1942

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

Dr. G. L. F. PHILIPS †

The founder of our concern, Dr. Ir. G. L. F. Philips, passed away at The Hague on January 26th 1942. Since 1891, the year in which he founded, with his father as sleeping partner, the firm of Philips & Co., till 1922, when he resigned as Managing Director of the N.V. Philips' Gloeilampenfabrieken, he employed his talents as man and engineer for one sole aim, that of making ever better incandescent electric lamps. Thus he built from the ground up and brought to perfection the manufacture of four kinds of lamps, *viz.*, the carbon lamp, the sprayed tungsten wire lamp, the drawn tungsten wire lamp and the gas-filled lamp. In this way he has made the Philips lamp famous in every part of the world. Under his stimulating leadership many ingenious machines for the manufacture of incandescent lamps and parts thereof were designed and made, which even to this day still form the foundation of modern incandescent lamp and radio valve manufacture.

The significance that he attached to science as a powerful aid to the development of the industry manifested itself in 1914 in the establishment of the Philips physical laboratory, where since that time new sources of light, the gas discharge lamps, have been developed and where the cradle has been and still is of many an extension of our company's activities.

THE DEVELOPMENT OF BLENDED-LIGHT LAMPS

by J. FUNKE and P. J. ORANJE.

621.327.9

Blended-light lamps are sources of white light composed of a mercury tube in series with a filament, which can be connected to the A.C. mains without any auxiliary apparatus. In this article several problems are discussed which are connected with the construction of these lamps. Special attention is paid to the measures which must be taken to obtain the desired blending ratio, namely equal amounts of mercury light and incandescent lamp light. In conclusion information is given about a newly developed blended-light lamp of only 300 Dlm, 160 W.

The blending of different kinds of light is at present an important aid in lighting technique. It is particularly important in connection with the employment of high pressure mercury lamps, since their colour rendering requires improvement when they are used because of their high efficiency for general lighting purposes. This improvement in colour was realized chiefly in two ways: the admixture of a certain amount of fluorescence light which is excited by the ultraviolet radiation emitted by the mercury lamp itself, or the admixture of ordinary electric light.

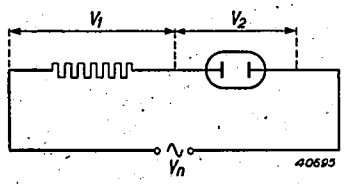


Fig. 1. Diagram showing the principle of the blended light lamp: a filament and a discharge tube in series

In addition to the simple combination of separate mercury and filament lamps in one fitting, a nice solution of the latter problem has recently been devised, which consists in connecting to the mains in series a mercury discharge tube and a filament, housed in the same envelope (see *fig. 1*). The advantages of this system have already been discussed in this periodical [1]): a lamp is obtained with a pleasant colour and with a considerably higher efficiency than the filament lamp, while the auxiliary apparatus necessary with mercury and other metal vapour lamps is eliminated, since the filament itself takes over its function. The lamp may therefore be connected directly to the A.C. mains [2]).

In designing such lamps a number of problems arose which will be discussed in the folowing. We shall first consider the behaviour of a mercury discharge tube burning in series with a resistance.

The variation of the tension on a mercury discharge burning in series with a resistance on an A.C. mains is represented in *fig. 2*. The gas discharge is extinguished twice per period, namely when the mains voltage falls below the value V_b in fig. 2, and it remains extinguished until the mains voltage reaches the re-ignition value v_{rign}. The time interval between extinction and re-ignition of the tube, the "dark period" δ, as may be seen from the figure, is dependent on the size of v_{rign}, and this in turn depends upon the value of the arc tension. At higher arc tensions the time at which the discharge is extinguished is shifted to an earlier moment, which results in an increase in δ. An increase of δ, however, results in a higher re-ignition voltage, because the number of free electrons and ions present falls when no current is flowing, so that a higher voltage is required to initiate the discharge again. For the arc tension we can choose a maximum value such that the re-ignition voltage is equal to the peak value of the mains voltage, since otherwise the discharge no longer re-ignites. For the sake of reliability a considerable margin will be allowed for surges in the mains voltage such as may occur in practice, for example upon switching on large machines.

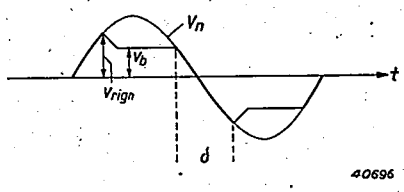


Fig. 2. Variation of the tension on a discharge tube which, in series with a resistance, is connected to a sinusoidal A.C. voltage V_n. At each current alternation the lamp is extinguished, and it is only re-ignited when the mains voltage has reached a certain value v_{rign}, which is higher than the constant value V_b of the arc tension in the remainder of the half period. This causes the occurrence of a dark period (δ).

[1]) Philips Techn. Rev. 5, 341, 1940.

[2]) An advantage of alternating current is that with a given value of the effective voltage the peak voltage is a factor 1.4 higher, so that the discharge tube is easily ignited. On the other hand direct current would have the advantage that the tube would only need to be ignited once, while with alternating current this must be done anew every half period. Blended-light lamps for direct current, however, have not yet appeared on the market.

The difference between the arc tension and the mains voltage must be taken up by the filament. Since the arc tension of the discharge tube is not constant, but begins with a very low value upon switching on the lamp, considerable overloading of the filament may occur when the lamp is switched on. In order to limit also this overloading it is desirable that the arc tension should not be chosen too high.

The whole problem becomes still more complicated because of the fact that we are not only interested in the reliability of the mercury discharge tube and the life of the filament, but we also make certain requirements as regards the efficiency and the colour. In the following therefore we shall go somewhat deeper into the problem of the voltage distribution between filament and mercury discharge in order to demonstrate how a satisfactory compromise was reached between the various requirements.

The voltage distribution

With a discharge tube of given dimensions it is in general possible to choose different values for the arc tension. The arc tension of a mercury discharge is very low at the beginning when the tube is still cold, and the final value is only gradually reached as the mercury pressure increases until all the mercury is evaporated, or, if there is an excess of mercury, until the final stationary state is reached. The larger the series resistance is chosen, the lower the current remains and thus the temperature of the mercury lamp. Its arc tension therefore also remains lower. In this way a very wide range of arc tensions can be realized practically.

The arc tension is made up of a steep voltage drop at cathode and anode (cathode drop and anode drop) and a gradual voltage drop along the tube. This gradual voltage drop increases very much as the mercury pressure increases, while on the contrary cathode drop and anode drop decrease slightly in that case.

Since the energy dissipated in the cathode and anode drops furnishes practically no contribution to the luminous flux, it is immediately clear that the efficiency of the mercury lamp increases with increasing arc tension. Actually the improvement is even more significant than would be concluded from this, since the efficiency of the light excitation itself is also found to increase with increasing voltage drop per cm length of the column [3]). In *fig. 3* an example is given of the combined action of the two effects. If, finally, one does not consider the mercury discharge alone, but the combination of discharge tube and filament in series, a relatively still greater improvement of the efficiency is observed. With increasing arc tension the ratio of mercury light to filament light is shifted in favour of the mercury light, and this is advantageous for the overall efficiency, since the mercury lamp possesses a considerably higher efficiency than the filament lamp. We may thus conclude that it is extremely important, as far as the efficiency is concerned, to have the discharge tube take up as large a part of the mains voltage as possible.

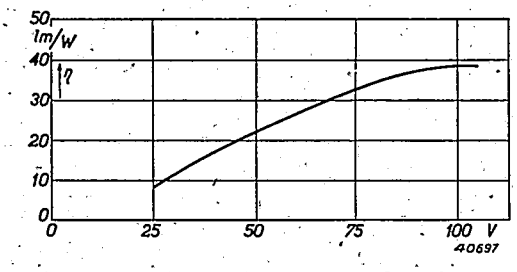


Fig. 3. Increase of the efficiency η of a high pressure mercury discharge with increasing arc tension.

In order to obtain a better and more numerical idea of the conclusions reached in the foregoing, let us consider the energy consumed by discharge tube and filament. The currents are the same for both at every moment, so that the division of the energy dissipation at each moment is given by the division of the voltage. If we pass from momentary values to effective values we may write the following for the energy of the filament:

$$W_1 = I \; V_1,$$

while the energy of the discharge tube must be expressed in the form

$$W_2 = a \; I \; V_2.$$

The factor a, which may have a value of 0.7 for example, expresses the fact that the energy consumed by a mercury discharge is smaller than the product of the effective values of current and voltage. Superficially this is the same phenomenon which occurs with chokes and condensers where the phase difference between current and voltage is the cause; a is then equal to the cosine of the phase angle.

In the case also of the mercury lamp one speaks of a kind of phase shift, since the current in each direction only begins to flow when the voltage in that direction has reached a certain minimum value (the re-ignition value). This apparent phase shift, however, has no well-defined significance, since the variation of current and voltage deviates very much from the sine form, which of itself also results in a decrease

[3]) A theoretical consideration of the efficiency of the mercury vapour discharge may be found in Philips Techn. Rev. **1**, 2, 1936.

of a. Therefore we shall designate a by the name of power factor.

If the power factor of the mercury tube is known and also the efficiencies η_1 and η_2 of filament and mercury discharge, the overall efficiency η of the lamp and the blending ratio M (luminous flux of mercury discharge divided by luminous flux of filament) can be given as a function of the voltage division. It is found that

$$\eta = \frac{V_1\eta_1 + a V_2\eta_2}{V_1 + a V_2} \quad \ldots\ldots \text{(1a)}$$

$$M = a \frac{V_2\eta_2}{V_1\eta_1} \quad \ldots\ldots\ldots \text{(1b)}$$

From this it follows that the blending ratio increases not only with the arc tension V_2 but also with the factor a. Since $\eta_2 < \eta_1$, this means at the same time an increase of the total efficiency η. In order to obtain as high an efficiency as possible, it is therefore desirable to make not only the arc tension but also the power factor as large as possible.

The influence of the power factor on the efficiency and the blending ratio is actually less simple than would be concluded from equation (1). When the arc tension V_2 has been chosen, the voltage drop V_1 on the series arc tension at a given mains voltage V_n is not yet determined but depends upon the power factor. For sinusoidal voltages with equal phase ($a = 1$) the following is naturally valid:

$$V_1 + V_2 = V_n, \quad \ldots\ldots \text{(2)}$$

i.e. the sum of the effective voltage on series resistance and mercury discharge is equal to the effective mains voltage. If, however, there is a phase difference ψ ($\cos\psi = a$) between the voltages on series resistance and mercury discharge, then

$$V_n^2 = V_1^2 + V_2^2 + 2\, V_1 V_2\, a, \quad \ldots\ldots \text{(3)}$$

from which it may easily be deduced that the sum of the effective voltages V_1 and V_2 begins to be greater than the mains voltage. If the value of a is not determined by a phase shift, but by the distortion of current and voltage, as is the case when a filament is employed as series resistance, equation (3) is also found to be valid.

The relation between V_n, V_1, V_2 and a results in the fact that at constant mains voltage and arc tension the power supplied to the mercury discharge not only becomes greater with increasing power factor but at the same time the power supplied to the filament decreases (V_1 becomes smaller). The influence of the power factor on efficiency and blending ratio is thus thereby reinforced.

Considerations from the point of view of lighting engineering and conclusions

Until now in discussing the voltage and power factor we have considered only the economic factors. The question of the blending ratio of mercury light and filament light was also only considered in connection with its effect on the overall efficiency. If now on the basis of the factors discussed we reach a certain compromise, we must still consider whether the result satisfies our requirements from the standpoint of lighting engineering. The mixing of mercury light and filament light was indeed done in order to improve the colour.

On the basis of experience with numerous installations in which filament lamps and mercury lamps are installed side by side, a blending ratio of equal amounts of mercury and filament light may be recommended for general applications. By far the majority of all blended-light installations are designed on this basis. It is thus obvious that in the case of the blended-light lamp the goal should be a blending ratio $M = 1$.

As we have seen from equation (1b), the blending ratio is determined mainly by the arc tension V_2 and the power factor a. The question thus arises whether the desired blending ratio can be realized with an arc tension which is high enough to give a reasonable efficiency but not so high that difficulties may occur in re-ignition. It is now found that in practice the latter requirement is difficult to satisfy: if for example a discharge tube with a power factor $a = 0.7$ is chosen, a blending ratio of only about $M = 0.85$ can be obtained. A further increase of M is only possible by improving the power factor, and it was along this line in fact that the desired goal was attained.

In order to improve the power factor a it was important to investigate the factors upon which a depends. In general it may be said that those factors which facilitate the re-ignition will also improve the power factor. The phase shift between current and voltage becomes smaller, the shorter the dark period lasts.

In agreement with anticipation it was found that with increasing distance between the elec-

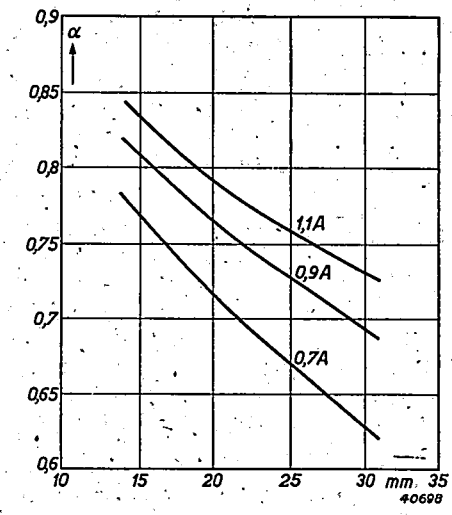


Fig. 4. Power factor a of a mercury discharge tube as a function of the distance between the electrodes at different values of the current, measured on a series of tubes with the same arc tension and diameter.

trodes the power factor improves, as may be seen from the measurements reproduced in *fig. 4*. Furthermore in *fig. 5* the experimentally found relation between the power factor and the diameter of the discharge tube is given. This relation has a less simple character. Different factors probably act in opposition to each other here, as may be concluded from the fact that for a certain diameter an optimum is reached. Since this diameter is fairly large we reach the conclusion that it is desirable to make the discharge tube relatively short and wide. By using this method a blending ratio of 1 : 1 was indeed attained.

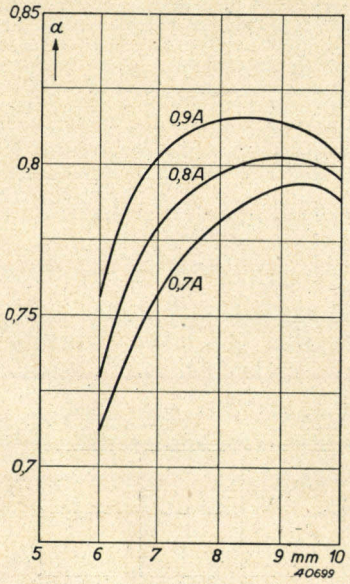


Fig. 5. Power factor a of a mercury discharge tube as a function of the tube diameter at different values of the current, measured on a series of tubes with the same arc tension and distance between the electrodes.

Construction of the blended-light lamp

In addition to the factors already discussed: arc tension, efficiency and blending ratio, in the practical construction of the lamp the power consumed is also important, In general it may be said that mercury lamps of high power can be more easily made than low power units. This peculiarity is illustrated in the fact that mercury lamps for street lighting could be realized much earlier than mercury lamps for interior lighting. In blended-light lamps it is even more difficult to obtain sufficiently small units, since an equal quantity of filament light is added to the mercury light.

The first type brought on the market, which was described in the article cited in footnote [1]), had a power consumption of 250 W and a luminous flux of 500 Dlm. In the meantime a considerably smaller lamp with practically the same efficiency has been successfully developed,

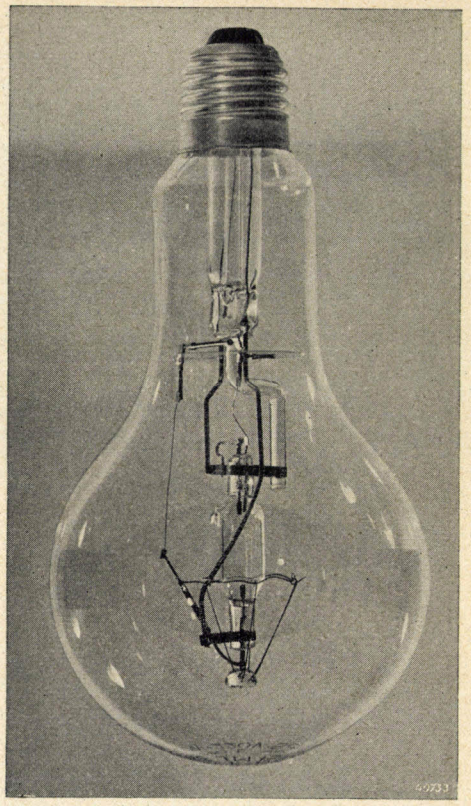


Fig. 6. The blended-light lamp ML 3000. The filament is in the form of a horizontal ring at the height of the arrow.

namely one with a power consumption of 160 W and a luminous flux of 300 Dlm. In *table I* the data of the two lamps are given side by side, while in *fig. 6* the construction of the blended-light lamp ML 300 may be seen.

Table I

Data of the blended-light lamps ML 300 and ML 500

	Unit	ML 300			ML 500		
		Discharge	Filament	Total	Discharge	Filament	Total
Luminous flux	Dlm	150	150	300	250	250	500
Wattage	W	45	115	160	70	180	250
Current	A	0.73	0.73	0.73	1.14	1.14	1.14
Tension	V	78	158	225	78	158	225
Power factor	—	0.79	1.0	0.98	0.79	1.0	0.98

Other properties of the blended-light lamps

As may be seen from the table, the blended-light lamps are designed for a voltage of 225 V. This has been done because of the fact that the effective voltage of so-called 220 V mains usually lies between 220 and 230 V with 225 V as the most frequently occurring value. Filament lamps also are designed for 225 V for various countries where 220/230 V is given as nominal value [4]). In the case of the blended-light lamp, which may be overloaded even less

than a filament lamp [5]), it is of still greater importance that the voltage for which the lamp is designed should be chosen to correspond to the actual voltage of the 220 V mains.

In the interval from 220 to 230 V the blended-light lamps may be used without hesitation. At higher voltages the life of the lamp rapidly decreases, while at lower voltages the efficiency decreases appreciably. If the lamp is burned on a mains voltage which differs from 225 V, current, luminous flux, power consumed and efficiency vary in the manner indicated in *fig. 7*.

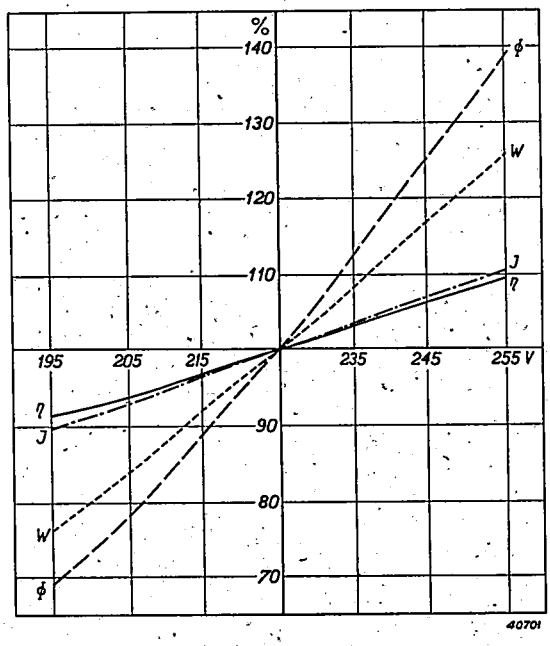


Fig. 7. Variation in current I, luminous flux O, power consumed W and efficiency η upon variations of the mains voltage for the blended-light lamp ML 300.

Under certain circumstances at mains voltages below 200 V, especially in the case of older lamps, difficulties with re-ignition may occur. If (due for instance to the switching on of a heavy load) a temporary large voltage drop occurs on the mains to which blended-light lamps are connected, it is possible that the lamps will be extinguished. The margin allowed for this drop in voltage, however, is large and amounts to about 50 V.

Upon switching on the lamp the applied voltage must be taken up almost entirely by the filament. As the mercury in the discharge tube evaporates the voltage on the filament decreases and the luminous flux from the discharge tube becomes greater, so that the blending ratio, beginning with practically zero, increases as the tube warms up in the manner indicated in *fig. 8*. In *fig. 9* the variation of current, wattage and luminous flux during the heating-up period are plotted.

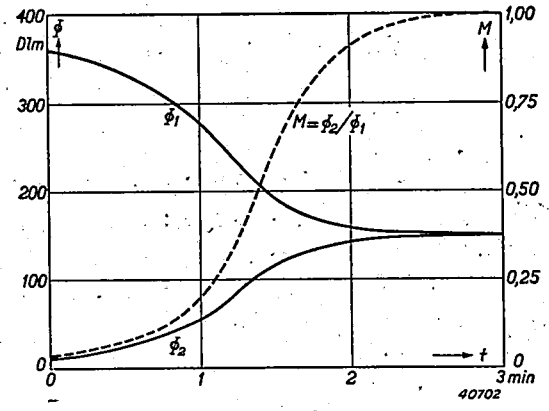


Fig. 8. Variation of the luminous flux Φ_1, resp. Φ_2 of the filament and the discharge tube, and the variation of the blending ratio $M = \Phi_1/\Phi_2$ for the blended-light lamp ML 300 during the heating-up period.

The heavy overloading of the filament each time the lamp is switched on has a strong effect on the life, so that the latter depends upon the time during which the lamp burns between

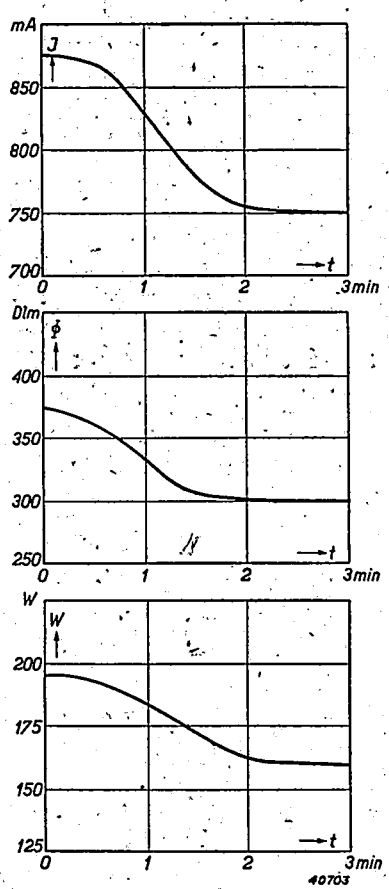


Fig. 9. Variation of the current I, the luminous flux Φ and the power consumed W during the heating-up period for the blended-light lamp ML 300.

[4]) See Philips Techn. Rev. 6, 334, 1941.

[5]) Upon slight fluctuations of the mains voltage the arc voltage of the mercury tube remains practically constant, so that the whole voltage variation must be taken up by the filament. The voltage variation on the filament is thus increased to a relatively higher degree.

two switching operations. In designing the lamp the calculations were made on the basis of an average life of 2000 hours with 700 switching operations. The overloading of the filament has the advantage that directly after being switched on the lamp gives a satisfactory light output; the luminous flux is even greater than that provided during normal use.

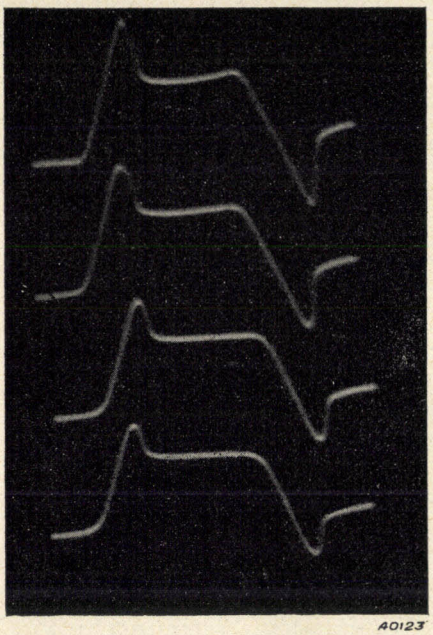


Fig. 10. Oscillograms of the variation of the voltage on a mercury discharge tube which is connected to a constant mains voltage *via* different resistances. It may be seen that the reignition voltage increases with increasing resistance (decreasing current) and that the dark period therefore also becomes longer.

As already noted, the light from the discharge tube is extinguished for a short time (dark period) at each alternation of the current direction. The length of this dark period is determined, among other factors, by the value of the resistance in series. This relation is illustrated by the oscillograms given in *fig. 10*. With increasing resistance the dark period as well as the re-ignition voltage increases.

The occurrence of the dark period gives a strong ripple in the current, which is also manifested in the variation of the light intensity of the mercury discharge. The temperature of the filament follows the fluctuations of the power applied with a retardation (due to its heat capacity) such that the light intensity fluctuates only slightly. The blended-light therefore has a ripple which is considerably smaller than that of the light of a mercury lamp alone. In *fig. 11* oscillograms are given of the light of the discharge tube, the filament and the blended-light lamp.

Due to the fact that the discharge tube is normally used in a vertical position, the greatest

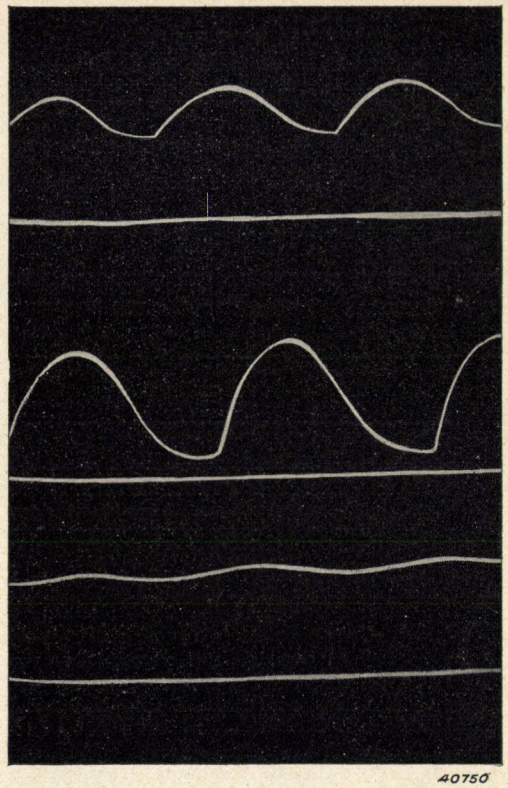


Fig. 11. Oscillograms of the light of the blended-light lamp. *a*) Light from the filament; *b*) light from the mercury discharge; *c*) blended light vertically beneath the lamp. The light from the filament exhibits a smaller ripple than that of the mercury discharge.

light intensity is produced in a horizontal direction, while the filament in the form of a horizontal ring possesses the maximum intensity

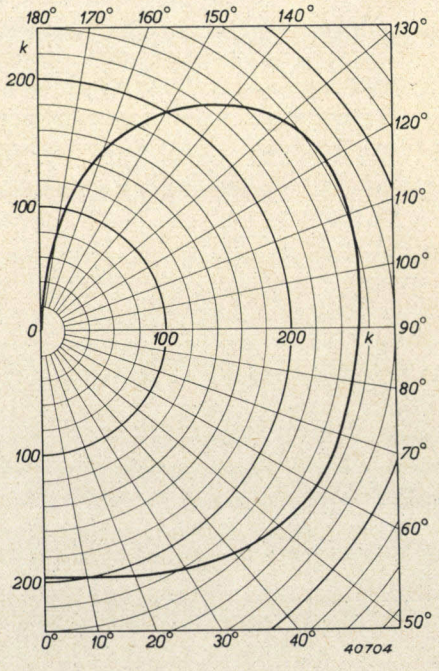


Fig. 12. Candle power distribution of the blended-light lamp ML 300.

in the vertical direction. The candle power distribution of the combination is given in *fig. 12* for a frosted-bulb blended-light lamp, type ML 300.

The possibilities of application of the blended-light lamp were briefly discussed in the article referred to in footnote [1]). Examples given were the lighting of factories, schools, offices, shops and other interiors where large lamp units are desired for illumination. With the new type ML 300 developed since then the possibility of application is appreciably extended in the direction of those cases where smaller lamp units are customarily used.

THE FUNCTIONING OF TRIODE OSCILLATORS WITH GRID CONDENSER AND GRID RESISTANCE

by J. van SLOOTEN.

621.396.615.1:621.396.621.5

In this article an analysis is given of the action of a triode oscillator which is provided in the usual way with a grid condenser and leakage resistance (grid resistance). In particular a study is made of the way in which the A.C. voltage produced and the average anode and grid current vary upon continuous change in the value of the leakage resistance. It is found that the final working state of the oscillator upon excitation of the grid D.C. voltage by means of the circuit elements mentioned may be quite different from that when the same grid D.C. voltage is excited by means of a battery. At the same time the cause of "blocking" is logically revealed.

In the majority of radio receiving sets now in use and in practically all of those now on the market, the superheterodyne principle is applied. The modulation of the high-frequency oscillations of the transmitter received is converted into a constant intermediate frequency by means of a local oscillator, whose frequency is changed from station to station [1]). The very frequent use of small oscillators resulting from this has naturally led to a careful study of their properties and of the special phenomena which are encountered in their use.

A phenomenon which is familiar to all builders of receiving sets is the so-called "blocking" of the oscillator, which has already been discussed in this periodical [2]). As was there shown, the blocking consists of a periodic interruption of the oscillation produced. By itself this phenomenon was already familiar before superheterodyne receivers were built, but for years no really satisfactory information was available as to the circumstances under which it occurs and its actual cause.

This may be explained from the fact that the actual cause has a rather complicated nature and only becomes clear upon a careful consideration of "normal" oscillation. In this article we shall consider this normal oscillation. For reasons which will later become apparent, we shall set about it in two steps, by first considering an oscillator with a control-grid bias which can be permanently set and then discussing the case where the control-grid bias is obtained in the usual way with grid condenser and leakage resistance.

An oscillator with grid bias which can be set

As point of departure we have chosen the connections given in *fig. 1*. In the diagram may be seen a triode, an oscillation circuit consisting of a condenser C and a self-induction L, with which a resistance r must be imagined to be in series, and finally a back-coupling coil in the anode circuit (M here denotes the coefficient of mutual induction with respect to L). The grid voltage V_g is taken from a potentiometer and is measured by the measuring instrument indicated. Further, we assume that the following quantities are measured: the average grid current $\bar{i}_g$, the average anode current $\bar{i}_a$ and the peak value W of the A.C. voltage which occurs on the oscillation circuit as a result of the oscillation.

When we vary the grid voltage by moving the potentiometer, $\bar{i}_a$, $\bar{i}_g$ and W will also change in a definite way. Their dependence is given in *fig. 2*a as measured in a special case. In this figure the static triode characteristic: i_{a_0} as a function of V_g (in the non-oscillating state), is

[1]) On the subject of the superheterodyne principle see Philips Techn. Rev. **1**, 76, 1936.

[2]) Philips Techn. Rev. **3**, 248, 1938 and **5**, 315, 1940.

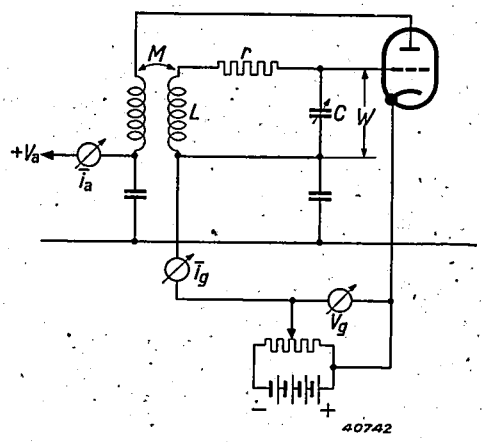


Fig. 1. Customary oscillator connections with triode. On the oscillation circuit, consisting of L, C and r, the A.C. voltage W is developed, M is the coefficient of mutual induction between the back-coupling coil in the anode connection and the circuit self-induction L. On this arrangement the grid of the triode receives a permanently fixed D.C. voltage by means of a potentiometer.

also given, as well as the static grid-current characteristic: i_{g_0} as a function of V_g.

The vertical line on the left indicates that when the negative grid voltage is raised above 40 V oscillation ceases. The anode current then falls to zero and the negative grid voltage V_g must be reduced to 4 V to obtain oscillation once more. For this purpose it is necessary that some anode current i_{g_0} should begin to flow in the non-oscillating state, so that the slope of the valve will reach the minimum value necessary for oscillation.

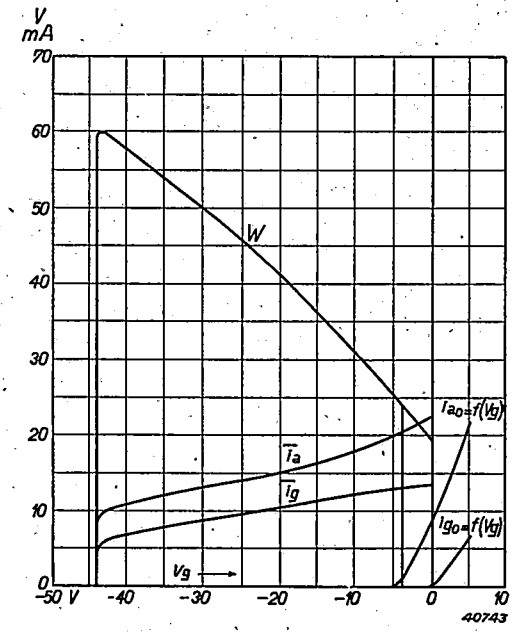


Fig. 2. Characteristics of a triode oscillator. In this diagram the behaviour is represented for a given case of the A.C. voltage W (peak value), the average anode current i_a and the average grid current i_g upon variation of the potentiometer voltage in fig. 1. The oscillation is seen to break off at a definite negative bias (about 44 volts).

We shall not attempt to explain the behaviour of $\bar{i}_a$, $\bar{i}_g$ and W in fig. 2 in detail. We shall only go more deeply into the character of the variation of these quantities in the neigbourhood of the maximum negative grid bias. It is remarkable here that W always continues to increase with increasing grid bias, except for a slight fall just before the interruption of the oscillation; $\bar{i}_a$ and $\bar{i}_g$, on the contrary, become steadily smaller with increasing grid bias. At the grid bias at which the interruption takes place they have not yet, however, decreased to zero, but still have finite values.

These facts may be explained as follows. As may be seen from the static characteristic, the valve is working in a region of grid voltages where the anode current would be zero if there were no A.C. voltage present on the oscillation circuit. In order to cause an anode current to flow there must be a certain A.C. voltage and the amplitude of this required vol age increases with the negative control-grid voltage. Due to this oscillation a finite amount of energy is dissipated in the oscillation circuit; the maintenance of the oscillation thus requires a finite anode current. The fact that the grid current thereby also remains finite will appear from the following considerations.

The oscillator with grid resistance and grid condenser

We now alter the oscillator diagram by substituting for the potentiometer a grid resistance R (often called leakage resistance) shunted by a condenser K, which we shall call the grid condenser. We then obtain the diagram of *fig. 3*. The grid D.C. voltage V_g here is thus formed by the average grid current i_g causing a voltage drop over the resistance R. The condenser K, acts more or less as a buffer, which transforms the irregular current impulses from the grid into a uniform current through the resistance. The current impulses from the grid are thus taken up by K, while a fairly

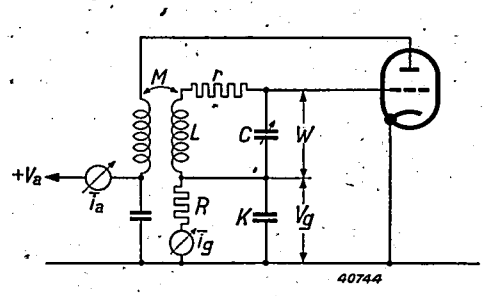


Fig. 3. Oscillator connections in which the fixed potentiometer voltage of fig. 1 is replaced by a variable bias obtained by means of a grid condenser and leakage resistance.

uniform direct current flows through R, which current maintains the level of the negative voltage of K. It is advisable to have this process clearly in mind, especially in connection with the discussion to follow. In the oscillator diagram of fig. 3 we can now obtain different grid voltages V_g by varying the value of the grid resistance R. We shall set about this systematically and vary R continuously from the value zero to the value infinity. We then include at the same time the average values of the anode and grid current i_a and i_j, and, as in the case of fig. 1, the peak value W of the oscillator voltage.

The result of this is reproduced in *fig. 4*.

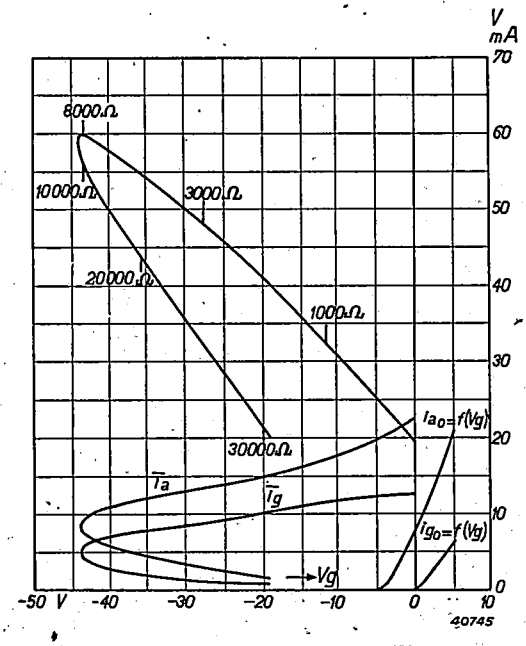


Fig. 4. Characteristics of a triode oscillator according to the diagram of fig. 3. It may be seen that the same curves are obtained as in fig. 2, which represented the behaviour of the circuit of fig. 1, but with an extension which returns to smaller values of V_g.

V_g is here determined, therefore, by taking the product of i_g and R. As may be seen from fig. 4, we first obtain the curves of fig. 2 once more, *while for values of R greater than the value which corresponds to the point where oscillation breaks off in fig. 2, we find an extension of the curves which returns to smaller values of V_g.* As functions of R the A.C. voltage W and the negative grid voltage Vg thus exhibit a pronounced maximum, while i_a and i_g decrease steadily.

Due to this receding character of the curves two adjustments of the oscillator are possible within a given interval for every value of the grid voltage, one of which can be realized with the help of a battery or a leakage resistance smaller than a certain value (9000 Ω in the case in question), while the other is obtained with a leakage resistance larger than 9000 Ω. These two possibilities of adjustment are especially different with respect to the damping of the oscillation circuit due to grid current. If, for example, the two adjustments are considered in the diagram for a negative grid voltage of 20 V, a grid current of $i_j = 10$ mA is found for an adjustment to the upper branch of the curves, while on the returning branch at the same grid voltage a grid current of less than 1 mA is obtained.

The fact that the grid current here is so small is in agreement with the fact that the oscillator amplitude W is only slightly larger than the negative grid bias (namely, $W = 21$ V, when $V_g = -20$ V), so that the total grid voltage swings only slightly into the region of positive values. In the case of the adjustment obtained with the help of a battery, as we have seen, much larger grid currents occur, so that much more energy is necessary to maintain the oscillation. The anode current i_a on the upper branch is then much larger than on the lower branch for a given value of W.

We shall now attempt to explain this double possibility of adjustment on the basis of a diagram. It will then become evident at the same time why with a fixed bias in fig. 1 we could not find the adjustments on the receding part of the curves in fig. 4, and why regular oscillation is no longer possible after a certain value of the leakage resistance R.

Effective slope, required slope and grid current damping

In the discussion of the oscillator with fixed bias (fig. 2) it was pointed out that at grid voltages greater than 4 V negative (*i.e.* to the left of the vertical line) an oscillation is no longer built up automatically when it has been interrupted by some cause or other. For the sake of brevity we shall call this region of the grid bias the C-region, the region to the right of it the A-region, to correspond with the terminology customary in the case of amplifiers: class C and class A.

Oscillation with the grid D.C. voltage in the C-region is thus only possible when the grid voltage is first taken in the A-region and subsequently shifted to the C-region. In the case of fig. 1 this was done by the variation of a potentiometer, in the connections of fig. 3 this process occurs automatically due to the fact that with increasing oscillator voltage the negative grid voltage increases because of the grid current.

The way in which the oscillation is maintained when the grid D.C. voltage lies in the C-region can be understood in the following

way. We assume a constant grid bias of for instance 30 V in fig. 2, and desire to know how the effective slope of the triode depends upon the grid A.C. voltage W. By the effective slope S is then meant the quotient of the first harmonic in the anode current and the grid A.C. voltage (assumed to be sinusoidal). The way in which S_{eff} behaves as a function of W is sketched in *fig. 5*.

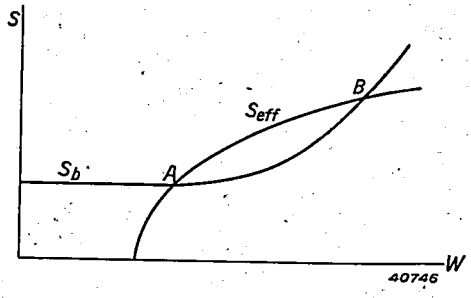


Fig. 5. The slope S_b necessary for oscillation and the effective slope S_{eff} as functions of the A.C. voltage W at a given negative grid D.C. voltage. In the case of this diagram the latter is chosen so large that the valve is in the overbiased state (C-region). For small values of the A.C. voltage W the effective slope is zero.

For small values of W, as long as the peaks of the A.C. voltage do not yet enter the anode-current region of the grid voltage, S_{eff} is zero. S_{eff} then takes on a certain value which grows gradually in order finally to reach a maximum value and then to decrease again if the triode exhibits saturation phenomena (this is not indicated in fig. 5).

Further, it is indicated in fig. 5 how the slope required for oscillation varies as a function of W. This required slope S_b is constant at first, since it is determined by the (constant) resistance. If, however, the A.C. voltage W becomes so large that grid current begins to flow, the oscillation circuit is damped by this grid current and the required slope increases. This explains the shape of curve S_b in fig. 5.

The curves for S_b and S_{eff} in fig. 5 in the case of the behaviour described will in general intersect twice (or not at all). At these points of intersection A and B the required slope is equal to the effective slope, and the oscillator voltage could thus adjust itself to the corresponding values of W.

It is now easy to understand that in the case of the circuit with fixed bias (fig. 1) the oscillator voltage will adjust itself at point B. The point A is an unstable adjustment. If the A.C. voltage becomes slightly greater at point A for example, the effective slope increases more rapidly than the required slope, so that there is an excess of exciting forces over the damping forces, with the result that the A.C. voltage will increase still more and finally reach the operating point B which, it is easy to see, forms a stable adjustment of the oscillator. Conversely, with a slightly too small A.C. voltage the oscillator would cease to generate. We thus reach the result that upon the use of a fixed grid bias, of the two possibilities of adjustment A and B, the adjustment with the larger oscillator amplitude will be chosen.

If we now pass on to the oscillator connections with grid resistance we can explain the character of the curves reproduced in fig. 4 by assuming that on the receding branch of these curves we are concerned with an adjustment of the nature of point A in fig. 5. The question then naturally arises as to the way in which this adjustment just described as unstable has now obtained stability.

Before answering this question, however, we shall consider the stability as a given property and ask what happens to the point of intersection A when the leakage resistance R is made steadily larger.

By changing R the grid voltage is altered and this changes the shape of the curves S_b and S_{eff}, while at the same time there will be a relative displacement of the curves. The highest negative bias which can be realized with the help of a leakage resistance amounts to about 44 V; at this voltage the curves S_b and S_{eff} lie so far apart that instead of two points of intersection there is only point of contact between the two curves [3]).

With decreasing value of the negative bias curve S_{eff} will be displaced towards the left, whereupon the point of intersection A is also displaced towards the left, so that the oscillator amplitude W becomes smaller. If W is drawn as a function of V_g the relation represented in *fig. 6* is obtained for the oscillator under consideration.

In order to draw conclusions from this about the behaviour of the oscillator, the quantity $W - V_g$ is also plotted in fig. 6 as a function of V_g. In the state of adjustment actually reached this quantity must always be larger than zero. If this were not so no grid current would flow, so that the grid voltage acting on the leakage resistance R could not continue to be maintained. This shows that with every leakage resistance the negative grid bias must adjust itself to a value which lies to the left of V in

[3]) If a battery is used for the excitation of the grid bias instead of a leakage resistance, the bias can of course also be made greater than 44 V. The point of intersection or contact, as the case may be, of the curves then disappears entirely, with the result that the oscillation is broken off (see fig. 2).

fig. 6 (in the example investigated about 15 V) and which approaches the limit V_∞ with increasing value of R.

The variation of W as a function of V_g reproduced is actually nothing else but the receding branch of the curve for W in fig. 4. In that case it was impossible to record the whole curve, since upon an enlargement of the leakage resistance beyond about 30 000 Ω the stability at the point of intersection is lost and blocking occurs. The receding branch therefore does not continue to $V_g = V_\infty = -15$ V, but ends already at −19 V.

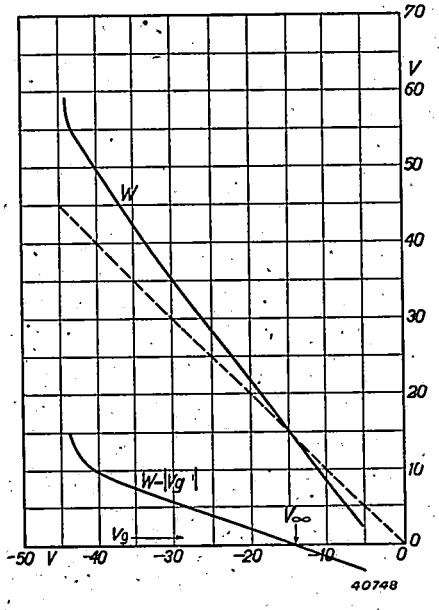


Fig. 6. If the diagram of fig. 5 is drawn for different values of V_g, the abscissa W of the point of intersection A takes on different values. The above curve gives this value of W as a function of V_g. Below is the difference between W and the absolute value of V_g. Stationary states are only possible when $W-|V_g|>0$, *i.e.* for grid voltages to the left of V_∞.

The stabilizing action of grid condenser and leakage resistance

Until now we have considered only those oscillator adjustments at which the negative grid voltage was so high that in the absence of a grid A.C. voltage only very little or no anode current would flow (C-region). The essential point in this is that the effective slope increases with increasing A.C. voltage W (see fig. 5).

If, however, we choose the grid D.C. voltage less strongly negative, so that an anode current already flows even without grid A.C. voltage, the slope then decreases with high A.C. voltage, due to the fact that a continually larger part of the grid A.C. voltage varies in a region with low anode current or an anode current of zero. In this way we obtain the diagram of *fig. 7* instead of that of fig. 5. We then have only one point of intersection of the curves for the required and the effective slope and this of itself already represents a stable adjustment. Thus when the grid bias lies in the A-

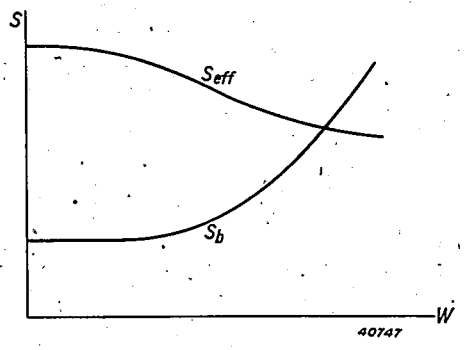


Fig. 7. Diagram analogous to fig. 5. The negative grid D.C. voltage, however, is chosen much lower, so that maximum slope occurs at low values of the A.C. voltage W (A-region).

region there is no essential difference between oscillation with fixed bias and oscillation with grid condenser and leakage resistance. Therefore the discussion of adjustments with the grid D.C. voltage in the C-region is sufficient.

The explanation of the fact that the unstable adjustment A of fig. 5 may become stable upon the use of an R-C circuit is the following. Let us again assume a slight deviation, for instance an increase in the A.C. voltage W. In the first instance this causes an increase of the effective slope, which would lead to a further increase of W. On the other hand a larger value of W means more grid current and, as a result, an increase in the negative grid voltage. As a result of this, however, the slope decreases again, so that the unstable growth of the oscillator amplitude is opposed.

It is now the question whether such a reaction may have as a result that an adjustment which is of itself unstable becomes stable. It may reasonably be expected that this will depend upon the speed with which the reaction takes place and thus on the magnitude of K. When the condenser K is made so large that upon short-lived fluctuations of the working state it acts practically like a battery with constant voltage, there is no question of a stabilizing influence.

In the opposite case of a very small value of the capacity K, on the other hand, the stability of the working state is easily demonstrated. In this case we may assume that every change in the oscillator voltage W is immediately accompanied by a certain change in the bias V_g. The equilibrium between the required slope S_b and the slope actually present S_{eff} can now be described by a diagram analogous to that of fig. 5 or fig. 7, but with the difference that the grid D.C. voltage does not have a constant

value but at each oscillator amplitude W is equal to the value which is automatically adjusted for this. With increasing value of W we then find a regularly decreasing value of S_{eff}, since the negative grid voltage becomes steadily greater while the peaks of the A.C. voltage extend only relatively slightly into the grid current region. The required slope S_b is now practically constant, since the grid current damping is slight and, moreover, almost constant.

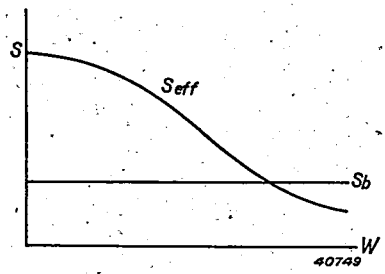


Fig. 8. Diagram analogous to fig. 5 and fig. 7. Now, however, the grid D.C. voltage is not considered to be constant, but a function of the A.C. voltage W, namely that D.C. voltage which is automatically attained by means of the grid condenser and leakage resistance at each corresponding value of W.

We thus obtain the diagram for S_b and S_{eff} which is shown in *fig. 8* and which qualitatively is analogous to fig. 7. Just as in fig. 7, we may also conclude that the point of intersection represents a stable state. It must be repeated that the stability of the adjustment is not yet proved for the general case by this consideration, since the diagram of fig. 8 is only correct when at each value of W the grid voltage is immediately equal to the value which would be reached in the course of time. If this adjustment proceeds too slowly, a state on the receding branch of the curves of fig. 3 finally becomes labile. Fig. 8 does indeed prove that at a given value of R there is only one possible stationary adjustment. This single adjustment can therefore still be unstable.

For the present we may summarize the foregoing discussion in the form of the following conclusion:

When in the case of an oscillator with grid condenser and leakage resistance the back-coupling is increased so much that the grid D.C. voltage lies in the C-region (see the preceding definition), with the customary large values of the leakage resistance the adjustment of the oscillator is one which would be unstable with a fixed grid voltage, but which has become stable due to the action of the grid condenser and leakage resistance. If this stabilizing action is insufficient, due for instance to the fact that the grid condenser has been chosen too large, the only possible adjustment becomes unstable and blocking occurs.

A careful study of this "wild" oscillation or blocking falls outside the scope of this article, but its actual cause will have become sufficiently clear from the foregoing. In a following article we shall supplement this qualitative discussion by a quantitative treatment of the problem of stability, in which at the same time more light will be shed on the behaviour of the oscillator in the unstable state. In addition the measures will also be studied which may be taken against blocking.

THE TEXTURE OF NICKEL-IRON STRIP

by J. F. H. CUSTERS.

620.18 : 669.15.24

After a discussion of the method of investigation and graphical representation of texture in a previous article, the texture of nickel-iron strip for loading coils is here discussed in several states of working as a first example of the study of textures.

Many properties of polycrystalline metals are considerably affected by their so-called preferred orientation or texture, *i.e.* by the way in which the crystallographic axis directions of the single-crystalline grains of which the metal is built up are oriented. In a previous article which appeared in this periodical [1]) it was explained how the texture of a piece of metal can be determined by means of X-rays and how it can then be represented graphically. In this article we shall consider as an example the texture of nickel-iron strip in various conditions of working. Nickel-iron strip which has undergone certain manipulations and heat treatments has, as previously mentioned in this periodical, an important technical application as core material for loading coils [2]).

[1]) J. F. H. Custers, A consideration of the texture of metals, Philips Techn. Rev. **7**, 13, 1942.

[2]) J. L. Snoek, Magnetic cores for loading coils, Philips Techn. Rev. **2**, 77, 1937. See also W. G. Burgers, Philips Techn. Rev. **2**, 93, 1937. Further: G. W. Rathenau and L. J. Snoek, Physica **8**, 555, 1941.

The manipulations referred to are briefly the following: Nickel and iron are alloyed and an ingot with coarse-grained structure is obtained. By several deformation and heating processes a fairly fine-grained strip is obtained from this ingot, which is about 1 cm thick and exhibits no texture at all: The strip is then further cold-rolled in many steps until the thickness of the strip is finally only 0.1 mm. As a result of this rolling the strip has undergone enormous deformations and has become even more fine-grained. Many properties have been very much altered compared with the original strip, which was 100 times as thick. The hardness has increased considerably, the resistance to bending and folding has been changed not only in value but also in dependence on the direction, while the tensile strength has also undergone similar changes.

Before the strip is ready for use it must still be subjected to two operations, namely heating and rolling once more. It is, however, interesting, and it was found also to be of practical importance, to make an investigation of the texture of the material in this intermediate stage of working. Upon projection of the reference sphere on a plane parallel to the plane of the strip, the pole figure given in *fig. 1* is then found for the cube planes of the crystal lattice.

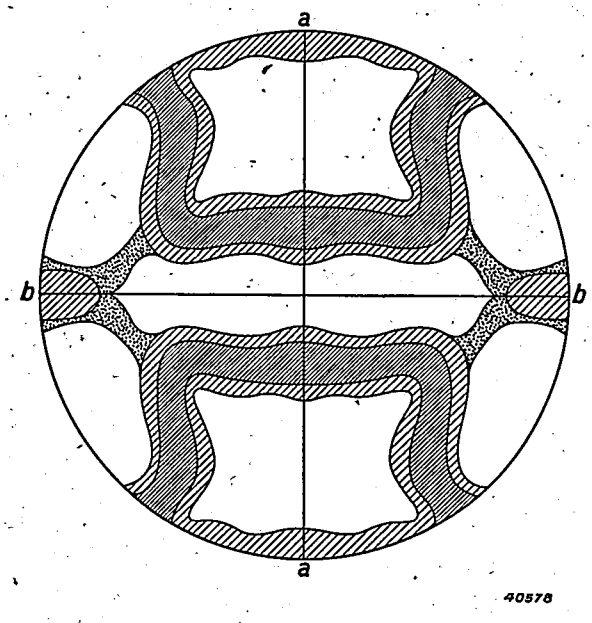


Fig. 1. Pole figure for the cube planes in the case of the rolling texture of nickel-iron strip. The projection plane of the pole figure (plane of the drawing) is parallel to the plane of the strip; *a-a* is the direction of rolling. The blackened parts represent the statistical distribution of the directions of the normals to the cube planes in all the cristal grains of the strip. The density of "occupation" is here indicated in four steps. The closely-spaced shading indicates the greatest, the wide shading a less density, the dots the smallest density of occupation which can be observed. The white parts are unoccupied.

As was described in the article refered to in[1]), a texture is represented graphically by considering a certain kind of lattice planes — for instance the cube planes — of the piece of metal, drawing for each crsytal grain the normals to those lattice planes and allowing them to pierce a sphere drawn about the specimen, the reference sphere. The more or less pronounced pattern of points obtained on the sphere, which furnishes an image of the texture of the piece of metal, is represented graphically, by stereographic projection of the reference sphere, in a pole figure, In the case of a given object with an unknown texture the pole figure can be determined by means of a series of X-ray diffraction photographs with monochromatic radiation which are taken in a series of slightly rotated positions of the object under investigation. In *fig. 2* such a series of photographs is given of the rolled nickel-iron strip, whereby the beam of X-rays was directed perpendicular to the direction of rolling of the strip and perpendicular to the surface of the film, while the strip, starting from a position parallel to the surface of the film, was turned 5° farther each time around the direction of rolling. On each photograph a heavy blackening can be seen (white on the positive) of certain parts of the different Debije-Scherrer circles, which circles correspond to the different lattice planes of the crystal lattice. If in all the photographs we consider only the circle for the cube planes — in fig. 2 this is in each case the second circle from the centre, the first is that for the octahedron planes — the pole figure can be constructed from the blackened segments in the successive exposures in the manner previously described.

The pole figure in fig. 1 is closely related to that found for other metals after rolling which, like nickel-iron, crystallize in a face-centred cubic lattice, aluminium and copper for example. It is therefore called a "rolling texture". The preferred positions which this texture involves and about which the separate grains are arranged with a certain scattering can be described in this case in different ways. A good approximation is obtained when the two positions of the cube faces indicated in *fig. 3* by a model of a cube with respect to the direction of rolling *a-a* are considered preferred positions [3]). To these positions correspond the poles drawn in *fig. 4*, and this shows that upon the assumption of a certain scattering in the actual positions of the grains about two ideal preferred positions, so that more or less extensive regions occur in the pole figure instead of sharp points, the pole figure of fig. 1 can be obtained. According to our own experience, however, considering the phenomena which occur upon heating the strip[4]), it is better to consider the rolling texture as characterized by the preferred position indicated in *fig. 5*. The latter as well as the positions symmetrical to it in the strip are represented in the pole figure by the dots drawn in *fig. 6*, and it may be seen that upon the assumption

[3]) See Frhr. v. Göler und G. Sachs, Z. Phys. **41**, 873, 1927.

[4]) See J. F. H. Custers and G. W. Rathenau Recrystallization in rolled nickel-iron, Physica **8**, 759-770, 1941 and J. F. H. Custers, Über die (111)-Reflexe im gewalzten und rekristallisierten Nickeleisen, Physica **8**, 771-788, 1941.

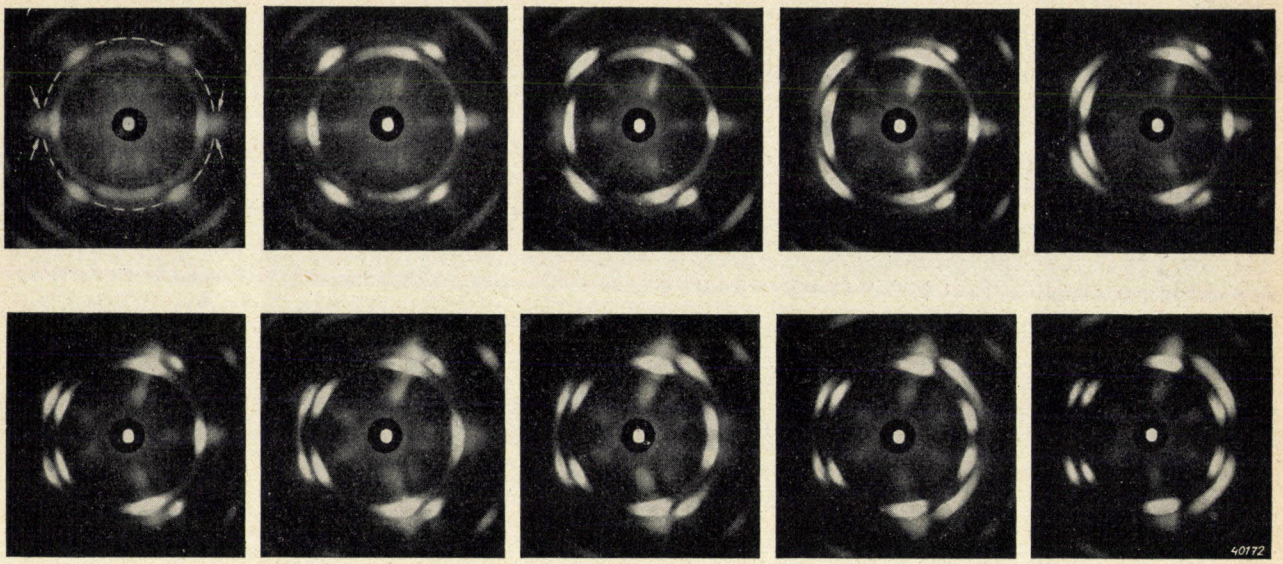


Fig. 2. X-ray diffraction photographs of rolled nickel-iron strip. In the first exposure the plane of the strip was parallel to the plane of the film, in the following exposures the strip was turned 5° farther in each case about the direction of rolling (direction from top to bottom). The blackened parts of the different Debije-Scherrer circles indicate where the so-called reflection circle which corresponds to each position of the strip (see the article referred to in footnote [1])) passes through regions on the reference sphere of the piece of metal which are covered with points of intersection. If in all the exposures the Debye-Scherrer circle for the cube planes (indicated by a dotted circle in the first photograph) is considered, the pole figure of the cube planes given in fig. 1 can be derived from the whole series. The spots indicated in the first exposure by arrows, which are also visible in part of the following exposures, are due to grains in the so-called cube position.

here also of a certain scattering the pole figure actually found for the rolling texture is satisfactorily approximated. The position indicated in fig. 5 is of itself rather difficult to describe in words [5]). It must, however, be kept in mind that for describing the texture it is not actually necessary to be able to give

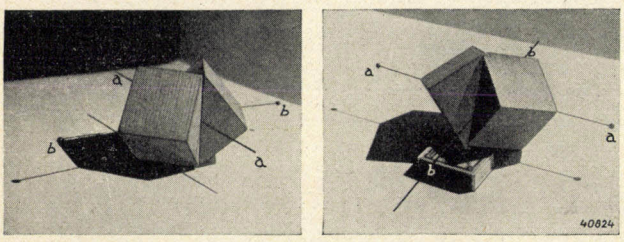


Fig. 3. According to Sachs [3]) the rolling texture of nickel-rion strip may be considered as originating from the two ideal preferred positions of the cube planes here represented in a model with respect to the direction of rolling *a-a* and the transverse direction *b-b* of the strip.

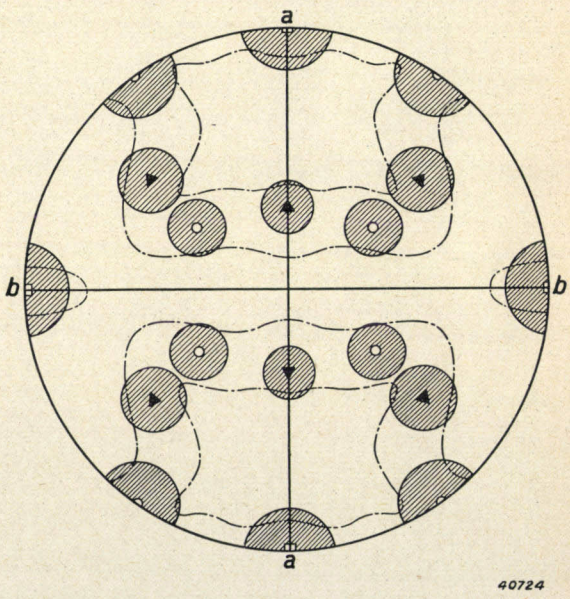


Fig. 4. The poles of the positions of the cube planes shown in fig. 3 form such a pattern in the pole figure that, with the assumption of a certain scattering around these positions, the pole figure of the rolling texture (fig. 1) is quite closely approached. The poles indicated by small triangles belong to the left-hand position in fig. 3, the small circles to the right-hand position. Around each pole a circle is drawn which corresponds to a scattering of 10° around the position in question. The rectangles with their circles do not belong to the rolling texture proper, but to the cube position.

[5]) To do this one must begin with a crystal whose three axes coincide with the direction of rolling, the direction of the normal and the transverse direction of the strip perpendicular to the first two directions, Let a twin be formed on such a crystal (in the so-called cube position), *i.e.* in the growth of one of the octahedron planes the crystal "by mistake" builds up an atomic layer in hexagonal closest packing instead of in cubic closest packing, then proceeding with cubic closest packing again the crystal is thereby rotated 60° about the normal to the octahedron plane. The preferred position in question now occurs when this twin is further rotated through a small angle (about 8°) about the normal mentioned.

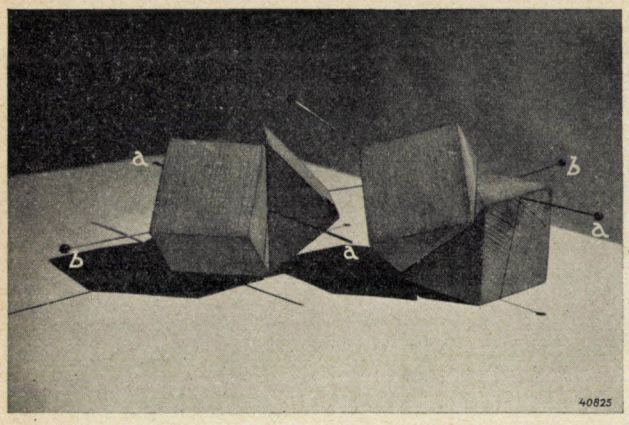


Fig. 5. According to our own investigations the rolling texture can best be described by the preferred position of the cube planes here illustrated.

names to the "ideal" preferred positions ; the positions in which the grains actually stand and the statistical distribution of their positions is indeed already completely represented by the experimentally determined pole figures. The recognition of the ideal preferred positions is only important when it is desired to make a further investigation of the mechanism of the occurrence of a given texture, or when a relation is sought on the basis of crystallographic data between the texture and quantitative differences in properties of different materials. In this respect, however, little can yet be said about the rolling texture.

We have already stated that the rolled strip is not yet ready for use, *i.e.* the rolling texture is not yet the desired texture. If the strip is now heated to a sufficiently high temperature and the texture then examined anew, a complete change is found to have taken place. The grains which we now encounter in the strip are much larger than in the rolled strip and nearly all of them are in such positions that the axis directions coincide with the rolling direction, the direction of the normal and with the transverse direction of the strip respectively (*fig. 7*). The scattering of the positions of the grains about this so-called cube position is relatively slight, so that the strip may be considered almost as a single nickel-iron crystal. The appearance of this new texture must be due to the growth of certain crystal nuclei thanks to the thermal agitation of the atoms (recrystallization). Due to the rolling, germs have indeed been formed in the nickel-iron strip which are in the cube orientation and which may function as nuclei for recrystallization. This may clearly be seen in the photographs of fig. 2 : the Debye-Scherrer circle for the cube planes in the first five or six exposures in which the strip is approximately parallel to the film contains a visible blackening at the points of intersection with the horizontal diameter of the circle, which is what occurs in the case of grains in the cube position. The reason why germs in the cube orientation have such a capacity for growth that in the recrystallization they consume all the grains not in this position, is not yet fully explained ; there is probably some connection with the magnitude of the energy of deformation which is stored up in the grains in the rolling process. Whatever the case may be, the cube orientation resulting from the recrystallization is found to be very sharp, as is shown by the X-ray diffraction photographs *fig. 8* and the pole figure *fig. 9* obtained from them. At the same time the properties have also changed compared with those of the rolling texture ; for example the material has become much softer again. In particular, however, it is found to begin to exhibit a property which is especially desirable for its use in loading coil

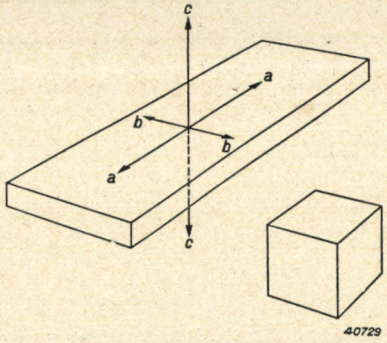


Fig. 7. In the so-called cube position the three cube axes of the nickel-iron crystals are parallel to direction of rolling *a-a*, transverse direction *b-b* and direction normal to the strip *c-c*, respectively.

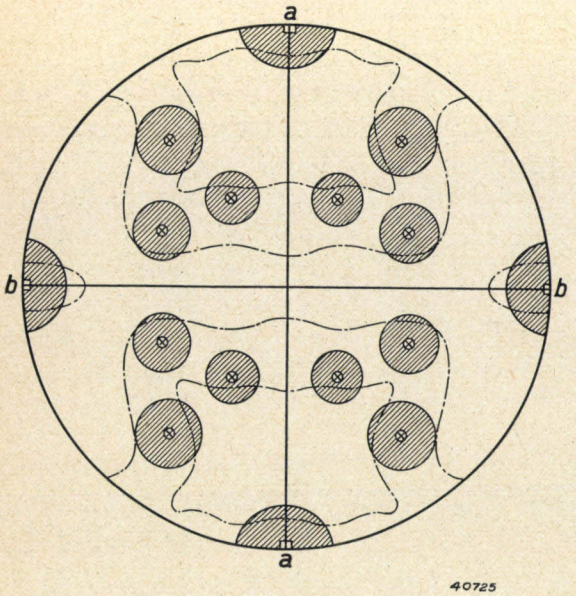


Fig. 6. The poles of the cube planes in the position of fig. 5, as well as the corresponding symmetrical positions, together with their circles of scattering, coincide satisfactorily with the pattern of the pole figure fig. 1.

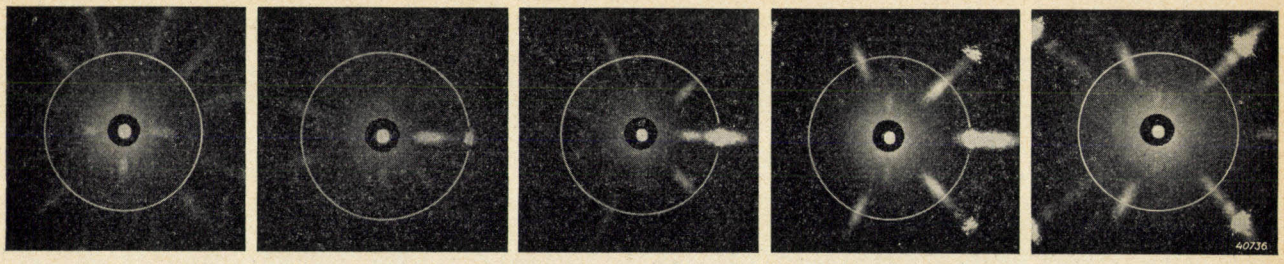


Fig. 8. X-ray diffraction photographs of recrystallized nickel-iron strip taken in the same way as in fig. 2. The texture is entirely changed : almost all the grains are in the cube position, as is evident from the intenseblacken ing on the circle of the cube planes in the third and fourth exposures of the rotated strip. The fact that in the first exposures (perpendicular incidence of the X-rays) the blackenings of the cube orientation, which were indicated in fig. 2 with arrows, are lacking, indicates precisely that the texture is very sharp : in the case of the position of 0° the reflection circle does not yet pass through the thickly occupied regions in the reference sphere and only moves through their middle at the position from about 5 to 15°, while with the less sharply pronounced cube position of fig. 2 the blackened regions on the reference sphere are so extensive that at the position of 0° the reflection circle already passes through them.

and which appears fully after a further final operation, namely rolling out to about half the thickness. The strip then has a strong mono-

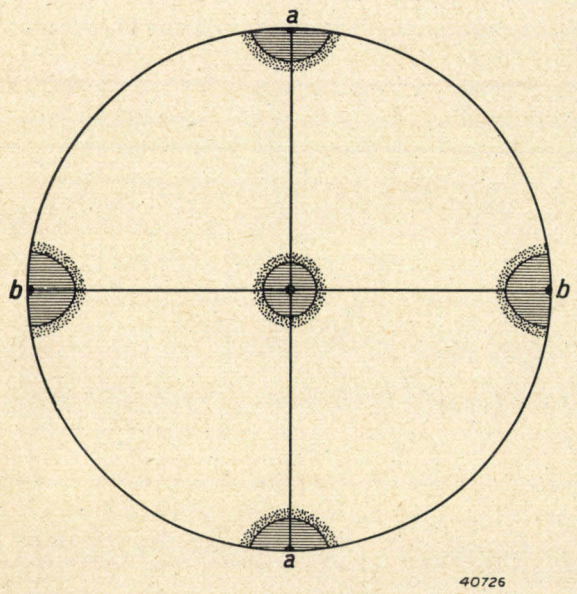


Fig. 9. Pole figure for the cube planes derived from fig. 8. The texture (cube position texture) is very sharp, the material behaving almost like a single crystal.

axial magnetic anisotropy : it can be magnetized with much more difficulty (*i.e.* only by stronger fields) in the direction of length (and that of the normal) than in the transverse direction. In consequence of this, with suitable assembly in the self-induction coil, the hysteresis and the so-called instability of the core, which are undesirable for loading colis, can be kept very low (see the first article referred to in footnote [2]).

The part actually played by the last operation in the appearance of the strong magnetic anisotropy has not yet been explained. The texture is not thereby appreciably changed, as may be seen upon comparison of the two X-ray diffracton photographs of *fig. 10*. The scattering in the positions of the grains about the cube position has only become somewhat greater.

Although it may be seen from the above that in this case — and in other cases it is often the same — there are still many questions to be answered, it will, nevertheless, have become clear that theoretically and practically it may be of importance in working a metal to study its texture at different stages in the proceedings and to attempt to connect the texture with the properties of the material.

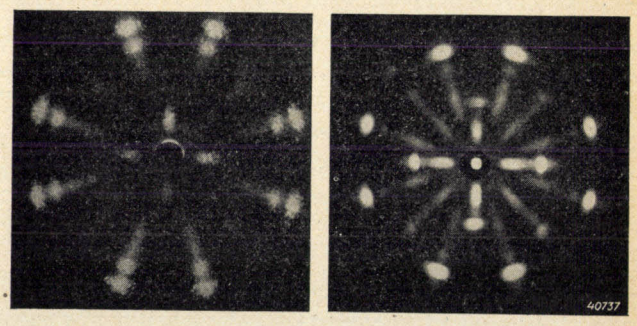


Fig. 10. After the last operation (final rolling) the texture of the strip is still the same as after the recrystallization, but somewhat less sharp, as may be seen upon comparison of these exposures made with perpendicular incidence: left before, right after rolling. The left-band exposure is in fact identical with the first of fig. 8. Due to the greater scattering after rolling, the blackenings for the cube orientation are already visible upon perpendicular incidence of the X-rays.

A RECORDING APPARATUS FOR THE ANALYSIS OF THE FREQUENCY OF RAPIDLY VARYING SOUNDS

by H. G. BELJERS.

534.441.2

By the use of a series of band filters whose transmission regions are distributed over the whole range of acoustic frequencies an accurate and rapid analysis of the frequency of speech and other rapidly varying sounds is possible. An apparatus based upon this principle is described here, with which the Fourier spectrum of the sound to be investigated is made directly visible on the screen of an electron ray oscillograph, so that the variations of the spectrum can be filmed. The properties of the apparatus, especially the resolving power and recording speed attainable are discussed, as well as a number of particulars of construction and use. In conclusion several spectrograms of vowels filmed with the apparatus are reproduced as examples of its use, and they are briefly discussed.

In the case of electro-acoustic apparatus for the transmission of speech or music it is customary to describe the behaviour of the whole or of the separate elements by the way in which they react to sinusoidal oscillations of different frequency. This custom is based upon the fact that every sound vibration can be resolved into a number of sinusoidal components (Fourier components), so that from the frequency characteristics of an element (amplifier, cable or the like) it is possible to deduce immediately the form in which the complex vibration will be transmitted.

There are many numerical, graphic and instrumental methods for the separation of a vibration into its Fourier components, *i.e.* methods of harmonic analysis. For the most part, however, these methods are based upon the use of functions which are given in the form of a diagram or a table. If, therefore, it is desired to analyse sound vibrations by these methods, the sound vibrations must first be recorded. Moreover, the frequency spectrum of sound vibrations, that of speech for example, continually changes, and it is often just these changes in which we are interested. In order to investigate these changes a large number of strips of the recorded sound would each have to be analysed separately, which is a very laborious method scarcely deserving practical consideration.

With these objections in view special methods of separation have been developed for electro-acoustics which correspond better to the existing requirements and possibilities in this field.

A very obvious method is to make use of a band filter as analysing element. If by means of a microphone the sound vibration is converted into an electrical A.C. voltage and this is fed successively to a number of band filters with different transmission regions, then from the occurrence or absence of an output signal it can be deduced in what frequency regions components of the vibration being investigated lie, and how strong they are. Obviously so many filters must be used that the desired fineness of structure of the whole acoustic spectrum is obtained.

In order to avoid the necessity of a large number of band filters, which make the apparatus complicated and expensive, the following device is usually employed. A sinusoidal voltage whose frequency can be continuously varied (auxiliary frequency), is fed, together with the signal to be investigated, to a mixing valve followed by which is a single band filter. If the difference frequency (or sum frequency) of the auxiliary frequency and one of the frequencies present in the signal falls exactly in the narrow transmission region of the band filter, an output signal is observed. From the values of the auxiliary frequency at which this occurs the Fourier spectrum of the signal can therefore be deduced.

Only one band filter is needed here, but two disadvantages are also involved: the measurement takes some time and as a result a fairly lengthy constancy of the voltage to be investigated is required. Therefore the auxiliary frequency method can indeed be used, for example, for determining the deformation which a given constant input signal undergoes in an apparatus, but not for the analysis of speech or other rapidly varying sound.

For an analysis of these sounds, therefore, recourse must be had to the fundamentally simpler method in which separate band filters are used for the different components of band filters required.

Equipped with such a set of analysing elements, it is now also possible to make the Fourier spectra visible directly in a simple way. For this purpose the outputs of the whole series of band filters are connected successively with an electron ray oscillograph by means of a rapidly rotating switch. By means of suitable connections the spectrum of the sound to be analysed can then be observed directly on the fluorescent screen, and its variations seen or recorded on a film.

An apparatus working on this principle and constructed in this laboratory will be described in this article [1]), while in conclusion some results obtained with the apparatus will be dealt with.

Description of the apparatus

In *fig. 1* the whole apparatus is shown diagrammatically. It contains 79 band filters (F), whose transmission regions are distributed over the range of frequencies from 90 to 8000 c/s. The signal to be investigated is fed to the inputs of all the band filters connected in parallel. Since each filter can only transmit a very small part of the energy of the signal and the output voltage of each filter must, nevertheless, project sufficiently above the ordinary level of interference, the signal is amplified to the necessary level (4 W) in the pre-amplifier V_1, which because of inverse feed-back causes only a very slight distortion.

Behind each filter there is a diode rectifier which rectifies the output voltage of the filter. Each of these rectifiers is connected to one of 79 contacts on a collector, which contacts are successively scanned by a rotating switch E_1. In this way the rectified voltage of each band filter in turn, which voltage is a measure of the corresponding component of the Fourier spectrum, is fed to the vertical deflection plates of a cathode-ray oscillograph (the elements G and M in fig. 1 will be disregarded for the present). At the same time a D.C. voltage increasing in steps is fed to the horizontal deflection plates. This voltage is taken from 79 taps of a potentiometer by the switch E_2 rotating in synchronism with E_1. The potentiometer is supplied from a source of D.C. voltage. When this is done a definite horizontal deviation of the fluorescent spot on the oscillograph screen corresponds to each band filter, and therefore a frequency spectrum is traced directly on the screen.

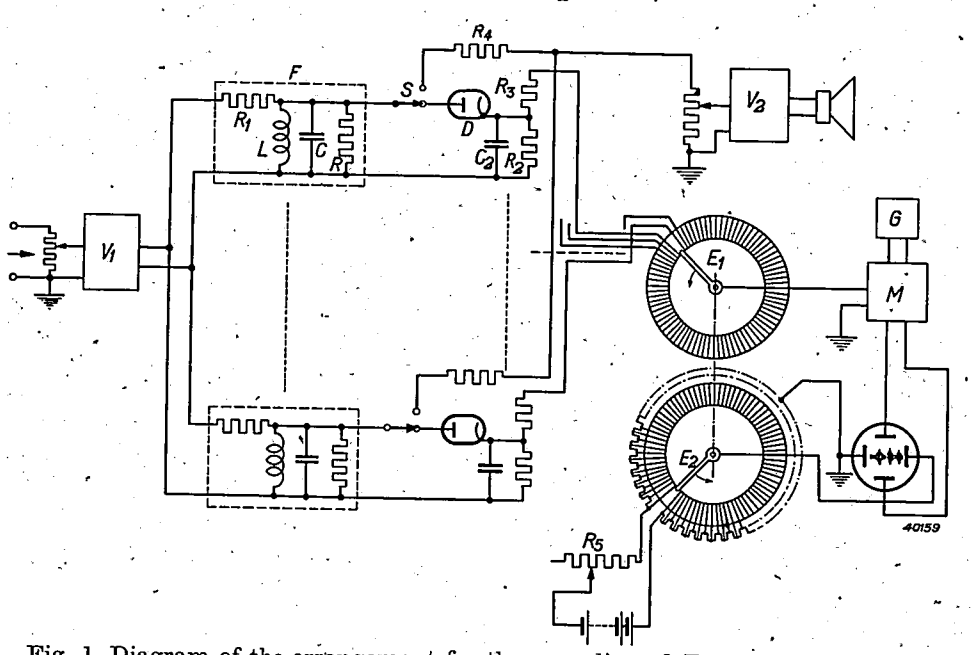


Fig. 1. Diagram of the arrangement for the recording of Fourier spectra. The sound to be analysed reaches the inputs of the 79 band filters *via* the pre-amplifier V_1; the part of the signal transmitted by each band filter is rectified with a diode D and fed *via* the rotating switch E_1 to the modulator M, which supplies a modulated carrier wave of high frequency to the vertical deflection plates of an electron-ray oscillograph. With the rotating switch E_2 a D.C. voltage increasing in steps is fed at the same time to the horizontal deflection plates. The length of the spectrum which appears in this way on the oscillograph screen is regulated with R_5; V_2 is a monitor amplifier which can be connected *via* resistances R_4 by 79 switches S to the output of each band filter.

If, as here described, the rectified output voltages of the band filters were fed directly to the oscillograph, the frequency spectrum would become visible in the form of a series of points. This would be difficult not only for the measuring of the frequency (*i.e.* the number of the band filter), but also for the measuring of the intensity of the components. Therefore an A.C. voltage of high frequency (50 kc/s) is actually fed to the vertical deflection plates, which frequency is excited by a generator G, and its

[1]) An acoustic spectrometer constructed by Siemens & Halske is based on the same principle (E. Freystadt, Z. techn. Phys. **16**, 533, 1935). In that apparatus, however, the division of the spectrum was less fine than in our case (namely 3 filters per octave).

amplitude is modulated in a modulator M with the rectified output voltages of the band filters. In this way vertical lines are traced on the fluorescent screen instead of separate points. In *fig. 2* such a spectrum is shown. This method also has the advantage that the apparatus could be made less sensitive to interferences. Behind the modulator there is another band filter which transmits only a narrow frequency region around the generator frequency of 50 kc/s. Interfering voltages of other (not too low) frequency, which may be induced on the connections between the rectifiers and the modulator, are filtered out in this way, while any distortion products and the noise are rendered practically harmless.

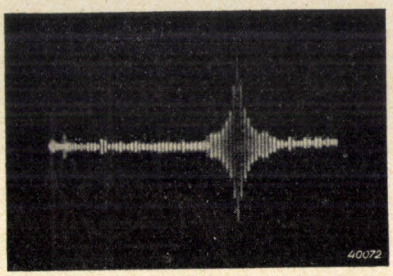


Fig. 2. Picture obtained on the oscillograph screen upon the application of a sinusoidal voltage to the input of the apparatus, when the frequency coincides with the resonance frequency of one of the filters.

The spectrum obtained on the screen, which in the case of speech for example continually varies, can be observed visually, or it can be photographed at short intervals on a moving film.

The action of the band filters can be checked by means of a monitor amplifier V_2 with loud speaker, which can be connected to each band filter in the place of the rectifier by means of a set of 79 switches S. By reversing several switches or whole groups of them the sound in different frequency regions can be heard and and the influence of the lack of certain frequencies on intelligibility can be studied.

After this brief description we shall now go into several important characteristics and structural details of the apparatus.

Resolving power and recording speed

It would be desirable to be able to determine exactly not only the frequency of each component but also the variations of its intensity with time, *i.e.* to be able to follow accurately the growth and disappearance of each component. The accuracy which can be attained here, however, is fundamentally limited by a kind of "relation of uncertainty". The frequency of a sinusoidal vibration can only be determined precisely when the vibration lasts for an infinite time. With shorter duration it is impossible to speak of one definite frequency of the vibration, but it must be ascribed to a spectrum of finite width, as appears from the theory of Fourier integrals. This width becomes greater the shorter the vibration lasts. Therefore the more rapidly a spectrum varies, the less sharp will the frequencies be determined.

In measuring, of course, only a section of finite duration of the vibration in question can be considered, so that even with an infinitely long vibration we cannot determine the spectrum perfectly sharply. But, moreover, the time interval must expressly be chosen short when rapid variations of a spectrum are to be observed; the organ reacting to the vibration (ear, filter or general measuring instrument) must "forget" again the preceding effects quickly enough. From a consideration of band filtersit is clear how in this case the antagonism mentioned occurs between the accuracy of the measurement of the frequency on the one hand (resolving power) and the recording speed on the other.

As indicated in fig. 1, the filters consist of single L-C circuits which are so damped by a resistance R in parallel that at resonance the impedance is 20 000 ohms. This is also the value of the preceding resistance R_1. The ratio a between output and input amplitude (transmission factor) of a vibration of any given frequency f with such a filter is

$$a = \frac{\mathrm{I}}{2 + jq\left[\left(\frac{f}{f_0}\right)^2 - 1\right]} \quad . \quad . \quad . \quad . \quad 1)$$

In this expression $f_0 = 1/2\pi\sqrt{LC}$ is the resonance frequency of the filter and $q = R/2\pi f_0 L$ is the so-called quality factor of the circuit (the loss resistance of the self-induction L is accounted for in the resistance R). The larger q, the steeper the resonance curve given by (1) falls away on each side of f_0 (*fig. 3*).

Let us suppose that a sinusoidal voltage is suddenly applied to the input of such a filter. In addition to a forced oscillation whose intensity can be calculated from (1), there then occurs a free oscillation with the frequency f_0 which gradually dies out according to

$$e^{-t/RC}.$$

The time $\tau = 2RC$ after which the intensity of the free oscillation has fallen by a factor $1/e$ is called the decay time of the filter. It will obviously be useless to measure the output voltage with a varying input A.C. voltage of the filter at intervals which are not at least equal to τ (preferably still much longer). Since the following is true:

$$\tau = 2RC = 2\frac{R}{2\pi f_0 L} \cdot 2\pi C L f_0 \cdot = \frac{q}{\pi f_0} \qquad (2)$$

the more rapidly it is desired to record, *i.e.* the smaller the decay time r is to be made, the smaller the quality factor q of the filters must be made, *i.e.* the flatter the resonance curves of the filters must be. Therefore the less accurately is the frequency of the transmitted signal determined.

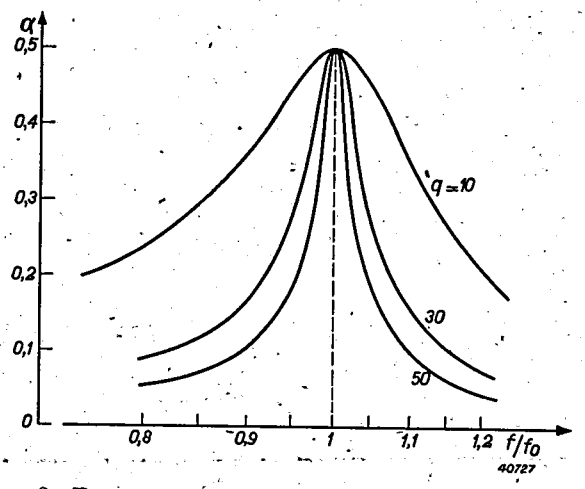


Fig. 3. Resonance curve $a(f)$ of the band filter for different values of the quality factor q.

From equation (2) it may at the same time be seen that at a given value of q the band filters have a shorter decay time for the high frequencies than for the low. For the high frequencies therefore a more rapid recording is possible than for the low frequencies. Since, when the whole spectrum is recorded with a single film, the same recording speed must be used for all frequencies, it would seem reasonable to construct all the filters with the same decay time, so that in the case of the filters for high frequencies it is possible to work with a high value of q, *i.e.* a sharp resonance curve and correspondingly great resolving power. We have not done this, however, because the sharper the resonance curves of the filters are made, the greater the chance that a Fourier component whose frequency lies just between the resonance frequencies of two neighbouring filters will remain unobserved. Since this must be avoided as far as possible the intervening regions between the filters must be made smaller, *i.e.* with a sharper resonance curve a larger number of filters is necessary.

In order not to be compelled to use too large a number of filters we chose the relatively small value of 32 for the quantity q. If we use a recording speed of 20 pictures per second, which will be sufficient for most investigations, the decay time will only be longer than the recording time for the filters with $f_0 < 32.20/\pi \approx 200$ c/s ; in the case of the lowest frequency with which we are concerned $f_0 = 90$ c/s, $\tau = 1/9$ s, so that rapid variations of the components in this neighbourhood will not be quite adequately brought out on the film. In practice, however, this is not a serious objection.

With the 79 filters already mentioned, whose resonance frequencies are distributed according to a geometrical series over the frequency range from 90 to 8000 c/s which is of importance for speech [2]) (there are then 12 filters to an octave and the frequency relation between two adjacent filters is therefore the same as between two adjacent notes on the piano (about 1.06)), it may be calculated according to equation (1) that the transmission factor of the filters and those frequencies f_m where two adjacent frequency curves intersect ($f_m/f_0 = 1.03$, see *fig. 4*), amounts to $\alpha = 1/2\sqrt{2}$. At resonance ($f = f_0$) $\alpha = 1/2$. Thus if a component lies just between two filters, both of these filters give a certain output signal from which the intensity of the component can be found by adding the squares. But when a component falls exactly at the resonance peak of a filter, several neighbouring filters to the left and right also give an output signal which in magnitude is 44, 22, 14, 10 per cent, etc. respectively, of the centre filter. This case — "excitation curve" in the case of a truly sinusoidal input voltage resonating with a filter — is illustrated in fig. 2. The spectrogram has a character similar to that of the excitation curve of the basilary membrane of the human ear, which can be represented as consisting of a similar series of damped resonators [3]).

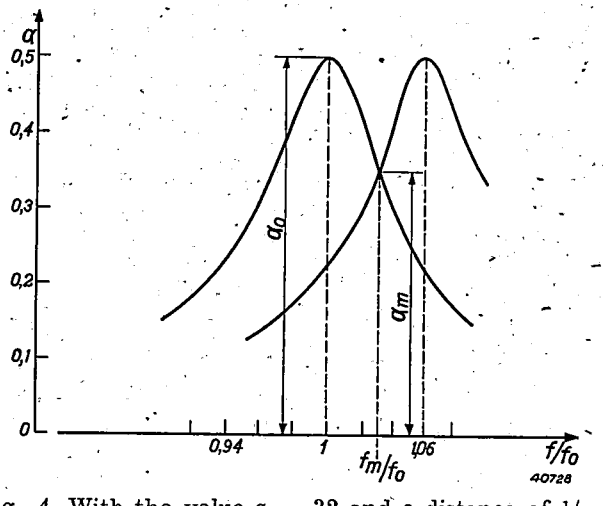


Fig. 4. With the value $q = 32$ and a distance of $^1/_{12}$ of an octave between the resonance frequencies of successive band filters, the transmission factor a_m at the point of intersection of two adjacent resonance curves is $1/\sqrt{2}$ times as large as that at resonance (a_0).

[2]) See for example R. Vermeulen, Octaves and decibels, Philips Techn. Rev. 2, 47, 1937, graph on p. 49.

[3]) See J. F. Schouten, The perception of pitch, Philips Techn. Rev. 5, 286, 1940, where on p. 290 such a model for the basilary membrane is discussed.

If the construction of the filters is based on a recording speed of 20 times per second, it is of course desirable that there should be no other elements in the apparatus which prevent recording at such a speed. Critical points in this respect are the outputs of the rectifiers, the input of the modulator and the electron ray tube. The rectified voltages are taken from a resistance R_2 inserted in every rectifier circuit, through which resistance the rectified current flows, while a condenser C_2 in parallel with R_2 serves to smooth the voltage to the required degree. The time constant R_2C_2 of this circuit must be so small that the condenser C_2 can be practically discharged between two recordings. R_2C_2 cannot be made indefinitely small, since R_2 must be large enough to obtain a high voltage and also in order not to affect the band filter by too high a consumption of current, and also since C_2 must be large enough to obtain the desired smoothing effect. Nevertheless, it was found possible to give the product R_2C_2 exactly the value $^1/_{20}$ sec, so that the recording of 20 pictures per second was not hindered by this. In fact, apart from the impossibility of making R_2C_2 indefinitely small, a much smaller value of R_2C_2 would not even be desirable, since then the consequences of a short-lived change in the signal disappear again so quickly that there would be a great chance that it would be unnoticed with a time interval of $^1/_{20}$ sec between successive recordings.

Similar considerations are also valid for the input of the modulator. As soon as the switch E_1 makes contact with a certain lamella of the collector, the input capacity C_3 of the modulator is loaded *via* the resistance R_3 in the connections between rectifier and collector. It was now found necessary to make C_3 large enough to limit the effects of the switching impulses occurring; at the same time, however the charging time C_3R_3 of the modulator input must be made extremely small. For the recording of 20 spectrograms per second, each consisting of about 80 measured points, only $\frac{1}{20}.\frac{1}{80} = 0.0006$ sec is available for the scanning of each lamella; since there must also be sufficient space between the lamellae of the collector, the actual time of contact is only about half as long. In order to record the correct voltage value, the time C_3R_3 must therefore be chosen appreciably smaller than 0.0003 sec. With the values chosen of $R_3 = 0.22$ MΩ and $C_3 = 500$ $\mu\mu$F, R_3C_3 became 0.00011 sec, which is small enough.

Finally there is the electron ray oscillograph. For tracing one vertical line in the spectrum, with the desired recording speed of 20 pictures per second, 0.0003 sec is available according to the above. With a maximum length of the lines of for instance 4 cm the tracing speed of the oscillograph [4]) therefore amounts to 4 cm/0.0003 sec = 120 m/s. Although in ordinary cases, for instance with the electron ray oscillograph GM 3152, exposures can easily be made with such a tracing speed, in this case this is not immediately true, since we must also pay attention to the resolving power. With a length of the spectrogram of about 8 cm the width of each of the 79 vertical lines may not be greater than about $^1/_2$ mm in order that they may not overlap. With the required very fine fluorescent spot the necessary light intensity for exposures with the tracing speed mentioned could only be obtained by the use of an electron ray tube with post-acceleration [5]).

[4]) Actually, of course, the fluorescent spot moves 16 times as fast, since it makes a vibration with a frequency of 50 kc/s along the line. It comes to practically the same thing, however, for photography, whether the spot describes the line once with a given velocity or 16 times with a velocity 16 times as great.

[5]) See J. de Gier, An electron ray tube with post-acceleration, Philips Techn. Rev. 5, 245, 1940

Relation between the length of line on the screen and the amplitude of the Fourier component

The rectified output voltage of the band filters can be modulated on the "carrier wave" of 50 kc/s in different ways. There are, however, two requirements: firstly that with a modulating voltage of $v_m = 0$ the amplitude V_d of the carrier wave should also become equal to zero; secondly that the relation between V_d and vF should have a certain character. By itself a linear relation would seem most obvious. There is, however, the objection that because of the great differences occurring in the intensity of the sound, the weak Fourier components would quickly become insignificant compared with the strong ones. If V_d increases less than proportionally with v_m, as for instance in the case of a logarithmic relation, then all intensities are dealt with equally, but there is the disadvantage that a peak in the Fourier spectrum is even more flattened than already results from the damping of the band filters (see fig. 2). An intermediate way, in which the relation between V_d and v_m begins approximately linear and then curves off towards a sort of logarithmic relation, is the most suitable.

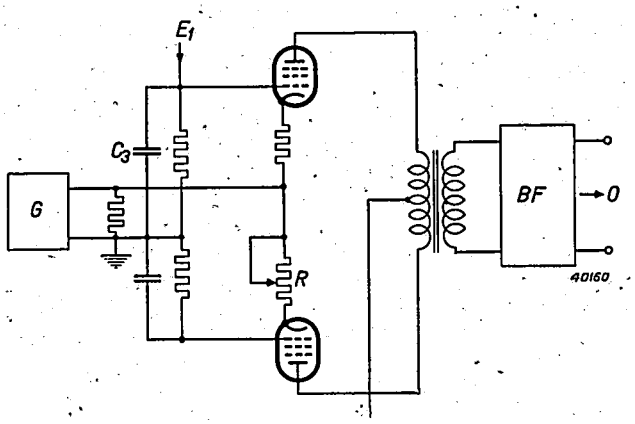


Fig. 5. Connections of the modulator. The anode alternating currents of the two valves which are excited by the carrier wave voltage of the generator G are adjusted to equal magnitude with the resistance R, so that the band filter BF receives no input voltage. By means of a D.C. voltage v_m which is fed from E_1 to the control grid of one valve, the equilibrium is disturbed and a carrier wave of a certain amplitude is therefore passed on to the oscillograph O.

Such a relation is now realized by the modulator connections shown in *fig. 5*, in which at the same time the first condition mentioned, namely that $V_d = 0$ when $v_m = 0$, is also satisfied. The voltage of 50 kc/s of the generator G is fed in the same phase to the control grids of two valves in push-pull connection working with inverse feed-back. Each valve furnishes a certain anode alternating current, but due to the compensation of the two currents

in the output transformer — which compensation can be precisely adjusted by the regulatory resistance R — no carrier wave is ordinarily transmitted to the band filter BF ($V_d = 0$). If by means of the rotating switch E_1 a D.C. voltage v_m is applied to the control grid of one of the two valves, the operating point of this is displaced on its characteristic to a point with a steeper slope, the anode alternating current of this valve becomes larger, the compensation is sufficient and the excess is transmitted as output signal (with 50 kc/s) to the band filter. If the matter is considered more carefully, in the case of the valve with inverse feed-back the relation between the anode alternating current i_a and the grid A.C. voltage v_g is given by

$$i_a = \frac{s}{1 + sR} v_g,$$

where s stands for the slope and R for the resistance in the cathode connection which effects the inverse feed-back. The slope s is chiefly determined by the grid D.C. voltage v_m and (at least in the beginning) increases proportionally with v_m. The expression $s/(1 + sR)$, however, increases less rapidly than in proportion to s and finally approaches the constant value $1/R$. From the cooperation of these two functions, upon suitable choice of the point on the characteristic at which the valve operates with $v_m = 0$, exactly the desired relation between V_d and v_m is obtained, approximately linear at first and later curving.

Actually it is not a question of the relation between V_d and v_m, but of that between the length of line on the fluorescent screen and the intensity of the corresponding Fourier component. Now the whole apparatus is about linear,

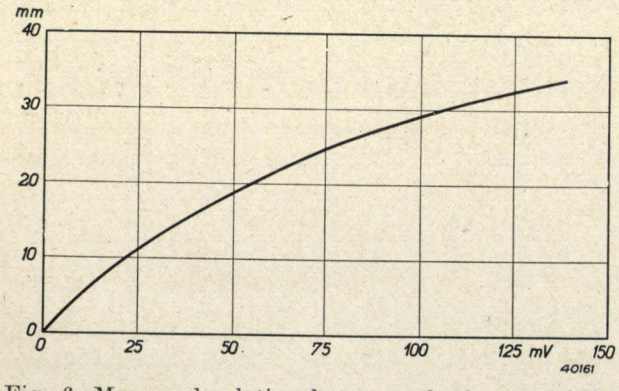


Fig. 6. Measured relation between the length of line on the fluorescent screen and the strength of a sinusoidal input A.C. voltage from the apparatus at a frequency which lies just in the centre of the transmission region of one of the band filters. The attenuator shown in fig. 1, which is inserted in front of the pre-amplifier V_1 in order to be able to regulate the input voltage of the band filters, was in this case in its extreme position (no attenuation). Furthermore, the maximum amplification of the oscillograph amplifier was operative.

thanks in part to the use of diodes (EA 50)in the rectifiers. Because of the large number of rectifiers (79) it would have seemed preferable to use blocking-layer valves, but owing to the linearity mentioned this was not done (moreover, the diode has the advantage of being less sensitive to overloading). Since the relation between the length of line on the screen and the deflecting voltage V_d on the cathode ray tube is also satisfactorily linear, the relation between the length of line and the input voltage of the whole apparatus has practically the same form as the relation between V_d and v_m. This relation as determined by measurement is reproduced in *fig. 6.*

The recording of the spectrograms

In order to make an "instantaneous exposure" of the changing sound the spectrum on the screen of the oscillograph may be photographed with a camera whose shutter must be opened for exactly one revolution of the rotating switches E_1, E_2. If it is desired to make a series of

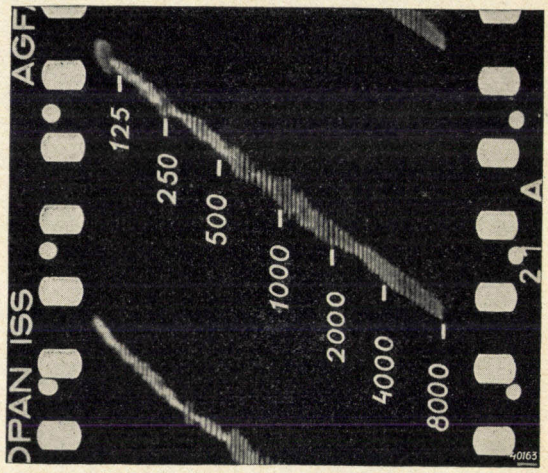


Fig. 7. Spectrogram recorded with continuously moving film. Due to the superposition of the film movement on the motion of the fluorescent spot the spectrum is recorded along an oblique line, descending towards the right. When the last spectral line has been traced the spot jumps to the left again and begins anew. The frequency scale in c/s is indicated on the film.

such exposures, and thus to "film" the sound, the film should be shifted by the height of one picture after each revolution of the switches. With the construction of the collectors here chosen, in which only as much space is allowed between the first and the last of the 79 contacts as is necessary to prevent the occurrence of short-circuiting of the source of voltage for the horizontal deflection of the spot (see fig. 1), the time available for the shift is too short, so that in each exposure the beginning of the spectrum would have to be missed or every other

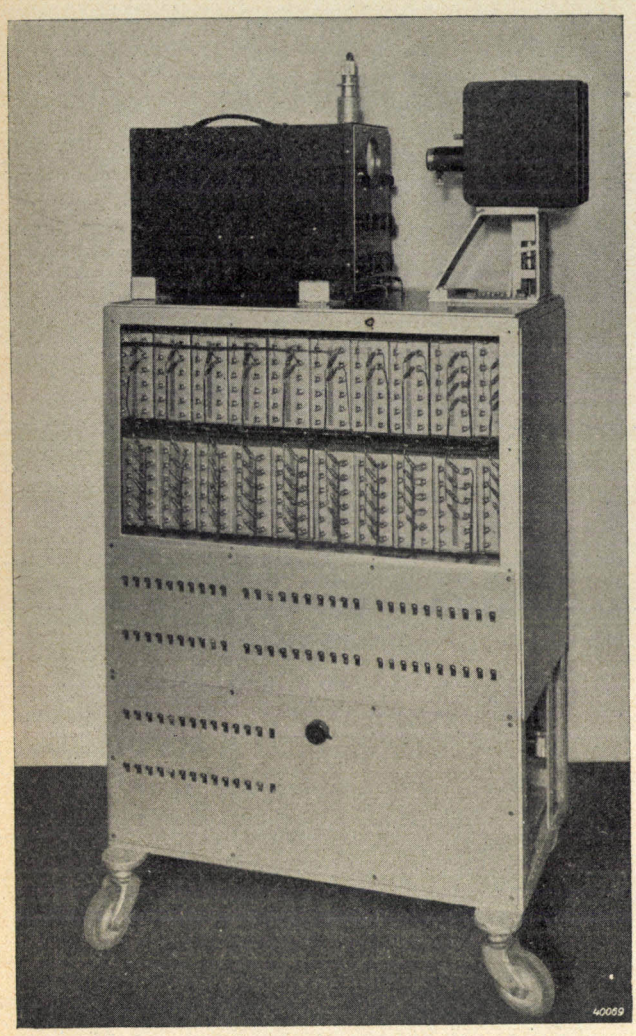


Fig. 8. In the upper part of the wheeled cabinet may be seen a large number of boxes in which the 79 band filters are housed. Underneath are the rows of switches S for listening to the filtered-out spectral parts of the sound being investigated. On top of the cabinet on the left is the electron ray oscillograph (GM 3152), on the right the film camera.

complete revolution of the switches would have to be omitted. A better and simpler solution was therefore to allow the film to move continuously in the direction of the vertical spectrum lines with permanently open camera shutter. When this is done of course a uniform movement in the direction of the film is superposed on the motion of the fluorescent spot, so that the spectra traced are pulled into an oblique position (see *fig. 7*), while, moreover, the ends of each spectral line are slightly less sharp. These features, however, constitute no disadvantage in the analysis of the diagrams.

Photographing with continuously moving film sets a sharp limit on the time of phosphorescence of the fluorescent screen. While with a discontinuously shifted film the permissible time of phosphorescence is determined by the recording time of a complete spectrogram (1/20 sec), with a continuously moving film the only permissible time of phosphorescence is that which is determined by the lack of sharpness to be tolerated at the ends of the lines, and which is therefore of the same order as the time necessary for tracing one vertical line in the spectrum (*i.e.* 1/1600 sec). In the case of the screen of the electron ray tube used by us the phosphorescence time was found to be sufficiently short.

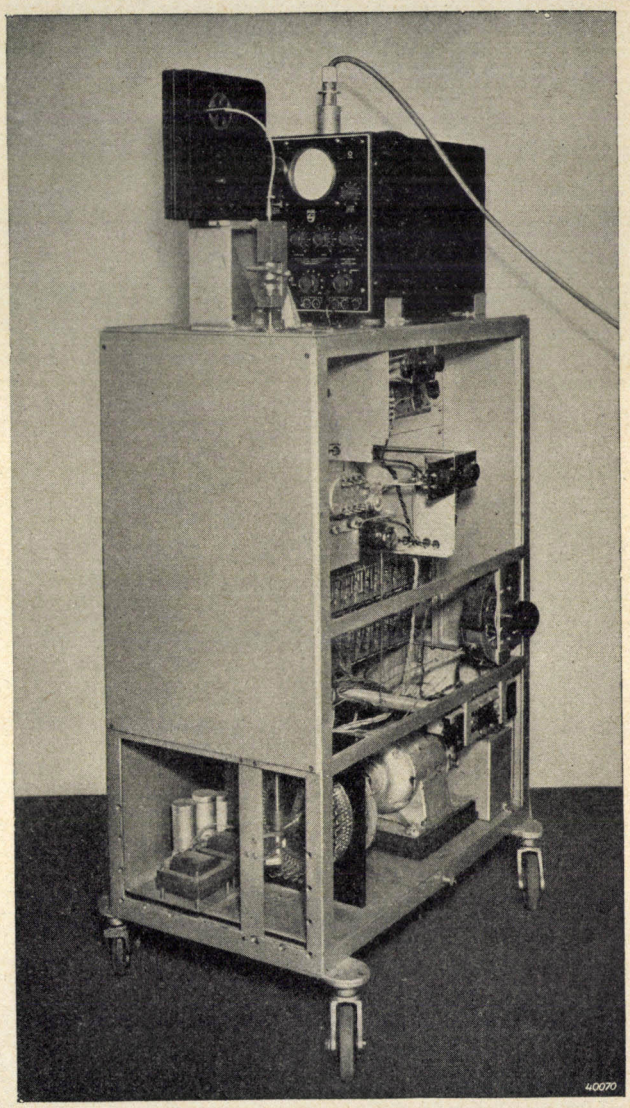


Fig. 9. On the bottom of the wheeled cabinet may be seen the motor which drives the two collectors E_1 and E_2 and, *via* a tooth-wheel transmission and a vertical axis, the film camera. In order to prevent acoustic disturbances the motor is mounted on a thick rubber plate. The collectors consist of a ring of high-frequency "Philite" in which a ring of conical copper pins is driven; with this construction wear is restricted and at the same time with little upkeep only a very low transition resistance at the contacts is obtained. In addition, on the bottom of the cabinet are the supply apparatus for the amplifiers, the modulator, the plates for horizontal deflection of the oscillograph, etc. Above are the 79 diode rectifiers. On top of the cabinet may be seen in the foreground the coupling in the driving shaft of the film camera, which may be connected and disconnected. The lead on top of the electron ray oscillograph supplies the post-acceleration voltage.

During one revolution of the rotating switches the film movement must be such that even with the largest amplitudes occurring successive spectra must not overlap. In order to ensure this once and for all at different speeds of revolution of the switches, the camera is driven, *via* a tooth-wheel transmission and a flexible shaft, by the same motor (a D.C. motor with variable speed) to the axle of which the switch arms are fastened. The camera is set in motion or stopped by means of a simple coupling mechanism.

As described in detail above, a maximum recording speed of 20 spectrograms per second can be attained. For many purposes a lower speed, for instance 10 pictures per second (thus half the speed of revolution of the switches) will suffice, and the recording speed then corresponds to the longest decay time of the filters occurring (1/9 sec with the filter for 90 c/s).

As a conclusion to this description, in *figs. 8* and *9* two photographs of the apparatus constructed are given. Several structural details are pointed out in the text below the figures.

Several results obtained with the apparatus

In *fig. 10* a number of filmed spectrograms are given which were made with the apparatus. They are the spectra of several sounds of speech, namely the vowels a, e, i, o, u, recorded at a speed of 16 spectra per second. The frequency scale is the same as that of fig. 7.

While it is generally known that the consonants are characterized for the most part only by certain introductory and transitional vibrations (most of them cannot be "held"), it is clearly apparent from the sections of film reproduced that vowels also fail to represent a completely periodic vibration. Each vowel is in fact characterized by a series of harmonic components which lie in certain formative regions, relatively independent of the pitch of the fundamental tone of the sound.

The fundamental tone, which may be quite different for male and female voices, in the case of the subject of the spectrograms in fig. 10 is about 250 c/s (peak on extreme left, band filter No. 10) for the e, i, o and u. In the case of the a during its pronunciation the fundamental tone is seen to rise slightly from 175 to 225 c/s, and in the case of oo it falls slightly from 250 to 175 c/s. Such variations in the fundamental tone are of little importance for the separate sounds of speech, but they do

[1]) The subjects of these experiments spoke the language of the Netherlands. The vowels have approximately the same sound as in French, the double o (oe in Netherlands) is the English oo as in root.

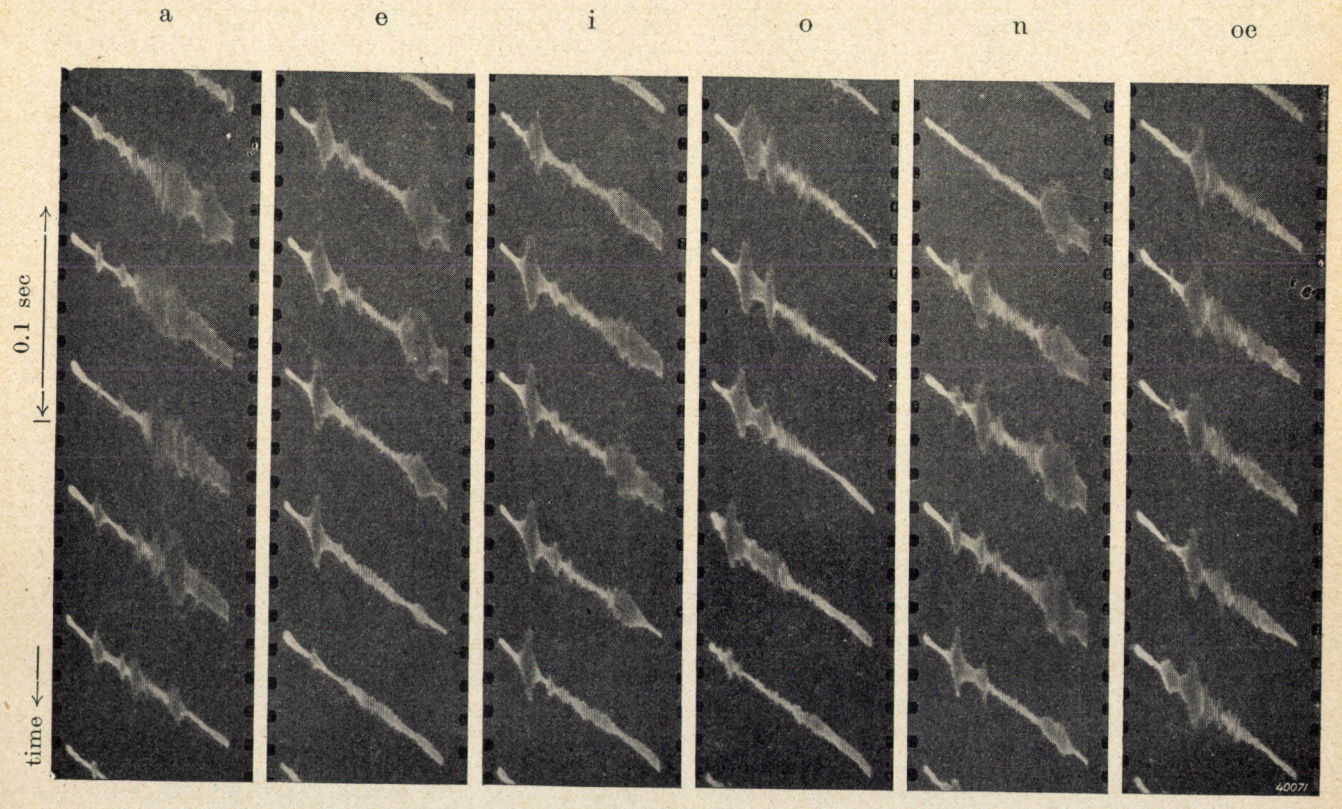


Fig. 10. Fourier spectra of several spoken vowels filmed with the apparatus. Frequency scale of each spectrum as in fig. 7. The recording speed was 16 pictures per second, so that in a vertical direction the time scale indicated on the left is valid.

play an important part in the capacity of the language for expression. This is an important point for investigation in phonetics. If the position and intensity of the other peaks in the spectrum are measured, the following data *(table I)* are found for instance for the a (third spectrum from the top) (applying the calibration curve of fig. 6).

Table I

Filter No.	Frequency c/s	Voltage mV
16	210	8
28	421	14
35	630	19.5
44	1060	75
56	2120	40
64	3360	30.5

As may be expected theoretically [6]), these frequencies are multiples of the fundamental tone (210 c/s). The strongest overtone is found at 1060 c/s, in which region lies the most important formative of the a. The weaker maxima usually occurring as well, which in this case for instance occur at 2120 and 3360 c/s, are less essential to the character of the vowel.

Upon investigation of a series of different voices we found for the most important formative of each of the vowels mentioned the frequency regions given in *table II*, column *A*. For the sake of comparison the values given by Stumpf on the basis of detailed investigations are given in column *B* [7]). It is evident that the individual differences may be quite considerable and that the frequency regions by which, according to the current conception, a vowel is characterized are anything but sharply defined. This is the more striking since in the synthetic production of vowels with the help of a number of band filters it is found that fairly accurately determined formatives are necessary for the recognition of a vowel; for example in our experiments on this subject a formative which deviated only little from 1500 c/s was necessary for the production of a clearly "pronounced" a.

Table II

Most important formative regions of six vowels, according to measurements with the apparatus described. For the main formatives our own measurements (column *A*) are to be compared with values given in the literature [1]) (column *B*).

Vowel	Subordinate formative c/s	Main formative c/s	
		A	*B*
oe	150-300	—	—
o	—	400-500	366-548
a	—	950-1498	548-1304
u	150-300	1888-2520	1843-2926
e	300-600	2378-2996	2069-2762
i	150-300	2828-3786	2192-3687

[6]) The spectra of the vowels are formed due to the fact that certain harmonics of the vocal chord vibration, which contains the fundamental tone with a large number of harmonics, are amplified by resonance of the cavity of mouth, throat and nose. See also J. de Boer and K. de Boer, The Laryngophone, Philips Techn. Rev. 5, 6, 1940.

Such experiments with synthetic vowels also furnish a good illustration of the above-mentioned fact that the filmed spectra clearly show a fluctuation of the components during the pronunciation of a vowel. In the case of artificial vowels where this fluctuation is lacking the vowel can indeed be recognized, but the sound is not "human" in character. For the appearance of the human character the natural vibration which is given to the sound by the muscles which contract the vocal chords and determine the shape of the cavity of the mouth, throat and nose (see footnote [6]) seems to be essential [8]).

In conclusion we should like to point out that the possibilities of application of the acoustic spectrograph described here are not limited to the investigation of speech. The apparatus undoubtedly offers important possibilities for the investigation of musical instruments, of the acoustics of halls (reverberation), of interferences on connection lines or of interfering noises of machines and the like.

[7]) According to C. Stumpf, Die Sprachlaute. Experimentell-phonetische Untersuchungen, J. Springer, Berlin 1926, p. 66. His values are for vowels sung with a fundamental tone of 183 c/s. In comparing it must be taken into account that the language of Stumpf's subjects was not the Dutch but the German.

[8]) This was also confirmed in unpublished experiments by J. F. Schouten in this laboratory. Vowels were imitated with the help of the optical siren (Philips Techn. Rev. 4, 167, 1939), in which stencils with a cut-out vibrational form were placed in a beam of light. When the stencil was fastened rigidly the sound obtained was not "human" but lifeless in character; it became much more human however, when the stencil was held in the light beam with the hand. The "vibrato" of the hand is then a substitute for the vibrato of the vocal chords and of the resonating cavities.

THE TESTING OF POWER CABLES WITH DIRECT CURRENT VOLTAGE

by W. HONDIUS BOLDINGH. 621.319.52:621.317.33:521.315.2

The requirements are discussed which must be satisfied by an apparatus for testing power cables with high voltage. A testing apparatus for 20 kV is described.

Cable testing

Newly laid high-tension cables are generally tested by applying to them for some time a D.C. voltage which is much higher than the A.C. voltage normally used on the cable. Moreover, if during use there is any indication that a defect in the insulation is beginning to make itself felt, a high D.C. voltage is also used; in this case with the purpose of burning through the insulation at the defect, so that it can be more easily localized.

The fact that D.C. voltage can be used for testing, but not A.C. voltage, is connected with the capacity of the mains being investigated. If A.C. voltage were used the capacity, which in the case of underground cables may amount to 0.2 μF per km for instance, would be charged to this high voltage twice per period. A very high wattless power would be necessary for this, which can scarcely be compensated by means of choking coils, as is ordinarily done in testing short lengths of cable during manufacture. If for example it is desired to test a high-tension cable 10 km long with an A.C. voltage of 50 c/s and 30 kV—, 250 kVA would have to be provided ; with 20 km and 40 kV— this would even be 2000 kVA. These are values of quite a different order of magnitude than the power which is necessary and sufficient to burn through a flaw (namely only about 1 kW) and for which it is still possible to construct an easily transportable cable-testing apparatus. The last mentioned power, moreover, is more than sufficient to charge a high-tension cable in several seconds to the high D.C. voltage required for the testing.

Until now for the purpose just mentioned more or less improvised apparatus was often used which was put together by the various electrical services for their own use. Since in other respects safety devices, method of operation and reliability in high power technology have been perfected to the utmost in the course of years, at first glance it may seem surprising that this has by no means been the case with cable-testing apparatus. This can be explained by the fact that the apparatus in question is used exclusively by very expert personnel, with experience in electro-technology, who easily adapt themselves to the danger of high voltage and switching complications, so that the demand for greater safety and technical convenience has remained in the background.

In recent years, however, in addition to the incidental cable examination upon the appearance of disturbances, it has become more and more customary to carry out a periodical inspection of all the cables of the high-voltage mains in order to detect insulation flaws before they lead to disturbances in the mains. As a matter of fact such routine work makes much higher demands on the safety, compactness and simplicity of the apparatus. While in the case of newly laid cables, to which according to the usual requirements the test voltage must be applied for one hour per conductor, easy operation of the testing apparatus is of practically no importance for the total working time, transportability and ease of manipulation become much more important in a periodical control of the cable network where each conductor can only be tested for a few minutes.

Requirements made of a cable-testing apparatus

We shall now examine in more detail the requirements which should be satisfied by a modern apparatus for testing high-tension cables.

1) The D.C. voltage on the cable must be able to be read off immediately under all circumstances.
2) The charging and discharging of the cable must be able to take place safely and without loose pieces of apparatus under high voltage.
3) It must be possible to protect the cable against voltages higher than necessary and desirable for the investigation.
4) It must be possible to connect the apparatus at the place where the examination is held, even with cable terminal connections which are not easily accessible, without complications caused by unprotected connection lines; up to within several metres around the apparatus any contact with high voltage must be impossible.
5) The apparatus should form a unit which requires no putting together before it can be used ,and which is as easy as possible to operate.
6) It must preferably be of such weight and dimensions that it can without difficulty be

transported in an ordinary motor-car to any point of the cable network.

With the present position of technology it is possible to construct an apparatus for voltages up to about 50 kV and a power of about 1 kW which satisfies these requirements.

In this article we shall confine ourselves to the description of a 20 kV apparatus, which is now being manufactured in series.

Apparatus for the testing of cables with 20 kV D.C. voltage

Although 20 kV is relatively low for tests as prescribed for newly laid cables (for instance according to VDE 0255), this value is, however, often used for periodical control. In this case only very little overvoltage is desired, in order not repeatedly to overload the cable heavily, which might cause it to exhibit defects prematurely. For "10 kV" multiphase-current cables which normally carry 10 kVeff between two phases, *i.e.* 5.9 kVeff and 0.3 kV peak voltage with respect to earth, a D.C. voltage of 20 kV is found to be more than enough to detect flaws in time with regular control.

The different types of connections which are used in practice for the high-voltage circuit of a cable-testing apparatus are mainly determined by the type of rectifier valve used.

If, for example, one has a valve which can withstand a voltage more than twice as high as the maximum charging voltage of the cable, one valve is needed, as indicated in *fig. 1*. A high-voltage generator for cable inspection then consists in principle of a connection in series of one rectifier valve V and a high-voltage transformer T, which in this case is loaded asymmetrically. A cable-testing installation as a rule further contains as components carrying high voltage a spark gap for measuring voltages, an earthing switch with water resistance and if necessary damping resistances.

If a larger current is to be used than the available rectifier valve can normally carry, double-phase rectification with two valves V must be used, for example according to the centre-tapped connections (*fig. 2a*). If smaller valves are preferred which can tolerate little more than the normal cable voltage, two

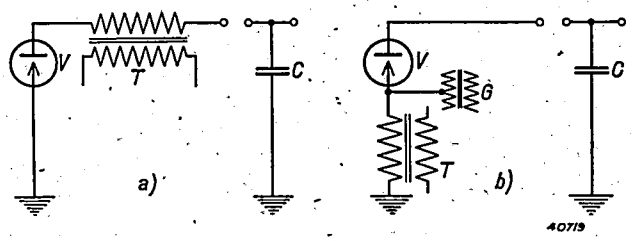


Fig. 1. Simplest connections for the excitation of a high D.C. voltage. Single-phase rectification with one rectifier valve V whose permissible counter-voltage is greater than twice the D.C. voltage to be furnished. The high-voltage transformer T is asymmetrically loaded.
a) The cathode of the valve V is earthed and two terminals of the high-voltage transformer are insulated.
b) The rectifier valve V is insulated at two terminals, while the cathode is fed by a heating current transformer G insulated for high voltage. The high-voltage transformer T is now earthed at one terminal.
C represents the capacity of the cable to be tested.

such valves V should be connected in series, and the Greinacher connection is usually employed (*fig. 2b*).

In the case of the last two connections mentioned a high-voltage transformer T with two high-voltage terminals is required. In the apparatus to be described here we have used a connection less common for cable testing in which a transformer with one high-

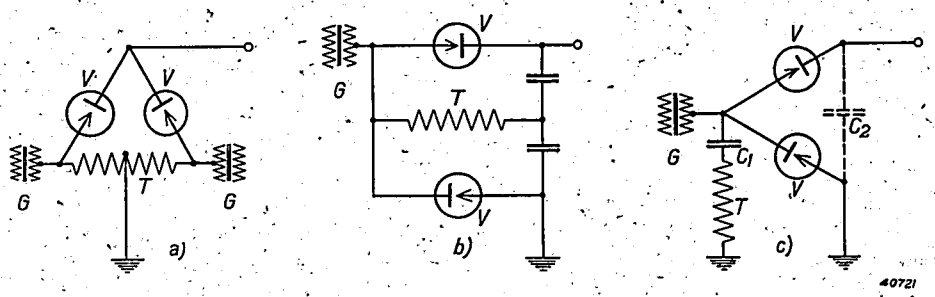


Fig. 2. High-voltage connections with two rectifier valves V.
a) Centre-tapped connections. Current doubled; high-voltage transformer T with doubled voltage, two terminals insultated and symmetrically loaded; the rectifier valves V receive the doubled counter-voltage and are fed from two heating-current transformers G.
b) Greinacher connections. The valves V receive the single voltage; high-voltage transformer T is now insulated at two terminals, has half the voltage and is symmetrically loaded.
c) Cascade connections. The rectifier valves V have the single voltage; high-voltage transformer T is here earthed at one terminal and is nevertheless symmetrically loaded.
C_1 is the cascade condenser and C_2 the smoothing condenser.

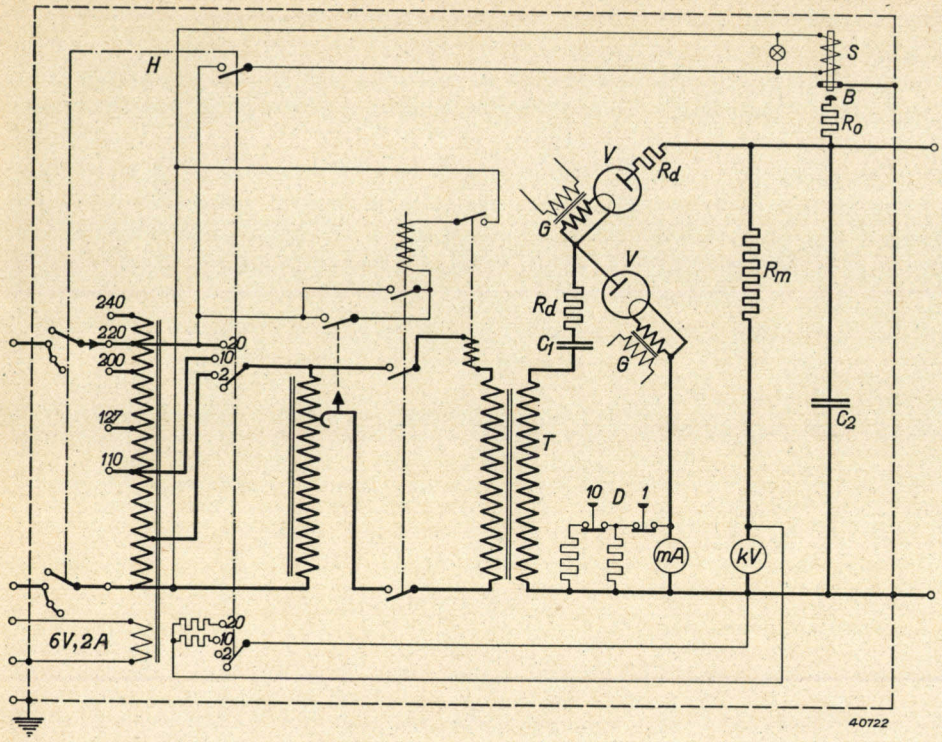


Fig. 3. Simplified diagram of the connections of the apparatus for testing high-tension cables with 20 kV D.C. voltage. V rectifier valves, T high-voltage transformer, G heating-current transformers, R_m resistance in series with the kilovoltmeter kV, mA milliammeter, H main switch, D push buttons for commutation between the measuring ranges, R_0 discharging resistance, R_d damping resistance, B adjustable voltage limiter, S magnetic switch with contact pin, C_1 cascade condenser and C_2 smoothing condenser added to permit the testing of objects with little capacity.

voltage terminal can be used. This is the so-called cascade connection (fig. 2c), which is also used for generating much higher D.C. voltages, as already described in this periodical [1]). These connections make it possible to use for a 20 kV generator two rectifier valves V with a maximum counter-voltage of 28 kV and a transformer T insulated at only one terminal and nevertheless symmetrically loaded.

In *fig. 3* a more complete diagram of these connections is given, while in *fig. 4* the complete apparatus is shown.

On the basis of the diagram we shall now discuss in more detail how the requirements given in the foregoing are satisfied by this apparatus.

Measurement of voltage, current and insulation resistance

The voltage is measured with the help of the kilovoltmeter kV indicated in fig. 3, which is provided with a series resistance R_m and placed on the top of the apparatus. It is thus possible under any circumstances during the testing to read off directly the actual voltage on the cable on an earthed instrument. This method of measurement is far preferable to the measurement with a spark gap or the measurement of the primary transformer voltage. Since the primary voltage regulation of the transformer T possesses three regions of regulation (2, 10 and 20 kV) the kV-meter is constructed for three corresponding measuring ranges which are automatically chosen with the selector (H) of the voltage steps. H is at the same time the main switch (*cf.* fig. 3).

Because of the direct reading of the voltage (kV) it is possible, in combination with a sensitive ammeter (mA), to determine the insulation resistance. While the mA-meter does in fact possess a measuring range up to 100 mA (upon exceeding this limit the current is automatically

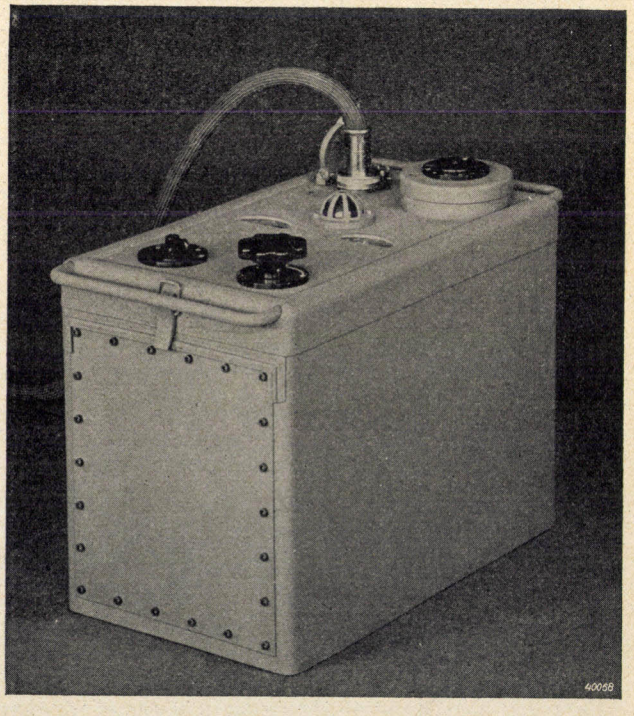


Fig. 4. The apparatus, type No. 11680, for testing high-tension cables with D.C. voltages up to 20 kV.

[1]) Philips techn. Rev. **1**, 6, 1936.

switched off), the measuring range can be reduced to 01 or 1 mA with the help of two push buttons D. In the latter case (1 mA) 20 μA per scale division is still read off, so that with negligible corona losses an insulation resistance of 100 megohms can still be observed. In order that such measurements of the insulation resistance with this apparatus need not be limited to high-tension cables with large capacity, but may in general also be carried out on insulation materials for high voltage but without appreciable capacity, a smoothing condenser C_2 is added which also guarantees a constant D.C. voltage when a current of several mA flows during the measurement.

With the measuring range up to 1 mA a combination of resistance and capacity prevents the occurence of disturbing fluctuations of the indication of the mA-meter due to fluctuations in the mains voltage. At the same time these connections form an efficient protection against damage to the meter by sudden current surges which may occur upon unexpected breakdown of a cable flaw during a measurement on the sensitive measuring range.

The two sensitive measuring instruments mentioned, here which have about 100 μA as the smallest measuring range, are made mechanically strong to resist rough usage. In order to make it possible to read them easily in the dark they are provided with an illuminated scale.

Damping and discharging resistances

In order to avoid large current impulses a water resistance is often used both for the charging and the discharging of cables, Such a more or less provisional component is in fact unsuitable for power technology. It is breakable, it cannot be used in all positions, it is difficult to fasten firmly and it may unexpectedly assume a low value due to contamination. The advantage of a water resistance is that in spite of its high resistance it possesses a large heat capacity, while that of a wire resistance, when the length of the wire is kept within reasonable bounds, becomes steadily smaller with increasing value of the resistance. For this reason a wire resistance is unsuitable for charging a spark gap, where at the moment of breakdown strong current impulses occur which would cause an intolerable heating of the wire. In the cable-testing apparatus, however, the spark gap is replaced by the series resistance R_m already mentioned, so that this advantage of the water resistance would not be decisive.

Just as for charging, so for discharging the water resistance is not necessary. For this purpose a resistance can just as well be made of wound wire with reasonable dimensions and sufficient heat capacity, especially if it need not have too high a resistance. The latter is indeed the case: if the self-induction may be neglected the size of the discharging resistance could theoretically be lowered to the wave resistance of the cable, *i.e.* to about 50 ohms, without fear of oscillations and reflections upon discharging. It is, however, unnecessary to use such low resistances ; a wire resistance of several thousand ohms with reasonable dimensions already has a heat capacity large enough to take up the heat developed upon discharge of a long cable at full voltage.

This may easily be seen on the basis of a simple calculation.

Let the specific resistance equal r ohm cm.
the cross section of the wire be q cm^2,
the length of the wire l cm,
the specific weight s g/cm^2 and
the specific heat c cal/g.

Then the resistance becomes: $R = \frac{lr}{q}$ ohm.

the weight of the wire: $G = lq\,s$ gram and
the heat capacity: $W = l\,q\,c\,s$ cal per degree.

If we now assume that the cable is so quickly discharged that the resistance first takes up all the electrical energy $E = {}^1/_2\,CV^2$ before it has time to give off a part of it by convection, the material of the resistance undergoes a short-lived temperature increase T which is given by $WT = 0.24\,E$. We can now express the cross section q and the length l of the wire in terms of the other quantities introduced above, and we find that

$$q^2 = \frac{0{,}24\,r}{s\,c}\,\frac{E}{TR}\ \text{cm}^4 \text{ en}$$

$$l^2 = \frac{0{,}24}{r\,s\,c}\,\frac{ER}{T}\ \text{cm}^2.$$

For the different resistance materials in practical use the value of s and c varies only very slightly, so that in general wire with the highest possible specific resistance r is the best. If for example we use chrome-nickel wire $r = 10{\cdot}4$, $s = 8.37$ and $c = 0.1$, we find for the diameter

$$0.83\sqrt[4]{\frac{E}{TR}}\ \text{mm and for the length } 0{,}535\sqrt[4]{\frac{ER}{T}}\ \text{m.}$$

The net surface wound, that is without the insulation space between the windings, thus becomes

$$O = 4{,}45\sqrt[4]{R\frac{E^2}{T^3}}\ \text{cm}^2.$$

With a given heat capacity, therefore, the lowest resistance occupies the smallest surface. This is also clear without calculations, since with thick wire more material can be wound on a given surface than with thin wire. If for instance we have a resistance in which an energy E of 4000 joules (4 kW sec) must be taken up upon discharging a cable for 20 kV with a capacity of 20 μF where an increase in temperature of 200 °C is permissible [2]), we obtain a thickness of the wire $d = \frac{1{\cdot}76}{\sqrt[4]{R}}$ mm and a net surface wound

$$O = 42\sqrt[4]{R}\ \text{cm}^2.$$

[2]) With such a temperature increase the linear expansion is 0.3%, which is permissible without danger of the windings working loose.

For a resistance $R = 6000$ ohms, d then becomes 0.2 mm and $O = 370$ cm², while for $R = 96\,000$ ohms the wire must be half as thick and the surface wound wice as large.

The dimensions of such a resistance, which can be made practically free of induction, are small enough to allow it to be housed in the high-voltage generator (R_0 in fig. 3), so that difficulties with external assembly are avoided.

The earthing of the cable over this resistance R_0 on the earth side is by means of an electrically operated magnetic switch with contact-pin, indicated in fig. 3 by S. This switch uses zero-load current, so that the cable is automatically earthed as soon as the apparatus is switched off, while upon switching on the high voltage this earth connection is automatically broken.

Voltage limiter

The voltage limiter B in series with the discharging resistance R_0 is connected in parallel

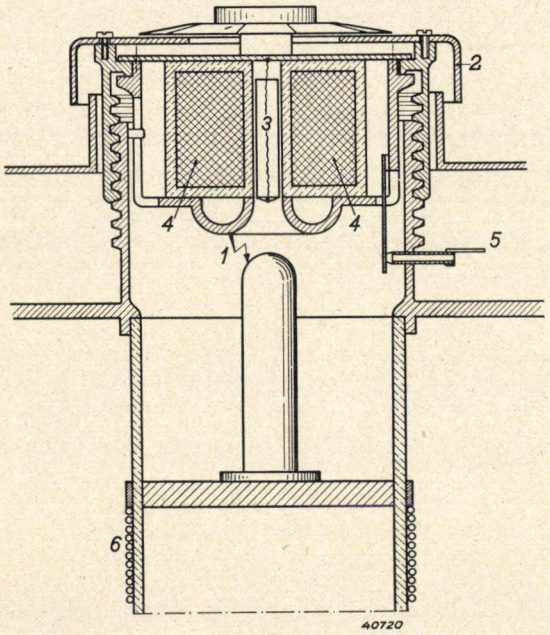


Fig. 5. Magnetic switch with contact pin. *1* Voltage limiter, *2* rotating cap for adjustment of the limiting voltage (calibrated in kV), *3* earthed contact pin, *4* solenoid for contact pin, *5* current supply for solenoid, *6* discharging resistance.

with the voltage to be measured (*cf.* fig. 3). It consists of a spark gap adjustable by hand which is calibrated in kV and by means of which too high a voltage on the cable can be avoided. The contact pin just mentioned is mounted concentrically in this spark as shown in *fig. 5*.

High-voltage protection

Since all the components of the high-voltage circuit are housed in the generator case, a single connection is sufficient to connect the generator to the cable to be tested. This connection is by means of a flexible rubber cable provided with a braided metal covering, one end of which is connected to the generator case to be earthed, and the other end to the lead covering of the cable to be tested. With sufficient length of connecting cable, therefore, the safety of anyone using the apparatus is fully ensured.

Operation

Fig. 6, which is a view of the top of the apparatus, shows that its operation has been kept as simple as possible. On the front left-hand corner is the main switch, which at the same time operates the contact pin and selects one of the three regulation regions for the voltage (2, 10 and 20 kV), while at the same time the kilovoltmeter is correspondingly adjusted (*cf.* diagram of fig. 3). In the front right-hand corner may be seen the voltage regulation knob with which the high-voltage can be increased continuously from 0 to 2, 10 and 20 kV, respectively. On the right hand in the middle is the voltmeter from which the voltage may be read off directly, while on the left of the voltmeter is the milliammeter with two push buttons for the switching

Fig. 6. View of the top of the 20 kV apparatus.

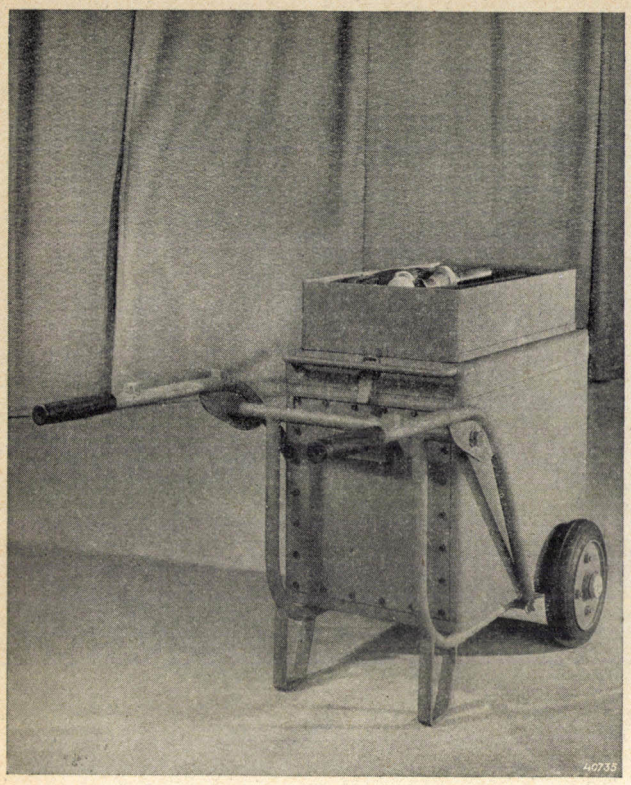


Fig. 7. The cable-testing apparatus ready for transportation.

between the measuring ranges. The voltage limiter is placed at the back and next to it the socket for plugging in the connection cable. The red signal lamp in the middle lights up when the contact pin is drawn up, so that it can immediately be seen that the cable is or will immediately be under high voltage.

Transportability

From the foregoing it is clear that the apparatus forms a closed unit which can on the one hand be connected with the low voltage mains and on the other, without further difficulties of assembly, with the cable to be tested. For easy transportation the whole can be placed on a car, which the connecting cable in a tray serving as cover *(fig. 7)*. The dimensions of the apparatus without the tray are 72 by 36 by 54 cm ; the total weight is 125 kg.

Characteristics.

In conclusion *fig 8* gives the behaviour of several electrical quantities as functions of the loading current. It may be seen that the voltage loss is relatively small, namely 0.12 kV/mA, so that there is no question of any large drop in the voltage with high currents ; this promotes the rapidity with which a flaw can be burned through. As may be seen in fig. 8, with a load of 50 mA for example the apparatus can still furnish a voltage of 16 kV, *i.e.* it can give off an energy of 800 W to a cable flaw with a resistance of 0.32 meghom, which will therefore be very quickly burned through to a lower value. Thanks to the directly readable kV-meter, it is in fact possible to follow the whole process of burning through much better than is usually the case in such investigations.

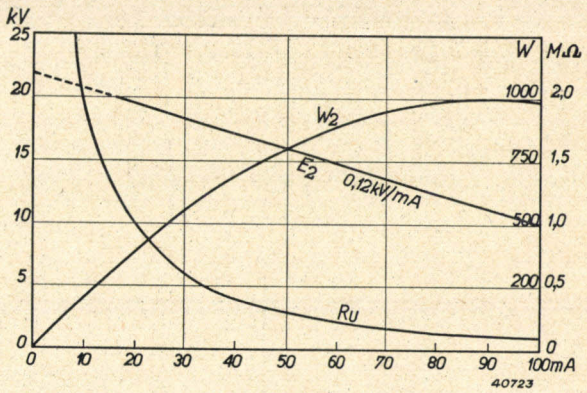


Fig. 8. The relation between various electrical quantities with the loading current, with a cable capacity of 20μF and at the highest position of the voltage adjustment. This adjustment is such that at a mains voltage 10 percent too low 20 kV zero-load can still be obtained, so that at normal mains voltage the highest zero-load voltage would be 22 kV. E_2 D.C. voltage on the cable, R_u resistance of the insulation flaw and W_2 energy for burning through the flaw. A load of 50 mA is permissible for one hour. At 18 kV the apparatus can furnish a current of 30 mA continuously.

Vol. 7 No. 3, pp. 65-96 MARCH 1942

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 AERIAL EFFECT IN RECEIVING SETS WITH LOOP AERIAL

by P. CORNELIUS.

621.396.677:621.396.62

A loop aerial can be used to advantage with portable radio sets because it is less sensitive to local interferences than a mains aerial or a movable capacitative aerial. By making use of the directional effect of the loop it is also possible in certain cases to diminish the interference by undesired transmitters. The condition for the satisfactory functioning of the loop in both cases is that the aerial effect should be kept small, a condition which requires special precautions when the receiver is supplied from the mains. Several methods of combatting the aerial effect are discussed in this article. One of these methods is the use of a loop of only one winding. This method is applied in the Philips receiving set type 902 A (superheterodyne receiver with a stage of high-frequency amplification).

In a previous article [1]) it was explained that a well-constructed loop aerial is the most suitable aerial for portable receiving sets with mains connection, and that the reason for this lies in the lower sensitivity to local interferences of the loop aerial compared with a capacitative aerial. In this article a study will be made of the problems which occur in practical cases when advantage is taken of these favourable properties of the loop aerial. In conclusion a receiving set will be described which is equipped with a loop aerial consisting of only one winding.

The reason why the behaviour of a loop aerial is favourable with respect to local interferences is briefly summarized in the following. If a station is being received which is a great distance away compared with the wave length, so that the receiving aerial is situated in the so-called radiation field of the transmitter, the intensities of the alternating electric and magnetic fields at the aerial are in a definite ratio to each other, namely the electric field strength F in V/m is a factor $120\,\pi$ larger than the magnetic field strength H in A/m. Now since a capacitative aerial reacts to the electrical component of the field, while the signal voltage is generated in the loop by the magnetic component, it is clear that a local source of interference will only have the same disturbing effect for these two types of aerials when the electric and the magnetic field strengths of the interference are also in the ratio $F/H = 120\,\pi$. It is now found, however, at least in the wavelength region above 200 m, that the ratio F/H is usually much greater than $120\,\pi$ for the local sources of interference, so that a capacitative aerial will receive relatively much more of the local interference than a loop.

[1]) See Philips techn. Rev. 6, 302, 1941.

The aerial effect

It is obvious that the advantage of the loop just described will only be fully realized when the loop does actually react only to the magnetic field, *i.e.* when no voltage or current is induced between the terminals of the loop by an alternating electric field whose magnetic component is negligibly small [2]). In general a loop will not satisfy this condition exactly, and the loop is then said to exhibit the "aerial effect".

The presence of the aerial effect can easily be demonstrated by comparing the directional action of different loop aerials which are situated in the radiation field of a transmitter. In the case of an electromagnetic wave which is propagated along a well conducting horizontal earth's surface in the open, the vector of the electric field is vertical, while the magnetic field is horizontal. At the same time both are perpendicular to the direction of propagation.

[2]) For the present we confine ourselves to frequencies at which the wavelengths are still long compared with the dimensions of the aerials considered. If this limitation is omitted a pure electric field (without magnetic field) cannot exist.

Now a loop is sensitive only to that component of the magnetic field which is perpendicular to the plane of the loop. This component disappears when the normal to the plane of the loop is directed towards the transmitter. Thus when the loop is turned about a vertical axis the direction diagram of *fig. 1* holds for the voltage excited in the loop by the magnetic field. This diagram is expressed by the formula

$$V_a = V_0 \sin \alpha, \quad \ldots \ldots \ldots \quad (1)$$

where α is the angle between the normal to the plane of the aerial and the direction of propagation. It is obvious that the maxima are flat and the minima sharp.

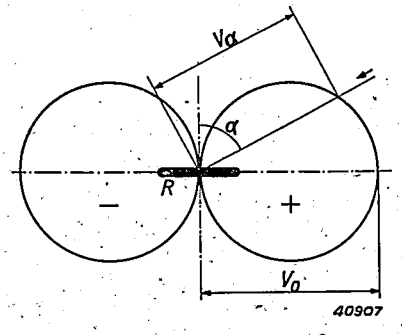


Fig. 1. Directional effect of the loop. The radius V_a indicates the amplitude of the voltage induced by the magnetic field of a wave in the direction of the arrow.

If a loop has an aerial effect a second term must be added to the right-hand side of equation (1), which term is independent of α and in general makes the minima much less sharp. The fact that the contribution of the electric field to the voltage on the loop is indeed independent of α follows immediately from the fact that the electric field has a vertical direction, so that the relative position of the loop with respect to the electric field does not change upon rotation about a vertical axis [3]).

The aerial effect is not only undesirable because of its effect on the sensitivity to local interferences, but also because, as we have seen, it makes the directional effect of the loop less sharp. The directional effect may be used to advantage in many cases to improve the selection between two transmitters operating on the same or almost the same wavelength: the undesired transmitter is simply "turned out". By this means, for instance, side-band interference can be suppressed without the necessity of lowering the quality of the reproduction by cutting off the high tones, as would be the case in reception with a capacitative aerial [4]).

Finally it is found, as will be discussed below, that the aerial effect is accompanied by a greater sensitivity to interferences which can be transmitted to the set *via* the power mains. Therefore for the sake of avoiding also this type of interference it is important to prevent the aerial effect. In this article we shall study the way in which the aerial effect occurs and the available means of combatting it. We shall then pass on to a description of a receiving set with a loop aerial developed by Philips in which special attention has been paid to the countering of the aerial effect.

The causes of the aerial effect

A loop aerial is a continuous conductor which connects the input terminals of the receiving set with each other. Thus with constant or slowly varying electric fields in space no potential difference is in general to be expected between the input terminals. In the case of high-frequency variations of the field strength, however, voltages may occur. In order to obtain a better insight into this phenomenon we shall go somewhat deeper into the behaviour of a conductor in an electric field.

If an originally uncharged conductor is brought into a constant electric field, there occurs in the conductor a separation of equal numbers of positive and negative charges which distribute themselves on the surface of the conductor in such a way that the electric field strength becomes zero in the conductor. We must therefore picture the separation of the charges in such a way that in the originally electrically neutral conductor a certain amount of negative electricity is transferred from one part to another, so that in the first part the corresponding amount of positive electricity remains in excess.

The amount of electricity transferred is proportional to the original electric field strength and further depends upon the form of the conductor and of its surroundings. As the simplest example, the charge distribution on a vertical rod situated in a vertical electric field of strength

[3]) In addition to the aerial effect there may be other causes which may spoil the directional effect. Such causes are: night effect, elliptical polarization of the radiation or the formation of rotating fields by short-circuited loops in the vicinity of the loop aerial. These causes have no connection with the nature of the aerial and will therefore not be considered in this paper.

[4]) It is obvious that this method of removing interference is only possible when the directions from the loop aerial towards the desired transmitter and towards the undesired transmitter (or source of interference) make a large enough angle with each other.

F is given in *fig. 2a*. If the field is directed upwards, positive charge is concentrated at the upper end and negative at the lower end. In the middle of the rod is a neutral zone where the transition from the negative to the positive surface charge takes place.

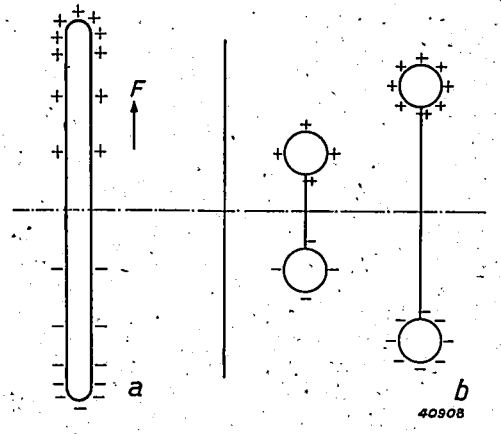


Fig. 2. *a*) Charge distribution on a rod situated in a homogeneous constant electric field F.
b) The same for an aerial consisting of two spheres connected by a wire. With a given value of the field strength the density of charge on the spheres increases with their distance apart.

The farther a surface element of the rod is from the neutral zone, the larger the amount of electricity which will be accumulated in this surface element with a given value of the field strength originally present. The density of charge at the top of a rod of a given thickness at a given field strength therefore becomes larger the longer the rod. Such a conclusion may also be drawn for other forms of aerials. If, for example, a vertical aerial consisting of two spheres joined by a vertical wire (fig. 2*b*) is considered, the charge which the spheres assume in a constant electric field will be larger the farther the two spheres are apart.

From the distribution of charge on a conductor in a homogeneous constant electric field, we may now draw conclusions about the behaviour of a conductor in an alternating field. By an up and down motion of the electric charge the conductor will try to keep the internal field always equal to zero. When the changes of the external field do not take place too rapidly, so that the corresponding wavelength is still large compared with the dimensions of the aerial system, it may be expected that the internal field will actualy be eliminated at every moment, at least when the inductive and ohmic voltage drop in the conductor with the currents resulting from the upward and downward motion of the charges are sufficiently small compared with the voltage which is compensated by the displacement of charge. In the case of the aerials customarily used for intermediate and long waves it may be assumed that this condition is fulfilled.

The charge distribution on an aerial in an alternating field is therefore similar at every moment to that in a constant field which has the same value at the moment considered. Due to the change in the charge with time we obtain an alternating current whose amplitude changes from point to point along the conductor. The largest current amplitude occurs in the neutral zone, since of course the whole negative charge situated on one side of the neutral zone must flow through this zone. Toward the ends of the rod (fig. 2*a*) the current becomes smaller and smaller; at the very ends it is zero. An analogous result is also found for other forms of aerial, for instance fig. 2*b*.

If we now pass on to the loop aerial, which in its simplest form can be represented by a closed loop in space, we obtain for a constant electric field a charge distribution such as that represented in *fig. 3*. The positive and negative charge in this case is mainly concentrated in points *1* and *2* respectively, while points *3* and *4* lie on the neutral plane. If we again consider a varying field strength we can again draw conclusions about the currents flowing in the loop from the change in the charge distribution in the branches.

In principle the current distribution in a loop cannot be completely deduced from the picture of the charges. Even though we know how much electricity must come to a certain spot on the conductor, the path along which it must come is not yet determined. Thus in principle

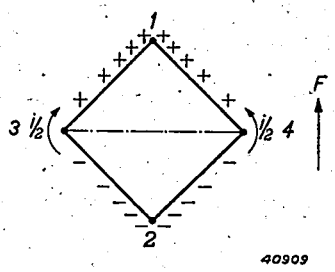


Fig. 3. Charge distribution on a short-circuited loop due to the field strength F. The arrows along the loop indicate the currents which flow in the loop when the field strength increases.

a negative charge in the neighbourhood of point *2* could reach its goal *via* point *3* as well as point *4*. Nevertheless, in the case in question for reasons of symmetry it may be concluded that the same currents will flow at points *3* and *4*, while points *1* and *2* are without current.

If an indicating instrument is placed in the loop which can be slid along its circumference,

the largest current will be measured at points *3* and *4* and the current at points *1* and *2* will be zero. If the loop is cut open at point *3* or *4* there will be a voltage across the open ends, while this would not be the case if it were cut open at *1* or *2*.

The occurrence of the voltage may be pictured as follows. As was already mentioned, the charge distribution is given at every moment by the external field, but the currents which flow as a result of the change in this charge distribution may flow in different ways in the closed loop. From a given current configuration such as that of fig. 3, another possible current configuration can be derived, by adding a current which has the same intensity at all points along the closed loop and thus does not change the charge at any point, because at every point the same amount flows to the point as away from it. This extra current may then change with the time in any way desired.

If it is now desired to describe the current configuration which occurs when the loop is broken at a given spot, the extra current must so be chosen that it compensates the originally occurring charging current at the spot where the loop is broken. The voltage at that spot then becomes equal to the product of the extra current and the impedance of the loop.

It may therefore be said that upon any given connection of the indication instrument a loop will in general exhibit aerial effect; only with certain methods of connection is this aerial effect absent.

The effect which we are considering here is relatively small for long and intermediate waves. It may become somewhat larger when there is a mass of metal in the neighbourhood of the loop, for instance the chassis of a receiving set. The charge which flows up and down in the loop may then become larger; the position of the points on the loop where the aerial effect is absent may also be shifted. Large effects are not, however, to be expected. If, therefore, an arrangement is chosen for which the loop aerial can be represented by a loop standing free in space, or, if necessary, even in the neighbourhood of a mass of metal, and even with any arbitrary connection of the set with the loop, little difficulty will be experienced from aerial effect.

The arrangement mentioned is realized in non-earthed receiving sets with battery supply; in the case of these sets the aerial effect is indeed found to be so insignificant that it exerts no disturbing effect on the directional action and the sensitivity to interference of the loop.

The situation becomes quite different, however, when receivers with mains connections are used. In this case the chassis is connected with earth for high-frequency currents *via* the capacity of the mains transformer through the mains connection. The loop replaces the tuning coil of the first circuit and in ordinary receivers is therefore connected at one end with the chassis, while the other end delivers the voltage received to the control grid of the first valve. In these connections the loop aerial can be represented schematically by a loop situated relatively far above the earth with one end connected to earth (see *fig. 4*).

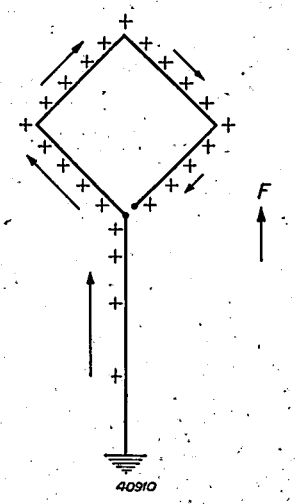


Fig. 4. Charge distribution on a loop earthed at one end due to the field F. The arrows indicate the currents which flow when the field strength increases. The currents are largest in the vertical earth connection and decrease gradually along the winding of the loop. The current is zero at the open end.

If we again consider the action of an electric field it is found that much greater displacements of charge must be expected than in the case represented by fig. 3. The type of aerial now exhibits certain similarity to that of fig. 2*b*, and, as in fig. 2*b*, we may conclude that the density of charge at the top will increase with increasing distance between the top and the neutral zone — in our case therefore with the height of the loop above earth. Thus it is no longer only the size of the loop but also its height above the earth which determines the electric charge of the loop in an electric field of a given strength. The magnetic receiving capacity on the other hand is independent of this height. It is thus clear that the aerial effect of a loop earthed at one end is affected by a quantity which has no connection at all with the magnetic receiving capacity, so that it may not immediately be concluded that the aerial effect is small compared with the signal received by magnetic induction.

There is still, however, a second reason why the loop earthed at one end exhibits much more aerial effect than the "floating" loop. As already stated, the charge Q_A, which in the case of the loop earthed at one end is drawn out of earth by an electric field, is in general much larger than the charge Q_Z, which is accumulated on one side of the neutral zone of the floating

loop under the influence of the same field. If we consider a floating loop and a loop earthed at one end, both with the same number of windings, then the same is true for the charges q_A and q_Z per winding.

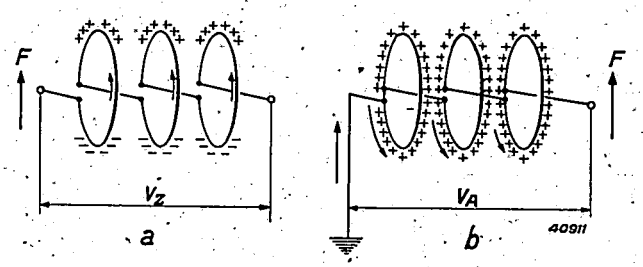


Fig. 5. Charge distribution *a*) on a floating loop, *b*) on a loop earthed at one end, due to the field F. The arrows indicate the local amplitudes of the currents which occur when F is an alternating field. In the case of the floating loop the charging current of each winding flows through that winding; in the case of the loop earthed at one end the charging current of the whole loop flows through the first winding.

In *figs. 5a* and *b* the current configurations are now given as occur in the case of the floating loop and the loop earthed at one end respectively. In the case of the floating loop the current in each winding is determined only by the charge on this winding itself: the current is equal to zero at the points indicated by dots, and at the diametrically opposite spots it has its largest amplitude ωq_Z. In the case of the loop earthed at one end, on the other hand, not only the current necessary to charge the first winding flows through this winding, but also the current which charges the following windings. If the aerial has n windings the current at the earthed side of the aerial has an amplitude of $n\omega q_A$; towards the open end of the aerial the amplitude gradually falls to zero.

The average strength of the charging current along the windings may be set equal to half the maximum strength, not only in the case of the floating loop but also in that of the loop earthed at one end. Thus for this average amplitude the following is valid :

$$I_Z = {}^1/_2\,\omega q_Z,$$
$$I_A = {}^1/_2\,n\omega q_A.$$

The voltage amplitude of the aerial effect is equal to this average current amplitude multiplied by the impedance ωL of the loop. We thus find that:

$$\left.\begin{aligned} V_Z &= {}^1/_2\,\omega^2 L q_Z \ldots \\ V_A &= {}^1/_2\,n\omega^2 L q_A \ldots \end{aligned}\right\} \ldots\ldots\ldots (2)$$

The difference between the aerial-effect voltage V_A of the loop earthed at one end and the voltage V_Z of the floating loop thus consists not only in the fact that q_A is larger than q_Z, but also that V_A is proportional to the number of windings n, and V_Z is not. The larger the number of windings, therefore, the more important the difference between the floating loop and the loop earthed at one end.

A further conclusion from equation (2) concerns the dependence on frequency. Since q_Z and q_A are proportional to F, it is clear that the voltage caused by the aerial effect is proportional to the electrical field strength and to the square of the frequency. The voltage induced by the magnetic field, however, as is of course well known, increases only with the first power of the frequency at a given value of H. From this it follows that with a given ratio of F to H the aerial effect increases proportional to the frequency.

Finally we should like to draw still another conclusion about the sensitivity of a receiving set with loop aerial to interferences from the mains. As may be seen from equation (2), the interference voltage V need not be the result of an electric field in space, but it may be caused by any phenomenon whereby an alternating electric charge q is induced on the loop. Such a phenomenon is also caused by the high-frequency interference voltages of the power mains.

We have shown that in practice an earth connection for high frequency occurs due to the supply of the set from the power mains. Because of this, interference voltages may often occur between the receiving set and earth, which are due to sparking motors and similar apparatus connected with the same mains at some other point.

In *fig. 6* these interference voltages are indi-

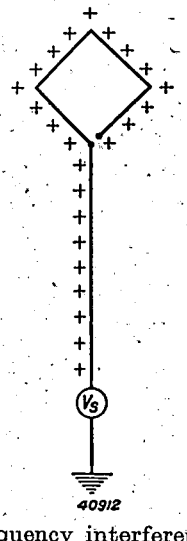


Fig. 6. A high-frequency interference voltage in the power mains can be represented in the case of loop receivers with mains connections as a source of voltage V_S in the earth connection of the loop. Such an interference voltage causes in the loop similar charge variations and consequently similar interference voltage to those caused by a vertical electric field.

cated by a source of voltage V_s. As may be seen, these interference voltages lead to an alternating charge on the loop, and this alternating charge will in turn result in an alternating voltage between the connections of the loop, which is proportional to the square of the frequency.

The relation discussed between aerial effect and sensitivity to interferences is confirmed experimentally. The less directional effect (the more aerial effect) a loop aerial exhibits, the more interference is experienced in reception, not only due to interfering electric fields acting on the aerial, but also due to interference voltages coming from the mains. Although these two interference phenomena are fundamentally of a slightly different nature, they thus usually have parallel results in practice.

Combatting the aerial effect

Different methods are used to combat the aerial effect, the three most important of which we shall discuss, namely:

1) Symmetrical construction.
2) Efficient shielding.
3) Diminishing the number of windings.

We shall confine ourselves to the case of a movable receiver which is connected with the mains, since in receivers with battery supply the aerial effect is not large enough to be disturbing.

1) *Symmetrical construction*

In the case of the symmetrical connections given in *fig. 7* the earth connection lies in the middle of the loop. The two halves have as nearly as possible the same surface area, the same self-induction and the same capacity with respect to earth and to their surroundings. The loading of the two halves is also made as equal as possible. In this way, for a vertically directed alternating electric field which thus lies in a plane of symmetry, the charging current flowing to the middle point of the loop is divided into two equal parts which flow in both directions, so that no voltage difference is generated between the ends of the loop. An alternating magnetic field, on the other hand, causes exactly the same alternating voltage as in the absence of the earth connection.

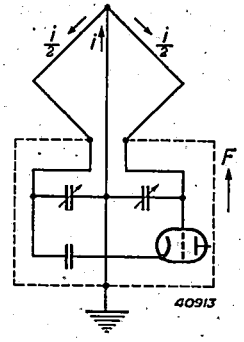


Fig. 7. Symmetrical connection of earth connection and input stage in a receiving set with loop aerial. An interference current i due to the electric field F causes no potential differences between the terminals of the loop, and consequently no variation of the control-grid voltage of the first amplifier valve.

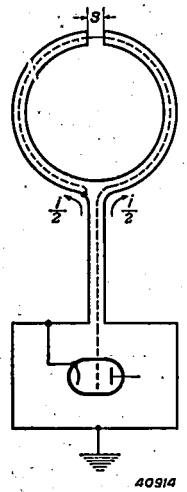


Fig. 8. Loop aerial (dotted line) surrounded by a shield. Thanks to this shield the loop winding is not exposed to electric fields. By means of the slit s circular currents are prevented in the shield. This is necessary, since otherwise the magnetic flux would also be shielded from the loop.

2) *Shielding*

If a loop is surrounded by a shield which makes the entrance of electric fields impossible, the aerial effect is of course completely eliminated. In doing this care must naturally be taken that only electric fields and not magnetic fields are screened off. This can be done by introducing slits in the shield in such a way that the current cannot flow in closed loops which are magnetically coupled with the aerial.

In *fig. 8* an example is given of such a construction. The loop windings are housed in a tube which is interrupted by the slit s. The connection wires between aerial and receiver are shielded in the customary manner.

Upon the occurrence of an external alternating field or an AC voltage in the earth connection, charging currents will indeed flow in the shielding, but there will be no electric field inside. From this it may not, however, be concluded that no AC voltage occurs on the loop, since the charging currents also have a magnetic effect on the windings of the loop. If, however, as shown in the figure, the slit is made in the middle and the two arms of the shielding tube are made equal, then the magnetic effects of the charging currents in the left-hand and the

right-hand tube compensate each other, so that no effect is exerted on the loop.

3) Limitation of the number of windings

With given dimensions of the loop aerial the voltage which is induced by an alternating magnetic field is proportional to the number of windings. The aerial effect, on the other hand, according to equation (2), is proportional to the self-induction, which increases with the square of the number of windings. From this it follows that the ratio between the aerial effect and the desired signal voltage becomes smaller as the number of windings is decreased.

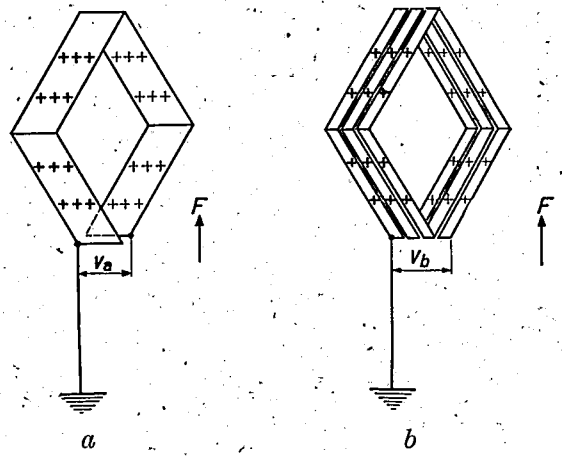


Fig. 9. Two externally identical loops which differ only in that *b*) is subdivided by a continuous slit in such a way that from a loop of originally one winding a loop of three windings is obtained. Its resistance and self-induction have thereby been increased by a factor 9.

In order to illustrate also this effect, let us consider *figs. 9a* and *b*, in which two loop aerials of equal dimensions are represented which are placed in a constant electric field *F* and earthed at one end. The only difference between the aerials consists in the fact that loop *b* has three windings while loop *a* consists of only one winding. With the same vertical field strength the charge on the two aerials will also be the same; if we then again pass over to an alternating field the same currents will flow in the earth connections. In loop *b*, however, this current must flow along a strip which is three times as long and 1/3 as wide as that of aerial *a*. From this it follows that the voltage drop along this path is 9 times as great. This holds for the ohmic as well as for the inductive component of the voltage drop.

Advantages and disadvantages of the different methods of combatting the aerial effect

In order to make a practical choice among the possibilities mentioned of combatting the aerial effect, we shall briefly discuss their advantages and disadvantages.

The *symmetrical construction* according to fig. 7 makes the input stage of the receiving set more complicated: it requires one more variable condenser, while the rest of the connections also become more complicated (floating cathode of the first valve or complete push-pull connections). The construction of the loop itself is also far from simple, considering the great precision required in the symmetry.

The *shielded loop* according to fig. 8 can be made much simpler and does not involve any complications for the input connections. The large capacity between the loop windings and the shielding in this case, however, forms a difficult problem. In order to decrease the effect of this zero capacity in the first tuning circuit, the self-induction of the loop can be chosen smaller than is desired for the first circuit and another coil can be connected in series with the loop. When this method is followed, however, the sensitivity and the ratio between signal energy and noise energy decrease.

If an attempt is made to decrease the capacity between loop and shielding directly by choosing constructions in which there is a large distance between the shielding and the windings, models are obtained which are unpleasant in appearance and which occupy relatively much space.

In the limitation of the number of windings the difficulty is encountered that the self-induction is decreased. For good matching of the tuning circuits it is desirable that the self-induction of the loop should be of the same magnitude as that of the other tuning circuits. This objection can, however, be met by connecting the loop not directly with the condenser but by means of a transformer (see *fig. 10*). If we assume that the transformer is ideal (no ohmic losses, no spreading, very large primary self-induction compared with the loop), the limitation of the number of windings results in no loss in sensitivity and no depreciation of the ratio of signal voltage to noise voltage, while a decrease of the aerial effect is obtained which is retained even after transformation. The last point is immedi-

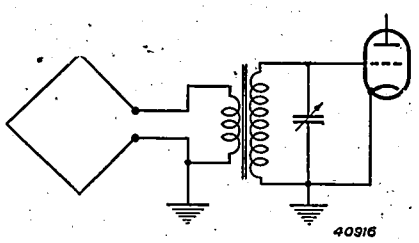


Fig. 10. Coupling of loop to first amplifier valve with the help of a transformer. By choosing a suitable value of the transformation ratio, the desired tuning frequency can be obtained with any arbitrarily chosen capacity of the rotating condenser and self-induction of the loop.

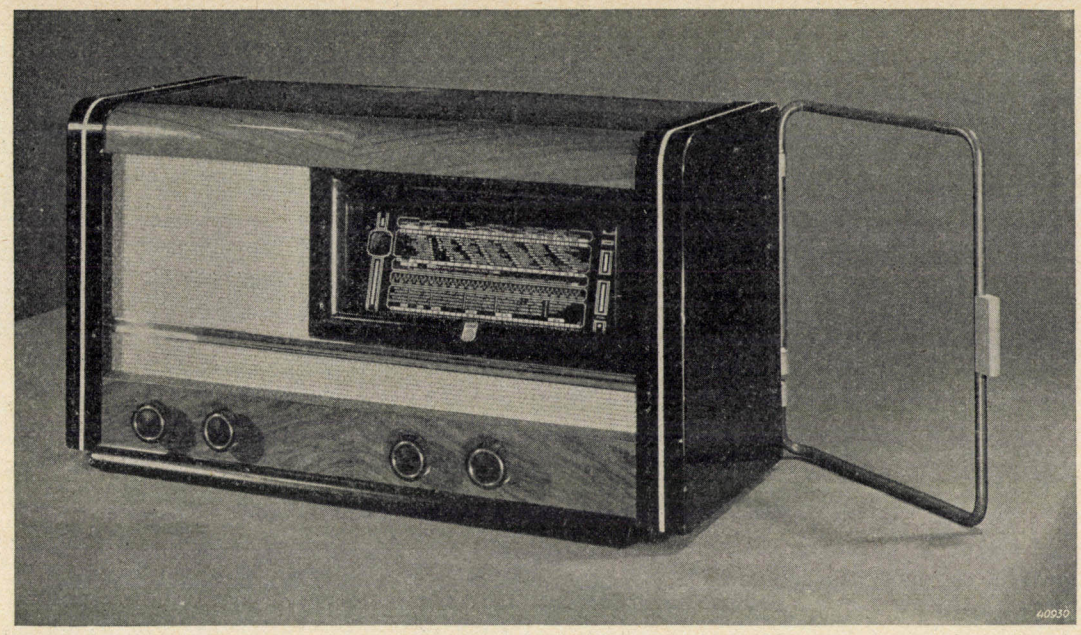


Fig. 11. The Philips receiving set 902 A with loop aerial.

ately clear, since the relation between the desired signal and the voltage which occurs due to the aerial effect is not altered by the transformation. Thus when a decrease in the aerial effect with respect to the desired signal occurs due to a decrease in the number of windings, this decrease is retained even after transformation.

The fact that the sensitivity and the noise actually do not change may be explained as follows. If L_1 is the self-induction of the loop and L_2 the desired self-induction the transformation ratio u is given by

$$L_2 = u^2 L_1.$$

On the other hand the self-induction could also be brought to the desired value by using no transformer, but by making the number of windings a factor n larger. The factor u is then determined by

$$L_2 = n^2 L_1,$$

so that $n = u$.

When the transformer is used the voltage is transformed upwards by a factor u, while upon increasing the number of windings the voltage which is induced by the magnetic field in the loop increases by a factor n. Since n and u are equal the voltage finally obtained is the same in both cases [5]).

It still remains to prove that an ideal transformer causes no change in the ratio between signal and noise.

If two connection terminals are introduced into any network consisting of resistances, capacities, self-inductions and mutual inductions, the mean square of the spontaneous fluctuation voltages between the terminals is proportional to the real part of the impedance measured between these terminals. If one now takes as connection terminals the secondary connections of the transformer, the relation between signal and noise is determined by an expression of the form V_2^2/R_2, where V_2 represents the open signal voltage between these terminals and R_2 the real component of the loop impedance as it is transformed in the secondary winding. Since

$$V_2 = u\ V_1 \text{ and}$$
$$R_2 = u^2 r_1$$

the expression $V^2{}_2/R_2$ is indeed independent of the transformation ratio.

Moreover, a change in the number of windings, with otherwise unaltered form of the loop, has no effect on V^2/R : the voltage V increases proportionally with n, while the resistance R increases with n^2, as mentioned in the discussion of fig. 9. V^2/R thus also remains constant.

In the practical application of the principle of limitation of the number of windings it is advisable to proceed so far that only one winding remains. The aerial can then be very simply constructed in the shape of a strong loop, which can easily be fastened to the set so that it can be turned. The directional effect can then be enjoyed without it being necessary to turn the set as well. From the aesthetic point of view also this is a satisfactory solution. It makes it possible to construct table sets with mains connection, which is not the case with symmetrical or shielded loop aerials, which for aesthetic reasons must be built into the set and can therefore only be considered for large cabinet models. The receiving set with loop aerial deve-

[5]) For a voltage which occurs as a result of aerial effect this does not hold. Such a voltage is, as we have seen, proportional to the self-induction of the loop and thus to n^2. With a given form and size of the loop, therefore, the increase in the number of windings results in a greater increase in voltage than a corresponding transformation ($u = n$).

loped by Philips, type 902 A, is constructed on this principle [6]).

The Philips receiving set with loop of one winding

The receiving set 902 A (see *fig. 11*) is a table set for connection with the AC mains. It is a normal superheterodyne receiver with one stage of high-frequency amplification. The single loop is fastened to the right side so that it may be turned, and it has about the same dimensions as the side to which it is fastened. If the directional effect of the loop is not desired the loop can be turned around so that is it out of sight against the rear wall. Since appreciable attenuation of the strength of reception occurs only in a small angular region in the neighbourhood of the minimum, with this permanent orientation of the loop most stations can still be received satisfactorily.

The most important problem of construction in a set with a loop of one winding is formed by the transformer by which the loop is connected with the tuning condenser.

The transmission of the aerial voltage to the input circuit will always be accompanied by certain losses, since an ideal transformer does not exist. These losses have an unfavourable effect not only on the sensitivity but also on the relation between signal and noise. The problem is therefore to construct the transformer in such a way that these unfavourable influences are kept as small as possible.

Now the set has a large amplification reserve, so that a certain diminution of the sensitivity is permissible. Therefore in designing the transformer the chief aim was to obtain as large a value as possible of the ratio between signal and noise (this need not amount to the same thing as obtaining maximum sensitivity). In this way a construction is reached which differs quite considerably from the customary one; in particular the self-inductions of the transformer windings are unusually small.

After a suitable transformer had been successfully constructed, the loop aerial of one winding was found in every respect to be a suitable solution of the problem of a movable receiving set with low sensitivity to interferences and supplied from the power mains. It is suitable for all wavelengths of the broadcasting region; upon passing from one wave region to another it is only necessary to replace the transformer by one with a different number of secondary windings [1]). In the case of systems without a transformer the switching over is much more complicated: a different loop must usually be taken for each wave region, with the accompanying possible difficulties due to the requirement that the mutual coupling of the loops must be small in order to prevent the harmful influence of self-resonances of the loops intended for long waves.

[6]) As may be seen from fig. 10, the loop is earthed at one end. The aerial effect is thus in principle still determined by the unfavourable arrangement of fig. 4, although the intensity of the unfavourable effect of the earth connection is considerably reduced by the decrease in the number of windings. If, however, it should be desirable to reduce the aerial effect still more, this can be achieved by earthing the primary winding of the transformer not at one end but in the middle. In practice this device has not been found necessary.

[7]) For the reception of short waves the transformer is omitted; then, compared with other aerial systems, the present system has no special advantages as far as the sensitivity to local interferences is concerned.

EXPERIMENTS ON THE PERMEATION OF GASES THROUGH METAL WALLS

by J. D. FAST.

620.193.29

The permeability of different kinds of metal for different gases is discussed on the basis of the theoretical conclusions stated in the previous volume of this periodical. It is found to be possible to point out one or more examples of each of the theoretically distinguishable cases. These examples are discussed and a description is given of the apparatus used by the author in the necessary experiments. The great influence is also shown of the nature of the surface in many cases on the permeability of the walls. Further conclusions are drawn about the avoidance of the difficulties to which the permeability of walls may lead in high-vacuum technique. In conclusion quantitative data are given on the way in which the permeability depends upon the temperature, while it is pointed out that the selective permeability of walls may be used for the purification and analysis of gases.

Introduction

In the previous volume of this periodical[1]) the phenomenon of the penetration of gases through metal walls was discussed. Briefly that discussion may be summarized as follows. In the most interesting cases the penetration does not occur as a result of the presence of "leaks" in the ordinary sense of the word, but as a direct result of the atomic structure of matter. Atoms (or ions) of gaseous elements are found in certain cases at not too low temperatures to be able to move between the atoms of the metal wall. Since these interstices are normally filled with electron "gas", this is true only of atoms which possess one or more unpaired electrons in the uncharged state[2]). The possibility of dissolving and diffusing therefore exists only for those gases whose molecules for the same reason consist of pairs of atoms outside the metal. As examples of gases whose atoms possess one, two or three unpaired electrons we may mention hydrogen, oxygen and nitrogen respectively.

As a result of their consisting of two atoms in the gas phase, the permeation of gases through metal walls requires at least five successive processes.

Aa): Splitting of the molecules into atoms (or ions) on the entrance surface.

Ab): Penetration of the atoms (ions) formed into the metal.

B): Diffusion in the metal.

Ca): Transition from the dissolved state into the adsorbed state on the exit surface, *i.e.* the opposite of Ab).

Cb): Recombination of the adsorbed atoms (ions) to molecules, *i.e.* the opposite of Aa).

[1]) Philips Techn. Rev. **6**, 365, 1941.

[2]) Two electrons are termed "paired" when they differ only in the spin quantum number. It is the electrons with non-compensated spins which are in the first instance responsible for the cohesion of atoms to larger aggregates in the form of the molecules of chemical compounds, or for instance of the abovementioned solid solutions.

The velocity of the total process of permeation will be mainly determined by the velocity of the slowest of these five processes. Five different cases can thus be considered in each of which a different process determines the velocity, and we shall show in this article that all five of these cases can actually be realized. Examples are ;

Aa: Hydrogen through an iron wall.

Ab: Hydrogen through an iron wall with a very rough entrance surface.

B: Hydrogen through a copper wall.

Ca: Oxygen through a zirconium wall.

Cb: Hydrogen through a palladium wall and oxygen through a copper or a nickel wall.

Aa) Permeation of hydrogen through an iron wall

One will in the first instance be inclined to hunt for the cases in which the reaction on the entrance surface determines the velocity of the permeation among those systems of gas and metal in which the energy level of the gas in the metal is higher than outside the metal. In *fig. 1* it is shown how the solubility of hydrogen at a pressure of 1 atmosphere in several metals depends upon the temperature[3]). It may be seen that there are metals whose solvent capacity for hydrogen increases with the temperature, and others of which the reverse is true. In the first group the heat of solution is negative, and one of these systems, namely the system iron-hydrogen, is indeed an example of a case in which the reaction at the entrance surface determines the velocity of permeation. Upon closer consideration it is found that the entrance surface may still exist in two states whereby in one case the splitting of hydrogen molecules into atoms and in the other the transition of

[3]) The figure is analogous to a figure in an article by G. Borelius, Ann. Physik **83**, 121, 1927, but elaborated according to the latest experimental data, which, like the older data, are mainly due to A. Sieverts and his co-workers.

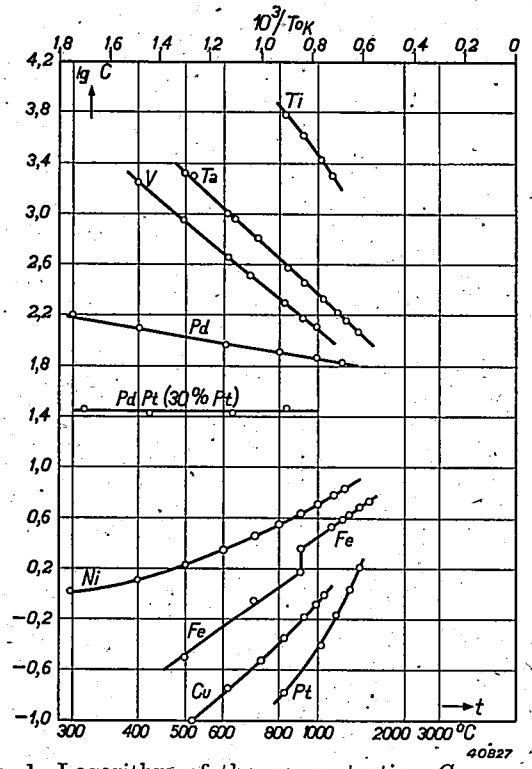


Fig. 1. Logarithm of the concentration C, expressed in atoms of hydrogen per 10^4 atoms of different metals at a pressure of 1 atmosphere, as a function of the reciprocal of the absolute temperature. For titanium, tantalum, vanadium and palladium this solubility decreases with increasing temperature, while for nickel, iron, copper and platinum it increases with the temperature. For an alloy of 70% palladium and 30% platinum the solubility of hydrogen is practically independent of the temperature.

the adsorbed atoms to the interior of the metal determines the velocity of permeation. In the case of walls which are not expressly made very rough we are concerned with the first case. We shall begin by discussing it in some detail.

It is found that iron walls only begin to transmit appreciable quantities of hydrogen above 200 to 300°C, when one side of them is brought into contact with molecular hydrogen. If on the other hand the gas, by the action of an acid or electrolytic development, is supplied to the entrance surface in atomic form, rapid permeation already occurs at room temperature. Since the conditions here are very complex, several experiments were carried out in this laboratory with gaseous atomic hydrogen [4]). One side of an iron wall was in connection with an evacuated space (H in *fig. 2*), the other with a space B in which hydrogen atoms were formed by thermal dissociation of hydrogen molecules on a glowing tungsten spiral. The dissociation took place under such circumstances that the formation of ions was out of the question.

Fig. 2 is a sketch of the apparatus used. The iron wall is situated at A and consists of a circular plate about 50 microns thick which divides a glass vessel into two parts (H and B). In B the tungsten spiral

[4]) J. H. de Boer and J. D. Fast, Rec. trav. chim. Pays Bas **58**, 984, 1939.

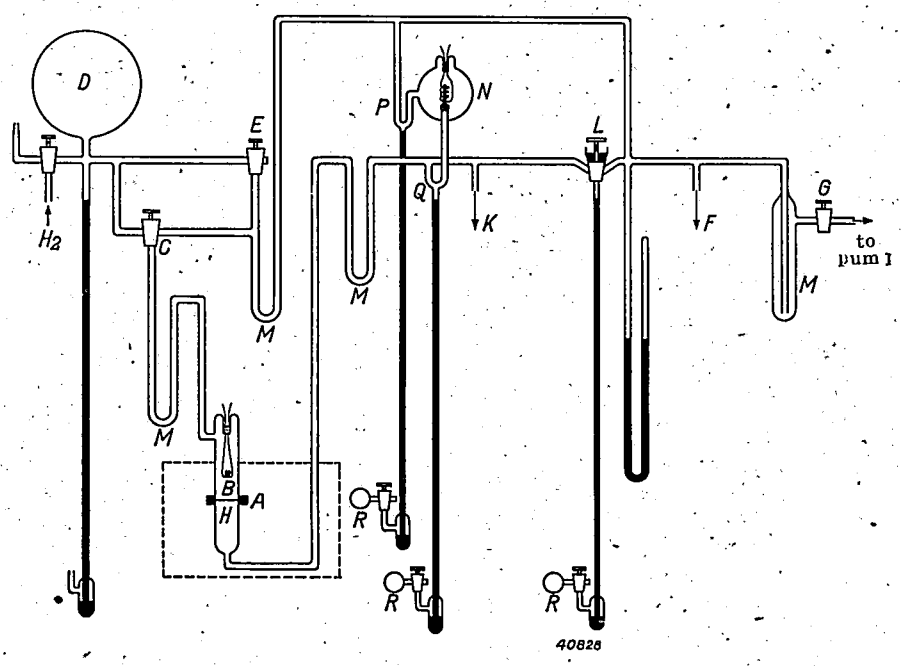


Fig. 2. Sketch of the apparatus for measuring the permeability of metal walls for hydrogen. A is an iron wall which separates the measuring vessel into two parts B and H; in B is a tungsten heating spiral. The part of the apparatus surrounded by a dotted line is placed in a thermostat. C and E stopcocks, D hydrogen reservoir. At F and K Mac Leod manometers are connected. L mercury seal and G stopcock leading to the high-vacuum pumps. M coolers for freezing out water vapour and stopcock grease. N bulb with a palladium tube with which the gas in H can be tested to ascertain whether it is really hydrogen and with which at the same time the gas can be purified. P and Q mercury seals operated with the rubber bulbs R.

is situated 4 to 5 cm away from the iron wall. The part of the apparatus between the dotted lines was placed in a water thermostat at 25° C during the experiments. *Via* stopcock *C* the space *B* could be connected with the hydrogen reservoir *D* in which the pressure of the carefully purified gas (the purification apparatus is not shown) could be varied as desired between 0 and 1 atmosphere. *Via* the stopcocks *C* and *E*, however, the space *B* could also be connected with a Mac Leod manometer *F* and at the same time *via* stopcock *G* with the high-vacuum pumps. The space *H* on the other side of the iron wall was connected with a Mac Leod manometer *K* and, *via* the mercury seal *L* and the stopcock *G*, could also be connected with the high-vacuum pumps.

Before an experiment was performed the iron membrane *A* was always tested to see whether it was tight for molecular hydrogen. For this purpose the space *B* was filled with hydrogen at 1 atmosphere and the space *H* was evacuated carefully for a long time. The requirement was then made that after this, with stopcock *L* closed, the pressure in *H* should remain too small to be measured for 24 hours ($< 10^{-4}$ mm). After this had been ascertained, the experiment proper could be begun.

In the experiment proper the hydrogen pressure in *B* was first brought to the desired value and the tungsten spiral was then heated to 2000° C. If the hydrogen pressure was greater than 1 mm of mercury, the pressure in *H* did not increase, apparently due to the fact that all the hydrogen atoms formed on the glowing surface recombine before they reach the iron wall. If on the other hand the pressure was of the order of 0.1 mm or less, the heating of the tungsten spiral resulted in a rapid increase of the pressure in *H*. *Fig. 3* gives as an example the pressure in *H* as a function of the time in an experiment in which the hydrogen pressure in *B* was kept practically constant (about 0.1 mm). If the heating of the tungsten spiral is only continued long enough the space *B* and that connected with it is completely evacuated due to the fact that all the hydrogen is finally transported through the iron wall to *H*. The permeation thus also takes place against an excess pressure because of the fact that the molecular hydrogen which accumulates at the exit side cannot penetrate into the iron and diffuse back at 25° C.

If hydrogen atoms are formed in hydrogen at a low pressure, they are also adsorbed on the glass walls. Several series of experiments were so performed that in addition to the increase in pressure in *H* the decrease in *B* and, moreover, the amount adsorbed could be measured. In one of these experiments a known amount of pure hydrogen was used, which upon opening the stopcock *L* was allowed to spread throughout the spaces *HKLFCB*. The tungsten spiral was then heated to 2000° C, while the spaces *B* and *H* were still connected with each other. Curve *1* in *fig. 4* shows the decrease in pressure as a result of the adsorption on the glass walls. After 3 hours *L* was closed, whereupon the pressure in *H* immediately began to rise with great velocity (curve *3*) and that in *B* to fall

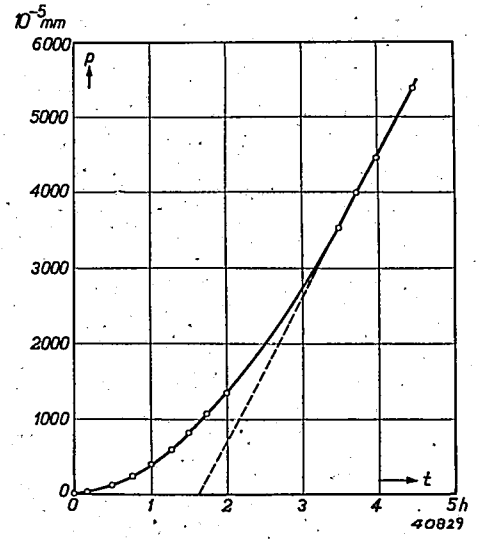


Fig. 3. The variation of the pressure p in space *H* of fig. 2 as a function of the time t at an approximately constant hydrogen pressure of 0.1 mm in *B* and a temperature of 2000° C of the tungsten spiral.

(curve *2*). As may be seen, the experiment was continued until space *B* was practically evacuated. *Fig. 5* gives the same curves as fig. 4, but recalculated by means of the known volumes on the basis of amounts of hydrogen at 0° C and 1 atmosphere. Curves *2* and *3*, as could be expected, are now exact mirror images of each other.

From the experiments it is found that already

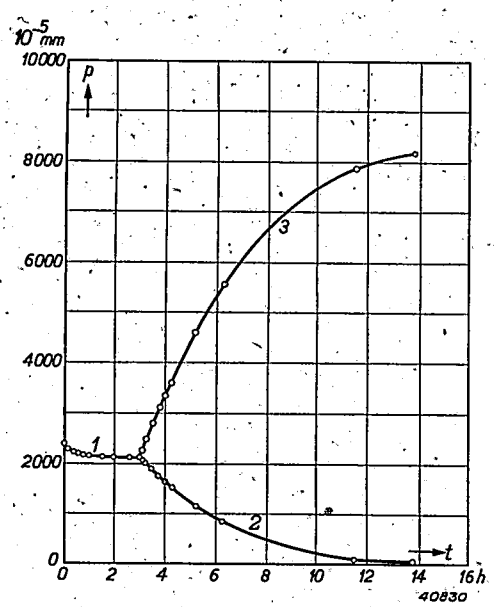


Fig. 4. The pressure p in the whole apparatus decreases slightly when the tungsten spiral is heated, as long as *L* in fig. 2 remains open, due to the adsorption on the walls (*1*). After 3 hours the mercury seal *L* is closed and the pressure in *H* begins to rise rapidly (*3*), while that in *B* falls nearly to zero (*2*).

at room temperature hydrogen atoms can penetrate into an iron wall, they possess great mobility in the metal and can easily leave the exit surface after recombination. The presence of an excess pressure of molecular hydrogen at the exit side constitutes no hindrance to this process. The process which determines the velocity upon supply of molecular hydrogen is thus the splitting of molecules into atoms on the entrance surface.

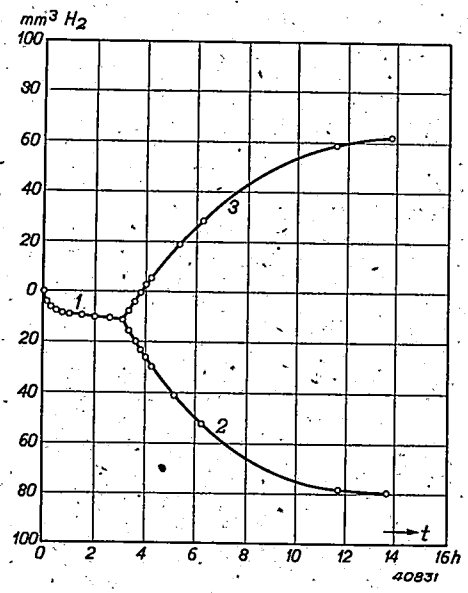


Fig. 5. The same phenomenon is shown as in fig. 4. The pressure in mm of mercury is here, however, recalculated into mm³ of hydrogen at 0°C and 1 atmosphere, whereupon curves *2* and *3* become mirror images of each other.

Ab) Permeation of hydrogen through an iron wall with a very rough surface

Remarkably enough, from measurements of the adsorption of hydrogen on specimens of iron in the powder state [5]) it is found that activated adsorption, *i.e.* splitting of molecules into atoms, already occurs at room temperature. On the rough irregularly shaped surfaces of the grains of which such powders consist the splitting thus already occurs at an appreciably lower temperature than is the case on the relatively smooth surfaces of iron walls. The phenomena are thus very closely connected with the nature of the surface, and in connection with this it is obvious that the entire surface of the iron powders cannot be assigned the same activity. It must rather be assumed that at certain "active spots" the energy of activation of the atomic adsorption is very small and that all possible conditions between slightly and very active spots exist [6]).

It may now be asked what will happen when by mechanical or chemical means one side of an iron wall is made very rough and molecular hydrogen is supplied to this side. Apparently even at room temperature activated adsorption will occur, and in connection with the foregoing one would perhaps be inclined to expect that the wall will now also already allow the passage of hydrogen. It is, however, more probable to expect that the hydrogen atoms (ions) formed on the surface are bound to sharp points and edges, and, from the point of view of energy, they are there so much more favourably placed than in the interior of the metal that the energy difference between these two conditions at room temperature cannot yet be overcome by an appreciable fraction of the atoms. If this is correct, the permeability of an iron wall upon being supplied with atomic hydrogen should become smaller upon making the entrance surface rough. On the other hand it may be expected that the permeability will become greater if the hydrogen is supplied in the molecular form. The activation energy which is necessary for the splitting of molecules into atoms has become much smaller due to the roughening of the surface. At higher temperatures the hydrogen atoms will be able to migrate over the surface to spots where they can more easily penetrate into the metal. This is in many cases, as represented diagrammatically in *fig. 6*, already equivalent to a partial penetration into the metal.

The facts observed are found to agree entirely with the expectations just mentioned. If an iron wall is made in the form of the cathode of a glow discharge in hydrogen [7]) the gas is supplied to the entrance surface in the form of ions. In ordinary cases part of the ions are to leave the metal on the other side. When however, regular scratches were made on the entrance side the permeability decreased, and with steadily decreasing distance between the scratches it finally fell to practically zero (at room temperature).

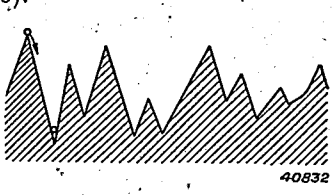


Fig. 6. Diagrammatic representation of a surface along which diffusion takes place. It is clear that this may be equivalent to a partial penetration into the metal.

[5]) For references to the literature, which are not given fully in this article, the reader is referred to: Chem. Wbl. **38**, 2 and 19, 1941.

[6]) This conception corresponds completely with experience gained in the field of heterogeneous catalysis, where often only a fraction of the total surface of the catalyst exhibits catalytic activity.

[7]) H. Betz, Z. Phys. **117**, 100, 1940.

On the other hand it has been observed [8]) that the permeability of iron and nickel when molecular hydrogen is supplied can be appreciably increased by making the surface rough (by chemical methods). A permeability already observable at room temperature cannot, however, be obtained under these conditions by making the surface rough.

In the example given it must further be expected that certain catalyst poisons will exhibit an effect which is the opposite of that of making the surface rough. Macroscopically smooth surfaces probably also still possess somewhat active spots which can be rendered inactive by such poisons. In complete agreement with this expectation it is found for instance that certain arsenic and mercury compounds in electrolytic experiments, and sulphur compounds in etching experiments cause the amount of hydrogen transmitted to increase considerably. Conversely, with the application of hydrogen in the molecular form a poisoning of the surface (in this case by a chemisorbed layer of oxygen) is found to cause a decrease of the permeability.

It must therefore be assumed that the surfaces will in general not be uniform. The activation energies of adsorption and desorption and the heat of adsorption will vary in many cases from point to point. The character of a rough surface will, however, be determined mainly by the most active spots. Upon application of molecular hydrogen the splitting into atoms takes place chiefly on these spots. Upon the application of atomic hydrogen also these spots play an essential part, as will appear from the following. The activation energy of the surface diffusion will be relatively small on the smooth parts of the surface, smaller than the energy necessary for the penetration into the metal and that necessary for recombination. A large number of the atoms which alight on these smooth parts therefore already move over the surface at room temperature. When two atoms meet, recombination takes place if they have sufficient energy to make up the required activation energy. Part of the atoms moving over the surface, however, disappear into the interior of the metal before they have the opportunity of recombining. If the smooth parts are small, most of the atoms reach the active spots and are fixed there at room temperature, due to the fact that the binding forces are much stronger there. Upon a continued application of atomic hydrogen these active spots will act as recombination centres. There is reason to assume that they lower not only the activation energy of dissociation, but also that of recombination. The fact that recombination always takes place more rapidly on rough surfaces than on smooth ones supports this assumption.

If this is correct, then upon the application of atomic hydrogen a change in the nature of the exit surface should have an effect on the permeability opposite to that of the same change in the nature of the entrance surface. The influence must, however, be much smaller, since it is clear from the foregoing that the permeability of an iron wall for hydrogen is mainly determined by the entrance surface. It is indeed found that the permeability can be increased by making the exit surface rough, but only by a factor 2. Upon the application of molecular hydrogen such parallel changes in the nature of entrance and exit surfaces will on the other hand change the permeability in the same direction.

The table gives a survey of the influence of different factors.

Permeability	by roughening	by smoothing (poisoning)	by roughening	by smoothing (poisoning)
	of the entrance surface		of the exit surface	
for atomic hydrogen	decreased	increased	increased	decreased
for molecular hydrogen	increased	deereased	increased	decreased

B) Passage of hydrogen through a copper wall

The case in which the velocity of diffusion in the metal determines the permeability also appears to exist. This can be deduced with fairly great probability from measurements of the permeability of complex membranes Cu-Pd-Cu and Pd-Cu-Pd for hydrogen [9]). At 400° C the permeability of palladium for hydrogen is 10^5 tot 10^6 times as great as that of copper. If the total thickness of the copper in both complex membranes is the same, the same permeability will be expected for both if the transmission of the copper is not determined by a process on one of the boundaries between gas and metal. If, however, the latter is true, very divergent permeabilities must be expected. The permeabilities were found to be the same.

Ca) Passage of oxygen through a zirconium wall

An example of a case in which the reaction at the exit surface determines the permeabiilty is the passage of oxygen through a zirconium

[8]) C. J. Smithells and C. E. Ransley, Proc. Roy. Soc. A, **150**, 172, 1935.

[9]) H. W. Melville and E. K. Rideal, Proc. Roy. Soc. A, **153**, 89, 1935.

wall. The heat of solution is strongly positive and the energy level of the oxygen in the metal is thus very low. For the chemisorption on the surface very little or no activation energy is needed. At room temperature and lower temperatures zirconium is already covered with a film of oxide upon exposure to gaseous oxygen. The penetration of the oxygen requires a much larger energy and only takes place with appreciable velocity above 600° C. At temperatures above 1000° C, for instance, the velocity of diffusion of the oxygen in the metal is fairly high. In spite of this it is impossible to drive the oxygen dissolved in zirconium out again by heating to a very high temperature in a high vacuum. At the exit surface therefore an insurmountably large energy difference must be overcome. There are indications that this large activation energy is required for the transition of the atoms (ions) from the dissolved to the adsorbed state.

The case discussed clearly demonstrates how necessary it is to make a sharp distinction between p e r m e a b i l i t y and v e l o c i t y o f d i f f u s i o n: the permeability of a zirconium wall for oxygen is zero notwithstanding the high velocity of diffusion in the metal.

Cb) Passage of hydrogen through a palladium wall and of oxygen through copper walls

For the passage of h y d r o g e n through walls of metals such as palladium, zirconium, tantalum, etc. the rate-determining process must be sought at the exit surface. In this case the energy level in the metal is lower than outside (see fig. 1). In contrast to the passage of oxygen through zirconium walls, however, the permeability at high temperatures is greater than zero. We shall consider the system palladium-hydrogen as an example.

From many experiments it is known that palladium in contact with m o l e c u l a r hydrogen can take up large quantities of this gas at room temperature or slightly higher temperatures, if the surface is in a favourable condition. Moreover, the velocity of diffusion in the metal under the same conditions is already quite high. Below 100° C, however, the gas is unable to leave the metal. In order for the hydrogen to leave the metal, therefore, a much larger activation energy is required than for entering it or for diffusion in the metal. This larger activation energy is apparently n o t required for the transition of the hydrogen atoms from the dissolved state into the adsorbed state, but for the recombination of the adsorbed atoms (ions) on the surface. At room temperature palladium does indeed lose hydrogen when it is situated in an atmosphere containing oxygen, apparently due to the fact that the hydrogen atoms o n t h e s u r f a c e react with the oxygen.

A second example in which the recombination on the exit surface determines the velocity of permeation is the passage of oxygen through copper walls. Upon heating copper which contains oxide to a high temperature, the larger oxide particles in the metal rapidly increase in size at the expense of the smaller ones, from which it is clear that the oxygen is able to diffuse in the copper. Upon heating copper containing oxygen in a high vacuum, however, it is impossible to drive the oxygen out at temperatures at which the copper itself does not evaporate appreciably. If on the other hand the heating is done in an atmosphere of carbon monoxide, which is insoluble in copper and which cannot therefore penetrate into it, the copper loses its oxygen fairly rapidly with the formation of carbon dioxide. If carbon monoxide is present then the recombination of the oxygen atoms is superfluous, since a CO molecule can react directly with an adsorbed O atom and form CO_2.

Several technical questions and quantitative data

Hydrogen-iron

The high velocity of diffusion which hydrogen exhibits at room temperature in several metals, especially iron, has led to various types of difficulties in technology. If the gas is developed on the surface of the iron in atomic form, for instance in etching or electrolytic development, it may, as discussed in the foregoing, penetrate into the metal and cause serious damage when small cavities or non-metallic occlusions are present in the metal. At those points the gas passes over into the molecular form (*i.e.* into a form in which it can no longer diffuse), and may cause enormous internal pressures. Because of this, intercrystalline cracks and bubbles filled with hydrogen (blisters) may appear on the surface. If for example electrolytic hydrogen[10]) is developed on the outside of a hollow iron cylinder, the pressure in the cylinder can be raised to 500 atmospheres. The high internal pressure does not have the least retarding effect on the speed at which the pressure increases, and it is, for reasons of safety only, advisable to interrupt the experiment at 500 atmospheres.

Even when there are no discernable cavities in the metal the hydrogen taken up is detrimental, because it causes a large depreciation in the ductility. This is only of a temporary nature,

[10]) C. A. Z a p f f e and C. E. S i m s, Metals and Alloys, **11**, 177, 1940.

since the gas is able to leave the metal at room temperature. The original condition however, is, never entirely regained, probably because the metal actually always contains submicroscopic cavities, perhaps in the form of points where only a few iron atoms are missing. Of this type of "holes" the largest can probably be expected at the grain and mosaic boundaries and in cold-worked metal at the slip planes (the specific weight of cold-worked metal is smaller than that of stress-free metal).

High-vacuum technique

In connection with the foregoing it will be clear that in the case of high-vacuum tubes whose walls are made partly or entirely of metal one must always be on one's guard against hydrogen, especially when this gas is present in the form of atoms or (non-hydrated) ions. The most striking phenomena again occur in the case of iron walls. Merely cooling the outside with water spoils the vacuum on the inside. Atomic hydrogen is formed on the outside surface by corrosion, and part of this, after diffusion, recombines on the inside surface to hydrogen molecules. In this laboratory the relation between permeation and corrosion was investigated by making use of solutions of different degrees of acidity. The apparatus was about the same as in fig .2 with the difference that it was now in addition possible to admit different solutions to B. Corrosion with the formation of atomic hydrogen was found to be the necessary condition for the occurrence of permeation. Neutral reacting and weakly basic liquids caused no corrosion and likewise no permeation; acid liquids had a corroding action already with a hydrogen-ion concentration of 10^{-6} ($p_H = 6$), which often occurs in water from the mains. With increasing degree of acidity the corrosion and the permeation increased. *Fig. 7* shows the increase of the pressure in the measuring space (H) as a function of the time for a solution with a hydrogen ion concentration of 10^{-6} (area of surface of iron 0.8 cm², thickness of wall about 50 μ, volume of measuring space about 1 liter). The amount passed through was of the same order of magnitude as the total amount developed. From the experiments it is evident that the hydrogen ions which are formed in the chemical reaction can penetrate into the iron, but that the presence of (hydrated) hydrogen ions in the liquid does not cause any taking up of hydrogen.

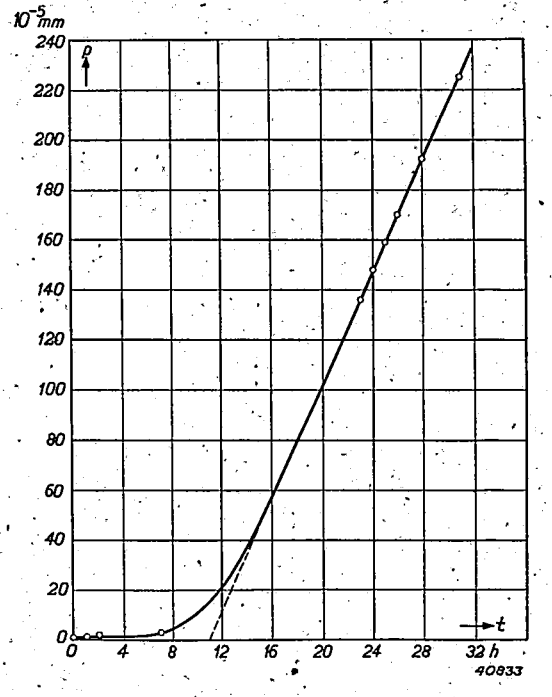


Fig. 7. Increase in the pressure in space H of fig. 2 with the time t when the wall A is in contact with a solution having a hydrogen-ion concentration of 10^{-6} g per litre which fills space B.

Not only by replacing the iron by another metal but also by the application of a different method of cooling, the difficulties described can be overcome. In the first case use is usually made of an alloy of iron with chromium or of pure copper. These metals are not only much more resistant to corrosion, but even when atomic hydrogen is expressly developed on the entrance surface a much smaller fraction of it permeates through the wall at room temperature than in the case of iron. Moreover, it will be shown (fig. 8) that upon exposure to molecular hydrogen the permeation only begins to become appreciable at much higher temperatures than with iron. In the second case use is made of cooling with non-corroding liquids or air-cooling. It is also possible to continue to use the combination of iron and water, but to cover the outside of the iron with a protecting metallic or non-metallic layer. This safety device is also necessary when the iron walls are not cooled, since an unprotected surface of iron is subject to corrosion in air.

Several quantitative data

The way in which the permeability m depends upon the absolute temperature T is found in all cases investigated to be satisfactorily represented by the formula

$$m = A\, e^{-Q/RT}, \quad \ldots\ldots\ldots \quad (1)$$

where A and Q are constants and R the gas constant.

The diffusion constant D, which indicates the velocity of transport of matter in a homo-

geneous medium, depends on the temperature in an analogous manner [1])

$$D = A'e^{-Q'/RT}, \quad \ldots\ldots\ldots \quad (2)$$

where Q' represents the heat of diffusion, *i.e.* the activation energy in diffusion.

Even in the case where in the permeation of a wall the diffusion in the metal determines the velocity (copper-oxygen) the significance of Q in formula (1) is not the same as that of Q' in formula(2),but may be described in the following way. The diffusion in the metal is in this case slow compared with the reactions on the boundary surfaces, so that it may be assumed that the equilibrium concentrations c_1 and c_2 are established in the metal immediately below entrance and exit surfaces. For a wall of unit thickness c_1—c_2 indicates the concentration gradient in the wall, and the amount which diffuses per cm² and per sec through the metal wall is then given by

$$m = D(c_1—c_2) \quad \ldots\ldots\ldots \quad (3)$$

If a vacuum is maintained at the exit side, c_2 becomes equal to zero and equation (3) becomes

$$m = Dc \quad \ldots\ldots\ldots \quad (4)$$

In a first approximation c varies according to $e^{-Q''/RT}$ with the temperature, where Q'' represents the heat of solution. The dependence of m on the temperature is thus given by a formula of the form

$$m = Ae^{-Q'/RT} \cdot e^{-Q''/RT} = Ae^{-(Q'+Q'')/RT}. \quad \ldots \quad (5)$$

Thus Q in formula (1) in this case is $Q'+Q''$, the sum of the heat of diffusion and the heat of solution. This sum is less than the heat of diffusion when the solution is an exothermic process and greater when it is an endothermic process.

If the boundary surfaces play an important part Q takes on quite a different significance. In the passage of hydrogen through an iron wall with smooth surfaces Q will be mainly determined by the activation energy required for the dissociation of a gram molecule of hydrogen into atoms (on the surface). In the passage of oxygen through copper and of hydrogen through palladium the activation energy of desorption plays this part.

In the table below the values of A and Q of formula (1) will be found for several combinations of metal and gas; as well as the values of m for 400, 500, 500, 700 and 800 °C. In order to obtain practical values m and A are not expressed in c.g.s. units but in cm³ per dm² per hour with a thickness of wall of 1 mm and at a gas pressure of 1 atmosphere at the entrance side and 0 atmospheres at the exit side. *Fig. 8* gives log m as a function of $1/T$ for these combinations of gas and metal.

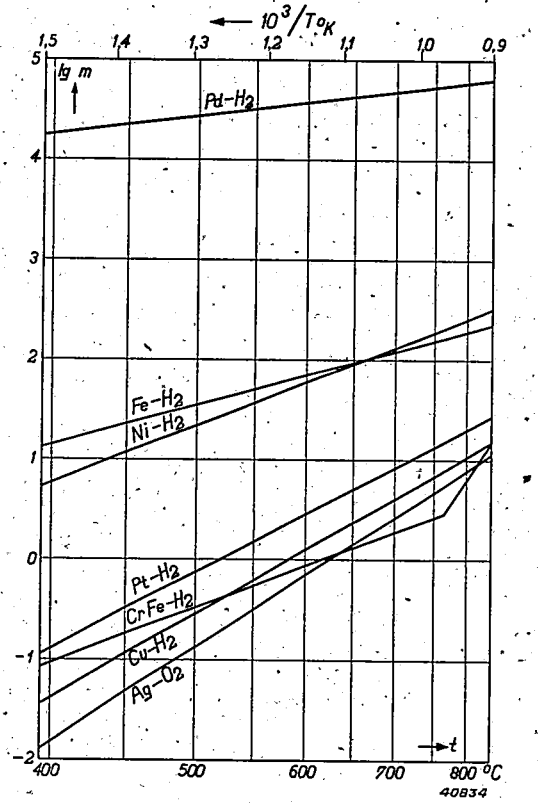


Fig. 8. Logarithm of the permeability m for different combinations of metal and gas, expressed in cm³ per dm² per hour with a wall thickness of 1 mm and gas pressures of 0 and 1 atmosphere, respectively, on the two sides of the wall, as a function of the temperature t. The reciprocal of the absolute temperature T (upper scale) is used as abscissa.

Purification and analysis of gases

Use is often made of the very high permeability of palladium for hydrogen in order to obtain the gas in a very pure state. If the palladium is heated to not too high a temperature (400 to 500° C for instance) only hydrogen and no other gas passes through it. The partial hydrogen pressure must, however, be kept low (not more than several cm mercury pressure), since otherwise the palladium very soon begins to exhibit macroscopic leaks. Use is also made of the selective permeability of hydrogen to determine the amount of hydrogen in a mixture of gases. Both possibilities were used with the apparatus of fig. 2.

In an analogous way silver can be used to obtain pure oxygen out of the air or from another gas mixture.

TABLE

Permeability $m = Ae^{-Q/RT}$ of different walls for different gases

System	A (cm³/dm² hr)	Q (kcal/mole)	m_t (t in °C) in cm³ per dm² per hr				
			400°C	500°C	600°C	700°C	800°C
Pd - H_2	4.1×10^5	4.2	18000	27000	37000	47000	58000
Fe - H_2	1.6×10^4	9.5	14	35	70	121	192
Ni - H_2	1.4×10^5	13.5	6	22	61	137	261
Pt - H_2	1.1×10^5	18.4	0.13	0.74	2.9	8.5	20
Cu - H_2	1.2×10^5	20.0	0.041	0.29	1.3	4.1	10.5
CrFe - H_2	1.8×10^3	13.3	0.09	0.33	0.88	1.9	—
Ag - O_2	3.0×10^5	22.6	0.015	0.13	0.72	2.7	8.1

MODULATORS FOR CARRIER-TELEPHONY

by F. A. DE GROOT and P. J. DEN HAAN. 621.396.619:621.395.44

The modulator of a carrier-telephone channel has as its function the modulation of the low-frequency speech vibrations on a carrier wave, while in most cases the carrier itself must be suppressed. In this article it is explained how this can be realized with a simple push-pull connection of two valves. By doubling the push-pull circuit (for instance in the form of a ring modulator) the efficiency of the modulator is improved. In addition to the desired side bands many undesired components also occur at the output of the modulator, and this is investigated here in different steps. The input signal is first assumed to be very small and a certain simple characteristic is assumed for the valves. The influence of stronger input signals is then investigated and also that of any given valve characteristic. Upon application of the double push-pull connections the last-mentioned generalization is found to have no effect. In conclusion the practical construction of the modulators is discussed with special emphasis on the effect of slight dissymmetries and the means of suppressing the carrier-wave leak caused thereby.

The fundamental features of carrier-telephony and the general construction of a carrier-telephone installation were recently dealt with at length in this periodical [1]). The principle may be summarized as follows. The speech currents which arrive in station *A* from the microphone of a subscriber (see *fig. 1*) are fed to a modulator over a fork connection and a low-pass filter. This modulator modulates the low-frequency speech vibrations on a carrier of higher frequency *f*, *i.e.* it causes the formation of two side bands of the carrier wave, one of which corresponds to a displacement of the frequency spectrum of the speech by an amount *f*, while the other is the mirror image of the first with reference to the carrier-frequency. The A.C. voltages thus obtained are fed to a band filter which suppresses one of the side bands. The side band passed is transmitted to station *B via* a telephone line with repeaters, where with the help of a demodulator it is brought back to the original low-frequency vibrations and supplied to the desired subscriber. By applying this process to the speech vibrations, using a different carrier-frequency each time, several conversations can be transmitted simultaneously over one pair of conductors.

In this article we shall discuss one of the most important components of such a system, the modulator. In order to make its action clear we shall first consider an apparatus which is used for other purposes in telephony and telegraphy, and which may be termed the blocking element.

[1]) D. Goedhart and J. de Jong, Carrier-telephony, Philips Techn. Rev. **6**, 325, 1941.

The blocking element

The blocking element consists of connections according to the principle represented in *fig. 2*. Two rectifier valves *a* and *b* are connected between two transformers T_1 and T_2, and it is assumed that with a voltage in one direction (blocking direction) they have an infinitely large resistance and with a voltage in the opposite direction (transmitting direction) they have no resistance or at least a resistance of a small constant value. The behaviour of such a valve

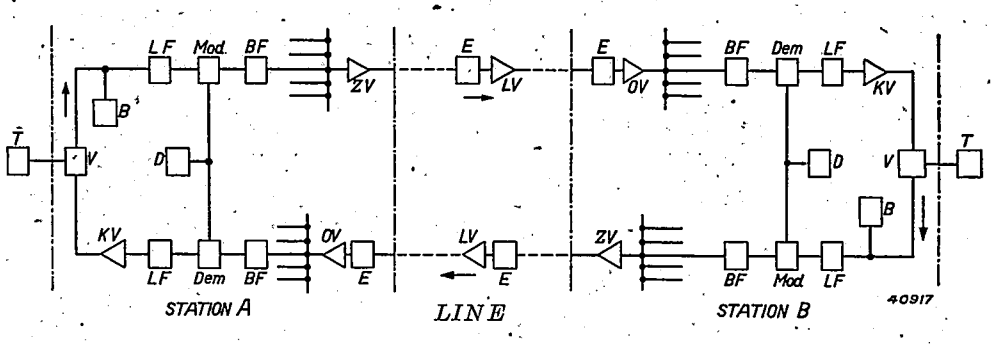


Fig. 1. Simplified diagram of a carrier-wave telephone channel. *T* subscriber's apparatus, *V* fork connection, *B* voltage limiter (for the avoidance of overloading of the repeaters), *LF* low-pass filter, *Mod* Modulator, *BF* bandfilter, *ZV* transmitting repeater, *E* equalization network, *LV* line repeater, *OV* reception repeater, *Dem* demodulator, *KV* channel repeater, *D* carrier-wave generator.

is thus described by a characteristic of the form sketched in *fig. 3*. If an A.C. voltage now acts on the input terminals *1—2* and a D.C. voltage greater than the amplitude of the transformed

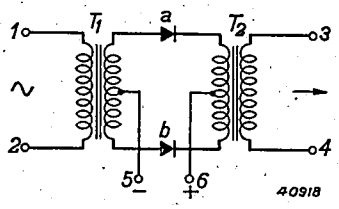


Fig. 2. Blocking connections. *1-2* input terminals, *3-4* output terminals; T_1, T_2 transformers; *a b* rectifier valves. The arrowhead of the symbol for the valves indicates the transmitting direction.

A.C. voltage on the terminals *5—6* (centre taps of the transformers), with the polarity of the D.C. voltage indicated a voltage in the blocking direction continually acts on each valve. Due to the blocking action of the valves, no current can flow, *i.e.* at the output terminals *3-4* of the blocking element none of the input voltage appears. If, however, we reverse the polarity of the D.C. voltage, a voltage in the transmitting direction always acts on the valves; the secondary voltage of T_1 now causes a current which in turn causes an A.C. voltage at the output. Depending on the polarity of the D.C. voltage on *5-6*, therefore, the connections will block or transmit an A.C. voltage acting on *1-2*.

It is clear that the blocking element acts as a switch, which is operated by reversing the polarity of a D.C. voltage. The same function could, for instance, also be entrusted to an electromagnetic relay. Compared with such a relay, however, the blocking element has not only the advantage that it requires practically no maintenance, — especially upon the use of blocking-layer rectifier valves — but also that it works entirely without time lag. This leads, for example, to the use of the blocking element as signal key for audio-frequency telegraphy. In this case an alternating current of a given frequency, for instance between 400 and 2400

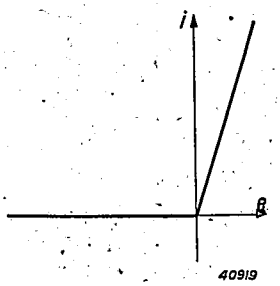


Fig. 3. Idealized characteristic of a rectifier valve. In the blocking direction, *i.e.* for negative voltage *e*, the current *i* is always zero, thus the resistance infinite; in the transmitting direction the resistance has a small constant value.

c/sec, must be interrupted in the quick rhythm of the telegraph signals (for instance Morse signals) [2]). To do this the A.C. voltage is applied to the input terminals *1-2* of a blocking element, while to the terminals *5-6* the D.C. signals of an ordinary telegraph apparatus for double-current telegraphy (*i.e.* in which the D.C. is not switched on and off, but reversed in polarity) are applied with the correct polarity. The oscillograms in *fig. 4* show the resulting appearance of the A.C. impulses transmitted. Due to the lack of inertia in the functioning of the blocking element no distortions of the telegraph signal occur even at high signalling speeds. Furthermore it may be seen that thanks to the balanced connections of fig. 2 the controlling D.C. voltage impulses themselves do not become observable at the output: due to the completely symmetrical connections they always cause equal and opposite currents in the two halves of the transformer T_2, so that the output voltages hereby induced compensate each other.

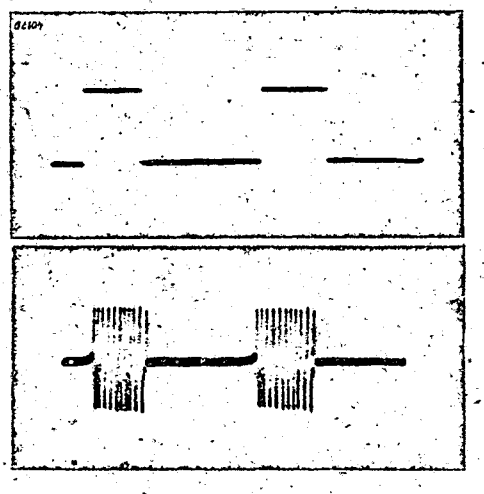


Fig. 4. Oscillograms of telegraph signals in audio-frequency telegraphy. Above: the controlling D.C. voltage impulses; below: the A.C. voltage impulses obtained at the output of the blocking element.

The blocking element as modulator

The blocking connections reproduced in fig. 2 and discussed above can now be used as a modulator without any alterations, when the low-frequency speech currents are applied to the input terminals *1-2* and the carrier wave on which the speech vibrations are to be modu-

[2]) The advantages of this system over ordinary D.C. telegraphy, where a direct current is interrupted or reversed, consist in the facts that the signals can easily be transmitted over long distances and that by the application of alternating currents with different frequencies in combination with band filters different telegrams can be sent simultaneously over the same line (thus a "carrier-wave telegraphy", entirely analogous to carrier-wave telephony except for the much smaller intervals between carrier waves).

lated is applied to the terminals *5-6*. The fact that the desired side bands actually appear on the output terminals may be explained as follows.

The carrier-voltage, which may have an angular frequency of p, will alternately open and close the blocking element due to the periodical reversal of sign. Let us now suppose that a sinusoidal voltage $Q \cos qt$ of low frequency acts on the input terminals (*fig. 5a*), and that

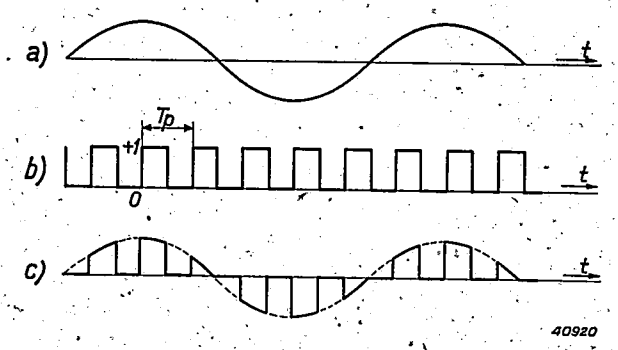


Fig. 5. At the input of the blocking connections a low-frequency A.C. voltage is applied (curve *a*). On the terminals *5-6* a much higher A.C. voltage acts with a higher frequency (carrier wave, period Tp), so that the blocking element is opened and closed periodically ("transmission") according to curve *b*). In this way an output voltage according to curve *c* is obtained.

Q is so small that the low-frequency voltage has no effect on the opening and closing of the blocking element. The blocking element then transmits only in the positive halves of the carrier wave, and thus the output voltage has the appearance shown in fig. 5*c*. The equation of this output voltage is found by multiplying the function $Q \cos qt$ by the equation of the block-form curve 5*b*, which in a Fourier series is as follows:

$$y = \frac{2}{\pi}\left(\frac{\pi}{4} + \sin pt + \frac{1}{3}\sin 3pt + \frac{1}{5}\sin 5pt + \ldots\right).$$

The equation of the output voltage thus becomes

$$e_0 = \frac{Q}{2}\cos qt + \frac{Q}{\pi}\sin (p-q)t + \frac{Q}{\pi}\sin (p+q)t + \frac{Q}{3\pi}\sin (3p-q)t + \frac{Q}{3\pi}\sin (3p+q)t + \frac{Q}{5\pi}\sin (5p-q)t + \frac{Q}{5\pi}\sin (5p+q)t + \ldots \quad (1)$$

In the output voltage the vibrations with the angular frequencies $p-q$ and $p+q$ are seen to occur. Further consideration shows that this also holds when a complex of vibrations of different frequencies $q_1, q_2, \ldots$ is fed to the input. In particular, therefore, upon the application of speech vibrations which occupy a whole band of frequencies q, two bands of the frequencies $p-q$ and $p+q$, respectively, which are the desired side bands, are obtained at the output.

At the same time it is evident that upon the presence of one frequency q at the input, in addition to the side-band frequencies $p-q$ and $p+q$ various other undesired frequencies occur in the output signal, namely the frequency q itself and, as by-products of the modulation, the side bands of higher harmonics of the carrier: $3p \pm q$, $5p \pm q$, ... These fall within the frequency band of another telephone channel and there cause cross-talk; they must therefore, like the one of the two side bands $p+q$ and $p-q$ which is not wanted, be suppressed by the transmitting-band filter (see fig. 1) following the modulator. This is especially true for the frequencies $3p \pm q$ which are only a factor 3 (*i.e.* 20 log 3 $\approx$ 10 db) weaker than the desired side band, and for frequency q which is a factor $\pi/2$ (*i.e.* about 4 db) stronger.

As in the case of the controlling D.C. voltage in the blocking element, the carrier itself does not appear at the output thanks to the balanced connections. This is important when, as is customary, it is desirable not to transmit the carrier wave itself over the line for the sake of keeping the load on the common repeaters for all channels as small as possible. The same carrier-frequency must then be added at the receiving end for the demodulation.

Modulator with double push-pull connections

If the efficiency of the connections of fig. 2 used as modulator is considered, the result is not very satisfactory. When, as has been done in equation (1), we assume the valve resistance in the transmitting direction to be equal to zero, the energy ratio of one of the side bands to the input signal amounts, theoretically, to: $(Q/\pi)^2/Q^2 = 1/\pi^2$, *i.e.* a difference of about 10 db.

A higher efficiency is obtained when the single push-pull connections of fig. 2, which are given again in *fig. 6a*, are elaborated into the double push-pull connections of fig. 6*b*. These consist of two similar parts, the first of which contains the valves a and b, the second the valves c and d. If the crossing point in the second part is considered not to be there, so that the two parts differ only in the direction of the valves, one part would give an output voltage according to *fig. 7a* and the other an output voltage according to fig. 7*b*. Upon the addition which takes place in the secondary winding of the output transformer T_2 these voltages would give the low-frequency vibration unaltered and nothing else, which is not the intention. By crossing the connections from c and d to T_2 as indicated

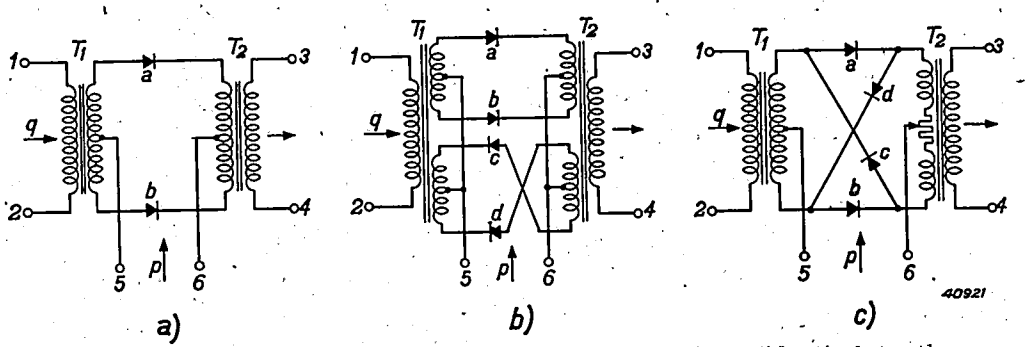


Fig. 6. *a*) Modulation with single push-pull connections (identical to the blocking connections of fig. 2).
b) Modulator with double push-pull connections.
c) Ring modulator (so called because the four valves are actually connected in a ring one behind the other).

in fig. 6*b*, however, the sign of the output voltage of fig. 7*b* is reversed, and upon addition the voltage represented in fig. 7*c* results. This voltage is the product of the low frequency

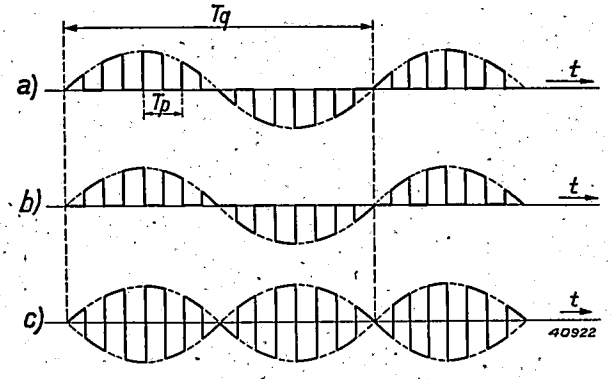


Fig. 7. *a*) Output voltage e_{ab} of the upper part of the double push-pull connections of fig. 6*b* (T_p period of the carrier wave, T_q period of the input voltage).
b) Output voltage e_{cd} of the lower part without the crossing.
c) Output voltage $e_o = e_{ab} - e_{cd}$ of the whole.

vibration $Q \cos qt$ and the block-shaped curve of *fig. 8*. If we again write the Fourier series for the latter, which is

$$y = \frac{4}{\pi}\left(\sin pt + \frac{1}{3}\sin 3pt + \frac{1}{5}\sin 5pt + \ldots\right),$$

and multiply this by $Q \cos qt$, after working out the equation for the output voltage we obtain

$$e_0 = \frac{2Q}{\pi}\left\{\sin(p-q)t + \sin(p+q)t + \frac{1}{3}\sin(3p-q)t + \right.$$
$$\left. + \frac{1}{3}\sin(3(p+q)t + \frac{1}{5}\sin(5p-q)t + \ldots\right\}. \quad (2)$$

The action of the double modulator connections, according to this formula, is in the main similar to that of the single connections of fig. 6*a*, except that the undesired low-frequency component has disappeared from the output signal, while the energy ratio of one of the side bands to the input signal is improved to $(2Q/\pi)^2/Q^2 = 4/\pi^2$, *i.e.* a difference of only 4 db. Compared with the single connections, therefore, the efficiency is a factor 4 higher, since the effective signal has become 6 db stronger with the same input voltage [3]).

The connections of fig. 6*b* can still be simplified to those of fig. 6*c*, the so-called ring-modulator, which is often used. This modulator is formed by superposing the two parts of 6*b*. It is easy to see that no change occurs in the functioning of the whole, since the superposed points of the two parts always have the same potential.

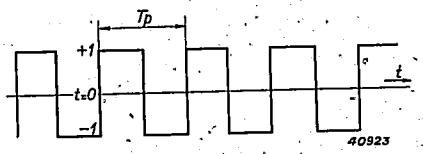


Fig. 8. Variation of the transmission of the double push-pull connections as a function of the time. T_p period of the carrier wave.

More detailed consideration of the action of the modulator

In the foregoing it has all along been assumed that the passage of the valves from the blocking state to the transmitting state and *vice versa* was controlled exclusively by the carrier wave. Actually it is not the carrier-voltage alone which is responsible for this, but the sum of the carrier and low-frequency voltages acting on each valve. Although the amplitude of the carrier voltage is much larger than that of the low-frequency voltage, the latter, depending upon its momentary value, will also contribute its bit by causing the valve to function slightly

[3]) In practice in both cases the loss is further increased by about 2 db, which may be ascribed to the losses in the valves and in the transformers.

sooner or later. The output voltage will therefore not have exactly the form shown in fig. 5*c* (or 7*c*), but the zero points in the curve will sometimes be shifted slightly to the left or to the right.

Although it is clear that this effect will cause the occurrence of new undesired components in the output signal, this cannot immediately be explained with the given method of representation. If it is desired to take this effect into account, it is no longer sufficient simply to consider the modulator as a high-frequency switch, but the behaviour of the voltage on each valve and the current caused thereby must be considered in more detail.

The relation sketched in fig. *3* between the current i and the voltage e on a valve can be represented by an exponential series with constant coefficients a_1, a_2 ...:

$$i = a_1 e + a_2 e^2 + a_4 e^4 + a_4 e^6 + \ldots \ldots \quad (3)$$

The fact that the odd terms e^3, e^5, ... do not occur here becomes clear when the function $i(e)$ is considered to be formed by the addition of the two functions i_1 and i_2 drawn as dotted lines in *fig. 9*. The first furnishes only the linear term $a_1 e$, the second, due to its symmetry, can only contain e v e n powers of e.

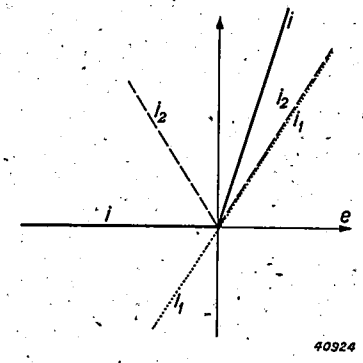


Fig. 9. The valve characteristic $i(e)$ according to fig. 2 can be regarded as the sum of the two functions $i_1(e)$ and $i_2(e)$ indicated by dotted lines.

If we now consider the simple modulator connections of fig. 6*a* we can write for the two valves a and b the following formulae:

$$\left.\begin{aligned} i_a &= a_1 e_a + a_2 e_a^2 + a_4 e_a^4 + a_1 e_a^6 + \ldots, \\ i_b &= b_1 e_b + b_2 e_b^2 + b_4 e_b^4 + b_1 e_b^6 + \ldots, \end{aligned}\right\} \quad (3a, b)$$

When, as is here assumed, the connections are completely symmetrical, and the valves are thus exactly alike, the conditions $a_1 = b_1$, $a_2 = b_2$, ... must be fulfilled.

Since i_a and i_b flow in opposite directions through the halves of the winding of transformer T_2, the output voltage e_{ab} will be proportional to the difference between i_a and i_b, thus:

$$e_{ab} \sim a_1 (e_a - e_b) + a_2 (e_a^2 - e_b^2) + a_4 (e_a^4 - e_b^4) + \ldots \quad (4)$$

In our case two A.C. voltages act on each valve, namely [4])

$$e_p = P \cos pt \text{ and } e_q = Q \cos qt, \ldots \ldots \quad (5)$$

due respectively to the carrier-wave signal and the input signal. The sum of these voltages acts on one valve, and on the other their difference:

$$e_a = e_p + e_q \text{ and } e_b = e_p - e_q.$$

This substituted in (4) gives

$$e_{ab} \sim a_1 \cdot 2e_q + a_2 \cdot 4 e_p e_q + a_4 (8 e_p^3 e_q + 8 e_p e_q^3) + + a_6 (12 e_p^5 e_q + 30 e_p^3 e_q^3 + 12 e_p e_q^5) + \ldots \quad (6)$$

Combined with (5) this becomes:

$$e_{ab} \sim a_1 \cdot 2Q \cos qt + a_2 \cdot 4\, PQ \cos pt \cos qt + + a_4 (8 P^3 Q \cos^3 pt \cos qt + 8 P Q^3 \cos pt \cos^3 qt) + \ldots (7)$$

Upon working out the products $\cos {}^m pt \cos {}^n qt$ it is found that the following frequencies occur in the output voltage:

q,		
$p \pm q$,	$p \pm 3q$,	$p \pm 5q$, ...
$3p \pm q$,	$3p \pm 3q$,	$3p \pm 5q$, ...
$5p \pm q$,	$5p \pm 3q$, ...	

while in the simpler discussion first given, where Q was temporarily assumed to be very small, only the frequencies in the first column of this table occurred. Indeed when Q is made so small that its higher powers can be ignored, all those terms in (7) which lead to the appearance of the new frequencies disappear.

Most of the new components can easily be suppressed by the band filters. This however, is, not true for the components with the frequencies $p \pm 3q$, $p \pm 5q$, ..., since when q is not too large they still fall in the side band to be transmitted ($p + q$ or $p - q$), and after demodulation at the receiving end they are heard as third and fifth harmonics, respectively, of the original signal. They thus cause a distortion of the speech. This distortion remains small when the low-frequency input voltage $e_q = Q \cos qt$ of the modulator is kept sufficiently small. The frequency $p \pm 3q$ for instance is due to the term $a_4 \cdot 8\, PQ^3 \cos pt \cos^3 qt$ in equation (7), whose magnitude rapidly falls with decreasing value of Q. By making the input voltage no larger than 20 to 30 percent of the carrier-voltage the distortion can be limited to less than 2 percent.

Formulae (6) and (7) are valid for the single push-pull connections (fig. 6*a*). If we now pass on to the double push-pull connections (fig. 6*b*)

[4]) Actually other voltages also act on the valve due to the reaction of the very output voltage to be calculated. We shall disregard the reaction here, since it makes the calculations very complicated without appreciably affecting the result.

or the ring modulator (fig. 6c) the output voltage can be found as before by adding the output voltages e_{ab} and e_{cd} of the two parts with the correct sign. Equation (6) holds for e_{ab}, and for e_{cd} an equation which, since the valves are reversed compared with a and b, can be obtained by substituting for e_p in equation (6) — e_p and for e_q, — e_q. When this is done, except for the first term, e_{cd} is found to be identical to e_{ab}, since in all the terms in $e_p{}^m e_q{}^n$ the sum $m+n$ is even; only the first term $2\,a_1 e_q$ has the opposite sign. Thus upon addition this term disappears while all the other terms are simply doubled. This agrees completely with the result of the simplified discussion first given: by the doubling of the push-pull connections the low-frequency component is suppressed and the efficiency of the modulator is increased by a factor 4.

Valve characteristics which are not straight

The double push-pull connections however, have, still another important advantage which appears when, instead of the simple valve characteristic (fig. 3) assumed, the actual characteristics of the valves in practical use are used as a basis of the calculation. In general these characteristics deviate considerably from the straight type sketched in fig. 3. In *fig. 10* for

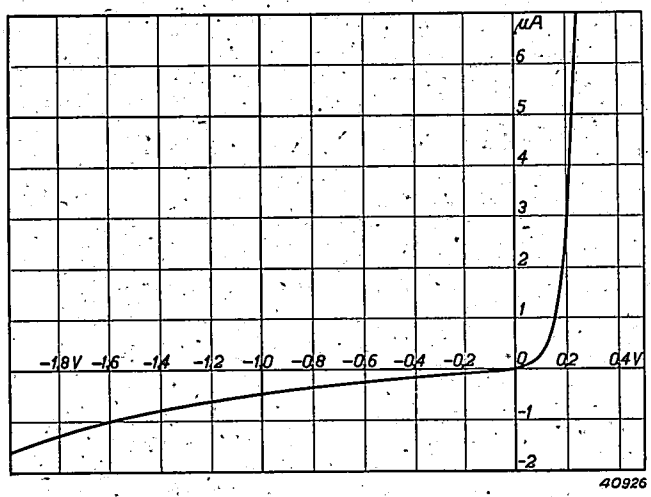


Fig. 10. Characteristic of a selenium valve as used for modulators.

instance the characteristic is given in the blocking and transmitting directions of a selenium valve [5]) as used by Philips in modulators for carrier-telephony, while in *fig. 11* the transmission characteristic is drawn for a still larger range of currents and voltages.

[5]) See W. Ch. van Geel, Blocking-layer rectifiers, Philips Techn. Rev. 4, 100, 1939.

The relation between the current i and the voltage e given by such an arbitrarily curved characteristic can be represented by a series quite analogous to equation (3), in which, however, all the odd powers of e must also occur:

$$i = a_1 e + a_2 e^2 + a_3 e^3 + a_4 e^4 + a_5 e^5 + \ldots\ldots \quad (8)$$

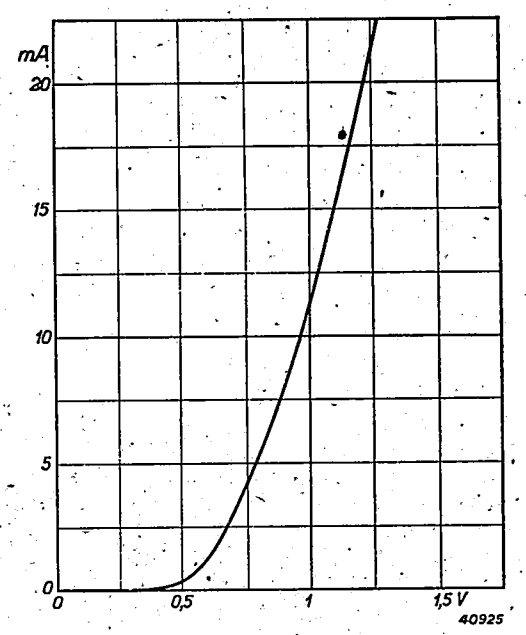


Fig. 11. Transmission region of the same characteristic as in fig. 10, but drawn on a smaller scale, so that the behaviour at the voltages actually used (order of magnitude 1 V) may be seen. The deviation from the idealized characteristic according to fig. 3 is considerable.

With this the calculation of the output voltage can be carried out in exactly the same way as above, and one then finds for the single push-pull connections instead of equation (6):

$$e_{ab} \sim a_1 \cdot 2e_q + a_2 \cdot 4\, e_p e_q + a_3\,(6\, e_p{}^2 e_q + 2 e_q{}^3) + a_4\,(8\, e_p{}^3 e_q + 8\, e_p e_q{}^3) + \ldots\ldots \quad (9)$$

The difference between this and equation (6) is that now terms in $e_p{}^m e_q{}^n$ also occur in which $m+n$ is *odd*. Because of this, as is found upon substitution of (5) and further working out, in the output signal there occur not only the side bands of the odd harmonics of the carrier-wave, *i.e.* $3p \pm q$, $5p \pm q$, etc., but also the side bands of the even harmonics, *i.e.* $2p \pm q$, $4p \pm q$, etc.

In *table I* (fifth column) this result is placed beside the frequency spectra found in the fore-

[6]) The even harmonics of the input frequency q are always missing, even in the side bands of the carrier wave and of the harmonics of the carrier wave. This is due to the balance of the connections of fig. 6a, just as in push-pull amplifiers, where the even harmonics are also found to disappear.

going, for the sake of comparison. Especially the components $2p \pm q$ (and $2p \pm 3q$, $2p \pm 5q$, ...) are a very undesirable "gain", since in order to suppress them new and more rigorous requirements are made of the transmitting-band filters.

Table I

"Straight" characteristic $Q \ll P$		"Straight" characteristic $Q \approx P$		Curved characteristic $Q \approx P$	
Single push-pull	Double push-pull	Single push-pull	Double push-pull	Single push-pull	Double push-pull
q		q		q $3q$ $5q$ ...	
$p \pm q$	$p \pm q$	$p \pm q$ $p \pm 3q$ ...	$p \pm q$ $p \pm 3q$ ...	$p \pm q$ $p \pm 3q$ ...	$p \pm q$ $p \pm 3q$ ...
				$2p \pm q$ $2p \pm 3q$ ...	
$3p \pm q$	$3p \pm q$	$3p \pm q$ $3p \pm 3q$ ...	$3p \pm q$ $3p \pm 3q$ ...	$3p \pm q$ $3p \pm 3q$ ...	$3p \pm q$ $3p \pm 3q$ ...
				$4p \pm q$ $4p \pm 3q$ ...	
$5p \pm q$	$5p \pm q$	$5p \pm q$ $5p \pm 3q$ ...	$5p \pm q$ $5p \pm 3q$ ...	$5p \pm q$ $5p \pm 3q$ ...	$5p \pm q$ $5p \pm 3q$ ...

If we now again consider the double push-pull connections and calculate the output voltage by writing out e_{ab} and e_{cd} and adding them, it will be seen that upon substituting $-e_p$ for e_p and $-e_q$ for e_q in equation (9), it is just the new terms with odd values of $m + n$ which change their signs and thus disappear upon addition. By the employment of the double push-pull connections, therefore, the disadvantage of the curved instead of straight valve characteristic is eliminated, as far as the number of components in the output signal is concerned. This may clearly be seen from the table. This advantage, together with that of the higher efficiency, leads to the common use of the double push-pull connections, in the form of the ring modulator or not as the case may be.

Practical construction of the modulators

In the practical construction of the modulators the most important point is the choice of the valves. From the detailed considerations of the action of the modulator which were based upon equations (3) and (8) for the valve characteristics, it has been made clear that it is actually not the "valve" action of the circuit elements in question, which action was emphasized in the initial general discussion, which is essential, but in general the n o n - l i n e a r i t y of the relation between current and voltage. It is especially the square term $a_2 e^2$ in the exponential series which describes this relation (equation 3 or 8) which causes the apparence of the desired side bands $p \pm q$. Every non-linear element which possesses a characteristic with this term is therefore in principle suitable for the modulator.

In the early days of the development of carrier-wave telephony (around 1920) electronic valves (triodes for example), such as those used in radio technology, were used as non-linear elements. During the last ten years, however, in the modulators as well as in other parts of the carrier-system where non-linear elements are needed, the electronic valves have been replaced by b l o c k i n g - l a y e r cells (blocking-layer rectifier valves), since the latter offer important practical advantages:

1) They possess no electrodes which require separate supply like the cathode and anode of the electronic valve. The connections and assembly therefore become simpler, the supply apparatus of the whole installation becomes smaller and the installation and running costs thus become lower.
2) The blocking-layer cells occupy very little space. This is demonstrated clearly in the photograph of *fig. 12*, where a selenium cell (selenium rectifier valve) developed by Philips for the purposes of telephony is shown together with a unit of four such cells intended for a modulator with double push-pull connections [7]). In *fig. 13* it may be seen how the unit is mounted in a complete modulator.

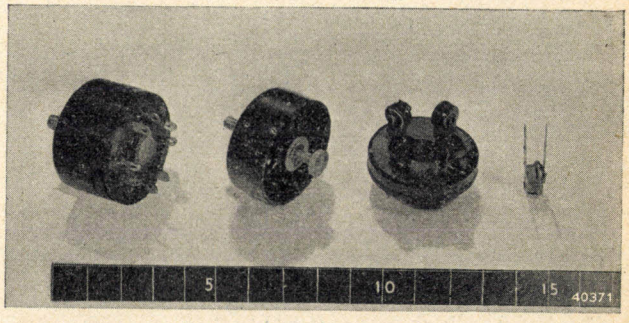


Fig. 12. On the right a single selenium cell as made by Philips for general purposes of telephony. On the left a "unit" of four cells for a modulator, in the middle the same unit with the cover removed. The centimer scale shows the small dimensions of the cell and the unit.

[7]) The small dimensions were made possible by the fact that these cells only need to carry very small currents, in contrast to selenium cells which are used for instance in supply or charging rectifiers. See in this connection D. M. D u i n k e r, Philips techn. Rev. **5**, 199, 1940.

3) The life of the blocking-layer cells with the light loading here involved is very long, so that the cells need not be placed at easily accessible spots, and greater freedom is thus obtained in the construction of the apparatus.

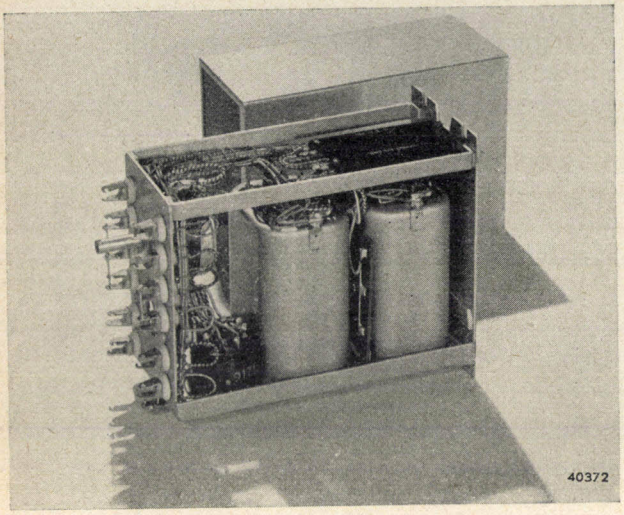


Fig. 13. Modulator of a carrier-telephone installation developed by Philips. The largest part of the chassis is occupied by two transformers. To the left below is the unit with the valves, above it a small potentiometer for balancing (see below).

In the foregoing complete symmetry of the modulator connections has been assumed: the two or four blocking-layer cells in the modulator had exactly the same characteristic, the halves of the windings of the transformers were exactly the same, etc. Actually this will never be precisely the case. In particular the characteristics of the blocking-layer valves, an example of which is given in fig. 11, always exhibit some difference; upon comparison of random examples the characteristics may even be quite different. The result of the dissymmetries hereby introduced is that in the calculation of the output signal given above equation (4) is already no longer valid, since $a_1 \neq b_1$, $a_2 \neq b_2$, etc., so that for the single (equation 6) as well as for the double push-pull connections terms in ep, e^2p, e^3p, e^2q, $epeq^2$, etc. still appear in the output voltage. Because of this, more or less intense components with the frequencies p (the carrier), $2p$, $3p$, $2q$, $p \pm 2q$, etc. also occur as products of the modulation. The oscillogram of *fig. 14* shows the effect of these on the form of the output signal.

Among these extra modulation products the carrier wave p could only with much difficulty be suppressed by the filters, while this is even fundamentally impossible for the frequencies $p \pm 2q$, $p \pm 4q$, etc. (as in the case of the components $p \pm 3q$, $p \pm 5q$, ... mentioned above), since they may fall within the desired side band itself. Now in order to keep these terms, and especially the so-called carrier-leak, small, the dissymmetries of the modulator must be limited as far as possible. For that purpose in the first place great care is devoted during manufacture not only to obtaining the greatest possible symmetry in the construction of the transformers, but also to selecting two or four valves as nearly as possible alike. Account must also be taken of the requirement that sufficient symmetry is maintained during use for a reasonable time. The dependence of the valve characteristic on the temperature and the small changes in the valve properties with time (ageing) must therefore be as nearly as possible identical for the four valves to be combined.

In the second place a small potentiometer may be connected between the halves of the windings of the primary of the transformer T_2, as was already indicated in the diagram of fig. 6*c*. By this means slight residual dissymmetries of the valves and transformers can be balanced, at least as long as the dissymmetries lie in the real part of the impedances. Since the blocking-layer valves possess a certain capacity and the transformers a certain leakage self-induction, which may also cause some dissymmetry, a complex resistance must actually be used for the balancing, for example a potentiometer in parallel with a differential condenser. In most cases, however, the carrier-leak can be adjusted to a sufficiently small value even without the last-mentioned element.

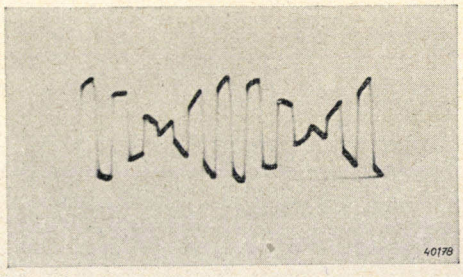


Fig. 14. Oscillogram of the output voltage of a modulator in double push-pull connection. The carrier-frequency here amounted to 6 000 c/sec, the low-frequency input voltage had a frequency of 600 c/sec. The deviations of the curve from the ideal form are to be ascribed mainly to dissymmetries in the modulator connections.

In conclusion a few words about the d e m o d u l a t o r, which serves to restore the frequency bands arrriving at the receiving end to their original frequency region. In principle it is similar to the modulator and the same apparatus can be used. If, for example, we apply to the modulator a signal with the frequency $p + q$ and a carrier wave with the frequency p, among other voltages those with the sum and difference

frequency $2p + q$ and q will occur at the output. Furthermore a number of undesired modulation products also occur again, which we can immediately derive from table I by replacing q by $p+q$. Of all the output components only those with the frequency q are desired, the others being suppressed as far as possible in the low-pass filters following the demodulator.

ON THE POROSITY OF WELDS

by J. TER BERG [1]).

621.791.056

In electric arc welding the weld is sometimes found to be porous. This porosity can be caused by the presence of sulphur. Certain welding rods are very sensitive to sulphur, while others are insensitive to it. In this article some experiments are described which illustrate the importance of the sulphur content of the material being welded on the resulting weld. In particular these experiments show that the non-uniformity of the distribution of sulphur in the work piece can affect the results obtained. It is also shown that it is advisable to use rods which are insensitive to sulphur.

One of the most important problems in welding technique is that of the measures which should be taken to obtain a weld with good mechanical properties. Especially in electric arc welding with coated electrodes it occasionally occurs that the weld is porous, which means a considerable decrease in the mechanical strength. It is clear that the occurrence of such a phenomenon or even the possibility of its occurrence is a strong impediment to the application of arc welding.

In the course of years many attempts have been made to discover the causes of the porosity. A successful result was only obtained when experiments were carried out in this laboratory with the object of investigating the relation between porosity and the presence of sulphur.

Fig. 1. Fillet weld made with a sulphur-sensitive electrode on a material containing considerable sulphur. The bead is porous and contains pocks and is, moreover, cracked.

[1]) Partially in collaboration with Ir. J. Sack †.

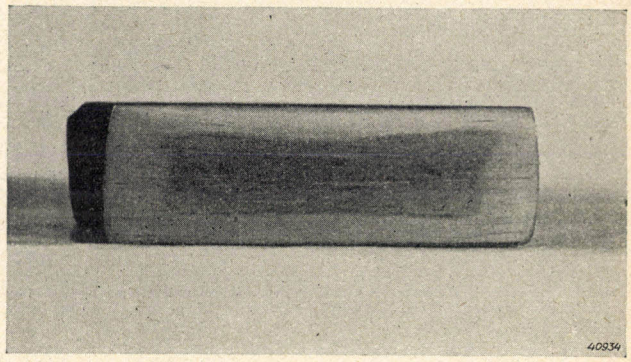


Fig. 2. Rolled and etched strip with sulphur segregation. This may clearly be seen on the polished surface.

For this purpose the element in question was expressely added to a large number of different kinds of welding rods. The addition of sulphur can be realized in a simple way by immersing the coated electrode in a saturated solution of sodium sulphate for a certain time and then drying it.

These experiments led to the surprising result that in the case of certain kinds of rods a very small amount of sulphur [2]) was able to cause the porosity [3]) under consideration, while with other types even a large percentage of sulphur had no effect on the soundness of the weld. We shall call the first type sulphur-sensitive electrodes and the second type sulphur-

[2]) The elements following sulphur in the periodic system, selenium and tellurium, exhibit this specific effect of producing porosity to an even greater degree. Phosphorus, however, which, is often classed with sulphur as far as its effect on metals is concerned, does not have this property.

[3]) By porosity in this connection must also be understood the phenomenon where the surface of the weld is covered with small depressions, usually called pock-marked.

insensitive rods. This sensitivity or insensitivity to sulphur is due to certain specific properties of the coating.

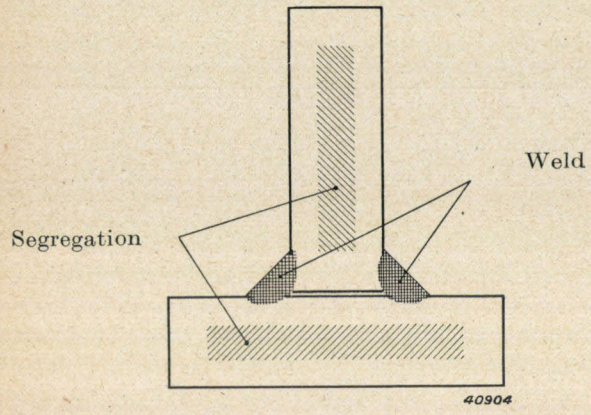


Fig. 3. Diagram of the cross section of a fillet weld. The fused zone penetrates only about 2 mm into the material and need not touch the internal segregation of sulphur.

A practical conclusion from the observations described is that the certainty of obtaining a sound weld with a sulphur-sensitive rod can be increased by choosing the sulphur content of core wire and coating very low, for instance lower than 0.015 percent in both cases. It is, however, also possible that the material to be welded contains so much sulphur that it is impossible to prevent porosity by this means. In that case it is necessary to use sulphur-insensitive electrodes. It will be clear that the use of sulphur-insensitive electrodes is far preferable for quality work, since then there need be no fear of an unfavourable composition of the material as far as the sulphur is concerned.

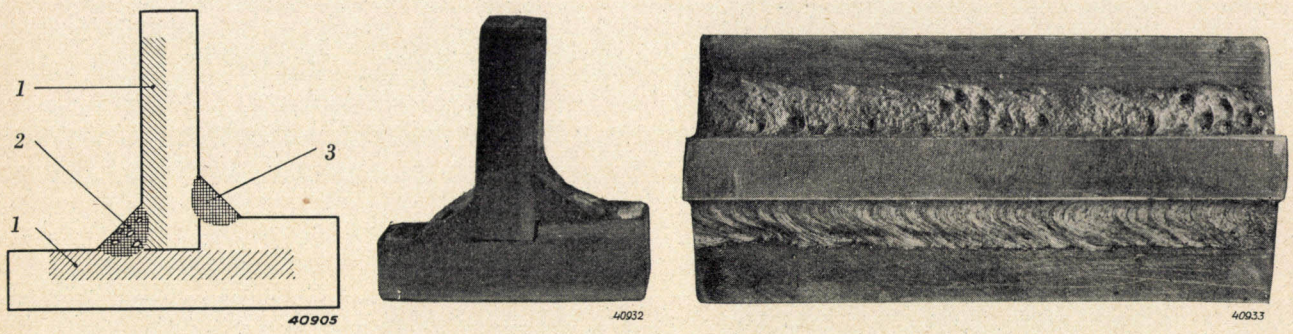
An example of an electrode which is insensitive to sulphur is the Philips 55, which has already been described in this periodical. [4]) Even free cutting steels which contain about 0.25 percent of sulphur can be welded with this rod without porosity.

In the following we shall describe several experiments which demonstrate the detrimental effect of sulphur upon the use of sulphur-sensitive electrodes, and which at the same time also give an idea of some phenomena which sometimes occur in welding practice. In this case we shall devote our attention exclusively to the sulphur content of the material to be welded.

In *fig. 1* a fillet weld is shown which results from the welding of material containing much sulphur with a sulphur-sensitive electrode. In addition to holes and pocks, cracking of the weld may also occur, as in this case. The cracking will not be discussed in this article, however.

Remarkably enough, no correlation could at first be found between the degree of porosity and the sulphur content of the work-piece as determined by a chemical analysis. This phenomenon could be explained, however, by taking into account the fact that the sulphur, like other impurities, is not as a rule homogeneously distributed throughout the material. It is known that impurities which lower the melting point of a metal become concentrated during the solidification mainly in the part of the metal that remains liquid for the longest time, thus in the interior of the block of metal. This accumulation of impurities continues to exist even after the rolling of the block to strip or other form and can be made visible by polishing and etching. In *Fig. 2* such segregations of sulphur in the interior of the strip are clearly observable. In making fillet welds in which only a relatively thin layer of the work-piece is fused (see *fig. 3*), the sulphur may not cause any difficulty and much better results may be obtained than would be expected from chemical analysis, which

[4]) P. C. van der Willigen, Philips Techn. Rev. **6**, 97, 1941.



a *b* *c*

Fig. 4. Fillet weld with a sulphur-sensitive rod in which on one side of the strips the upper layer was planed off until the segregation was reached. a) Diagram, 1 segregations, 2 porous weld, 3 sound weld. b) Cross section of the actual weld, c) view of weld from above. The difference in quality between the two welds is clearly visible.

gives the a v e r a g e sulphur content. Thus in judging the weldability of a kind of steel attention must be paid not only to the average sulphur content but also to the d i s t r i b u t i o n of this element [5]).

The correctness of the above conclusion can be demonstrated in a simple way by cutting away part of the surface of two strips until the sulphur-containing core is bared, in the manner ilustrated in *fig. 4a*. In laying a double fillet weld with a sulphur-sensitive electrode it is now found that one weld may be sound while the other, which penetrates into the segregation, exhibits many holes and is pock-marked. Fig. 4*b* shows an etched cross section of such a piece of work; in the left-hand weld one of the holes may be seen on the surface. Fig. 4*c*, a view from above, clearly shows the difference in quality between the two welds.

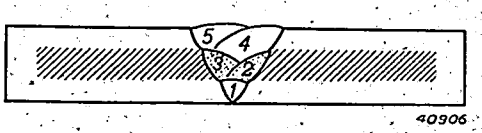


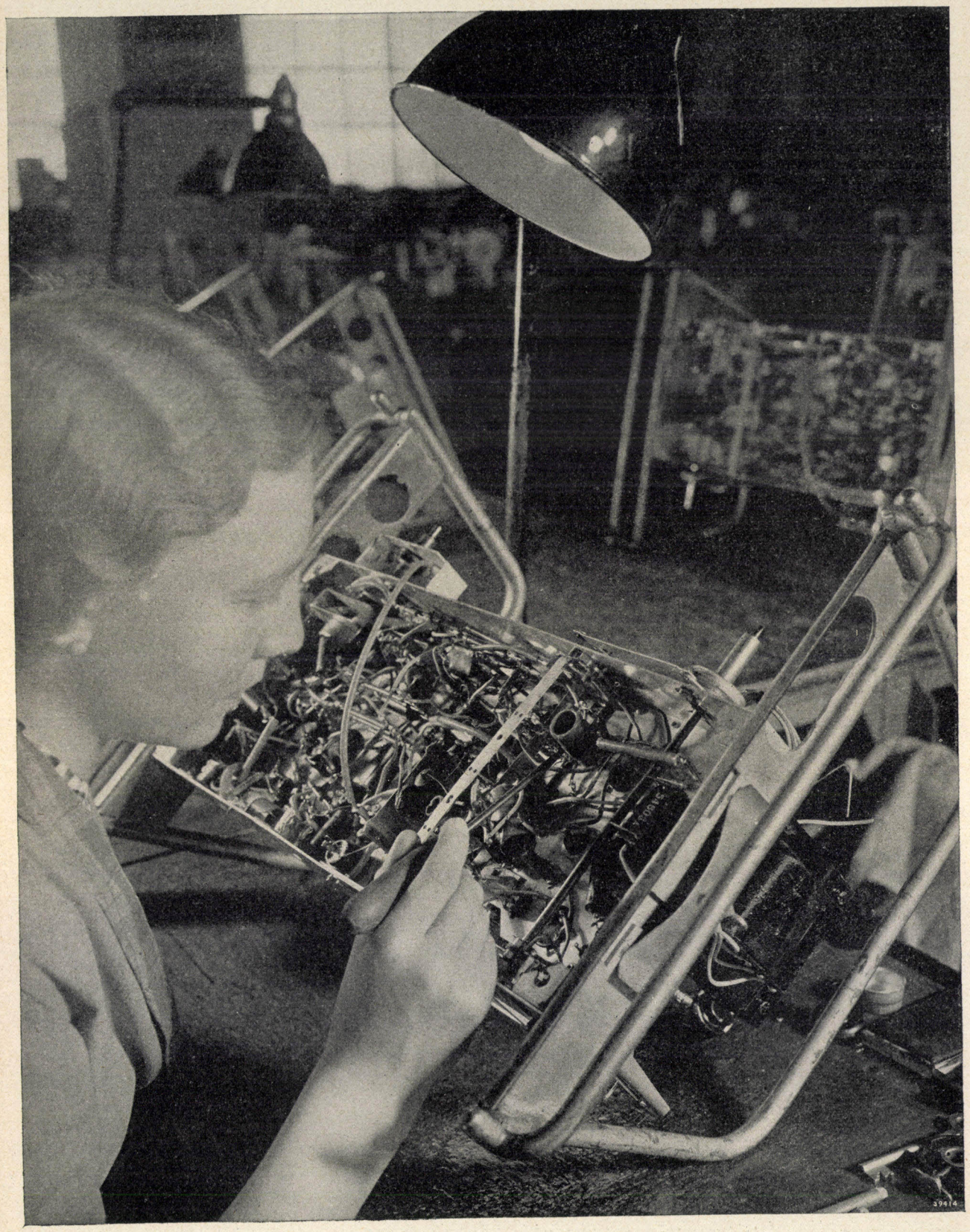
Fig. 5. Diagram of the cross section of a V-weld in a material with sulphur segregation. The beads 1, 4 and 5 are sound, the beads 2 and 3 may become porous upon the use of a sulphur sensitive electrode.

An important case where the distribution of sulphur plays a part in actual practice occurs in making a V weld as represented in *fig. 5*. Upon welding with a sulphur-sensitive rod porosity may occur in the beads *2* and *3* while the other beads are sound. Such phenomena have sometimes been observed also in other types of welds, and can thus be explained as due to the non-homogeneity of the distribution of sulphur.

In conclusion we present a general remark about the weldability of iron and steel. In various countries, including the Netherlands, committees have been appointed with the object of studying this question more closely. The function of these committees consists chiefly in formulating definite requirements, primarily relating to the chemical composition, which the material to be welded must satisfy. From the foregoing, however, it is obvious that a general solution of the problem cannot be found in that way, but that account must also be taken of the d i s t r i b u t i o n of the impurities in the work-piece and of the specific properties of the welding rod, like the sensitivity to sulphur discussed in this article.

[5]) This distribution can also be made visible, for example by pressing a polished surface of the work-piece on photographic paper. Upon contact with the sulphur the white silver chloride is converted into brown silver sulphide, and a picture of the sulphur distribution is thus obtained.

THE TESTING OF CONNECTIONS IN A RADIO RECEIVING SET



In a radio receiver there are numerous soldered and welded contacts and other connections, and a fault in any one of these may seriously disturb the working of the whole set. These contacts therefore have to be subjected to several tests and although these take considerable time they cannot be dispensed with.

SUMMARY OF RECENT SCIENTIFIC PUBLICATIONS ISSUED BY N.V. PHILIPS' GLOEILAMPENFABRIEKEN

Reprints of these publications will be gladly supplied free of charge upon request to the Administration of the Natuurkundig Laboratorium, Kastanjelaan, Eindhoven. Reprints of any publications marked with an asterik, however, are not available in sufficient number for distribution.

1561: J. L. Snoek: Effect of small quantities of carbon and nitrogen on the elastic and plastic properties of iron (Physica **8**, 711-733, July 1941).

The elastic after-effect phenomena both in pure iron and in commercial iron are discussed in so far as they are caused by nitrogen and carbon in solid solution. In all the materials examined, for so far as they are cubic body centered, the elastic after-effect appears to be of the same nature. At the temperatures examined the after-effect can be described with sufficient accuracy by one single relaxation time, which agrees with the magnetically determined value. As function of the measuring temperature the damping shows a maximum. This all refers to annealed materials. When they are cold worked this maximum disappears and at higher temperatures another maximum occurs. This latter maximum is explained as being a consequence of the diffusion damping, whilst for the first maximum an explanation is given which differs somewhat from what has previously been assumed to apply.

1562: J. L. Snoek: On the decarburisation of steel and related questions (Physica **8**, 734-744, July 1941).

The already known ferritic bands with a sharp boundary, arising from the decarburisation of steel, are attributed to the high rate of diffusion of carbon in α iron. This phenomenon is considered from the same point of view as the low affinity of carbon to α iron and the known extremely low degree of solubility. In all these respects there is a great similiarity between carbon and nitrogen. Also the plastic behaviour of iron is very characteristically influenced by very small quantities of carbon or nitrogen.

1563: J. L. Snoek: A mechanical counterpart to the Rayleigh law of ferromagnetic hysteresis (Physica **8**, 745-747, July 1941).

Measurements have been made to determine in how far the hysteresis damping of an annealed nickel wire is dependent upon the amplitude. As expected, it proves to increase linearly with amplitude. In a magnetic field of 300 Oersted such an increase disappears entirely and the damping becomes independent of amplitude. The hysteresis part of the damping proves to be rather strongly dependent upon temperature.

1564: J. F. H. Custers and G. W. Rathenau: Recrystallization in rolled nickel-iron (Physica **8**, 759-770, July 1941).

Rolled sheets of polycrystalline aluminium recrystallize in the deformation texture, whereas sheets of nickel-iron and copper recrystallize in cube orientation. This has been explained by the fact that crystals in cube orientation are found in the rolling texture of copper and nickel-iron, not in this texture of aluminium. A difference of the slip mechanism is held responsible for the difference in deformation texture. Small additions of some elements prevent nickel-iron and copper from recrystallizing in cube orientation. It is shown that contaminations of nickel-iron with phosphorus must be considered to affect not the slip mechanism but the process of recrystallization itself, since cube germs are present in the rolling texture of the contaminated samples. The recrystallization texture of pure nickel-iron itself depends on temperature. A study of the temperature-dependence with the aid of microstructures shows that after recrystallization at about 900-1000° C the cube crystals contain twins, which vanish at still higher temperatures. On recrystallizing nickel-iron and copper at high temperatures the cube orientation becomes absolutely dominating, although the majority of the deformed crystals had different orientations. By etching nickel-iron strips to small thickness before recrystallization at high temperatures, it is shown that the recrystallization texture in cube orientation gives way for an ideal rolling texture as the thickness of the sample decreases. An explanation is put forward geometrically.

1565: J. F. H. Custers: Uber die (111)-Reflexe im gewalzten und rekristallisierten Nickeleisen (111 Reflexes in rolled and recrystallised nickel-iron) (Physica **8**, 771-788, July 1941).

The rolling and recrystallization textures of a nickel iron with 48% nickel have been examined with X-rays to investigate the so-called (111) reflexions caused by the presence of crystal grains in the annealed strip, which differ from the dominating cube orientation. Part of the crystal grains are already in cube orientation in the rolled strip. The small crystals, which are not in cube orientation after heating at a temperature below 900° C, cannot be ascribed to twin formation, nor to a superposition of (111) and (112) orientations, which occur most in the rolling texture. The formal definition of these so-called Z grains, as twins turned 8° around an axis perpendicular to the twin plane (111), fits entirely with the experimental findings. In the recrystallization texture of very thin strips of nickel iron there are many of these Z grains. After recrystallization at 1100° C, however, one finds twins having an octahedral plane in common with the cube orientation. With the aid of the abovementioned Z texture one can generally well understand the rolling texture of nickel iron. Both the rolling texture and the recrystallization texture differ in the internal and in the external layers. This explains in part the connection between rolling texture and non-cubic orientation of crystal grains after recrystallization. After recrystallization the cube orientation is noticeably sharper in an external layer than in an internal one. Further, the degree of perfection depends also upon the thickness of the material; in extremely thin strips the cube orientation after recrystallization is very imperfect.

1566: W. de Groot: The decay of the luminescence of zinc-sulphide-phosphors excited by X-rays (Physica, **8**, 789-795, July 1941).

The manner in which the luminescence of zinc-cadmium-sulphide decreases with the time after irradiation with X-rays follows a hyperbolical law; the intensity of the luminescence proves to be inversely proportional to the square of $1 + t/\Theta$, where the time constant $\Theta = 3 \cdot 10^{-4}$ sec. This is entirely independent of the intensity of the previous X-ray irradiation, which means that the individual quanta of X-rays of 50 kV each excite a small volume of the crystal independently of each other. By using a constant already found in the law of the bimolecular reaction for the recombination of "free electrons" and "holes" in the crystal lattice, one finds $6 \cdot 10^{-15}$ cm^3 as the upper limit for the total volume of an electron cloud released by a single X-ray quantum. In willemite and cadmium the luminescence appears to decrease exponentially with time with a decay period of about 0.005 sec.

1567: J. L. Meyering: Some relations concerning the compressibility of the solid elements (Physica **8**, 796-804, July 1941).

For various metals the relation between atomic volume and compressibility is compared with: 1) heat of sublimation, 2) the lattice energy, *i.e.* the sum of the heat of sublimation and the ionisation heat, and 3) the melting point. An expression given by Bomke has been verified. The parallelism proves to be best in the first and third cases; the given relations can then be applied also to non-metals.

1568: J. H. Gisolf, W. de Groot and F. A. Kröger: The absorption spectra of zinc-sulphide and willemite (Physica **8**, 805-809, July 1941).

Contradictions in the results reached by different authors in regard to the absorption spectra of zinc-sulphide and willemite are reduced to differences in thickness of the layers examined, resulting in different spectral regions playing a preponderant role in the absorption in each different case. Some new measurements are reported for the absorption of ZnS-Cu crystals (dimensions about 30 μ) in the wavelength region between 3000 and 4000 Å.

1569: E. J. W. Verwey: Properties of suspensions, especially in non-aqueous media (Rec. trav. chim. Pays Bas **60**, 618-624, July-Aug. 1941).

The properties of suspensions of finely ground materials in various dispersion media, as a function of their degree of colloid chemical stability, are discussed. Special attention is given to unstable suspensions showing a phenomenon similar to coacervation (found with lyophilic colloids), and to stable suspensions with dilatant sediments.

Contents of Philips Transmitting News 8, No. 3, September 1941.

7e No. 4. pp. 97-128 APRIL 1942

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

EFFICIENCIES OF LIGHTING INSTALLATIONS

by H. ZIJL.

535.241:628.93

The empirical efficiency tables for lighting installations compiled by Harrison and Anderson more than twenty years ago are subjected to a critical examination. It is found that these fundamental empirical data can no longer be used directly for high-ceilinged rooms because of the present day use of a larger number of light points. Futhermore the examination shows that the data included in the tables probably give somewhat too low values for the illumination efficiency in the case of indirect lighting. Considering these discrepancies it may be desirable to repeat the measurements of Harrison and Anderson with modern apparatus.

The majority of the problems which arise in applied lighting technology relate to the determination of the light flux necessary to obtain a pleasing average illumination intensity E on the working surface in given circumstances. This working surface is a horizontal plane assumed to be at table height and having the same area S as the floor of the room to be illuminated. The ratio of the "effective light flux", *i.e.* the light flux which falls upon the working surface, to the nominal light flux Φ installed is called the efficiency η of the lighting installation

$$\eta = ES : \Phi \quad \ldots\ldots\ldots \quad (1)$$

This lighting efficiency depends upon the lighting system installed, upon the proportions of the room (the ratio of length to breadth to height) and upon the reflective properties of the surfaces bounding the illuminated space. The light flux to be installed can be calculated from the given illumination efficiency for a desired intensity of illumination. It must, however, be kept in mind that the illumination intensity obtained will decrease in the course of time. This may be ascribed to several causes: the light flux of the lamps decreases, dust is deposited on the light sources and on the reflecting or transmitting parts of the fixtures, the reflectivity of painted surfaces decreases with time. This is taken into account by means of a depreciation factor d, so that the nominal light flux to be installed is represented by:

$$\Phi = \frac{ES}{\eta d} \quad \ldots\ldots\ldots\ldots \quad (2)$$

It is impossible to give a fixed value for the depreciation factor; the only known quantity is the average decrease in light flux of the lamps during their lifetime. In efficiency tables therefore a value of the depreciation factor based upon normal conditions and regular upkeep is usually given. It is left to the insight and experience of the illuminating engineer to change this value according to the circumstances.

The investigations of Harrison and Anderson

More than 25 years ago Harrison and Anderson carried out fundamental investigations [1]) on the illumination efficiencies which were attained with the methods of illumination then in common use. On the basis of these measurements they compiled empirical tables [2]) which are still generally used in designing lighting installations.

Their measurements were carried out in a number of test rooms having different relations between length, breadth and height, while the reflectivity of ceiling and walls could be varied in steps from black to white. Use was made successively of naked lamps, silvered reflectors directed upwards and enamelled reflectors directed downwards. Three types of the latter were examined, namely a reflector with a "broad" light distribution with which 35 to 40 percent of the light flux emitted falls within a cone with an apex angle of twice 40°, a reflector with a "narrow" light distribution with which the proportion mentioned amounts to as much as 45 tot 50 percent, and a reflector with a light distribution lying between the other two. The last mentioned light source corresponds in the first approximation to a diffusely radiating sur-

[1]) Trans. I. E. S. **11**, 67, 1916.
[2]) Trans. I. E. S. **15**, 87, 1920.

face element with a light distribution according to Lambert's law which emits 41.3 percent of its total light flux within the above-mentioned cone.

For each of the lighting systems investigated the results of the measurement were represented graphically, and from the data obtained the empirical tables were derived by determining the efficiency for each combination of conditions separately by interpolation and extrapolation. Lighting systems which do not correspond to the standard cases investigated can be dealt with by combining the results of various tables in a suitable way. For those who have little experience in this field the use of Harrison and Anderson's tables is quite difficult, and for that reason various handbooks with more or less explanatory summaries have been published.

An objection to the data obtained in this way, which was not valid when the data were collected, is that in the measurements in a room whose length and width were about one and one half times the height of the light source only one central light source was installed. At the present time in such a room, even with indirect lighting, two to four fixtures would in general be used, and a lower efficiency would probably thereby be obtained. Only in the second step of their measurements, namely in a room whose length and width were two and a half times the height of the light source, did Harrison and Anderson use four fixtures.

A second objection which must be noted is a result of the fact that in the case of certain measurements the test rooms were abnormally small in dimensions compared with the fixtures used. As a result in one of the measurements for indirect lighting the distance from reflector to ceiling amounted to only 25 cm. With such a short distance an abnormally large part of the light flux is reflected by the ceiling back into the reflector, and this light must be considered as largely lost.

The enamel reflector already mentioned, which possesses a light intensity distribution approximately according to Lambert's law, furnishes a connecting link between the results of the measurements of Harrison and Anderson and theoretical illumination engineering. Cases in which the light distribution is according to Lambert's law can be used for exact calculations. The results of these calculations can then be compared with the results of measurements obtained with the last mentioned type of reflector.

Diffusely radiating ceiling and black walls

As test room let us consider the parallelopiped with edges a, b and h sketched in *fig. 1*. The upper surface $PQRS$, with direct illumination, corresponds to the plane in which the reflectors are situated, and with indirect illumination to the ceiling of the room. The basal plane $KLMN$ is the working surface. For the case which we shall deal with first, in which the walls and the working surface are black, the working surface receives only the light which comes directly from the fixtures or is reflected by the ceiling.

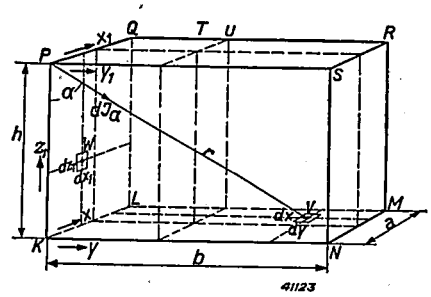


Fig. 1. Parallelopiped representing a test room. The upper surface is diffusely radiating.

If we assume that the ceiling $PQRS$ in fig. 1 is diffusely radiating with a uniform brightness, namely with a "light flux density" of F lm/m^2, and that the other surfaces are black, it is possible to calculate for each surface element of $PQRS$ the part e of the light flux emitted which will reach the working surface $KLMN$.

A surface element $dx_1\ dy_1$ lying in the angle P of the radiating surface possesses (since we assume diffuse radiation) a light intensity in the direction P to any given point $V(x, y)$ on the working surface equal to

$$dI_\alpha = \frac{F}{\pi}\cos\alpha\ dx_1\ dy_1.$$

According to the well-known laws of illumination, the intensity of illumination at point V of the plane $KLMN$ caused by the radiating of the surface is

$$dE_V \approx \frac{F\cos\alpha}{\pi r^2}\cos\alpha\ dx_1\ dy_1 = \frac{Fh^2}{\pi r^4}\ dx_1\ dy_1,$$

and the light flux which falls upon a surface element $dxdy$ at point V, when r is expressed in terms of x, y and h, is

$$d\Phi_{P\to V} = \frac{F}{\pi}\frac{h^2}{(x^2+y^2+h^2)^2}dx_1\ dy_1\ dx\ dy.$$

By integration, once between $x=o$ and $x=a$ and then between $y=o$ and $y=b$, the expression is obtained for the light flux from the surface element at P to the whole plane $KLMN$;

$$d\Phi_{P\to KLMN} = \frac{F}{2\pi}\left[\frac{b}{\sqrt{b^2+h^2}}\tan^{-1}\frac{a}{\sqrt{b^2+h^2}}+\frac{a}{\sqrt{a^2+h^2}}\tan^{-1}\frac{b}{\sqrt{a^2+h^2}}\right]\cdot dx_1 dy_1$$

The light flux emitted by the surface element $dx_1\,dy_1$ is $d\Phi_P = F\,dx_1\,dy_1$, so that the magnitude of ε is determined by the expression within the brackets divided by 2π.

If a surface element is considered which does not lie in an angle of the radiating surface, but for instance at any point T (fig. 1) the radiating surface is divided into four rectangles which have point T in common. If the receiving surface is divided in a corresponding manner, four figures are obtained to which our formulae can be applied. The total light flux received is then related to the light flux emitted as $\varepsilon = \varepsilon_1 + \varepsilon_2 + \varepsilon_3 + \varepsilon_4$, where ε_1, ε_2, ε_3 and ε_4 indicate the proportions of each of the subdivisions of the working surface. In the case of indirect illumination the contributions to the effective light flux for the whole upper surface must be added together. The arbitrary point T thus moves over the upper surface, and the general formula for the contribution to the light flux of a surface element $dx_1\,dy_1$ at T must therefore be integrated over the whole upper surface. If the upper surface is homogeneously illuminated and if it radiates according to Lambert's law, this manipulation gives for the total light flux from the upper to the lower surface

$$\Phi_{PQRS \rightarrow KLMN} = \frac{2F}{\pi} ab \left[\frac{\sqrt{b^2+h^2}}{b} \tan^{-1} \frac{a}{\sqrt{b^2+h^2}} + \frac{\sqrt{a^2+h^2}}{a} \tan^{-1} \frac{b}{\sqrt{a^2+h^2}} - \frac{h}{b} \tan^{-1} \frac{a}{h} - \frac{h}{a} \tan^{-1} \frac{b}{h} + \frac{1}{2} \frac{h^2}{ab} \ln \frac{(a^2+h^2)(b^2+h^2)}{(a^2+b^2+h^2)} \right].$$

The light flux emitted by the radiating surface is $\Phi_{PQRS} = F.\,ab.$ so that division of the right-hand member of the last formula by $F.ab$ gives the ratio ε.

For the sake of simplicity we shall confine ourselves now to the discussion of the case in which the length a and the width b of the testing room are equal: $a = b$. The space ratios can then be characterized by the so-called "form index". [3])

$$K = \frac{a}{h}.$$

A small value of K thus means a high-ceilinged room, while a room which is large in proportion to its height is characterized by a large value of K.

For the ratio ε of the light flux which reaches the working surface $KLMN$ to the light flux emitted by the radiating surface $PQRS$ we now find

$$\varepsilon = \frac{4}{\pi} \left[\frac{\sqrt{1+K^2}}{K} \tan^{-1} \frac{K}{\sqrt{1+K^2}} - \frac{1}{K} \tan^{-1} K \right] + \frac{1}{\pi K^2} \ln \frac{(1+K^2)^2}{1+2K^2} \quad \ldots (3)$$

If K is now allowed to vary, for each element of $PQRS$ the value of ε will change with it, just as for the whole radiating surface. For four spots on $PQRS$ which are indicated in *fig. 2* by I, II, III and IV, as well as for the whole surface $PQRS$ (curve V) the behaviour of ε as a function of K is given in *fig. 3* for diffusely

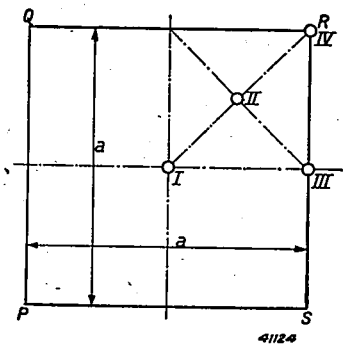


Fig. 2. Upper surface $PQRS$ of the test room with a square working surface.

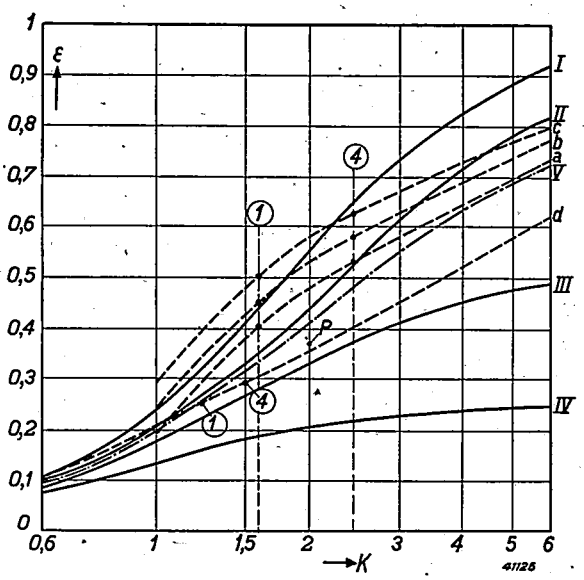


Fig. 3. The ratio ε in the case of a diffusely radiating ceiling and black walls as a function of the form index K for radiating elements lying at the points I, II, III and IV indicated in fig. 2. Curve V represents the behaviour of ε when the whole ceiling emits a homogeneous diffuse radiation which may be compared to an infinite number of similar sources of light. Curve I holds for a centrally hung light source, curve II for four symmetrically arranged elements. For large values of K (rooms which are large compared with their height), these two curves asymptotically approach 1, curve III approaches 0.5, curve IV approaches 0.25 and curve V approaches 1. Curves a, b and c hold for the behaviour of ε derived from the results of the measurements of Harrison and Anderson with three different types of enamel reflectors. For small values of K (high-ceilinged rooms) curve b follows the shape of curve I, which corresponds to a centrally placed source of light with light distribution according to Lambert's law. At $K = 3$ the curve takes on the form which is to be expected for four (curve II) and more (curve V) sources of light. The points indicated by *1* and *4* represent one and four light sources respectively. Curve d finally corresponds to indirect lighting, a loss-free reflecting ceiling being assumed.

[3]) This "form index" differs from the index used by Harrison and Anderson.

radiating ceiling and black walls. If we are concerned with a rectangular working surface it is found that curve V remains quite accurately valid when we assign to K the value

$$K' = \frac{0.2\ a + 0.8\ b}{h} \quad \ldots\ldots\ldots \quad (4)$$

where a represents the longer side of the working surface.

It may now be expected that direct illumination with enamelled reflectors will as a rule give a curve which lies between I and V. This is indeed true to a certain extent for the reflectors investigated by Harrison and Anderson for which the curves a, b and c are derived from their tables for the limiting case of absolutely black working surface and walls. As may be expected, curve a for the reflector with a wide beam is the lowest, because with the wide light distribution the loss on the walls is greatest. Curve c of the reflector with narrow beam is the highest. The light intensity curve of the reflector giving curve b lies between the other two and most nearly approaches the form of a circle (Lambert). The irregularity in curves a, b and c may be explained by the transition from results of measurements obtained with one centrally placed source of light (values measured marked *1*) to those with four (marked *4*) and more light points.

In fig. 3 the result (curve d) is also given of the measurements of Harrison and Anderson for indirect lighting in a room with diffusely reflecting ceiling, extrapolated to a reflection factor 1 and black walls and working surface. The surface $PQRS$ is here identified with the ceiling of the test-room, while in this case also r represents the ratio between the "effective" light flux falling on the working surface and the light flux emitted by the fixtures. Theoretically a correspondence should be found between d and the calculated curve V; the irregularity of the ceiling illumination with indirect lighting with fixtures should be expressed in a higher value of ε than the one for V. However, the circumstance that part of the light flux emitted by the fixture falls directly upon the black walls ("wall loss") leads to lower values for ε.

Three conclusions may be drawn from the foregoing:

1) For direct illumination there is no reason to doubt the reliability of Harrison and Anderson's measurements. Indirect lighting, however, gave experimental results for the efficiency which are lower than expected from the theory. An explanation of this may be sought in the effect already mentioned in this article of a reflection back to the fixture, which must occur in test rooms built on too small a scale, and perhaps also in the measuring instrument used.
2) The inevitable limitation in the set-up of the investigation of Harrison and Anderson does not permit the use of the tables without previous check for lighting systems which deviate very much from the basic cases.
3) There is need of a correction of the tables based upon the present-day more lavish method of illumination for lower values of the form index K (higher rooms), for which new measurements must be made.

Diffusely reflecting walls

In the foregoing we confined ourselves to an illumination by a diffusely radiating ceiling, since we have assumed that walls and working surface were absolutely black. Actually, however, the reflections at the walls furnish an important contribution to the effective light flux, so that for an anywhere near precise calculation it is necessary to take the wall reflections into account. Since an exact calculation is impossible, attempts will be found in the literature since 1920 to reduce the method of calculation into a usable form by means of certain simplifications. None of the methods developed in this way, however, gives sufficiently reliable results. Unfortunately the measurements of Harrison and Anderson have not been repeated since 1920. Their empirical tables were, however, in 1940 supplemented by measurements of installations with tubular luminescence lamps in various types of fixtures [4]).

For completely diffusely reflecting walls it is now possible to take into account the effect of the reflections on the efficiency by considering, as was done above for the ceiling, the action of a diffusely radiating vertical wall.

For a complete calculation of the illumination efficiency in a room with reflecting walls the point of departure is the light distribution of the fixture employed. With a given arrangement of the light points it is possible to deduce from this light distribution how the surfaces which bound the room are illuminated by the direct light flux. This can be done graphically, or idealized light distributions can be chosen for the general case, which are wholly or partly susceptible to exact calculation. Such a set of light intensity curves [5]) is reproduced in *fig. 4*.

When the light flux distribution on the surfaces bounding the room has been determined

[4]) Trans. I. E. S. 35, 759, 1940: W. M. Potter and W. G. Darley, The design of luminaires for fluorescent lamps.

[5]) For an indirect illumination by means of a cove along the wall the calculation is somewhat more complicated.

and thus also the direct component of the effective light flux, the light flux is then calculated which reaches the working surface after first being reflected by the ceiling. This is possible since with given proportions of the room the value of ε can be determined for every point on the ceiling. A practical method is to divide the ceiling into a number of rectangular areas

$$\Phi_{KLQP \rightarrow KLMN} =$$

$$= \frac{F}{\pi} ah \left[\tan'^1 \frac{a}{h} - \frac{\sqrt{b^2+h^2}}{b} \tan'^1 \frac{a}{\sqrt{b^2+h^2}} + \frac{b}{h} \tan'^1 \frac{a}{b} + \right.$$

$$+ {}^1/_4 \frac{b^2}{ah} \ln \frac{b^2 \ (a^2+b^2+h^2)}{(a^2+b^2)(b^2+h^2)} - {}^1/_4 \frac{a}{h} \ln \frac{a^2 \ (a^2+b^2+h^2)}{(a^2+h^2)(a^2+b^2)} +$$

$$\left. + {}^1/_4 \frac{h}{a} \ln \frac{h^2 \ (a^2+b^2+h^2)}{(b^2+h^2) \ (a^2+h^2)} \right].$$

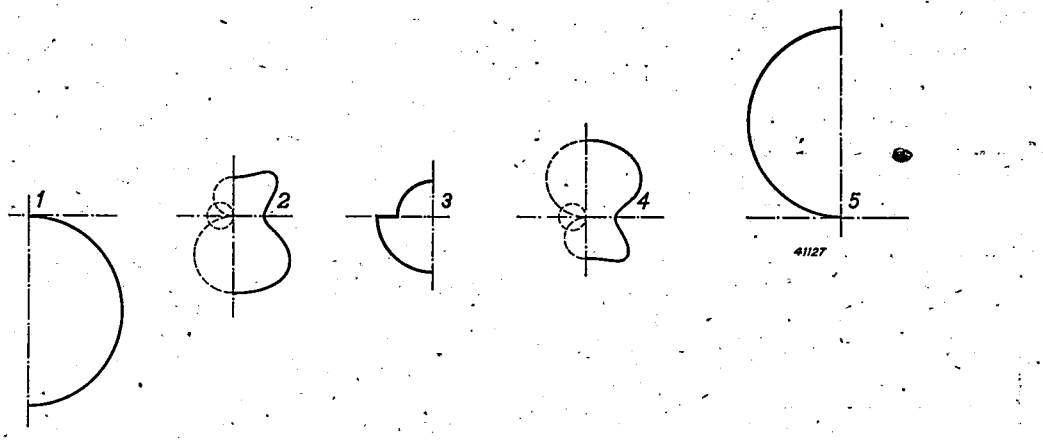


Fig. 4. Light intensity curves for five common lighting systems: 1) direct, 2) half-direct, 3) diffuser, 4) half-indirect, 5) indirect. The more complex light distributions 2) and 4) are composed of the spherical light ditributions of a direct and an indirect component, as well as the toroidal distribution of a lateral component. The five diagrams relate to the same total light flux.

to which in good approximation the same value of ε may be assigned as for the centre of each rectangle. The light flux reflected by the ceiling which is not directed toward the working surface strikes the walls, and in addition a part of the light emitted by the fixtures is radiated directly to the walls. By introducing suitable simplifications it is possible to calculate what part of the light reflected by the wall will be directed to the working surface and what part to the other walls and to the ceiling. Of these last two contributions the part is calculated which still reaches the working surface after the repeated reflection. This contribution is then multiplied by a certain factor to account for the irregularity of the distribution of the light flux over the walls in question.

In order to perform the calculations described we begin with a diffusely radiating surface element $dx_1 \, dy_1$, lying at any given point $W \ (x_1, y_1)$ on the vertical plane $KLQP$ of the parallelopiped of fig. 1, and determine the light flux $d\Phi_{W \rightarrow V}$ emitted by it which falls upon the horizontal surface element $dx \ dy$ at any given point V on the working surface:

$$d\Phi_{W \rightarrow V} = \frac{F}{\pi} \left\{ \frac{yz_1}{(x-x_1)^2+y^2+z_1^2} \right\}^2 dx_1 \ dz_1 \ dx \ dy.$$

When we integrate this light flux over the whole side wall $KLQP$ and over the whole working surface $KLMN$, assuming the illumination of the side wall to be homogeneous, we obtain [6])

[6]) This formula was already derived by L a m b e r t, although not quite directly (Photometria, 1760).

Actually the illumination of the side wall is by no means homogeneous. It is therefore desirable to divide this wall into horizontal strips and to carry out the calculation with the average value of F belonging to each separate strip. Division of the wall into strips gives sufficient accuracy.

For the calculation of the influence of the successive reflections we start for the time being with the assumption that the light fluxes concerned are entirely uniformly distributed over the areas on which they are incident. Of a light flux Φ_w which hits a wall, a fraction $R\Phi_w$ is reflected toward the ceiling as well as toward the working surface, while a fraction $W\Phi_w$ is reflected toward the other walls. Of the light flux $R\Phi_w$ falling on the ceiling a part $PR\Phi_w$ goes back again to the walls, while $QR\Phi_w$ falls upon the working surface. For the sake of simplicity we may continue to assume that the working surface is black. Due to the second reflection on the walls a light flux $PR^2\Phi_w$ is now furnished on the working surface and the same on the ceiling, while $PRW\Phi_w$ goes to the other walls.

In this way we now obtain an infinite series of light flux contributions whose sum for the total light flux furnished on the working surface by the uniformly reflecting walls is finally found to amount to:

$$\Phi_{w \longleftarrow\longrightarrow v} = \Phi_w \frac{R \ (1+Q)}{1-(PR+W)}.$$

The light flux which finally reaches the working surface from a uniformly reflecting ceiling is found in a similar way, and amounts to:

$$\Phi_{p \longleftarrow\longrightarrow v} = \Phi_p \frac{PR + Q \ (1-W)}{1 - (PR + W)}.$$

A correction must still be applied to the part of the effective light calculated in this way. Considering the fact that the light flux distribution which occurs upon repeated reflection is not uniform, the light flux

actually falling on the working surface will deviate from that calculated. Since, however, the part in question is small compared with the rest, the great accuracy which is only to be obtained by very laborious calculation is unnecessary.

As the result of a series of calculations *fig. 5* gives four curves for an indirect illumination with fixtures. The quantity ε is again the ratio of the effective light flux to the light flux emitted by the fixtures. For each curve the reflection factor of the ceiling (ϱ_p) and of the walls (ϱ_w) is indicated. The corresponding dotted line curves, taken from the tables of Harrison and Anderson, serve for comparison. The calculated curves show, as already mentioned, higher values than those found empirically. Moreover, they are considerably steeper because they are calculated for four points of light even for low values of K. If they were calculated for a single centrally placed light point up to $K = 2$, they would up to this point also be about parallel to the empirical curves at somewhat higher values of ε.

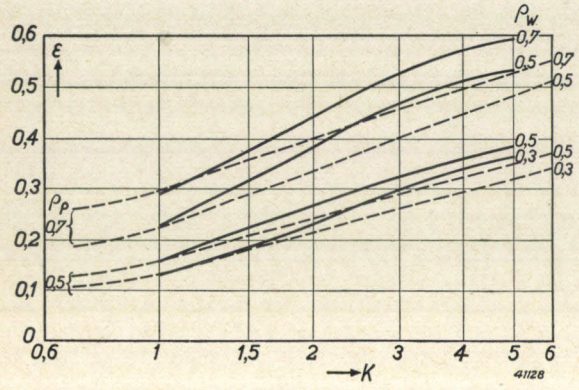


Fig. 5. The ratio ε for indirect illumination in rooms whose ceiling as well as the walls are reflecting; as in fig. 3 the form index K is plotted horizontally. The full-line curves are valid for the calculated values of ε, and the broken-line curves are derived from the tables of Harrison and Anderson. The reflection factors for ceiling and walls are indicated by ϱ_p and ϱ_w respectively.

Practical diagrams

Fig. 6 gives a completely worked out picture of the distribution of the light upon different reflections. A fixture was used with the following characteristic: 17 percent absorption in the fixture, 70 percent of the light flux emitted in the upper hemisphere and 30 percent in the lower hemisphere (see *fig. 7*). The interpretation of fig. 6 is extremely simple. On the left the reflectivity of the ceiling and of the walls is indicated. Then follows a column for the form index K for each reflection group and then the final efficiency η corresponding to every value of K. The graphs show how the light flux is distributed. Irrespective of the surroundings 17 percent of the light flux is lost by absorption in the fixture. In time this loss increases due to dust deposits. Moreover, the light flux of

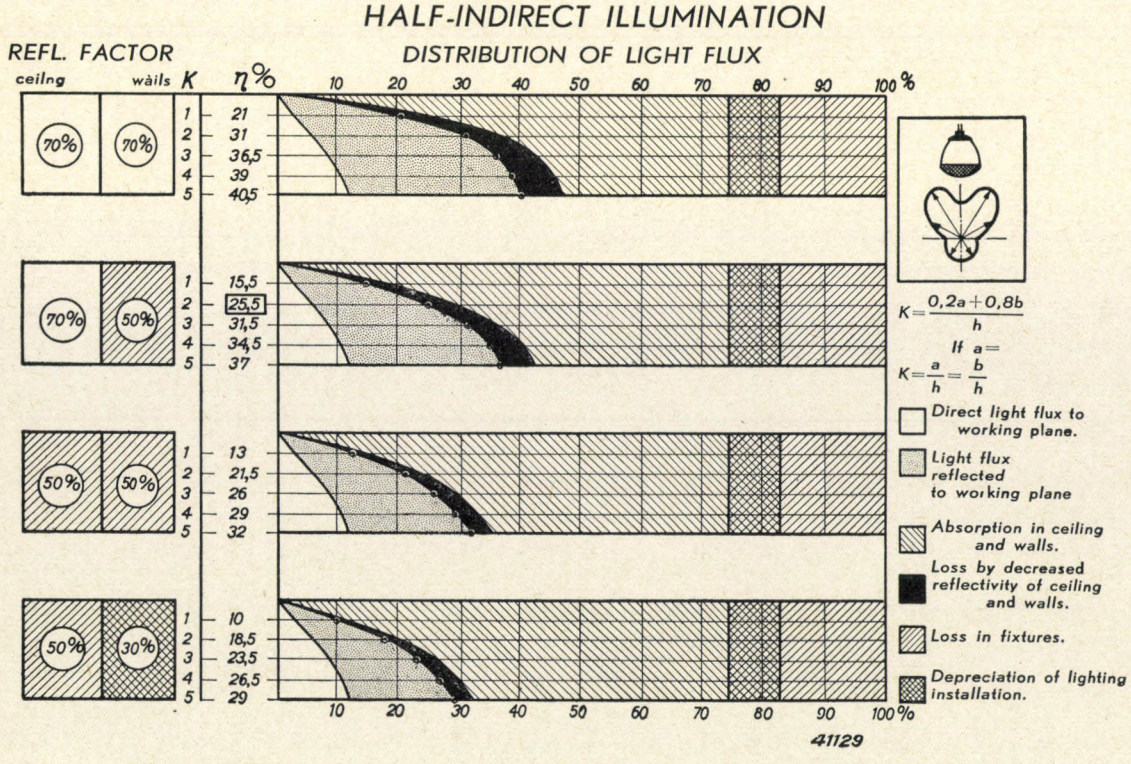


Fig. 6. Efficiency table and distribution of the light flux for a half-indirect illumination. A collection of such figures relating to average cases occurring with common lighting systems will be published in the book "Kunstlicht en Architectuur" (Artificial Illumination and Architecture) by L. C. Kalff, which will appear shortly.

the lamps decreases, so that in the ordinary case after the elapse of some time it may be expected that about 74 percent of the nominal light flux will leave the fixture. This 74 percent will be divided into the light flux which reaches

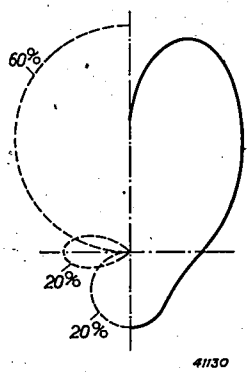


Fig. 7. Light intensity diagram for the half-indirect fixture for which the efficiency table is given in fig. 6. On the right the total light intensity diagram of the armature is drawn, while on the left the three components (direct, indirect and lateral) of which the light distribution can approximately be built up are indicated by a broken line.

the working surface directly (extreme left in the graph), the light flux which reaches the working surface after reflection on the ceiling and walls (between the left-hand and the two right-hand curves), and the light flux which is absorbed by the walls and ceiling. This last part, and with it the part before the last, changes with time, the reflecting walls become darker. If it is assumed that 70 percent reflectivity decreases by a factor 0.85, 50 percent by a factor 0.9 and 30 percent by a factor 0.95 of the original values, the transition region between the two righthand curves occurs. The two curves are calculated on the basis of 74 percent of the nominal light flux. In the beginning therefore an efficiency which amounts to 83/74 of the value indicated by the right-hand curve may be counted on. The efficiency gradually decreases to the values of the second curve from the right, which corresponds to the figures in the column beside the graph. It is evident that the parts of the effective light flux vary in the first approximation proportionally with the light flux emitted by the fixture, so that any necessary reduction offers no difficulties.

From the graph it is evident that the part of the direct light flux which is independent of the ceiling and wall reflections increases absolutely as well as relatively with the form index. The diffuseness of the illumination decreases upon decrease of K. The form index K (see equation (4)) introduced by us for a diffusely radiating ceiling is found to be sufficiently accurate for other cases as well, and has the advantage of being simpler than that of Harrison and Anderson, so that there is less chance of error. The results obtained with it differ only little from those of Harrison and Anderson and also agree satisfactorily with other known experimental results.

FILTERS FOR CARRIER-WAVE TELEPHONY INSTALLATIONS

by TH. J. WEYERS.

621.318.7:621.395.44

In carrier-wave telephony, since the conversations of different subscribers, modulated on different carrier waves, must be transmitted over the same line, care must be taken that only the desired frequency bands reach each receiving channel. This means that the frequency bands of the other channels which would cause cross talk must be suppressed by filters with adequate attenuation. The requirements which must be made of the total attenuation of these filters may be deduced from the intensity distribution of the speech, the sensitivity of the microphone, telephone and human ear and the permissible degree of cross talk. This deduction is explained in this article, using as example the installation of a 17-channel system designed by Philips. A discussion is also given of the way in which the required total attenuation must be divided among the different filters at the transmitting and receiving end, while in conclusion the construction of the filters and the results obtained with them are described.

In an earlier number of this periodical [1]) an outline was given of the equipment of an installation for carrier-wave telephony. It was found that the various filters thereby perform an important function. We shall here indicate more presicely the requirements which these filters should fulfil, and study the way in which the filters are built up.

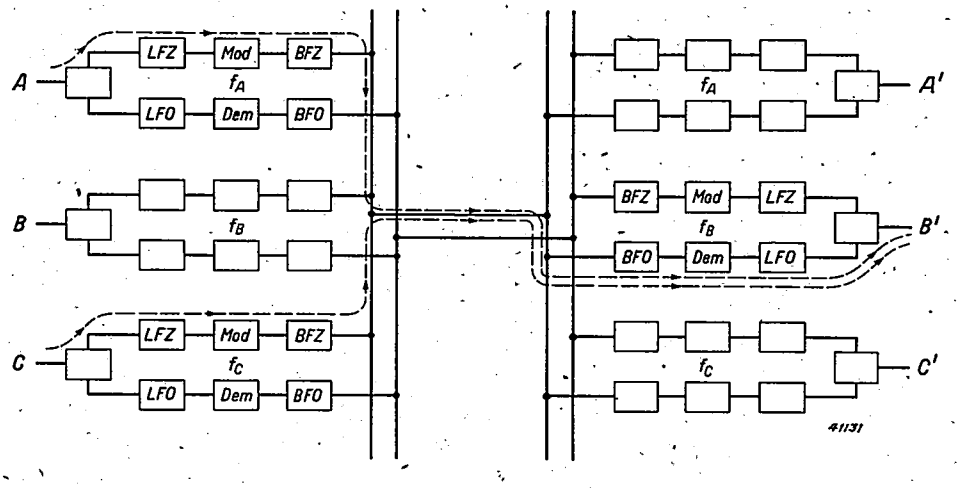


Fig. 1. Simplifed diagram of a carrier-wave telephone installation. Three channels are indicated. *LFZ* low-pass transmitting filters, *Mod* modulators, *BFZ* transmitting band-pass filters, *BFO* low-pass receiving filters. The cross talk between adjacent channels (*A* to *B'*, *C* to *B'*) is indicated by dotted lines.

Fig. 1 gives once more a simplified diagram of a carrier-wave installation for three speech channels in which *A* speaks with *A'*, *B* with *B'* and *C* with *C'*. In each speech channel, at the transmitting as well as at the receiving end, there is a low-pass filter and a band-pass filter. Let us now suppose that the two transmitting filters at *C* and the two receiving filters at *B'* are missing. In addition to the upper side band of *C*, which is cross-hatched in *fig. 2*, and which we actually wish to use for the transmission of the conversation, the lower side band is also transmitted by *C*. Futhermore the two side bands transmitted have a much greater width (indicated by the dotted line in fig. 2) than the band of about 300 to 3000 c/sec required for the transmission, since in speech itself frequencies up to above 10 000 c/sec occur. Finally

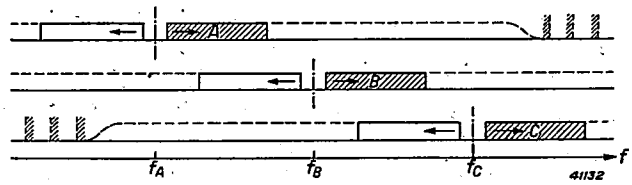


Fig. 2. Transmitted frequency spectra of the three speakers *A*, *B* and *C* in the absence of filters at the transmitting end. By the modulation of the speech spectrum on the respective carrier waves f_A, f_B and f_C, the side bands indicated are formed. Of these only the shaded portion in each case should be used for the transmission. The dotted lines with each spectrum represent the by-products of the modulation which actually, although with low intensity, also occupy whole frequency bands. The carrier waves themselves are not generally transmitted.

[1]) D. Goedhart and J. de Jong, Carrier-wave telephony, Philips techn. Rev. **6**, 325, 1941. See also: F. A. de Groot and P. J. den Haan, Modulators for carrier-wave telephony, Philips techn. Rev. **7**, 83 1942.

several kinds of weaker undesired frequency bands are also sent out by C, which are formed chiefly in the modulator and are indicated in fig. 2 by a few lines.

Due to the absence of the filters mentioned, subscriber B' receives not only the desired side band from B, but also the whole frequency spectrum sent out from C; cross talk occurs, and in this case it is unintelligible cross talk, since in the channel of B' the frequency spectrum of C is not demodulated with its original carrier wave f_C, but with the carrier wave f_B lying, for instance, 4000 c/sec lower. A speech frequency f, which by the modulation was converted into $f_C - f$ and $f_C + f$, after the "wrong" demodulation, now appears as frequency $f_C - f - f_B = 4000 - f$ or as $f_C + f - f_B = 4000 + f$. The lower side band of C is thus "inverted" for B', *i.e.* the low frequencies become high frequencies and *vice versa*. In the upper side band all the frequencies are increased by the amount $f_C - f_B$.

In the same way, when we suppose the transmitting filters of A to be absent, cross talk from A to B' occurs, but with the difference that now the upper side band of A is inverted and the lower band shifted upward. Furthermore, the more distant channels may also cause cross talk by means of the parts of the frequency spectra indicated in fig. 2 with dotted lines.

The function of the filters is now to prevent this cross talk.

It is quite clear that cross talk would be prevented if at the transmitting and receiving end of each channel band filters were introduced, each of which passed exclusively the side band belonging to that channel (shaded in fig. 2) and possessed an infinitely large attenuation on both sides of that frequency region. Such filters are impossible of realization in practice of course, and actually it is unnecessary to isolate the frequency bands so rigorously from each other, because the interference voltages always have only a finite amplitude and a certain small amount of cross talk does not prevent the carrying on of a satisfactorily intelligible conversation. For reasons of economy indeed it is desirable not to make the filters, which are made up of filter sections in series composed of coils and condensers, better than is absolutely necessary in connection with the permissible amount of cross talk. In order to design the filters for a given carrier-wave telephone installation, therefore, it is first necessary to study in detail the attenuation which each frequency must experience on the undesired path from C to B' or from A to B', etc. in order to keep the cross talk just within the permissible limits.

This must be done in different steps. For each separate frequency the following must be taken into account: the intensity of the speech to be expected at this frequency, the strength in which vibrations of this frequency are transmitted by the microphone, the sensitivity of the cross talk. By the combination of these data the desired attenuation curve is found.

In the following we shall explain more fully the steps mentioned for determining the attenuation requirements, choosing as basis the situation in the case of a 17-channel carrier-wave telephone system worked out by Philips. In this case the frequency difference of the successive carrier waves amounts to 4 kc-sec and the upper side band is transmitted, as was assumed above. In conclusion the method will then be discussed by which the attenuation curve found is given practical realisation in the filters.

The spectral distribution of the intensity of speech

It is impossible to reproduce the energy distribution of speech vibrations over all frequencies simply by a Fourier spectrum. The sounds of speech change continually and with them also the intensity of the different components in the spectrum. One may of course consider the average value of the intensity at each frequency, but this would lead to incorrect conclusions in the problem here presented. If for instance we gave the filters an attenuation curve such that at each frequency the average intensity of the interfering speech vibration was attenuated to below the threshold of hearing, then in the continual fluctuation of the intensities the peaks would still project above the threshold of hearing at times, and it is just these peaks which may cause an interference. The intensity distribution of the speech must therefore be characterized by the magnitude and the number of the peaks occurring at each frequency and the peaks may for instance be counted with a recording apparatus.

The term "each frequency" used above must still be defined more fully. If by means of a filter with an ideally sharp resonance curve a single frequency, *i.e.* an infinitesimally narrow frequency region, should be filtered out of the speech spectrum, no intensity of that frequency would ever be observed. In measuring therefore a small frequency region of a certain width must be investigated in each case, and in order to be able to record the correct size of the peak in the speech, the width must be greater than the spectral width which the peaks ordinarily possess, but still narrow enough so that in general two peaks cannot occur in it at the

same time to be recorded as a single peak of a greater height.

A similar problem is encountered in connection with the time which is available for the recording of each peak. This time must be long enough so that a peak is recorded as a whole, since otherwise instead of one peak two or more are recorded. On the other hand the time for recording must also not be so long that there is danger of counting two successive peaks as one.

Considering these elements of uncertainty which lurk in the characterization of speech it is not to be wondered at that the results of different investigators are somewhat divergent in these respects. Moreover, the results, which of course are averages in each case over a number of speakers, depend to a certain extent upon the language, and, related, to this, upon the articulation of the speakers. Nevertheless, it has been found possible with the data available in the literature [2]) to construct with sufficient accuracy a generally valid basis for our considerations, a basis which, according to experiments with the filters constructed, need not be subjected to modifications for different languages.

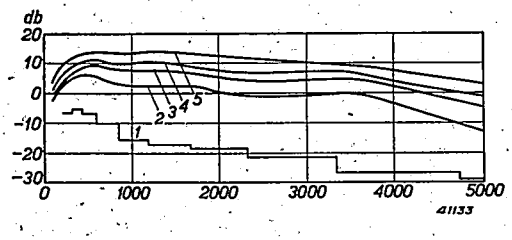


Fig. 3. Spectral distribution of the intensity of speech for an average speaker. The average over the time of the level of the whole sound is set equal to 0 db. The stepped curve *1* gives the average intensities which occur in the different frequency intervals (half an octave wide). Curves *2, 3, 4, 5* give the levels which are exceeded by the intensity peaks in speech once per second, once per 3 seconds, once per 6 seconds and never, respectively.

In *fig. 3* several curves are now given which were derived from the data mentioned, and which describe the intensity distribution in speech. As a basis for the measurement, in agreement with Fletcher and others, a frequency interval of a half an octave is assumed and a time interval of 1/8 second. For each frequency interval the average intensity level (in db) occurring therein is plotted and the level which is never exceeded, exceeded once in 6 seconds, once in 3 seconds and once per second, respectively. The average intensity level of the whole conversation is set equal to 0 db. This figure thus gives a picture of the height and the number of the intensity peaks occurring in each frequency interval for an average speaker.

The sensitivity of the microphone

The sensitivity of a microphone of the kind used in telephony is closely dependent on the frequency, as is shown in *fig. 4* for two different types of microphone. The form of the microphone currents occurring as functions of the frequency thus deviates from the form of the air vibrations which are incident upon the microphone when it is spoken into. The resulting intensity distribution can be constructed in a simple way by multiplying the intensities given in fig. 3 for each of the successive frequency intervals by the average sensitivity of the microphone over this interval. Now according to fig. 4, however, the frequency characteristics of different types of microphones vary quite considerably. In the construction of an installation for carrier-wave telephony the fact must be taken into account that the subscribers may use any given type of microphone and that in spite of this cross talk should remain below the permissible limit in every case. We have therefore carried out the construction described for curve 1 of fig. 3 for several types of microphone, among which are those whose characteristics differ most widely, and have then determined the uppermost "envelope" of the curves for the different microphones. This envelope, which is drawn as curve *1* in *fig. 5*, now gives, as a function of the frequency, the level above which the average over the time of the intensity of the microphone currents does not project, no matter what type of microphone is used. Here also the average level of the total microphone current is set equal to 0 db. With the help of the other curves of fig. 3 the curves *2, 3, 4* and *5* in fig. 5 are then also constructed, and these indicate respectively the levels which are exceeded once per second, once per 3 seconds,

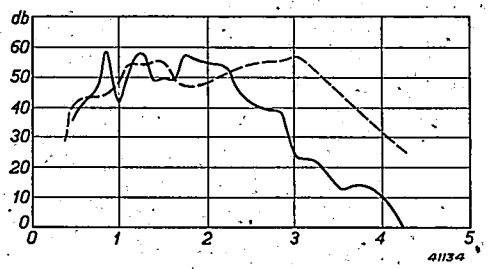


Fig. 4. Frequency characteristics of two different common types of microphones (the zero level is arbitrary).

[2]) See especially H. Fletcher, Physical characteristics of speech and music, Rev. mod. Phys. 3, 258, 1931. Further also publications by Crandall, Sivian, Fletcher and Munson and Colpitts in Bell System techn. J. 1925 to 1937, as well as F. Trendelenburg. Klänge und Geräusche, Springer, Berlin 1935.

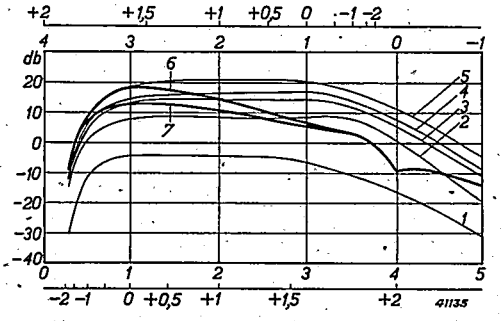


Fig. 5. Spectral distribution of the microphone currents averaged over different speakers. The average over time of the level of the whole conversation is set equal to 0 db. Curve *1* is the "envelope" of the curves found with different types of microphone for the average intensities which occur in the different frequency intervals. Curves *2*, *3*, *4*, *5* give the levels which are exceeded by the intensity peaks of the microphone currents once per second, once in 3 seconds, once in 6 seconds and never, respectively. The level which is exceeded once per 3 seconds serves as a basis for the permissible cross talk. Due to the inversion experienced by the speech spectrum upon demodulation with the "wrong" carrier wave (that of an unintended subscriber) a partial compression and partial expansion of the spectrum takes place which may be seen in the octave scale which is given for the original as well as the inverted spectrum (lower and upper scales respectively). As starting point the frequency 1000 c/sec is arbitrarily chosen. Due to this compression and expansion curve *3* must be replaced by curve *6*. The limiter present in every channel provides that the strongest peaks in the microphone currents are attenuated by about 6 db, so that curve *6* passes over into curve *7*.

once per 6 seconds and never exceeded [3]).

The permissible level of the interferences received

Which one of the curves in fig. 5 must now be used as a basis for the further discussion? We have already stated that it is unnecessary to give the filters so much damping that the conversation of A or C remains entirely below the limit of hearing of B'. Therefore we need not use the level curve which is never exceeded by the peaks (curve *5*). It has been found experimentally that the following of a telephone conversation is not appreciably disturbed when an interfering conversation causes an audible sound not more than once in three seconds. In the first instance therefore we begin with curve *3* in fig. 5. As this level is of course exceeded in each frequency interval of half an octave once in three seconds just as many interfering sounds would be heard every three seconds as there are half octaves in the whole frequency region. It will appear later that this is not actually the case, however, and that we may without hesitation work upon the assumption that in each half octave an audible sound may occur once in three seconds.

This condition holds for the receiving channel. Now the lower side-band of the frequency spectrum of C reaches the telephone of the unintended subscriber B' in an inverted form (see above). Thus for example the region from 300 to 1400 c/sec of the original spectrum, *i.e.* more than four half octaves, is transferred into the region from 2600 to 3700 c/sec, *i.e.* about one half octave. In order to hear an interference in the receiving channel within this single octave only once in three seconds, an audible interference must occur within the region from 300 to 1400 c/sec of the original spectrum only about once in twelve seconds per half octave, *i.e.* practically never. For corresponding reasons in the half octave from 2600-2700 c/sec of the original spectrum an audible interference may occur more than once in three seconds. In this way curve *6* of fig. 5 is found as the level of the speech currents which is exceeded as often as corresponds to the permissible interference in the receiving channel being interfered with.

What is plotted in fig. 5 is actually the difference between the peak intensities occurring and the average speech intensity. Now in every carrier-wave telephone installation so-called amplitude limiters are employed, which are placed in every channel between the microphone and the modulator and whose function is to limit the largest current maxima occurring. This precaution, which is taken mainly to combat overloading of the line repeaters, also has the result that with a given degree of cross talk larger average microphone currents can be used. It has been found experimentally that the average currents may be twice as large. For our curve this is the same as if the amplitudes of the highest peaks occurring (which cause the cross talk) were a factor 2 lower, *i.e.* as if they reached a level 6 db lower. According to curve *6* of fig. 5 the highest peaks which thus fall victims to the action of the limiter occur at about 1000 c/sec. Therefore we may allow the curve to fall about 6 db in this region and thereby finally obtain curve *7*, which now indicates the level to be considered or the permissible interference. This curve is drawn separately in *fig. 6*.

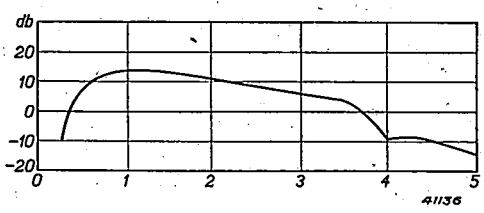


Fig. 6 Spectrum of the interferences which must be used as a basis for the damping requirements.

[3]) The C.C.I.F. (Comité consultatif international de communications téléphoniques à grande distance) propose a network which, as far as its frequency characteristic is concerned, may be considered as the equivalent of the microphone. We have not used this, in order to be able to take into account also the properties of better microphones which were not yet in use when the C.C.I.F. characteristic was determined.

The threshold of the disturbing effect caused by interference

The filters between C (or A) and B' serve to provide that an interference with the spectrum of fig. 6 falls just below the threshold of hearing of B', or more accurately below the threshold of the disturbing effect. In determining this threshold the sensitivity of the ear of B', as well as the sensitivity of the telephone which converts the electrical vibrations into audible air vibrations plays its part. Both sensitivities depend very closely upon the frequency, as is shown in *fig. 7*. In this figure the sensitivity of the ear (measured at the level of interference permissible for telephony) and the sensitivityof two different types of telephone are plotted, all three in db at an arbitrary zero level. In addition to these frequency characteristics of ear and telephone, various kinds of physiological effects also play a part in the disturbing effect of a

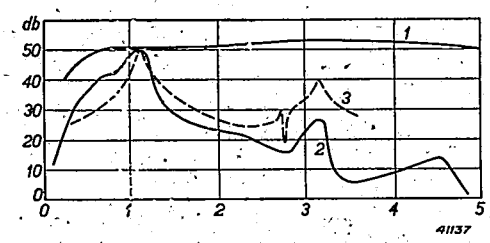


Fig. 7. Spectral sensitivity of the human ear (*1*) and frequency characteristics of two different types of telephone (*2* and *3*). The zero level is arbitrary.

sound, such as occur upon the simultaneous presence of a desired and an undesired sound. In order also to take these effects into account there is no other possible way but to determine the threshold of the disturbing effect directly by experiment. This is done in the following way. The interference voltage is measured with a voltmeter, and it has been found that the deflection observed may be considered as an objective measure of the disturbing effect independent of the n a t u r e of the interference voltage, provided the voltmeter is given a very definite frequency characteristic by means of a filter connected in series with it. Such a measuring instrument is called a p s o p h o m e t e r, and the psophometer characteristic mentioned, in which the ear sensitivity and the sensitivity of the telephone are already taken into account, furnishes the desired threshold curve of the disturbing effect when the empirical fact is also taken into account that an interfering sound in the telephone is only just not experienced as disturbing when the psophometer indication lies 62 db below the average level of the telephone conservation.

The threshold curve of the disturbing effect so obtained must be recorded with different common types of telephones in order to make it possible, in the same way as described above for the microphone, to draw the "envelope" of

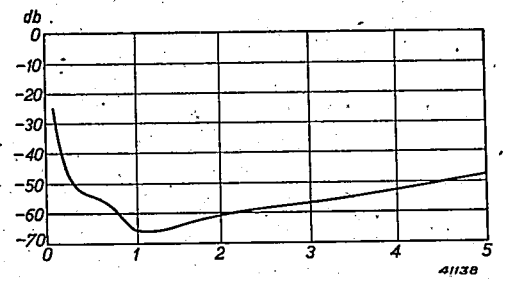


Fig. 8. Threshold curve of the disturbing effect obtained from the psophometer characteristic. At each frequency the interference must be attenuated to below the level indicated here.

the curves obtained for different telephones. Taking into account the possibility that the telephones now in common use may be replaced by newer and better types [4]), we have arrived at the threshold curve of the disturbing effect reproduced in *fig. 8*.

The required attenuation of the filters

In order to derive the required attenuation curve of the filters from the interference curve fig. of 6 and the threshold curve of fig. 8, we must still ascertain what takes place in modulation and demodulation. The interference with the spectrum according to fig. 6 is for instance modulated on the carrier wave f_A. The side bands on both sides of f_A indicated in *fig. 9* by the two branches of curve *1* are thereby formed. If we assume further that B' receives a sound which has at all frequencies exactly the intensity of the threshold curve according to fig. 8, this sound can occur by demodulation not only from the right-hand but also from the left-hand side band of the carrier wave f_B. In the frequency regions of the demodulated vibrations, therefore, the threshold curve *2* drawn in fig. 9, consisting of two branches to the right and left of f_B, must be taken into account. The difference between curves *1* and *2* gives the curve of the required total attenuation of the filters between A and B' (curve *3*).

In this curve is included not only the contribution of the band-pass filters but also that of the low-pass filters. The actual attenuation curves of the latter lie, of course, in the audio-frequency region and must therefore, like the curves of fig. 6 and fig. 8, be considered to be transferred by modulation on the carrier-waves

[4]) The C.C.I.F. also gives a psophometer characteristic in which, however, the use of modern telephones is not yet taken into account.

f_A and f_B, respectively, to the frequency region of the modulated vibrations considered in fig. 9, in order to be able to add them to the attenuation of the band-pass filters.

If we do not consider the cross talk from A to B' (to the higher channel), but from C to B' (thus to the lower channel), it is easy to see that the required attenuation is represented by a curve symmetrical to curve *3* of fig. *9* with respect to f_B.

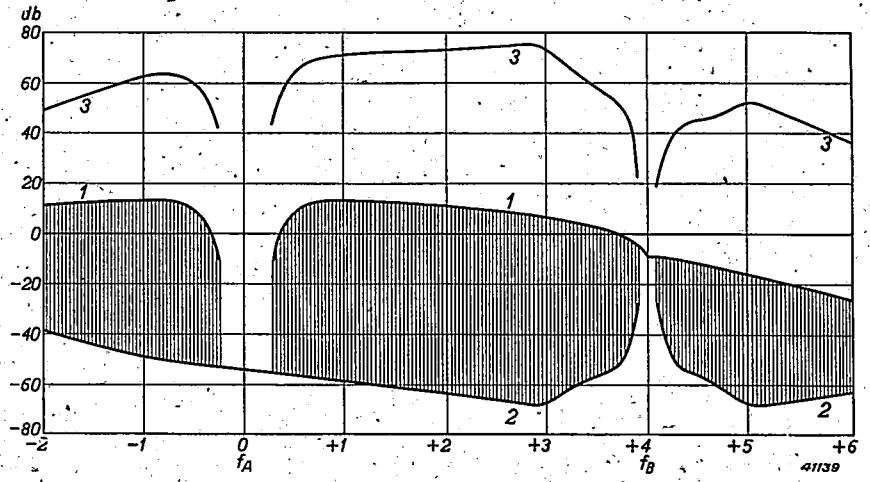


Fig. 9 In the frequency region of the modulated vibrations the curve of fig. 6, by modulation on the carrier wave f_A, gives the interference curve *(1)*, the curve of fig. 8 by modulation on the carrier wave f_B gives the threshold curve *(2)*. The shaded difference between these two curves is the attenuation required at every frequency between A and B'. This attenuation is also plotted (curve *3)*.

Division of the damping among the different filters

In the beginning it was pointed out that there are four filters along the path between A and B': low-pass filter and bandpass filter at the transmitting end and the same at the receiving end. As soon as the total necessary attenuation is determined the question arises as to how this attenuation must be divided among the four filters.

In order to obtain some insight into this question we first ask whether four filters are actually needed. The fact that at least two are needed is immediately clear when it is kept in mind that not only is a certain attenuation required between A and B' but also a certain transmission between A and A' as well as between B and B'. A filter at A can furnish no contribution to the attenuation in the frequency region of the upper side band of f_A, since this side band is just the one used for the transmission from A to A' (*fig. 10*). In this frequency region therefore the attenuation required according to fig. 9 can only be furnished by a filter at B'. For the same reason, however, a filter at B' can furnish no contribution to the attenuation in the upper side band of f_B, so that a filter is necessary at A for this purpose.

If we now try to get along with two filters it becomes clear in a similar way that these two filters cannot be low-pass filters. With the above-described transference of the attenuation curves lying in the region of audio-frequencies to the frequency region of the modulated vibrations every low-pass filter furnishes two symmetrical branches, and according to fig. 10, since no low-pass filter may have any attenuation up to about 3000 c/sec, no attenuation can be obtained in the largest part of the region between the neighbouring carrier waves f_A and f_B.

At least two band-pass filters are therefore needed. The fact that actually two low-pass filters are made to cooperate in the total attenuation is due mainly to economic considerations. The same fact which makes it impossible to get along with low-pass filters only, namely the occurrence of two symmetrical branches of the attenuation curve in the frequency region of the modulated vibrations, also makes it possible to gain double advantage from each attenuation contribution of the low-pass filters,

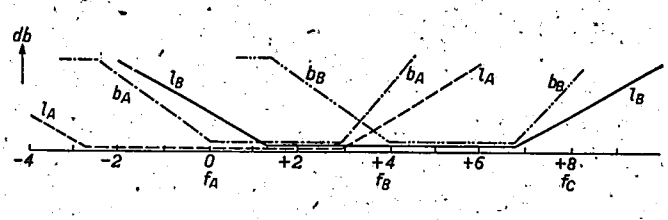


Fig. 10. Diagrammatic representation of the contributions which the low-pass filters (l_A and l_B) and the band-pass filters (b_A and b_B) can make to the total attenuation on the traject between A and B'.

to the right and to the left of the corresponding carrier wave. Moreover, in the case of low-pass filters it is much easier than in the case of band-pass filters to make the attenuation curves sharp, *i.e.* with only slight rounding off in the neighbourhood of the limits of the regions transmitted, a very desirable property for securing the uniform transmission of all the frequencies necessary for the conversation. These facts make it desirable from the economical point of view to have as much as possible of the damping provided by low-pass filters, and to use the band-pass filters mainly for the frequency region mentioned lying between the carrier waves, where the low-pass filters cannot offer any help.

There are, however, other considerations, only secondarily connected with cross talk, which limit the applicability of this conclusion. In the first place in the desired transmission from A to A' (or from B to B') the whole of the lower side band which is not used must be adequately suppressed. If parts of this side band still come through with appreciable intensity then they will amplify the demodulated desired side bands for some frequencies and for others weaken it, according to the phase rotation which the vibrations with the freqencies in question experience in the filters. The fidelity would suffer thereby. For the suppression of the undesired side bands, in which as a matter of fact the low-pass filters cannot help, a damping of at least 35 db is necessary at every frequency in this region, which damping can be distributed equally over the two collaborating band-pass filters at the transmitting and receiving ends.

Another consideration which makes for entrusting a large share of the total attenuation to the band-pass filters is the fact that the energy of the incoming signal at the receiving end is divided among all the channels in ratios which are determined by the attenuation of the different band filters. It is of course desirable to allow as little as possible of the available signal energy to be taken from the desired channel by the undesired channels, and this is ensured by giving the band-pass filters a large attenuation in the attenuated bands.

Practical construction of the filters

After an idea has been formed on the basis of these and other considerations about the way in which the required attenuation must be divided among the filters, it must be ascertained how each filter must be built up in order to obtain the attenuation curve in question. No other practical method for doing this can be suggested than that of t r i a l based upon a knowledge of the properties of a number of simple types of filter sections and the changes which these properties undergo upon variation of the elements of the sections (self-inductances, capacities, resistances). In this process of trial the projected division of the attenuation among the filters will be somewhat modified by several practical considerations: the smallest possible number of filter elements is desired, the number of sections in each filter must of course be the whole number [5]), and it is an advantage in manufacture to reduce to a minimum the number of types of filters in the installation.

In this way the filter connections given in *figs. 11* and *12* were derived for the above-mentioned 17-channel system of Philips. The low-pass filters are identical for all channels but

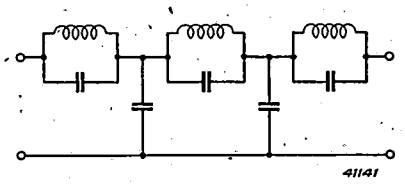


Fig. 11. Structure of the low-pass filters in the 17-channel system. At the transmitting and receiving ends filters of the same configuration are used but with slightly differing values of the elements and thus also with different attenuation characteristics.

differ at the transmitting and receiving ends, which is to be understood considering the asymmetrical shape of the required attenuation curve according to fig. 9. The band-pass filters at the transmitting and receiving ends are made similar in the main, and as far as possible they are always built up of the same types of sections, but of course with different values of the capacities and self-inductances for each channel (each carrier-wave frequency).

We shall here give a few particulars of the construction of the filters. The band-pass filters for the various channels are connected in parallel at the transmitting end as well as at the receiving end. The end branches of these filters must therefore be so composed that the various filters connected in parallel do not affect each other unfavourably. For this purpose the correction element I indicated in fig. 12. is added. Furthermore, the bandwidth $f_{max} - f_{min}$ is the same for all band-pass filters; the *relative* band width

$$(f_{max} - f_{min}) / \sqrt{f_{max} \cdot f_{min}}$$

is therefore smallest for the channels with the highest carrier-wave frequencies. As a result the rounding off of the attenuation curves near the limiting frequencies is in general the most pronounced for these channels, and thus the transmission in the pass-band is less uniform. In order to obtain a uniform transmission of all speech frequencies for the higher channels also, a correction element II is now connected in parallel at the input terminals of the filters at the transmitting end of these channels. This element causes a certain extra attenuation in the middle of the pass-band and gives practically no attenuation at the edges. Due to this device the total attenuation of the channel in question is of course several db greater, but this can easily be compensated in the corresponding channel repeater.

The attenuation curves realized with the four filters for one of the middle channels are shown in *figs. 13a* and *b*, while in *fig. 14a* the total attenuation curve between A and B obtained from them by addition is drawn, together with the curve derived above for the minimum required attenuation (curve *3* of fig. 9). In the same way in fig. 14*b* the total attenuation ob-

[5]) H a l f c e l l s are actually also possible, but as a rule their use is not very economical.

tained between C and B' is drawn. The fact that this is not exactly the same as that between A and B' follows simply from the fact that the low-pass filters are not identical at the transmitting and receiving ends. Upon transference of their attenuation curves to the frequency region of the modulated vibrations, in the first case (A to B') the contribution of the transmission low pass filter to the total attenuation lies at the lower limit of the receiving channel, that of the receiving low pass filter at the upper limit. In the second case (B' to C') it is just the reverse. Moreover, the curves for the required minimum attenuation, as already stated, are different for the two cases, *i.e.* they are symmetrically located with respect to f_B.

In both cases, *i.e.* in fig. 14*a* as well as in *b*, it is immediately evident that the attenuation obtained is considerably greater in practically the whole frequency region than the required minimum attenuation. There is only one small frequency region where the surplus is small. The obvious question is whether this does not mean a waste. The answer must be that on the one hand it is impossible with simpler filters to satisfy the requirement that the attenuation obtained should never fall b e l o w the desired attenuation curve at any frequency, while on the other hand an advantage is obtained with the extra attenuation. It may be recalled that in fixing the attenuation requirements we began with the intensity level which was exceeded in speech once every 3 seconds i n e a c h h a l f o c t a v e. It is now evident that with the attenuation curve obtained this actually leads to the result that, as was considered permissible, a t o t a l of one interference is heard no more than once in three seconds. Only in a frequency region about one half an octave wide do the desired and the actual attenuation curves approach each other so closely that at this point

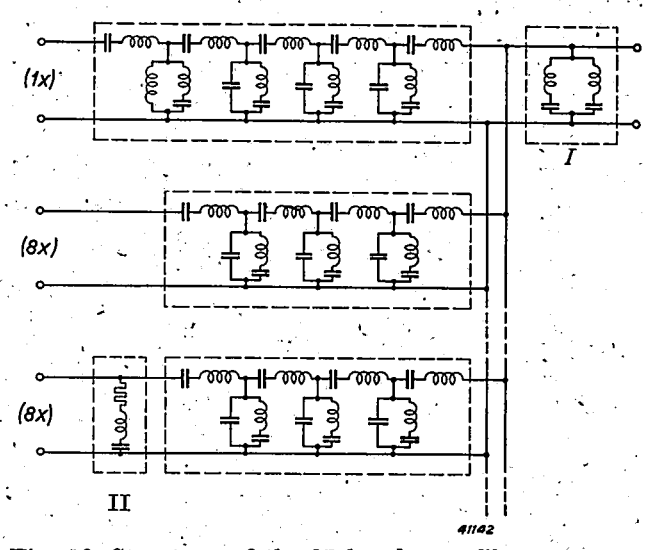


Fig. 12. Structure of the 17 band-pass filters at one end connected in parallel. The first band-pass filter (for the channel with 4 kc-sec as carrier-wave frequency) has a different structure, the other 16 (with 8 to 68 kc/sec as carrier-wave frequencies) are composed of the same types of filter sections. The topmost 8 transmitting filters also contain a correction element *II*; this element is unnecessary in the case of the receiving filters. In parallel with the common end of the band-pass filters a correction element *I* is introduced for the correct termination without mutual effects of the filters.

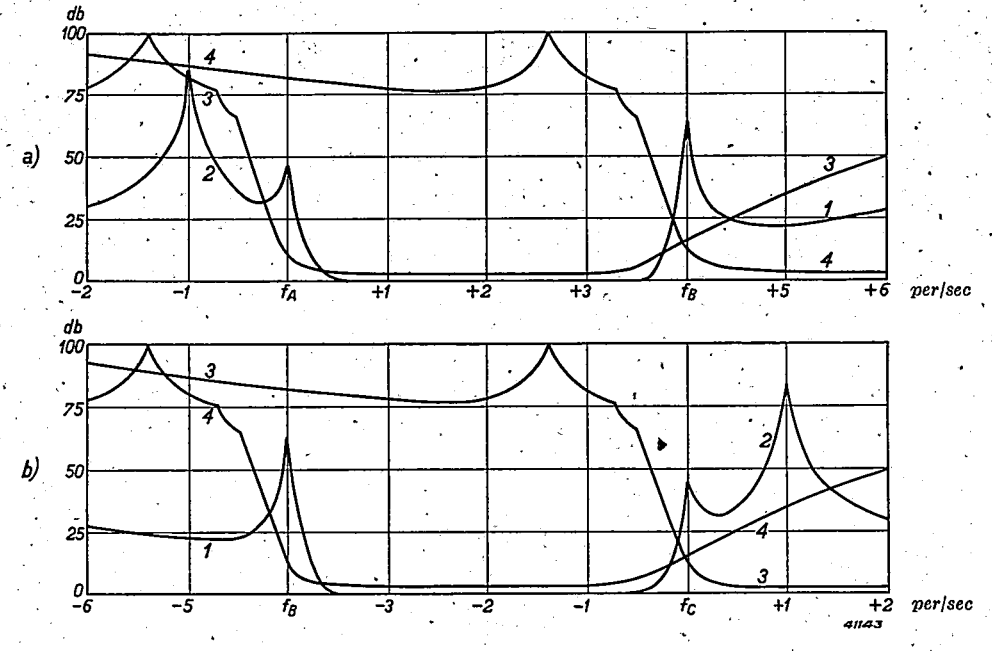


Fig. 13. The attenuation of the filters between two neighbouring channels obtained with the finished installation plotted in the frequency region of the modulated vibrations. *1* and *2* contributions of the low-pass filters, *3* and *4* of the band-pass filters at transmitting and receiving ends, respectively.
a) Attenuation on the traject from A to B ($f_{B'} < f_A$).
b) Attenuation on the traject from C to B ($f_{B'} > f_C$).

the permissible interference reaches the unintended subscriber almost in full. Where this half octave would lie in the whole frequency region could not be foretold in advance, but we now see that indeed only a single half octave, instead of every half octave, of the total frequency region had to be taken into account.

Moreover, we based our derivation on the cross talk between neighbouring channels, while we stated above also that more distant channels, due to all kinds of by-products occurring upon modulation [6]), may cause cross talk. For the suppression of just these interference contributions the extra attenuation of the filters is very convenient. The fact that it is actually able to suppress this interference adequately was found to be true in measurements of cross talk with the psophometer in the case of the finished installation.

These measurements were carried out for different speakers in different languages. The level of interference found agreed extraordinarily well with what was assumed as permissible level of interference in our derivation of the damping requirements, while the transmission in each channel was found to be sufficiently uniform for all frequencies between 300 and 3400 c/sec. This shows not only that the method of derivation indicated is correct in spite of several not absolutely sure assumptions which were made, but also that in this way the installation has been successfully made to comply with the requirements: it is good enough, but not unnecessarily good, *i.e.* not more complicated and expensive than necessary.

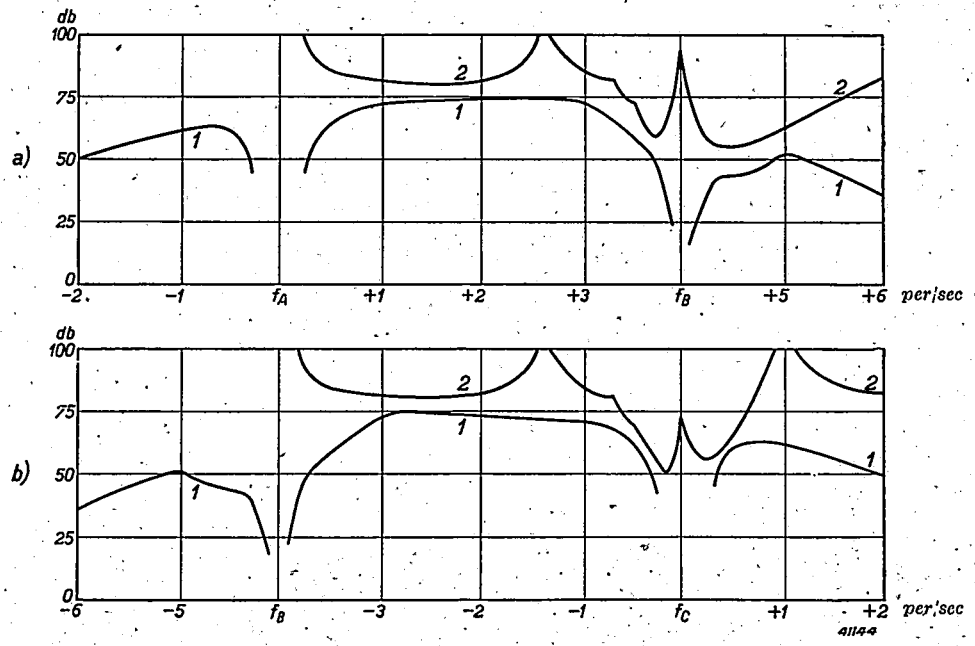


Fig. 14. Comparison of the required attenuation characteristic (*1*) and that actually obtained (*2*) between two neighbouring channels: *a*) from *A* to *B*', *b*) from *C* to *B*'.

[6]) See the second of the articles cited in note [1]).

THE LOCALIZATION OF CABLE FLAWS

by W. HONDIUS BOLDINGH. 621.317.733:621.317.333.4:621 315.2

A description is given of a measuring bridge for 50 kV protected against high voltage, which, in combination with a high voltage generator for DC voltage, is used for the localization of cable flaws. The conditions are examined which must be satisfied by the connections and especially by the galvanometer in order to combine great strength with great accuracy.

Introduction

The way in which power cables can be tested with high DC voltage has already been described in this periodical [1]). In testing a cable by this method when an abnormally low insulation resistance is found the problem is to find the position of the defect in the insulation as accurately as possible. The usual method of doing this is to carbonize the insulation at the point of the flaw by means of the current provided by the DC voltage generator, so that the defective spot can be localized with a low-voltage measuring bridge.

The latter is usually done by the loop method of Murray: the defective conductor is connected with an intact conductor at one end of the cable, a measuring bridge is connected with the other end and the source of voltage connected between measuring bridge and earth. In this way the ratio can be determined of the resistance x of the part of the defective conductor on one side of the flaw to that of the other part increased by the resistance r of the intact conductor (see *fig. 1*). This method has the advantage that the resistance of the flaw itself does not occur in the localization calculations.

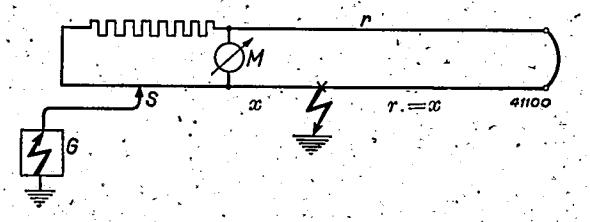


Fig. 1. Localization of a cable flaw by the loop method of Murray. The resistance of an undamaged cable conductor is r, while the resistances of the two sections of cable at either side of the flaw are x and r-x. G is the high-voltage generator and M the galvanometer. In the left-hand part of the bridge, by means of a slide wire, the ratio of the resistances of the two branches is made equal to that in the right-hand part. On the slide wire S the distance of the flaw is read off as a percentage of the length of the cable.

It is clear that the accuracy with which the position of the flaw can be determined depends upon the current which can be sent through the insulation flaw. There will be a certain minimum current at which satisfactory measurement is just possible. It therefore depends upon the resistance of the flaw how much *voltage* is needed for the measurement.

Now it is often found to be very difficult, or to require much time, to burn the insulation through to such a degree that reliable measurement with a low-voltage measuring bridge is possible. The case may even occur in which, indeed an abnormally low insulation resistance is measured, but upon long continued testing at the highest permissible or available voltage the current remains low. A sensitive measuring bridge is then needed which must be able to remain under high voltage during the measurement.

For the localization of flaws under these circumstances measuring bridges have been improvised and even offered for sale in a more technical form, which, in combination with a cable testing apparatus, make possible measurements under high voltage. These measuring bridges consist chiefly of an adjustable resistance (decade resistance or slide wire) and a galvanometer, which are insulated together for the full voltage and placed for instance on a tripod of insulation material, and which can be adjusted and read while under tension.

The remarks in our previous article [1]) about the danger of high voltage in the case of cable testing generators is even truer for the measuring bridges: in this case as a rule there is not the least protection against high voltage for the operator. It was therefore important, in addition to the cable testing generator with protection against high voltage, to design a measuring bridge offering the same degree of safety.

We shall now describe such an apparatus which was constructed for a voltage of 50 kV.

Galvanometer

The centre of the cable measuring bridge is of course the galvanometer. In the first place the accuracy of the localization depends upon the sensitivity of this instrument, and in the second place the form of the apparatus is determined by the manner of reading the galvanometer under high tension.

The advantages of a reflecting galvanometer are immediately obvious: it is very sensitive because the moving indicator is a beam of light and a beam of light is an ideally

[1]) Philips techn. Rev. **7**, 59, 1942.

insulated, long and weightless "pointer" between the galvanometer under high tension and the earthed reading scale. Disadvantages of the reflecting galvanometer are its fairly large dimensions and its low resistance to mechanical and electrical shocks. Although in recent years so-called light-spot galvanometers which are transportable have appeared on the market, which constitute a considerable improvement in this line, they cannot immediately be used for the localization of cable flaws.

A galvanometer has now been successfully constructed which, because of its small dimensions, great sensitivity and high mechanical and electrical resistance, is particularly suitable for the object in view. Advantage was taken of the use of a modern magnet steel of high magnetic power [2]), with which a strong field in a relatively wide air gap can be obtained with small weight and dimensions of the permanent magnet. This makes it possible to give the galvanometer coil a large number of windings with a given resistance. In this way, in spite of the use of a strong suspending wire, the power necessary for a deflection of one scale division could be reduced to several tenths of a picowatt (10^{-12} W).

Fig. 2 is a photograph of the galvanometer. The rotating coil together with the mirror is suspended on a stretched wire; the wire has the cross section of a ribbon. To a large extent it is the properties of this ribbon which determine the quality of the galvanometer: the ribbon must submit easily to torsion in order to produce a high sensitivity, but at the same time a relatively thick ribbon is desirable for the sake of mechanical strength and electrical resistance to current impulses. As for the latter, the ribbon form is especially favourable, and we shall discuss this point later.

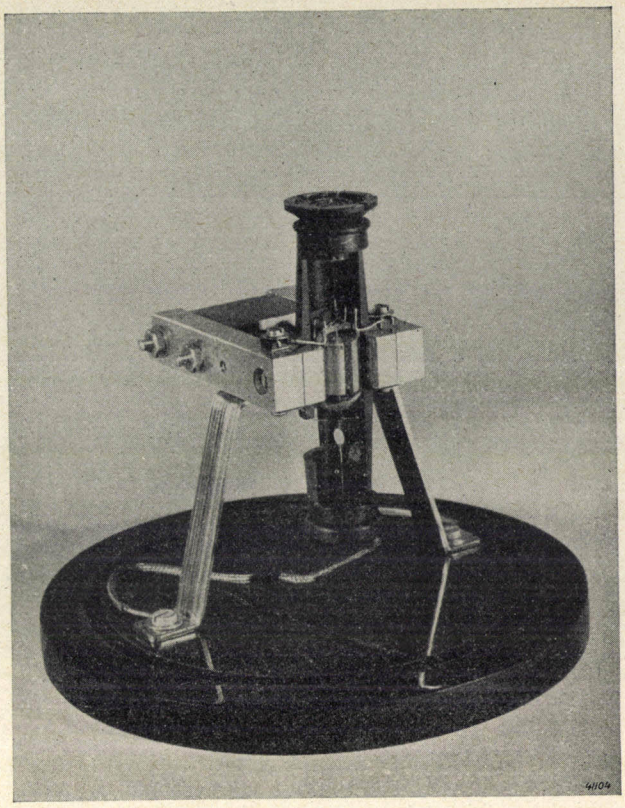


Fig. 2. View of the galvanometer, whose total height is 8 cm.

The accuracy of the localization

Considering the fact that the opening of a cable sleeve or the digging up of a section of cable costs considerable time and money, and that a cable must only be out of use for as short a time as possible, the rapid and accurate localization of a flaw is of the greatest importance. Various sources of error outside the measuring bridge may affect the accuracy of measurement. Thus, for example, the tolerance of the copper cross section of the cable sets a limit to the accuracy with which the position can be deduced from a measurement of the resistance of a cable conductor. Thermo-forces in the measuring circuit, stray currents in the ground or corona losses in the switching apparatus at the end of the cable may disturb the measurements. Finally the normal insulation resistance of the cable may cause an error in the localization.

In this article we shall leave such sources of error out of consideration and confine ourselves to the theoretical accuracy of the localization. Assume that the measuring bridge can be adjusted continuously (with a slide wire) and that the scale can be read with an accuracy of one millimeter. The question is then what error, expressed in meters of cable, will be made in localization if the galvanometer is not set exactly at zero but has a deflection of 1 mm. If we express this length λ, which we shall call the inaccuracy of localization, in quantities pertaining to the measuring circuit, we can study how the influence of this error in reading can be limited to a minimum, and thus how the localization can be made as accurate as possible.

Let us call

the direct current through the insulation flaw i mA,
the resistance of the slide wire R_s ohm,
the resistance of the galvanometer R_g ohm,
the voltage constant of the galvanometer u_g μV/mm,
the resistance per meter cable length r_1 ohmm,
the cable resistance (one lead and return) R_k ohm.

[2]) Philips techn. Rev. **5**, 29, 1940 and **6**, 8, 1941.

By application of Kirchhoff's laws we then find that

$$\lambda = \frac{0.001\ u_g}{i r_1}\left(1 + \frac{R_k}{R_g} + \frac{R_k}{R_s}\right) \quad (1)$$

From this it appears in the first place that λ is independent of the position x of the flaw. In the second place it is clear, as it was also without the formula, that the inaccuracy of localization is inversely proportional to the current i. In the third place it follows from formula (1) that R_s must be large compared with R_k. This last condition can easily be satisfied, as will appear later.

When the galvanometer resistance R_g is also large compared with the cable resistance R_k, which is usually the case with the instruments used for this purpose, the formula for the inaccuracy of localization *for 1 mA*, λi in the case of cables with a copper cross section of Q mm² becomes:

$$\lambda i = \frac{0.001\ u_g}{r_1} = 0.057\ Q\ u_g \quad (2)$$

It is found that in this case ($R_g \gg R_k$) the product λi depends exclusively upon the copper cross section Q, so that the accuracy of the localization is independent of the length of cable.

The conclusion must not, however, be drawn that the greatest accuracy is obtained with $R_g \gg R_k$: I_g and u_g are not independent variables and must therefore be considered in their mutual relation.

For this purpose we disregard in formula (1) the quantity $R_k/R_s \ll 1$ and write the formula as follows:

$$\lambda = \frac{0.001}{i r_1}(u_g + i_g R_k), \quad (3)$$

where $i_g = R_g/u_g$ is the current constant of the galvanometer in μA per scale division. Since $u_g \cdot i_g = W_g$, the so-called power constant, is a more or less constant quantity of the galvanometer independent of the resistance R_g, λ is a minimum for

$$u_g = i_g R_k, \quad \text{thus} \quad R_g = R_k.$$

The most favourable conditions for the measuring circuit are therefore when

galvanometer resistance = cable resistance ≪ comparison resistance.

When in this way the galvanometer has its optimum adaptation to the cable resistance ($R_g = R_k$), and taking into account that $R_g = R_k = 2000\ L r_1$, where L is the cable length in kilometers, one finds that

$$\lambda_{\min} \cdot i = 0.68\ \sqrt{W_g Q L}.$$

The minimum inaccuracy of localization is thus proportional to the square root of the volume of copper QL.

It must, however, be taken into account that the cable resistance varies between wide limits, so that R_g is not always equal to R_k. In practice

Q will vary from 25 to 150 mm², and
L ,, ,, ,, 0.2 to 10 km.

Of course the cables, as a whole, are always longer than 0.2 km and sometimes longer than 10 km, but for the localization of a flaw the shortest possible length will always be chosen for measurement. When in addition we take into account the fact that thin cables as a rule are not very long, we can calculate that the resistance of the length measured, out and back, will vary between about 0.05 and 5 ohms.

The most unfavourable combination is that with the largest copper volume: $L = 10$ km, $Q = 150$ mm². In this case $R_k = 2.3$ ohms. If we choose approximately this value of the resistance for the galvanometer we find for the most unfavourable practical case:

$$\lambda i = 0.68\ \sqrt{10 \cdot 150\ W_g} = 26\ \sqrt{W_g}.$$

A galvanometer with $W_g = 0.5$ pW/scale division², $R_g = 2$ ohms, *i.e.* $i_g = 0.5$ μA and $u_g = 1.0$ μV per scale division, with 1 mA measuring current will give a theoretical inaccuracy of 18 meters for a cable of 150 mm² and 10 km length. To calculate this for other cases we use formula (3), which is reduced to:

$$\lambda i = 0.057\ Q u_g + 2\ L i_g \quad (4)$$

For a shorter length or a smaller cross section the inaccuracy of localization for 1 mA may fall to 9 m, for a shorter length and in addition a thinner cable, for instance 200 m, 25 mm², it may even fall to 2.2 m (see *fig. 3*).

If the galvanometer resistance were 200 ohms instead of 2 ohms with the same power constant $W_g = 0.5$ pW, u_g would equal 10 μV and the inaccuracy of localization would be 5 to 10 times as large.

In *fig. 4* with different system resistances from 2 to 200 ohms λi is given as a function of the cable length for a meter of 0.5 pW and a cable of 150 mm². From this figure the advantage can clearly be seen of the optimum adaptation of the galvanometer resistance.

With a galvanometer of the type described above, with a well adapted resistance, therefore, with no more than 1 mA through the cable flaw a very satisfactory theoretical accuracy of localization can be obtained.

Mechanical and electrical resistance

Notwithstanding its high sensitivity the galvanometer is very resistant to the mechanical and electrical shocks which may occur in cable

measurement. The complete measuring bridge in all positions with short-circuited but unlocked galvanometer can be carried on a car with massive rubber tyres over a very uneven pavement

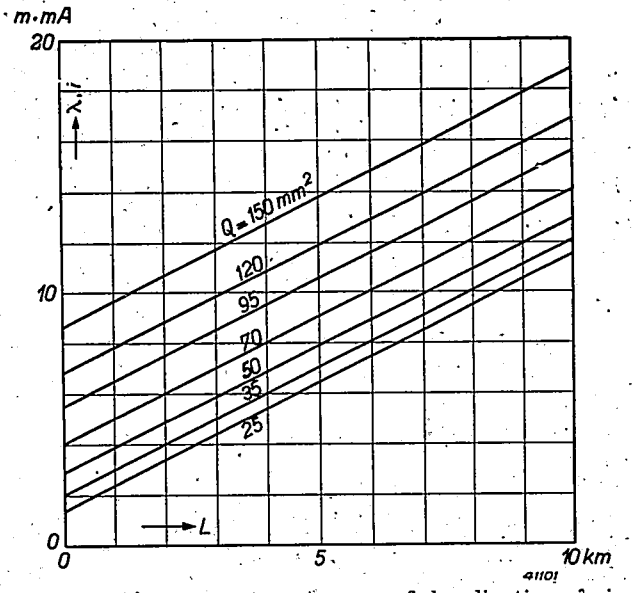


Fig. 3. Theoretical inaccuracy of localization λ in metres times the current i in mA as a function of the cable length L in km with different areas of copper cross section Q in mm². The resistance R_g of the galvanometer is 2 ohms, the current constant is i_g = 0.5 μA per scale division and the voltage constant u_g = 1.0 μV per scale division.

without the zero position undergoing appreciable change. An absolutely horizontal position during the measurement is not required.

The electrical dangers to which the galvanometer of a cable measuring bridge is exposed are

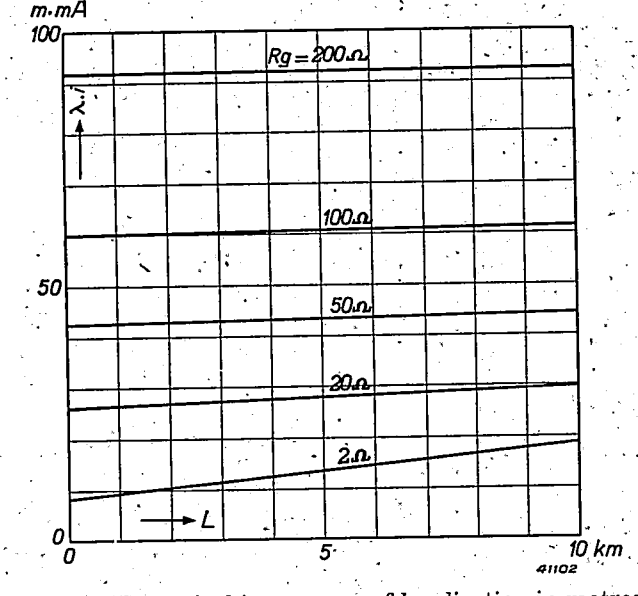


Fig. 4. Theoretical inaccuracy of localization in metres at 1 mA (λi) for a cable of 150 mm² cross section as a function of the cable length L in km and for a galvanometer with a power constant W_g = 0.5 pW per scale division². The resistance R_g of the galvanometer is taken at 2, 20, 50, 100 and 200 ohms, respectively.

many. Under various circumstances the galvanometer may receive current impulses which are many thousand times as great as those which cause a normal large deflection. It should be required of a measuring bridge for the localization of cable flaws that it be proof against all the catastrophes which may occur during a measurement and which cannot be avoided by reasonably good operation. On the basis of the connections in *fig. 5* we shall now study the different incidents which may occur.

The most serious case is that in which during a measurement on a long cable at full voltage the flaw suddenly breaks down. In the first place a voltage surge may thereby occur on the terminals G_1 and G_2 of the two conductors between which the galvanometer is connected, and in the second place the DC voltage generator will suddenly begin to give a high charging current, which *via* S flows through the measuring bridge and, before it is switched off by an automat, may be several hundred times as large as the original measuring current. If the bridge is already in equilibrium this will not be so serious, but the breakdown may occur before the bridge is balanced.

Another case which must be considered is a discharge of the cable due to earthing of the high-tension pole of the generator (point S), either expressly, in order to earth the cable, or unexpectedly upon the breakdown of a safety spark gap. As a rule this takes place over an earthing resistance, in the previously described cable testing apparatus over a built-in fixed resistance with magnetic switch or spark gap. Upon such a discharge the discharging current, which may amount for a short time to several amperes, flows through the measuring bridge. Considering the fact that the total comparison resistance amounts to 4000 ohms, a high voltage would occur between S_1, S_2, and S. This is limited by two rare-gas cartridges across S-S_1 and S-S_2. Further protection is, however, impossible without unfavourable effect on the accuracy of measurement; the galvanometer must therefore be able to tolerate the remaining difference between the voltages on the two branches of the comparison resistance in all points of balance of the bridge, and even when the bridge is completely unbalanced and preferably then also at maximum sensitivity.

It has been found that even with a cable capacity of 12 μF charged to 50 kV, *i.e.* an energy of 15 kW sec, the above-mentioned requirements can be satisfied even with a discharge over only a few thousand ohms. Only in the case of a complete breakdown of the cable in the neighbourhood of the measuring bridge is the galvanometer inevitably ruined when it is

connected in its most sensitive state. The chance of such a combination of unfavourable circumstances, however, is very slight.

In other cases, however, the amount of energy which flows through the galvanometer circuit may still be of such a nature that the meter must have an unusually high resistance against

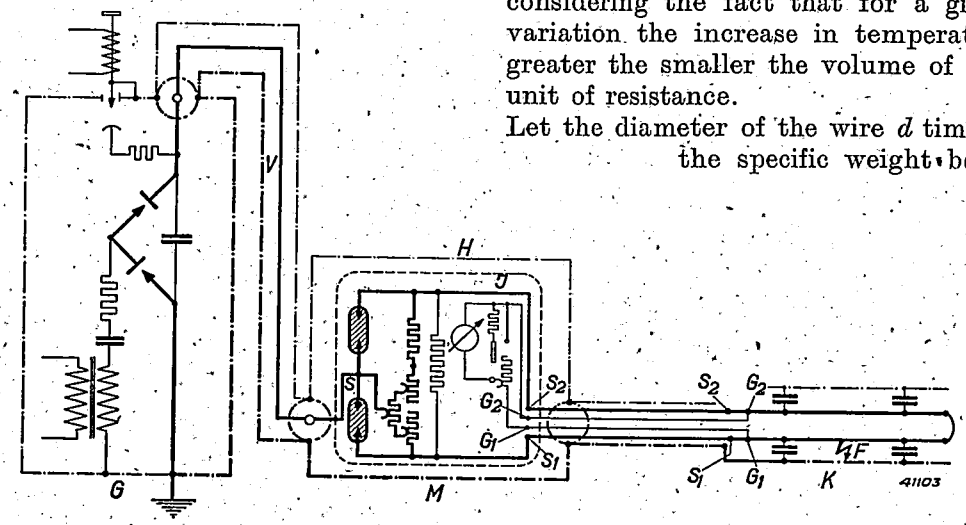


Fig. 5. Diagram of the connections of the installation for the localization of cable flaws, consisting of the previously described [1]) cable testing generator G which is joined by a short connection cable V to the cable measuring bridge M described in this article. Surrounding the whole a covering H proof against high voltage is indicated, part of it with a thick line to indicate the high-voltage circuit. All parts within the broken line I are insulated for 50 kV toward earth. K is the cable in which a flaw F must be localized. The high-voltage terminal of the generator is connected with point S. The voltages between point S and the points S_1 and S_2 of the cable conductors to be measured are limited by rare-gas cartridges; the galvanometer is connected between the terminals G_1 and G_2 of the measuring conductors. The connection between the measuring bridge M and the cable to be measured K is by means of a long connection cable. This cable has 4 conductors in order that the resistance of the connection lines and the transition resistance of the terminals may have no effect on the accuracy of the localization.

the electrical shock which a current impulse causes in the system. A mechanical impulse then occurs and a heating of coil and suspension wire. In addition to the fact that the electrical damping is very strong due to the strong magnetic field and the low ohmic resistance in the circuit, the rotating coil can be mechanically damped after a small deflection, since it is here a question of a zero instrument where large deflections are unnecessary and unwanted. The mechanical impulse can be taken up without danger.

The thermal resistance of a rotating coil of about 2 ohms is very large. When 1 A flows continuously through the coil only 2 W. are taken up, which the coil can easily tolerate. A current impulse may still attain a much higher value without overheating the coil. It is clear that in the last case the suspension wire is the weakest point. This can, however, withstand much more than would be expected in such a sensitive instrument. Thus a ribbon of the galvanometer used here can conduct 1 A continuously without reaching the melting temperature, since the heat developed can be adequately dissipated. In the case of a current impulse which lasts for less than for instance 0.1 sec, however, cooling plays no part. In that case it is a question of the *heat capacity per ohm*, considering the fact that for a given current variation the increase in temperature will be greater the smaller the volume of the wire per unit of resistance.

Let the diameter of the wire d times b equal q,

the specific weight be s,

the heat capacity c,

the specific resistance r_1.

Then, if we also take into account the fact that $Wg \sim d^3b$, the heat capacity per ohm becomes proportional to

$$\frac{Wg^2 \cdot s \cdot c}{d^4 \cdot r_1}.$$

In contrast to the case with a discharging resistance, the specific resistance must here be low; silver or copper-silver is therefore chosen. It is found further that a flat suspension ribbon is more important here than in galvanometers with continuous loading; *with a given sensitivity the thermal resistance is inversely proportional to the fourth power of the thickness of the ribbon.*

The application of these conclusions has made it possible to satisfy the requirements of strength without sacrificing part of the sensitivity.

Time to reach a steady deflection

It is mainly the time necessary to obtain a steady deflection which sets a practical limit to the sensitivity of a galvanometer. Due to the light weight the characteristic vibration time of the system is indeed always relatively short, but due to the low internal resistance (about 2 ohms) and an external resistance which sometimes almost amounts to a short circuit, the adjustment to a steady deflection becomes of the dead beat type.

Although the above-mentioned quantities for the galvanometer (0.5 μA and 1.0 μV) are not the most extreme values attainable, this construction is chosen in order to obtain a dead beat deflection of not more than a few seconds. The sensitivity is still so great that increasing it would bring no appreciable improvement due to the unavoidable external sources of disturbance.

With a switch which can short circuit the galvanometer the sensitivity can also be made 10, 100, 1000 and 10 000 times as small. Upon decreasing the sensitivity the galvanometer circuit is made almost aperiodic with a time for reaching steady deflection of less than 1 sec.

Comparison resistance

For a continuously regulable zero adjustment of the galvanometer in bridge connections a stretched resistance wire with sliding contact is theoretically the best method. When, however, the value of the resistance must be large compared with galvanometer and cable resistance, for instance 100 ohms, the length of the wire must be very great or the wire so thin that its cross section cannot be considered constant, so that a special calibration is necessary. Still worse is the fact that the wire will be worn more at certain spots than at others, so that the calibration becomes valueless and a precise measurement complicated.

It is much better to choose a resistance in steps of the type of a decade resistance (see fig. 5). Because of the fact that 10 times 10 positions are too few for an accurate balancing, and that it is undesirable to have more than two adjusting knobs, 20 times 20 positions were chosen. Moreover, an equally large fixed resistance is connected in series with the variable resistance, so that the balancing can be done with only one half, corresponding to the resistance of the defective cable conductor; the fixed resistance corresponds to the undefective return conductor.

In this way one stage for fine regulation corresponds to 2.5 per thousand of the cable length, *i.e.* to 2.5 meters per km. Suppose that the measuring length of a cable is 10 km and the theoretical inaccuracy of localization amounts to 2.5 metres. Then one stage for fine regulation will correspond to 25 metres of cable, thus with a galvanometer deflection of 10 scale divisions. By interpolation the position of the flaw can be found to within 2.5 metres. With a cable length of 1 km this would even be possible without interpolation.

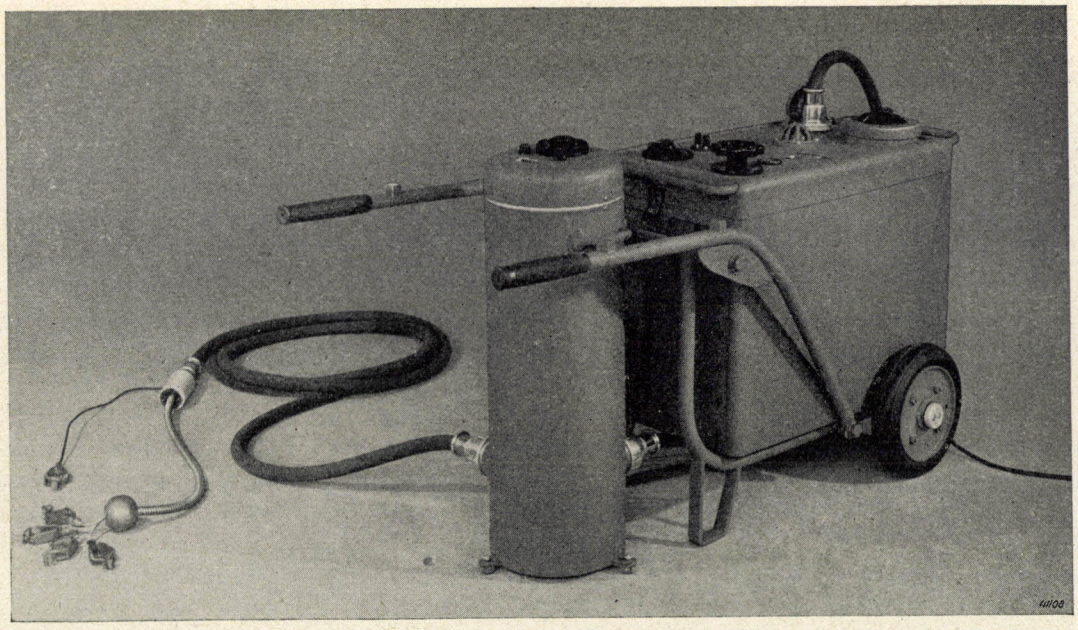


Fig. 6. The complete apparatus for the localization of cable flaws, consisting of the cable testing generator for 20 kV and the cable measuring bridge.

Construction

Fig. 6 shows the combination of the measuring bridge with the generator for 20 kV. *Fig. 7* shows the high voltage part of the measuring bridge and *fig. 8* the interior of the same.

It may be seen that the whole measuring bridge including safety elements and sensitivity adjustment of the galvanometer is housed in a well rounded can which is insulated from and fastened to a cover plate upon which the light source and scale of the galvanometer are mounted. The comparison resistance is varied from above by means of two concentric insulated shafts. Beside them are two other insulated shafts, one for the sensitivity regulation and one for zero point correction.

The whole galvanometer circuit (with the exception of the connection terminals) is enclosed in the above-mentioned can, which thus shields the circuit as a Faraday cage and which is itself connected with point *S* of the high-voltage circuit (see fig. 5). By this means any corona losses and leakage currents are kept outside the measuring circuit, so that they can have no unfavourable effect on the accuracy of the localization.

The high voltage part is placed in a metal cylinder, see fig. 7, into the lower part of which two connections for high-voltage-proof connection cables are screwed for the connection with the high-voltage generator and the cable to be measured, respectively. By this means the whole exterior is connected *via* the braided metal cable covering with the mass of the generator to be earthed, so that there is nowhere any part under tension which can be touched. The insulation between the high-voltage part and the earthed housing with operation knobs is calculated for 50 kV.

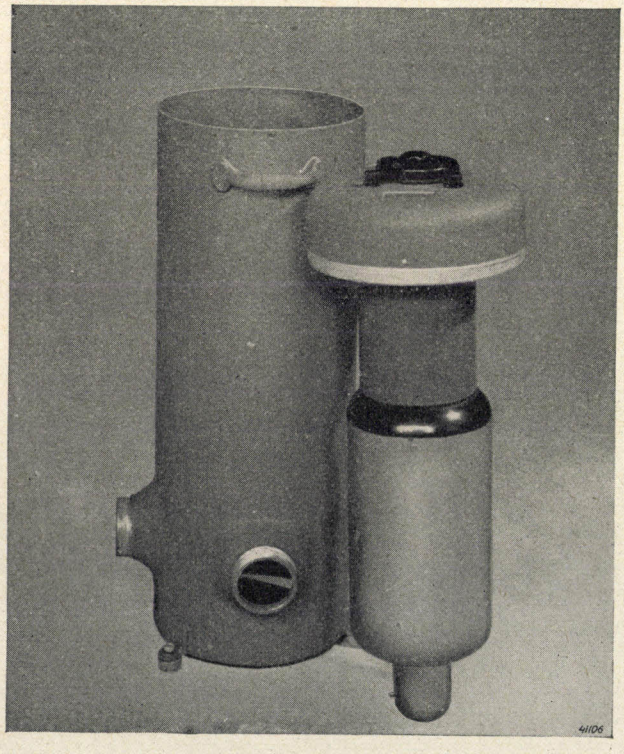


Fig. 7. High voltage part of the cable measuring bridge.

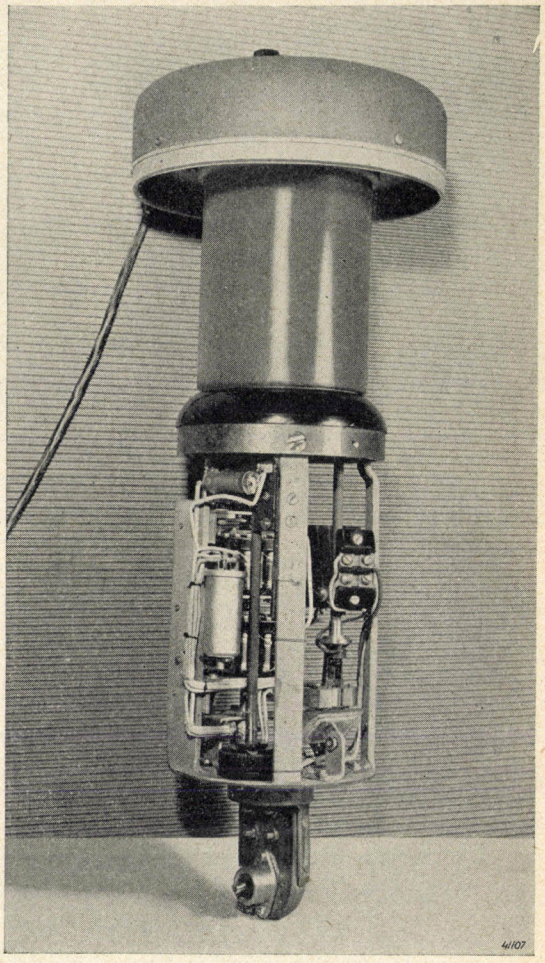


Fig. 8. View of the interior of the high voltage part.

Measuring voltage

Although cables which must be tested with voltages higher than 50 kV are no exception, and cable testing apparatus are therefore also necessary for higher voltages than 50 kV, in general an attempt will be made to limit the voltage for the localization of flaws to 50 kV, since at higher voltages the corona losses at the point of connection with the cable to be measured may be disturbing.

In this respect a sensitive measuring bridge which makes it possible to carry out the measurement with a lower current and therefore a lower

voltage, is of particular value: with a measuring bridge which is five times as sensitive as another the same theoretical accuracy of localization can be attained with five times as low a voltage. The limit to this decrease of the voltage is set only by certain sources of error which are *independent* of the measuring voltage. In general, with the measuring bridge here described the best results will be obtained with measuring currents between 1 and 10 mA.

THE TEXTURE OF CROSS-ROLLED MOLYBDENUM

by J. F. H. CUSTERS.

620.18

When cold-rolling molybdenum which crystallizes in a body-centred cubic-lattice and especially during cross-rolling employed practically, a typical texture may appear which differs entirely from the previously discussed rolling texture of certain face-centred cubic crystalline metals. In the case of molybdenum strip for technical applications measures are taken to prevent this texture, while in an investigation of the elastic behaviour of molybdenum the texture of the cross-rolled material could be used to advantage.

Various metals when rolled in the cold state acquire a texture, *i.e.* the crystal grains of the polycrystalline metal, which originally are arranged at random as far as the directions of their crystallographic axes are concerned, due to the rolling take up more or less sharply distinguished preferential positions. In a previous article attention was called to the rolling texture of nickel-iron alloys and other metals which crystallise in a face-centred cubic lattice (aluminium, copper, etc.). Quite a different rolling texture is found in the case of certain metals which crystallize in the body-centred cubic lattice (see *fig. 1*), such as molybdenum,

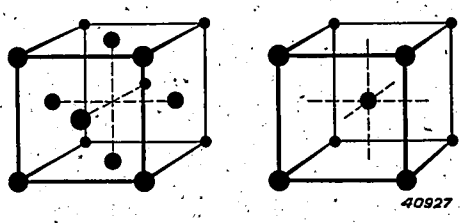


Fig. 1. Positions of the atoms in a face-centred (left) and a body-centred (right) cubic lattice.

tungsten, α-iron. In this article we shall study in particular the texture of rolled molybdenum which is used for instance for anodes of transmitting valves and the like, and we shall consider several questions connected with their texture.

The occurrence of rolling textures in general must be pictured in the following way. Due to the mechanical stresses caused by the rolling, displacements take place in each crystal grain along definite lattice planes (slip planes) and in certain directions (slip directions), the grain being very much hampered in its movements by the neighbouring grains. The result is that each grain undergoes not only a complex deformation and a possible splitting up into smaller grains, but also a rotation about definite axes.

The slip planes and directions referred to are different for a face-centred lattice from those for a body-centred lattice, and this constitutes one of the most important reasons for the differences in the textures which occur upon rolling such metal lattices. While upon the rolling of certain face-centred cubic lattices, as was explained in the article cited [1]), the crystal grains mainly take up a preferred position whose correlation with the direction of rolling is quite complicated, the texture which is formed when cold-rolling molybdenum and other metals crystallizing in a body-centred cubic lattice can be characterized rather simply: most of the grains take up such a position after the rolling that one of their six sets of rhombododecahedron planes [2]) is perpendicular to the direction of rolling [3]). The position of these grains is not in this way fully determined, since a rotation about the direction of rolling as an axis is still possible, as illustrated in *fig. 2*. Indeed in this respect the grains still actually show various orientations, as may be concluded from the X-ray diffraction diagram *fig. 3*. Upon further investigation, however, it is found that the statistical distribution among these orientations is not entirely uniform, but that the grains show a

[1]) J. F. H. Custers, Philips Techn. Rev. **7**, 45, 1942.

[2]) A rhombododecahedron plane is a plane passing through diagonally opposite edges of the cube of the crystal lattice.
See also Philips Techn. Rev. **7**, 15, 1942 (fig. 4).

certain preference for a position in which the set of rhombododecahedron planes which lie in the rol ing direction is perpendicular to the surface of the rolled strip This preference now

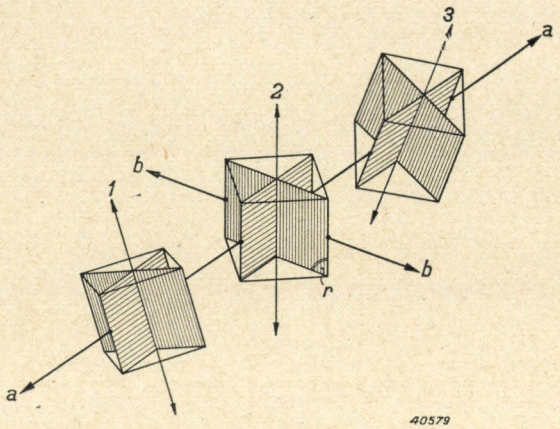


Fig. 2. After rolling, the crystal grains of the polycrystalline molybdenum specimen take up such a position that a rhombodecahedron plane (*r*) is perpendicular to the direction of rolling *a-a*. The rhombododecahedron plane which is parallel to the direction of rolling may still be at different angles with the transverse direction *b-b* of the rolled strip, as is here shown diagrammatically for three grains *1*, *2* and *3*.

becomes almost exclusive when the rolled strip is further rolled in a direction perpendicular to the direction of the first rolling. The strip is then said to be cross-rolled. Upon

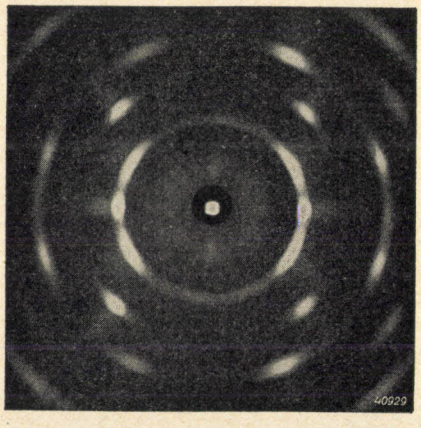


Fig. 3. X-ray diffraction diagram of rolled molybdenum upon normal incidence. The direction of rolling is from the top toward the bottom.

consideration of fig. 2 it is immediately clear that the freedom in the orientation of the grains remaining after the first rolling will indeed be lost upon the second rolling. In that case the second set of rhombododecahedron planes mentioned must take up a position perpendicular to the new direction of rolling, whereupon the final position appears [3]). As may be seen in the diagram of *fig. 4*, in each grain a cube plane is now parallel to the surface of the rolled strip, while the two other cube planes are perpendicular to this surface and make an angle of 45° with the direction of rolling.

In *fig. 5* an X-ray diffraction photograph is reproduced of cross-rolled molybdenum with normal incidence of the X-rays on the strip. The four-sided symmetry of the texture compared with the only two-sided symmetry of

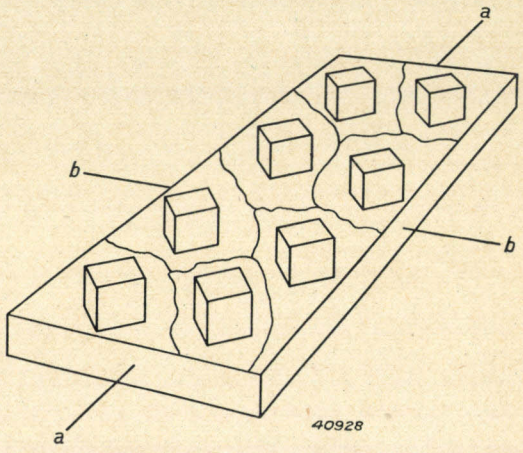


Fig. 4. Position of the crystallographic axes (cube edges) of the separate grains in cross-rolled molybdenum. *a-a* first, *b-b* second rolling direction.

fig. 3 is immediately noticeable. The innermost Debije-Scherrer circle of fig. 5, which here clearly contains four spots, corresponds to reflections at the rhombododecahedron planes. If the ideal position of the grains were fully attained, these four spots would not occur on the photograph, since the X-ray beam would then have a purely glancing incidence along the rhombododecahedron planes, while for reflection a finite angle of incidence (to be derived from Bragg's equation) is required. The fact that the spots, nevertheless, occur thus indicates

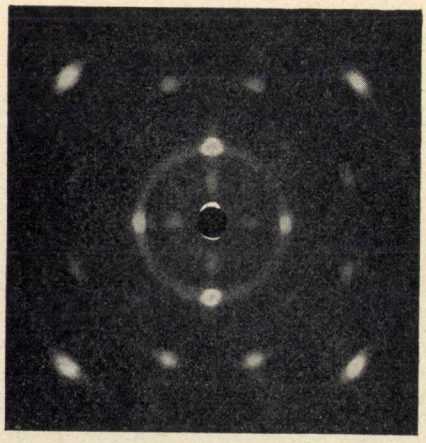


Fig. 5. X-ray diffraction diagram of cross-rolled molydbenum, to be compared with fig. 3.

[3]) C. E. Ransley and H. P. Rooksby, Inst. Met. **62**, 205, 1938.

that the grains exhibit a certain scattering about the ideal position. The minimum value of this scattering can be deduced directly from the photograph. Moreover, there are also a number of grains which take no part at all in the preferred orientation described, as may be seen from the fact that in the photograph of fig. 5 the complete Debije-Scherrer circles can still be seen faintly.

If this and the slight scattering mentioned are neglected, cross-rolled molybdenum strip may be considered as a pseudo single crystal: there are indeed crystal boundaries between the different grains, but in all manipulations in which the influence of crystal boundaries may be neglected the strip will behave almost exactly like a single crystal. In particular it will exhibit an anisotropy in some of its mechanical properties. A very striking example of such anisotropy occurs in the case of tungsten, which is closely related to molybdenum. If rolled tungsten strip is annealed somewhat unevenly at a high temperature the stresses occurring in the strip may cause small rectangular bits to spring out of the strip [4]). It is found that the coherence between the crystal grains is least along the cube planes at 45° to the rolling direction. The fact that fissures may appear along such planes upon the occurrence of stresses is also often experienced in severe rolling of tungsten and molybdenum strip in the cold state, whereby cracks may occur in directions at 45° to the direction of rolling.

The texture of rolled, and especially of cross-rolled molybdenum, in the treatment which the strip must undergo for practical application (for anodes for example it must be cold folded and provided with grooves), may therefore lead to difficulties.

Now from an economical point of view cross-rolling is the most suitable method of preparing molybdenum strip. We shall illustrate this by a consideration of the preparation of molybdenum.

Since molybdenum has an unusually high melting point, the metal is not obtained technically from a melt, but as in the case of tungsten the methods of powder metallurgy are applied[5]). The metal is first obtained in the form of a very fine powder by reduction from one of its compounds, the powder is compressed in a matrix to a rod which is then carefully presintered at about 1300° C. Then, when it has somewhat more mechanical strength, it is heated by a very high electric current to the true sintering temperature of about 2000° C. The sintered rod must now be rolled to strip or sheet with a width of for instance 12 cm, while the desired thickness may be for instance 0.2 mm. The largest possible rods are preferred as a starting point. The dimensions of the rod, at least its length, are, however, limited by the low mechanical strength before sintering, and its thickness is limited by the available rolling pressure. These factors lead to the use of rods of the dimensions 12 × 40 × about 200 mm. Such a rod is first hot-rolled transversally, so that, with the length of 20 cm remaining the same, the desired width of 12 cm is obtained. In order to be able to employ not too great a rolling pressure in spite of the work-hardening of the deformed and cooled material the strip of 20 × 12 cm obtained is then further rolled in the direction of length until the desired thickness of 0.2 mm is obtained.

The problem is to prevent the formation during this cross-rolling of a product with the above-described texture, which is unfavourable for further working. This can be done in various ways. The rolled strip already possessing the texture can be subjected to an annealing process above 1300° C. The strip then recrystallizes to give a coarse-grained material entirely without texture, as may be seen from the Laue diagram [6]) of *fig. 6*. In this diagram there is only a limited number of spots, which means that the X-ray beam, which is about 1 mm wide, has encountered only few crystal grains. The spots ares cattered quite at random, which indicates that there is no longer any texture. Such a coarse-grained material, however, has the disadvantage of being very brittle. A different method is therefore usually applied to prevent the occurrence of the rolling texture. After the transverse rolling and during the longitudinal rolling the strip is annealed at a relatively low temperature between successive stages of the rolling. It is found that in this way the rolling texture fails entirely to develop, and at the same time there is the advantage that in the heated state the material can more easily be rolled. This method is used in the practical manufacture of molybdenum strip.

While for practical purposes the object is to avoid the occurrence of the texture during cross-rolling, a case occurred in which advantage could be taken of the texture. It was a question of a comparative study of the elastic constants of several metals carried out in this laboratory [1]). While in measurements on polycrystalline specimens of metal only the average values of the elastic constants (averaged over all the crystallographic directions) are found, for theoretical insight into the elastic behaviour it is desirable

[4]) W. G. Burgers and J. J. A. Ploos van Amstel, Physica 3, 1064, 1936.

[5]) See for example J. D. Fast, Philips Techn. Rev. 4, 321, 1939.

[6]) See for example W. G. Burgers, Philips Techn. Rev. 5, 161, 1940.

to determine the constants in the different directions. To do this a single crystal of reasonable dimensions must be available, in this case of molybdenum. Since, however, the preparation of a large molybdenum single crystal is a very laborious task, use was made here of the fact that cross-rolled molybdenum possesses a texture which gives it the character of a pseudo single crystal. The photograph of fig. 5 has already been referred to in this connection, and it was pointed out that a number of crystal grains still occupy positions which deviate entirely from the preferred positions described. It was now found that these grains could also be brought practically into the ideal position by annealing the rolled strip for several minutes at 1200° C (upon annealing at a temperature 100° higher, on the other hand the texture disappears, as we have seen above). In *fig. 7* an X-ray diffraction pattern of the strip thus treated is reproduced. The intermediate segments of the Debije-Scherrer circles which were still visible in fig. 5 have here practically disappeared. The piece of metal has actually become almost a pseudo single crystal. The elasticity measurements could also be carried out satisfactorily with this strip. Incidentally it may be stated that these measurements showed that with molybdenum, in contrast to the cubic crystalline metals of lower melting point, the modulus of elasticity is smaller in the direction of the space diagonal of the cubic lattice than in the direction of the edge. The ratio of the two moduli in the case of molybdenum is 0.81, while with the alkali metals, for example, a value of approximately 5 is found, with gold and silver approximately 2.5.

Fig. *6*. Laue photograph of cross-rolled molybdenum which was afterwards annealed at 1300° C. The texture has disappeared entirely. (Laue photographs are made with X-radiation having a continuous spectrum. This is necessary in this case since with monochromatic radiation the few randomly oriented crystal grains present will not in general satisfy Bragg's equation, so that no reflections at all will be obtained. With a continuous spectrum on the other hand there is for every lattice plane in every grain a ray with a wavelength which can be reflected).

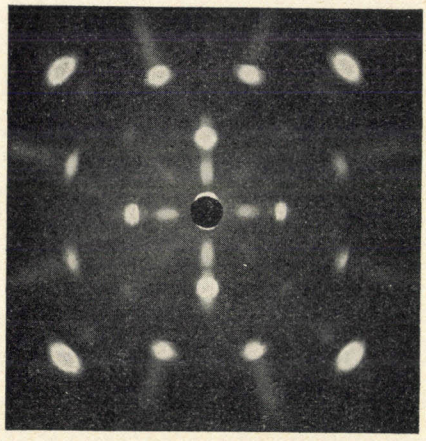


Fig. 7. X-ray diffraction diagram of cross-rolled molybdenum which was afterwards annealed at 1200° C. The texture in this case has become still sharper ; in particular the grains which did not yet occupy the desired orientation and which led to the appearance of the complete Debije-Scherrer circles have now disappeared.

[7]) M. J. Druyvesteyn, Physica **8**, 439, 1941. For the method of measurement see Philips techn. Rev. **6**, 372, 1941.

A DIODE FOR THE MEASUREMENT OF VOLTAGES

by M. J. O. STRUTT and K. S. KNOL.

621.385.2:621.3.029.6

A diode may be used to measure high-frequency AC voltages. In this article the problems are discussed which occur when this method is applied in the frequency region of decimetre waves. The problems are of the same type as those encountered in other applications of electronic valves in the region of decimetre waves. According to the principles which are applied in the construction of receiving valves for short waves a measuring diode was constructed, which can be used to a wavelength of 30 cm. In conclusion several other details of the measuring apparatus are discussed.

If an AC voltage is applied to the electrode system of a diode consisting of a hot cathode and an anode, current will flow mainly during that part of the period in which the anode is positive with respect to the cathode. Use can be made of this rectifying action in a very simple way to measure AC voltages. It is only necessary to introduce into the anode circuit a condenser with a large resistance in parallel (see *fig. 1*). The condenser is charged by the

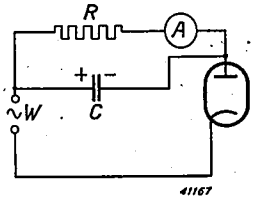


Fig. 1. Diagram showing the principle of the measurement of AC voltages by means of a diode.

rectified current in the circuit, while at the same time it is discharged by the resistance R. As the voltage V_c on the condenser increases, the discharging current becomes greater and the charging current smaller (see *fig. 2*). Finally the discharging current becomes equal to the average charging current so that the voltage no longer increases.

The stationary value of the DC voltage V_c thus obtained forms a measure of the amplitude W of the AC voltage. The relation between W and V_c is practically independent of the frequency, provided the latter is so high that the condenser loses only little voltage in the time interval between two maxima of the intermittent charging current ($\omega CR \gg 1$). The diode is there ore a suitable device for measuring AC voltages of high frequency. To do this one simply measures the DC voltage on the condenser, or, what amounts to the same thing, the DC current through the resistance.

We shall not at this point go deeply into the relation between W and V_c, since this relation will in practice be determined experimentally [1]). We shall state only that at very small amplitudes of the AC voltage ($W < 0.1$ V) the DC voltage is proportional to W^2, while for very large amplitudes a linear relation is found between W and V_c.

At the frequencies of broadcasting waves and even in the television region (about 7 m) the calibration of the measuring arrangement performed at low frequencies will usually still hold satisfactorily. At shorter waves, however, deviations are observed which may be ascribed to the same phenomena which also hamper the use of amplifier valves in this frequency region. These phenomena, which have repeatedly been dealt with in this periodical [2]), fall into two groups: firstly the influence of undesired impedances, mainly self-induction of connecting wires and mutual capacity of electrodes, secondly inertia phenomena in the electronic valve due to the transit time of the electrons between cathode and anode. The first phenomenon results in the fact that at high frequencies a different voltage acts between cathode and anode of the valve than would be expected on the basis of the diagram of the connections as drawn. The result of this is a deviation of the diode current and thus also of the final state of equilibrium from that expected, even if it were permissible to assume that the relation between anode current and anode voltage at

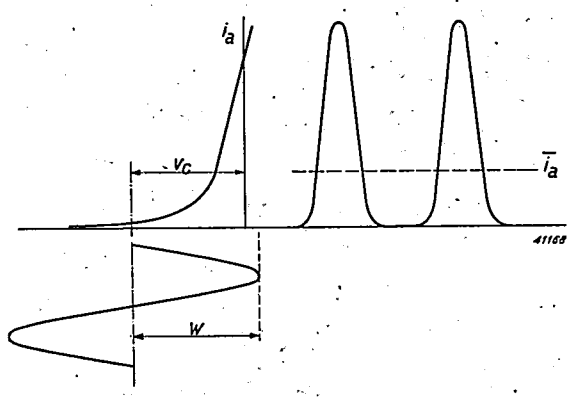


Fig. 2. If an AC voltage W with respect to the cathode acts upon the diode, the anode current i_a flows. The anode DC voltage V_c takes on such a value that the average anode current i_a is equal to the current which flows through the resistance R as a result of this voltage V_c.

---

[1]) A simplified theoretical treatment of the relation between W and V_c can be found in Philips Techn. Rev. **6**, 285, 1941.

[2]) Philips Techn. Rev. **1**, 161, 1936; **3**, 103, 1938.

these high frequencies were still given by the static characteristic.

The second phenomenon, the inertia of the electrons, results in the fact that the average anode current at very high frequencies becomes less than is calculated from the static characteristic. Part of the electrons which leave the cathode during the time when the anode is positive no longer have time to reach the anode, but when part of the way across are driven back to the cathode by the anode, which in the meantime has become negative. In *fig. 3* the behaviour of the anode current resulting from this is reproduced; for an explanation see the text under the figure.

The practical significance of these phenomena will be further explained later on in this article on the basis of apparatus actually constructed.

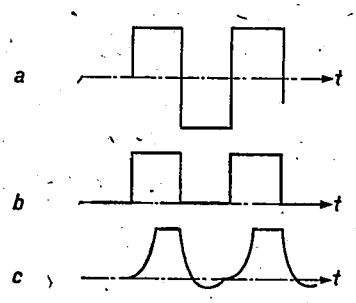


Fig. 3. Effect of the transit time of the electrons on the behaviour of the anode current.
a) Form of the anode voltage.
b) Form of the anode current when a current flows only at positive voltage.
c) Deformation of the anode current with a relatively long transit time of the electrons. The current, which is given by the number and average velocity of all the electrons present between cathode and anode, increases gradually to the stationary value after the commutation to the positive voltage, and after reversal of the voltage it decreases gradually to zero. At this point a negative current occurs temporarily due to returning electrons.

A new diode suitable for the measurement of voltages on decimetre waves

In order to be able to carry out voltage measurements in the frequency region of decimetre waves a diode was designed in this laboratory in which the disturbing factors above mentioned are avoided as well as possible. A common device for combating the effects of inertia is to diminish the distance between cathode and anode. Decreasing the distance between two electrodes, however, involves an increase in their mutual capacity. Since the capacity between cathode and anode forms one of the above-mentioned undesired circuit elements, its enlargement must be avoided. To do this it is necessary to decrease the surface of the electrodes with their distance apart. This of course means a decrease in the anode current. The limit of what can be attained in this direction is then found to be determined by the sensitivity of the available current meter.

It is obvious that full advantage can only be taken of the small dimensions of the electrode system when the connection wires to the electrodes are also kept as short as possible. Making the connection wires short is important not only to combat undesired impedances, but also to limit the external dimensions of the whole system. The momentary value of an AC voltage between two points is a well defined quantity only when the work necessary for the transportation of the unit of charge from one point to another is independent of the distance covered by the charge between the two points. In an electrostatic field this condition is always satisfied. In a high-frequency AC field, however, this is only the case when the available paths are short compared with a quarter wave length. If correct results are still required at a wave length of 50 cm, the external dimensions of the valve are limited to about 2 cm by this requirement.

The construction of the measuring diode

The measuring diode is shown in *fig. 4*. The system is mounted on a ring-shaped flange by means of which the base and the cover of the valve are fused together. The length of the connecting wires could be kept shorter than 1 cm in this way, while the dimensions of the electrode system amount to only a few millimetres.

As cathode a tungsten wire is used lying along the axis of the cylindrical anode. The cylindrical shape has a favourable effect on the transit times: the transit time between a wire and a cylinder is slightly less than that between two parallel plates the same distance apart. Moreover, the relatively high temperature of this hot cathode has the advantage that the electron shave a relatively high initial velocity, whereby the transit times are further decreased.

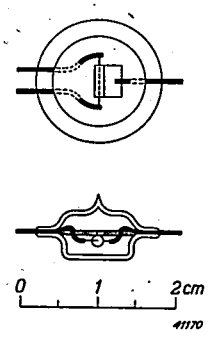


Fig. 4. Sketch of the measuring diode (external dimensions in actual size).

Results

In order to study quantitatively the influence of the various structural devices we shall consider the equivalent circuit given in *fig. 5*.

The connecting wires of cathode and anode have a certain resistance r which may be neglected at low frequencies but due to skin effect at the frequencies of decimetre waves may increase to such a degree that a disturbing damping occurs. Of still greater importance than this ohmic resistance of the wires, however, is the inductive resistance ωL, while finally the above-mentioned capacity c between cathode and anode must be taken into account.

At the frequencies of broadcasting waves and metre waves practically the entire externally applied voltage V acts on this capacity. At still higher frequencies, however, deviations begin to occur. The first phenomenon to be observed is a series resonance of the circuit L-c-L, which may cause the voltage V' on the condenser to become many times as high as the externally applied voltage V. For the resonance frequency an infinite voltage would even occur if there were no damping (resistances r and R_d) present in the circuit. If the resonance frequency is exceeded the voltage between cathode and anode rapidly falls to zero (with constant voltage between the external terminals). It is therefore only the region below the resonance frequency which has practical utility.

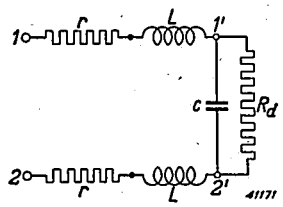


Fig. 5. Equivalent diagram of the parasitic impedances of the measuring diode which begin to exhibit an effect on the results of measurements at very high frequencies. r series resistance, L self-induction of the connection wires, c capacity between cathode and anode, R_d damping resistance caused by the transit of electrons.

The ratio V'/V is given in *fig. 6* for the new diode as a function of the frequency. For the sake of comparison corresponding curves are given for an older measuring diode and two other modern diodes. The values of L and c with which these curves are calculated are given in the following table *3*).

As the table shows, the improvement compared with earlier diodes is quite considerable. The resonance wavelength of the new diode lies at about 11 cm, which means that the error remains less than 10 percent to a wavelength of about 30 cm. In the case of the earlier measuring diodes this limit often lies above 1 m; it may therefore be said that of voltage measurements in the decimetre wave region have only been made possible by the new construction.

Table I

Selfi-nduction L of each conducting wire, capacity c between cathode and anode, and resonance frequency $f = 1/(2\pi\sqrt{2Lc})$ for the new measuring diode DA 50 compared with three other diodes designed earlier.

	L (10^{-8} henry)	c (pF)	f (10^8 c/sec)	$\lambda = 3\times10^8/f$ (m)
1) E A B	5	1,5	3,5	0,86
2) E A 50	1.8	2.1	5.8	0.52
3) Measuring diode 4674	1.5	1.65	7.1	0.42
4) Measuring diodeDA 50	0.75	0.24	26.5	0.11

When the source of the voltage to be measured has a high internal resistance, as is for example the case with a resonance circuit in tuning, errors in measurement may occur due to the damping effect of the diode connections. It is

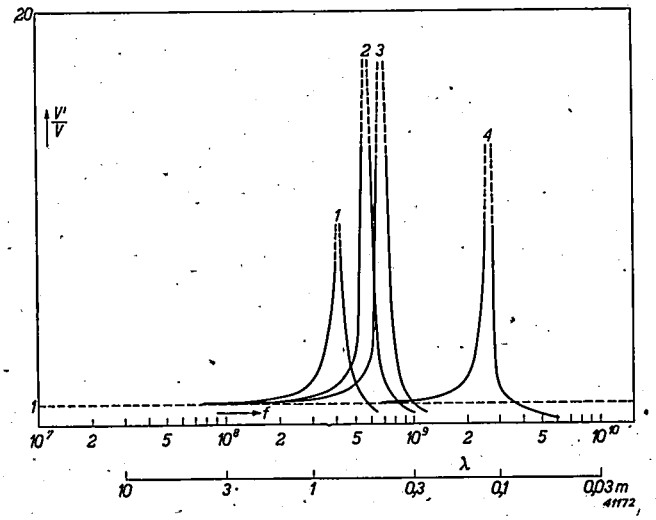


Fig. 6. Ratio of the voltage V' between cathode and anode of a diode to the voltage V between the external connection terminals of the supply wires as a function of the frequency. The variation is drawn for the equivalent circuit given in fig. 5. Curves *1* to *4* refer to the following diodes:
1) E A B, *2*) E A 50, *3*) 4674, *4*) DA 50 (new measuring diode).

[3]) It may be noted that the diode is often used as a relative meter to compare two voltages of the same frequency. In that case resonance phenomena of the parasitic impedances have no effect on the results of the measurement, since they have the same value in both cases. Errors may, however, occur due to the transit times of the electrons. The effect of the transit time depends not only on the frequency but also on the amplitude of the voltage applied. Nevertheless, under all practical conditions these errors remain small.

obvious that in the construction of the diode care has been taken to obtain as little damping as possible due to dielectric losses and leakages between the electrodes. Another property of the new diode which is favourable for the attenuating resistance is the extraordinary small anode current. The limitation of the anode current, which had to be accepted in order to be able to make the dimensions sufficiently small, although *per se* undesirable, is found to be an advantage in this connection. In the finished valve the average anode current in the operating region amounts to less than 1 μA; the diode has an internal resistance (attenuation resistance) of the order of 10^5 Ω. This resistance is sufficiently high for all practical cases.

At very high frequencies, as already stated, a further damping occurs as a consequence of the ohmic resistance of the connecting wires. If this attenuation is not expressed by a series resistance r, but by a new resistance R'_d in parallel between cathode and anode, the following is found:

$$\frac{1}{R_{d'}} = \frac{2r\,\omega^2 c^2}{(1-2\omega^2 Lc)^2}$$

Since in the new measuring diode the capacity c is unusually low (see table I), the attenuation $1/R'_d$ will also remain relatively low. In order to decrease this damping still further the resistance r was decreased by using lead wire for the connections (due to the skin effect it is only the outermost layer of the wire which determines the conductivity at high frequencies). In this way R'_d is made to amount to about 10^4 Ω at a wavelength of 40 cm, which may be considered very satisfactory.

A final cause of attenuation is formed by the transit time of the electrons between cathode and anode. Due to the low value of the anode current, however, this effect remains small and may be disregarded.

The other effect of the transit time of the electrons already discussed above, the decrease in the anode current at high frequencies, is also not disturbing and to a certain extent even an advantage. Due to this the increase in the sensitivity in the region before the resonance is partially compensated, so that the characteristic remains flat over still wider frequency regions than would otherwise be the case. The dimensions of the diode are so chosen that the rise in sensitivity due to resonance begins at about the same frequency as the fall in frequency due to transit time effects. In the earlier types the resonance frequency usually lies much lower, while the transit times do not vary appreciably. This compensation phenomenon was therefore previously not taken advantage of.

Application of the diode

It would lead us too far if we were to discuss the different possibilities of connection of the diode in detail. All connections may be considered as variations of the simple principle indicated in fig. 1. These variations depend chiefly upon the varying character of the sources of voltage to be measured. The high-frequency source of voltage may possess an infinitely high resistance for direct current or it may form a short circuit; the voltage source may be earthed at one end or in the middle; all these cases require slightly different arrangements of the connections. The points of difference, however, are very obvious and involve no fundamental principles.

Of greater interest than the connection is the technical construction of the elements which are used in it. We shall here briefly discuss three of these elements: the supply line for the voltage, the valve holder and the measuring instrument. When working with very high frequencies the supply line for the voltage will often be in the form of a Lecher system on which the connection terminals of the measuring diode form sliding contacts. This principle with which the voltage can be measured as a function of the position has the advantage that there is no uncertainty about the phase of any possible standing waves at the position of the measuring diode, provided the distance between the Lecher wires is small compared with a quarter wavelength. In *fig. 7* this arrangement is shown diagrammatically; it is also evident from this figure how the cathode can be heated without the necessity of introducing disturbing extra wires in the neighbourhood of the Lecher system.

In the part to the left of the shield K only DC voltages occur. In that part the operations necessary for the measurement can be performed with outfear of any reaction on the happenings in the high-frequency part.

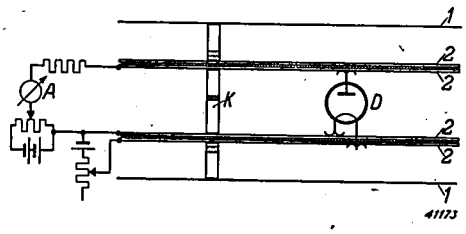


Fig. 7. Lecher system for measuring high-frequency AC voltages. *1*) external shielding, *2*) Lecher wires, K sliding bridge forming shield at the same time, D diode. Each of the wires of the Lecher system consists of two conductors separated by a layer of insulation. This division makes it possible to supply the voltage for heating the cathode *via* the Lecher system as well. Shaded parts consist of insulation material.

In the construction of the valve holder special care must be taken that the advantage of the low capacity between cathode and anode (0.24 $\mu\mu F$) is not destroyed by parasitic capacities. As few metal parts as possible will therefore be used and for the non-metallic parts a material with a low dielectric constant and as small dielectric losses as possible. In *fig. 8* a sketch is given of a valve holder constructed on this principle ; it is made of polystyrene, an artificial resin which is as clear as glass and which satisfies the requirements mentioned.

As a measuring instrument we use a reflecting galvanometer constructed in this laboratory, whose most important properties are described elsewhere in this number in connection with the discussion of a cable measuring bridge. This instrument excels by its sensitive yet sturdy system, which, in contrast to many other sensitive galvanometer systems, need not be placed in a horizontal position. Deviating from the description given on page 114, in our case a coil with an internal resistance of 500 Ω was used; moreover, in our case it was unnecessary to take special precautions for the insulation of the measuring system against high voltages. *Fig. 9a* gives a sketch of the construction of the instrument, while *fig. 9b* shows the external finish of an experimental model.

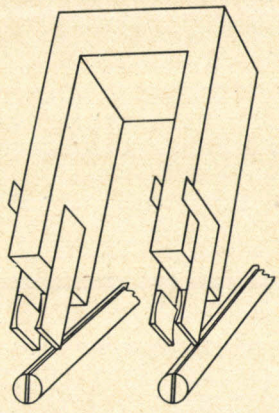
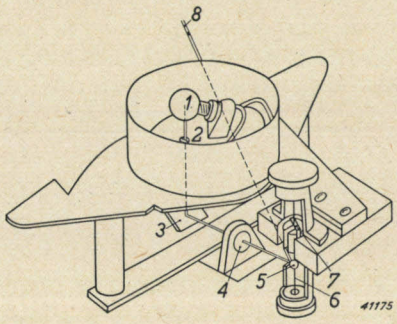


Fig. 8. Valve holder for the measuring diode. The non-metallic parts are mode of polystyrene.



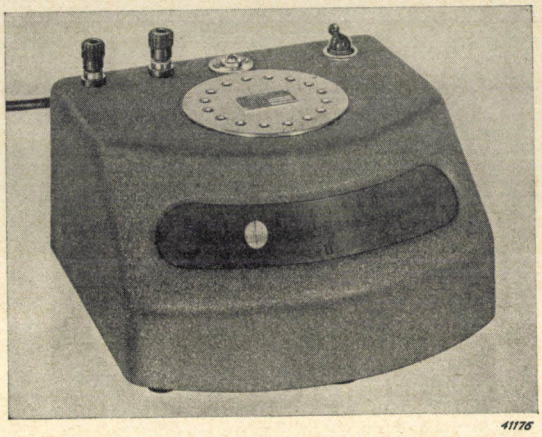


Fig. 9. *a*) Reflecting galvanometer suitable for measuring the DC of the measuring diode. *1*) lamp, *2*) diaphragm with black lines, *3*) mirror, *4*) lens which focuses the black line on the scale, *5*) rotating mirror which together with the coil *7*) is suspended between the wires *6*), *8*) beam of light toward, the scale. *b*) Form of construction of the reflecting galvanometer.

Vol 7 • No. 5, pp. 129-160 MAY 1942

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 NINE-KILOWATT EXPERIMENTAL TELEVISION TRANSMITTER

by M. VAN DE BEEK.

621.397.6

While in the case of a broadcasting transmitter the modulation spectrum has a width of only a few kc/sec, the spectrum of the video-frequency signals with which the carrier wave of a television transmitter is modulated occupies a region several Mc/sec in width. This fact, as well as the related fact that a television transmitter must work on very short wave lengths, causes the transmitting apparatus of a television transmitter to differ considerably in many respects from that of a broadcasting transmitter. The description here given of the powerful experimental television transmitter in Eindhoven attempts to present these differences clearly. Following a short description of the excitation of the carrier-wave frequency (wavelength 7 m) a detailed discussion is given of the method of modulation, the dimensions and construction of the output stage, which contains two water-cooled pentodes type PAW 12/15 and furnishes a maximum power of 9 kW to the transmitting aerial, as well as several particulars of the modulator.

The development of television technique has from the very beginning aroused considerable interest. It has been particularly the methods of recording and reproducing the pictures which have attracted the most attention and which have been dealt with in many technical and popular treatises. The problem of transmitting the electrical signals into which the picture is translated was usually less well described, since in principle the wireless transmission of signals was already familiar from broadcasting and therefore provided no new technical sensations. In this periodical also the subject has only been mentioned incidentally, while studio equipments and receivers have repeatedly been dealt with elaborately [1]). Notwithstanding all this, the actual transmitting apparatus of a television installation has its own specific problems which arise from the great width of the frequency band which must be faithfully transmitted as far as amplitude and phase are concerned, and from the short wave length connected with the former on which the transmitter must work. In the following article, in which we shall describe the experimental television transmitter in Eindhoven, this will become quite clear.

The transmitter in question originated from the experimental television transmitter which was developed here in 1935 [2]) and which furnished an output of several hundred watts. The output has been increased to approximately 9 kW (maximum output radiated by the aerial).

Survey of the complete transmitter

In *fig. 1* the electrical structure of the whole is represented in the form of a block diagram. The carrier-wave frequency is derived from the oscillation of a plate of quartz crystal which vibrates at a frequency of 5.4 Mc/sec (wave.

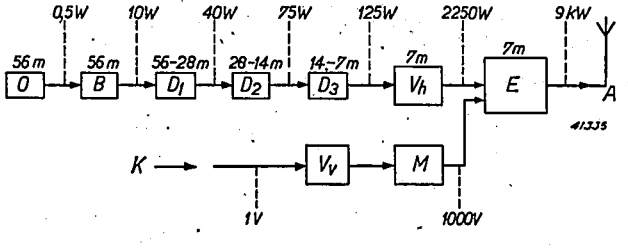


Fig. 1. Block diagram of the television transmitter. *O* crystal oscillator, *B* buffer stage, $D_1D_2D_3$ frequency doubling stages, V_h high-frequency amplifier stage, *E* output stage, *K* cable from studio, V_v video-frequency amplifier, *M* modulator, *A* aerial.

[1]) Philips Techn. Rev. **1**, 16, 1936 (An experimental television transmitter and receiver); **1**, 321, 1936 (Television); **2**, 33, 1937 (A television receiver); **2**, 72, 1937 (Television system with Nipkow disc); **2**, 249, 1937 (The enlarged projection of television pictures); **3**, 1, 1938 (A transportable television installation); **3**, 285, 1938 (Television with Nipkow disc and interlaced scanning); **4**, 42, 1939 (The Nipkow disc); **4**, 342, 1939 (Television receivers).

[2]) J. van der Mark, An experimental television transmitter and receiver, Philips Techn. Rev. **1** 16, 1936.

length 56 m). In a number of amplifier stages, which at the same time serve partially as frequency doublers, the high-frequency output is increased to about 100 W, at the desired wavelength of 7 m. The following amplifier stage furnishes the power necessary for the excitation of the large water-cooled transmitter valves which are included in the output stage. This output stage, in which at the same time the video-frequency signals furnished by the studio are modulated on the carrier wave, delivers to the transmitting aerial a maximum energy of 9 kW.

Oscillator and frequency-doubling stages

By the employment of a quartz plate cut from a quartz crystal along a so-called temperature-insensitive plane, a very constant carrier-wave frequency is obtained without it being necessary to place the oscillator stage in a thermostat. There are two reasons why the excitation of the desired carrier-wave frequency is begun with a quartz plate which vibrates with a much lower frequency. In the first place it is practically impossible to make quartz plates for wavelengths shorter than about 20 m, since they become too thin. In the second place with an initial wavelength which is not too short it is much easier to attain a satisfactory amplification, so that fewer amplifier stages are needed than when the oscillator stage already operates on 7 m.

In order to prevent loading variations in the last stages, due for instance to the modulation, from reacting on the oscillator stage (and thereby on the carrier-wave frequency), the first amplifier stage is constructed as a buffer stage, *i.e.* the control grid of the amplifier valve in question is given such a high negative bias as to prevent flow of grid current, so that there is practically no load on the oscillator stage. This is in contrast to the ordinary situation in transmitter stages, where in order to attain a reasonable efficiency the grid voltage is chosen so that grid current shall flow.

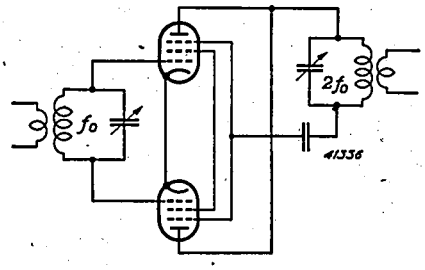


Fig. 2. Diagram of the connections of the doubling stages. The two valves are excited in opposite phase; the anodes are connected in parallel. The anode circuit is tuned to a frequency twice as high as that of the grid circuit. In order to give the free end of the anode circuit as constant a potential as possible, *i.e.* one which varies as little as possible at a high frequency, with the very short waves used here a connection with earth is not suitable, since the long connection wires may have considerable impedances for these waves. Therefore the anode circuit is connected here with the mutual connection of the screen grids of the two valves, which, with the valves used, leads also to a very simple structural solution, since anode and screen grid leads are placed at the top of the valve envelope.

Each frequency-doubling stage contains two valves which are connected in the manner shown in *fig. 2*. Each valve, since it operates in Class C arrangement, only furnishes anode current during less than half a period of the excitation voltage. Since the control grids are excited in opposite phase, where as the anodes are connected in parallel, an anode current of the form shown in *fig. 3* is obtained. The fundamental frequency of this current,, which is filtered out by the tuned anode circuit, is double the frequency of the excitation voltage. In this way the desired doubling of the frequency is obtained with the relatively high efficiency of about 55 percent. A pleasing property of the chosen system of amplification with frequency doubling is that, due to the difference between the frequencies of input and output voltage, no reaction is possible and therefore no self generation of the frequency-doubling stages can occur.

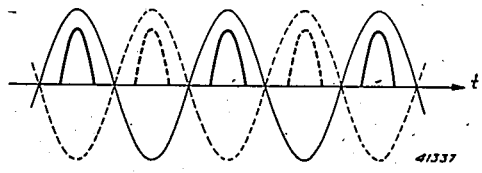


Fig. 3. Grid voltage (thin line) and anode current (heavy line) of the frequency-doubling stage. The full drawn lines are for one valve, the broken lines for the other. The total anode current has a fundamental frequency twice as high as the grid voltage.

Preliminary considerations in the design of the output stage

In order to discuss the connections and dimensions of the other transmitter stages, particularly the output stage, it is desirable first to go back to the video-frequency signal as it is sent from the studio to the transmitter.

When n pictures per second are transmitted, each having N lines, while the ratio of width to height of picture is b/h, the lowest modulation frequency [2]) occurring is $f_{min} = n$ c/sec, the highest $f_{max} = {}^1/_2\, n.N^2\, b/h$ c/sec. In our case $n = 25$, $N = 405$ and $b/h = 1.2$, so that one must count on modulation frequencies between 25 c/sec and 2.5 Mc/sec. When this video-frequency spectrum is modulated on a carrier wave there occur, in addition to the carrier-wave frequency, two side bands, each 2.5 Mc/sec

in width (*fig. 4*). From the point in the transmitter where the carrier wave is modulated up to and including the aerial the connections must therefore have a transmitting region

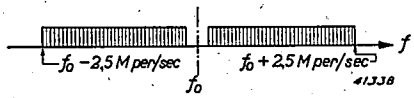


Fig. 4. Frequency spectrum of the carrier wave modulated with the video-frequency signal.

(band width) of at least 5 Mc/sec. This condition is made more imperative by the fact that all the vibrations within the frequency band mentioned must not only be transmitted in the correct amplitude relations, but also, in contrast to broadcasting transmitters, in a certain relative phase: the phase shift must vary linearly with the frequency so, that no relative displacement of points in the picture and thus distortion may occur. Since in every case such a linear relation between phase and frequency can only be obtained for phase angles which are not too large, the last requirement mentioned comes down to the condition that the phase rotations must only be very small [3]).

What consequences has this for the connections?

In the first place the requirement relating to the phase rotation implies that no band-pass filters can be used for the realization of the wide transmission region, since with such filters large phase shifts always occur. Recourse must therefore be had to single circuits, which may be represented by the diagram of *fig. 5*: a loss-free self-induction L and capacity C, with a loading resistance R. This loaded circuit is connected with a source

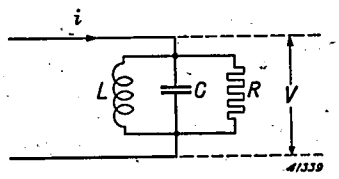


Fig. 5. Diagram of loaded tuned circuit.

[3]) A similar requirement is encountered in carrier-wave telephony, where wide frequency bands are also used; see for example F. de Fremery and G. J. Levenbach, Carrier-wave telephony on coil-loaded cables, Philips Techn. Rev. 4, 20, 1939. If the requirement of linearity is satisfied no "phase distortion" occurs, but the signal has a certain transit time which may sometimes be quite long. While in carrier-wave telephony certain limits must also be set to this, the finite transit time is no objection at all in television.

of constant alternating current i (a pentode for instance). The impedance $Z = V/i$ of the circuit, which in this case represents the behaviour of the voltage V on the circuit, is drawn in *fig. 6* (for certain values of L, C and R) according to magnitude $|Z|$ and phase φ, as a function of the frequency. If for the outermost side-band frequencies we allow a fall in the amplitude from 1 to $1/\sqrt{2}$, a frequency region of the width

$$2\,\Delta f = \frac{1}{2\pi\,RC} \quad \ldots\ldots\ldots \quad (1)$$

around the resonance frequency $f_0 = {}^1/_{2\pi}\sqrt{LC}$ is found to be useable. In this frequency region, according to fig. 6*b*, the phase remains below

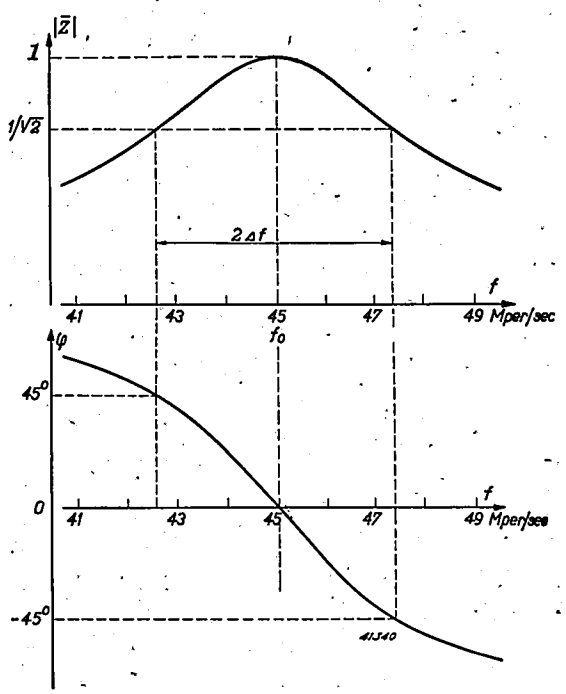


Fig. 6. *a*) Absolute value $|Z|$ and *b*) phase φ of the impedance of a tuned circuit according to fig. 5, as a function of the frequency f. For the circuit elements the following values are here assumed: $L = 0.395$ μH, $C = 31.8$ pF, $R = 1000$ ohms.

45° and varies practically linearly with the frequency.

In order to give the circuit the desired great band width of $2\,\Delta f = 5$ Mc/sec, according to eq. (1), R and/or C must be made very small (a decrease in C must be accompanied by an increase in L, in order that the resonance frequency may remain in its place). Now C cannot be made arbitrarily small, because even when no capacity at all is introduced into the circuits the capacities of the transmitting valves in parallel always act over them. In push-pull connections of two medium-sized valves there is in this way always a capacity of at least

30 pF. The result is that in order to obtain the required band width the resistance R may not be made larger than $^1/_4\pi\Delta fC \approx 1000$ ohms. This is very unfortunate, since the power output of the transmitter valve, of which the circuit forms the load on the anode, is proportional to R within certain limits.

If there are several modulated stages the total frequency characteristic of the transmitter is obtained by multiplying the amplification factors of all the modulated stages with each other at each frequency, and adding up the phase rotations. Therefore, for each separate stage, a stricter criterion than the above must be chosen for the useable band width, so that for each of these stages the anode impedance must again be considerably lower. It is therefore clear that a television transmitter should always be modulated in the output stage. For a given output of the transmitter one then obtains on the one hand a higher efficiency and output from each of the preliminary stages, so that the number of such stages may be smaller, while on the other hand the output stage may still retain a relatively high anode impedance and thus this stage, which is the most expensive one as far as valve and current consumption are concerned, can operate with the greatest possible efficiency.

Since the quality factor Q of a resonance circuit may be written as $Q = f_0/2\Delta f$, a great band width is always accompanied by a poor quality factor (poor building up), so that doubt may arise whether the introduction of the self-induction for the tuning of the anode circuits is actually worth the trouble. The tuned circuits do, however, justify their use for the suppression of the harmonics of the carrier-wave frequency — in our case, where $f_0 = 43.2$ Mc/sec and $2\Delta f = 5$ Mc/sec, the second harmonic is attenuated to about 1/20 and the third to about 1/30 — and to prevent the video-frequency signals from reaching the aerial directly, which would cause radio interferences in a wide region around the television transmitter on all wavelengths above 120 m.

Let us now consider the choice of a modulator system. Should the carrier wave be modulated by variation of the anode voltage or of the grid voltage of the output stage? Grid voltage modulation is generally employed in television transmitters for the following reasons. In the first place, from the point of view of simplicity of construction, grid modulation is preferable, since with anode modulation a very large modulation power is needed because of the high anode voltages of the output stage. This leads to a large and expensive modulator for the necessary wide modulation spectrum. In the second place, and this is even more important, with grid modulation the synchronization impulses can be better transmitted. While all the gradations of brightness of the picture to be transmitted are transmitted by varying the carrier-wave amplitude between 30 and 100 percent, each synchronization impulse consists of a short-lived suppression of the carrier wave (*fig. 7*). With anode modulation

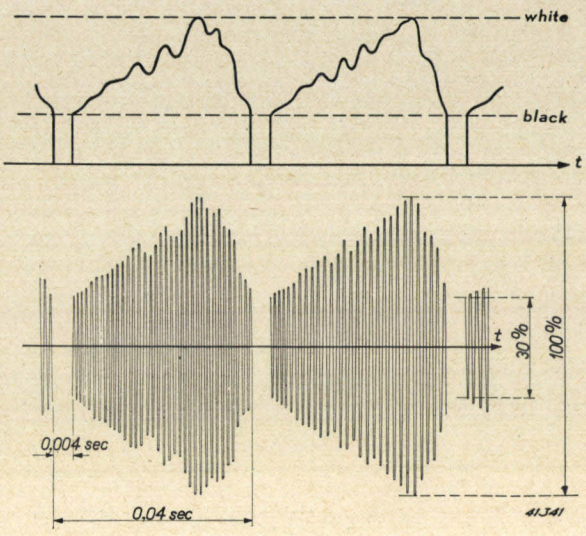


Fig. 7. *a*) Variation with time of the video-frequency signal, schematically. *b*) Variation with time of the high-frequency signal emitted by the television transmitter. (The line synchronization signals are omitted for the sake of simplicity).

the anode voltage may indeed become equal to zero during the synchronization impulses, but in spite of this, due to the induction effect, a certain anode alternating current continues to flow [4]) and there is still, therefore, some radiation by the aerial. With grid modulation, on the other hand, it is possible to make the control-grid voltage so strongly negative at those impulses that the output valve passes absolutely no anode alternating current. Sharper synchronization signals are thus obtained and thereby a more reliable functioning of the receiving sets.

It is in fact less easy with control grid modulation than with other methods of modulation to realize a strictly linear modulation [5]) (*i.e.* a linear relation between carrier-wave amplitude and momentary value of the modulation voltage). In the case of a television transmitter, however, this is much less important than in the case of a broadcasting transmitter, since the eye is much less sensitive to non-linear distortion than the ear. It is a familiar fact that large distortions in the rendering of brightness gradations, for instance in photography, can be tolerated without producing an "unnatural" impression on the eye.

[4]) See Philips Techn. Rev. **1**, 176, 1936.

[5]) See J. P. Heyboer, Five-electrode transmitting valves, Philips Techn. Rev. **2**, 257, 1937.

Dimensions of the output stage

The output stage contains two water-cooled pentodes, type PAW 12/15, in push-pull connection (these are the largest pentodes made). As was explained above, the capacity of the anode circuit must be made as small as possible. In anticipation of the description of the construction to be given below, it may be stated that this capacity could be limited to 32 pF. According to eq. (1) an anode loading with an impedance of about 1000 ohms is then permissible to obtain the desired band width.

In order to develop as much high-frequency energy as possible in this anode impedance, which is very low compared with broadcasting transmitters, the anode current must be modulated as much as possible, preferably to saturation. What anode DC voltage, excitation voltage and modulation voltage must be chosen for this, and what output can be obtained?

The valve PAW 12/15, with a normal tungsten filament for 22 V, 80 A, has an emission of 11 A. This, however, is not fully available, since, in addition to the anode, the control grid and the screen grid also require their part of it. If on the basis of experience we reserve 1.75 A for the control grid current and 0.75 A for the control grid peak current, there remains available for the anode peak current I 8.5 A.

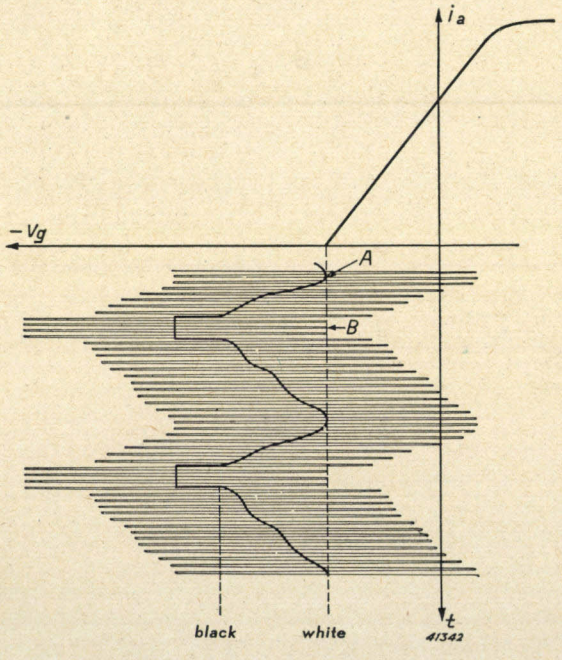


Fig. 8. Variation of the grid voltage of the output stage with control grid modulation. High-frequency excitation voltage and video-frequency modulation voltage are superposed in such a way that at the peak adjustment (completely white picture, point A) the valve just operates in class B (grid bias equal to overbiasing voltage), and that the carrier wave is suppressed (peak of excitation still below overbiasing voltage) during the synchronization impulses (point B).

In order to obtain as linear a modulation as possible in both valves in push-pull connection, the control grid modulation is so effected that upon adjustment to the largest carrier-wave amplitude (completely white picture) the valves just operate in "class B", *i.e.* the control grid DC voltage is equal to the overbiasing voltage, see *fig. 8*. In this way at this peak adjustment (A) the anode current takes on the form indicated in *fig. 9*. The first harmonic, with which

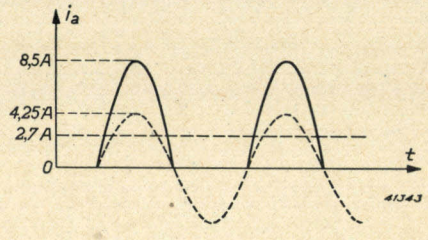


Fig. 9. Form of the anode current in one valve of the output stage, at the peak adjustment. The available emission permits a peak value of 8.5 A; first harmonic 4.25 A, average value (anode direct current) 2.7 A.

we are concerned, has, as the Fourier series shows, the amplitude $I/2 = 4.25$ A. Now the action of the push-pull connections may be conceived as if each of the two valves were loaded with half of the anode impedance, *i.e.* 500 ohms. The power contributed to this by the anode alternating current amounts to $\frac{1}{2} \cdot 4.25^2 \cdot 500 = 4.5$ kW. Thus at the peak adjustment the two valves give a total output of 9 kW. Since only in the case of a completely white picture do the valves operate at this peak adjustment, and with entirely black picture the carrier-wave amplitude is adjusted to 30 percent of the maximum, the average output radiated amounts to about 3 kW.

It is further evident from the above that the anode AC voltage at the peak adjustment amounts to $4.25 \times 500 = 2125$ V. Since it is known from the static I_a-V_a characteristics of the valve PAW 12/15 that the momentary value of the anode voltage may not be lower than about 2600 V, the anode DC voltage must be taken approximately equal to 4750 V. Since for a current of the form shown in fig. 9 the anode DC $I/\pi = 2.7$ A, a DC power of $4750 \times 2.7 = 12.8$ kW is applied to each valve. The difference between this power applied and the output, *i.e.* 12.8—4.5 = 8.3 kW, is dissipated in the valve. The permissible dissipation (12 kW) is thereby by no means exceeded. The efficiency of the anode at the peak adjustment is $100 \times 4.5/12.8 = 39$ percent.

The fact that the output of the valve in our case is limited by the emission of the filament, and not, as is often the case in broadcasting transmitters,

by the permissible dissipation, is directly connected with the great band width required, and the small anode impedance thus necessary. With larger anode impedance the calculated current would result in a higher anode AC voltage, as a result of which the anode DC voltage would also have to be chosen higher. The power applied would then increase, and it would increase more rapidly than the power output, so that also the dissipation would quickly approach the permissible limit and a limitation of the anode current would finally become necessary. At the same time the voltage modulation, which in our case is only $2125/4750 \approx 0.45$, then approaches the more normal values in the neighbourhood of unity (see the article referred to in footnote [5]).

Let us now calculate the required excitation power of the output stage. The average slope (static and dynamic) of the valve PAW 12/15 is about 6 mA/V. Thus for the two valves in push-pull connection we need a grid AC voltage with the peak value 8.5 A : 6 mA/V = 1400 V. In modulation the total grid voltage will vary as in fig. 8, from which with the help of the characteristics it may be deduced that the average grid current at the peak adjustment will amount to about 300 mA. The theoretically necessary excitation power for the two valves together is therefore $2 \times 0.3 \times 1400 = 850$ watts. However, due to the finite transit times of the electrons [6]) an extra grid damping occurs, as a result of which the excitation power is increased to about 1125 W. Moreover, it is necessary that during the modulation when the grid current is varying, the excitation power should remain constant. Since this is only possible with not too great loading fluctuations of the stage which furnishes the excitation voltage, the latter must be given a constant initial load independent of the grid current of at least the same magnitude as the excitation power, by means of grid damping resistances. This stage must therefore be able to deliver 2250 W of high-frequency power. This is obtained by means of four valves type PB 3/800 two by two in push-pull connection, each of which can deliver 625 W on 6.5 m. This stage itself in turn requires $4 \times 25 = 100$ W modulating power, which is furnished by the preceding frequency doubling stage (see the diagram of fig. 2).

Construction of the output stage.

The capacity of the anode circuit of the output stage, when no extra capacity at all is added, consists not only of the unavoidable grid-anode capacity of the two valves PAW 12/15 connected in push-pull, but also of various parasitic capacities: that between the two water-cooled anode jackets of the valves, between the cooling jackets and earth, etc. (see *fig. 10*). By placing the valves as shown in *fig. 11* at a good distance from each other and

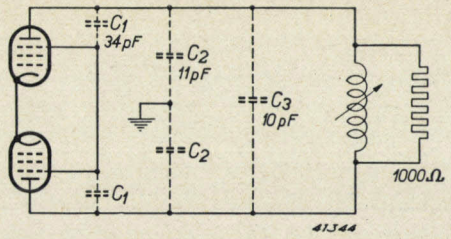


Fig. 10. The anode circuit of the output stage contains a capacity which is composed mainly of the anode-grid capacities C_1 of the two valves in series, the capacities C_2 between each anode cooling jacket and earth, in series, and the capacity C_3 between the two anode cooling jackets.

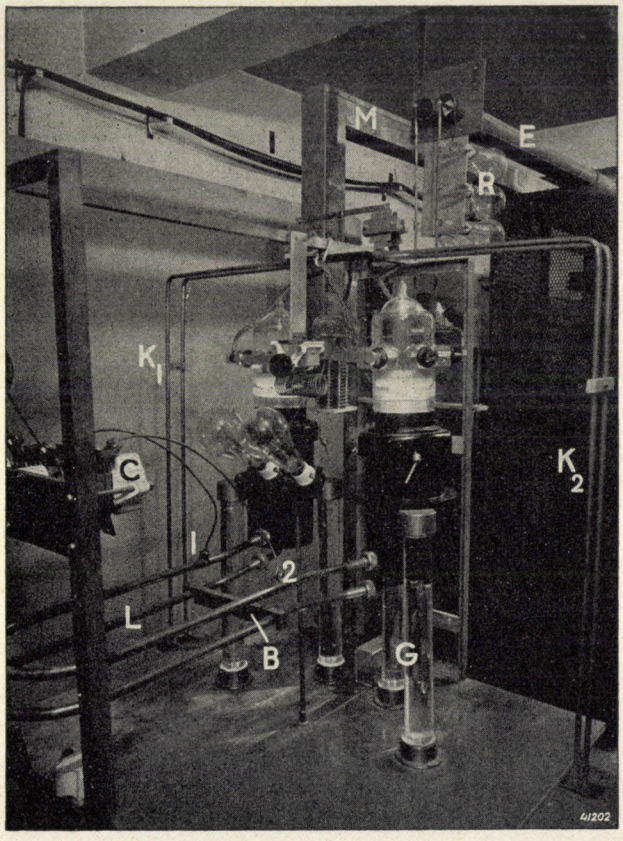


Fig. 11. Part of the output stage. The two water-cooled pentodes PAW 12/15 are mounted on glass columns. *G. L* = L e c h e r system, which functions as self-induction and is tuned with the short-circuiting bridge *B*. At *1* and *2* the supply line is tapped off which conducts the high-frequency energy *via* the high-voltage coupling condensers *C* to the aerial. In order to be able to use the wires of the L e c h e r system, which are constructed in the form of tubes, for the inlet and outlet of the cooling water for the anode jackets, this system is built up of two identical parts in parallel. For the sake of simplicity the taps are placed on one part and the short-circuiting bridge on the other. K_1 and K_2 cathode supply, *E* lead for the excitation voltage with carbon filament lamps *R* serving as grid-damping resistance, *M* lead for the modulation voltage. Between the two valves may be seen the elements for the series tuning of control grid and screen grid circuit.

[6]) See for example the article referred to in footnote [4]); also Philips Techn. Rev. **6**, 129, 1941, footnote [10]).

at a generous height above the earthed base plate by means of long glass columns, these capacities could be reduced to the values given in fig. 10 and in this way the total circuit capacity to the value of 32 pF already mentioned.

The self-induction of about 0.4 μH required for the tuning to the desired carrier-wave frequency (43.2 Mc/sec) is realized by a short Lecher system connected with the anodes and tuned by means of a short-circuiting bridge [7]). Because of the fact that the distance apart of the Lecher wires must be sufficiently small, the two valves could not be placed any farther apart. The Lecher wires are in the form of tubes which serve at the same time for the inlet and outlet tubes for the cooling water for the anode jackets.

The load on the output stage is formed by the impedance of the transmitting aerial, which is coupled with the anode circuit *via* a supply line. Since the supply line is terminated at the aerial end by its wave resistance in order to avoid reflections, considered from the anode circuit end, it also has an impedance equal to the wave resistance, which amounts to 425 ohms. Since, however, according to the above, the circuit may be loaded with 1000 ohms, the supply line is connected by a "tap" on the Lecher system. In order to prevent the high DC voltage of the anode from reaching the aerial, two mica coupling condensers are inserted in the supply line. These details are represented diagrammatically in *fig. 12*; the practical construction is shown in fig. 11.

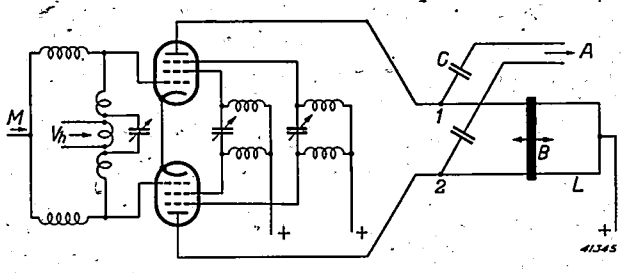


Fig. 12. Diagram of the practical construction of the output stage. At M the modulation voltage is applied, at V_h the excitation voltage. In all three grid circuits series tuning is introduced. L Lecher system, B short-circuiting bridge, C coupling condensers in the supply line to the aerial A.

Further particulars of the connections are the following. The two screen grids of the valves in push-pull connection are not connected directly with each other, but between them a tuning arrangement in series is introduced (see fig. 12). This was necessary since otherwise, due to the impedance of the external and internal connection wires, which cannot be ignored at these short wavelengths, high-frequency voltage still continued to act on the screen grids, so that they would not have fulfilled their purpose. With the control grid circuit the case was nearly the same. The internal capacities of this circuit are so large that even a direct connection of the grid terminals still possesses too much self-induction to give the desired high resonance frequency. Here also tuning in series was applied, and in this case the circuit capacity was so reduced by a series condenser that the self-induction necessary for the tuning took on a reasonably large value, so that a coil could be used, which at the same time makes possible a satisfactory coupling with the preceding stage.

The filaments are fed with AC voltage from the mains. In order to limit the hum percentage in the output signal, however, for the feeding of the filaments of the two valves in push-pull connection two voltages with a relative phase shift of 90° were used. In fig. 11 the two pairs of supply lines for this cathode feed may be seen to the left and right of the valves.

Dimensions and connections of the modulator

As we have already seen above, the largest carrier-wave amplitude is obtained by making the average control grid voltage equal to the overbiasing voltage of the valves (point A in fig. 8), while for the synchronization impulses the carrier wave must be entirely suppressed by making the average control grid voltage sufficiently negative (point B in fig. 8). From this it follows that the peak value of the modulation voltage must be equal to the amplitude of the excitation voltage, for which we found a value of 1400 V.

In order to give the modulator the necessary band width (from 25 to $f_{max} = 2.5 \cdot 10^6$ c/sec) attention must again be devoted to a relation similar to eq. (1), *i. e.* the loading impedance R must be smaller than $^1/_{2\pi f_{max} C}$ [8]). The modulator is loaded by the grid circuit of the output stage, which circuit contains in parallel the grid-cathode capacities of the two valves in push-pull connection. In our case this is a total capacity or 116 pF. The anode impedance of the modulator may not therefore amount to more than $R = 10^{12}/116 \cdot 2\pi \cdot 2.5 \cdot 10^6 = 550$ ohms. Since with the prescribed voltage the power is inversely proportional to R, this makes it necessary to have a large video-frequency

[7]) The fact that a Lecher system whose length is less than $^1/_4$ wavelength behaves in the main as a self-induction is explained in C. G. A. von Lindern and G. de Vries, Lecher systems, Philips Techn. Rev. **6**, 240, 1941.

[8]) See for example the discussion of a broad-band amplifier in J. D. Veegens, A cathode-ray oscillograph, Philips Techn. Rev. **4**, 198, 1939.

powe (more than 1500 W). This is obtained with two valves type PB 3/800 connected in parallel.

Since the video-frequency signal which comes from the studio amounts to only 1 volt across the wave resistance of 1000 ohms with which the connecting cable is terminated at the transmitter end, a number of amplifier stages must precede the modulator. In order to do with as few stages as possible, valves with the steepest possible slope were chosen, while in several stages two or even three valves were connected in parallel. The advantage in steepness gained by this means is of course partially compensated by the fact that the input and output capacities become correspondingly greater, so that the loading resistance of the stages in question had again to be reduced.

We shall not go into detail here about the further particulars of the pre-amplifier, such as measures for suppressing tendencies to generate, for improving the phase reliability at the low frequencies, for the suppression of interferences by mains voltage variations etc. One important difference, however, between the video-frequency modulator and the audio-frequency modulator in broadcasting transmitters will be discussed in somewhat more detail.

In contrast to the AC voltages which are used to transmit speech or music, the video-frequency AC voltages which are delivered by the iconoscope have an average value which differs essentially from zero. The magnitude of this average value (DC voltage component) is in fact a measure of the average brightness of the picture. Actually, therefore, the video-frequency amplifiers in the studio as well as in the transmitter should be constructed as DC amplifiers. There are, however, many objections to this in connection with the broad spectrum and the high power finally necessary. Therefore in the case of television transmitters, the DC voltage component is usually filtered out in the studio amplifier, so that all the remaining amplifier stages can be constructed as AC amplifiers, and only at the input of the last amplifier stage (*i.e.* the modulator stage proper) is the DC voltage component applied again. The modulator stage itself is therefore constructed as DC amplifier and connected with the output stage of the transmitter in the manner shown in *fig. 13*. In this way all the sources of supply (also those of the filaments) of the transmitter output stage are at the positive anode voltage (about 1500 V) of the modulator with respect to earth, which necessarily leads to a somewhat unusual construction of the supply installation.

As regards the final application of the DC voltage component in the modulator stage, that component can be transmitted through a separate cable from the studio to the transmitter, as is indeed done in certain transmitters. In the case of the transmitter in Eindhoven, however, a much simpler but nevertheless very

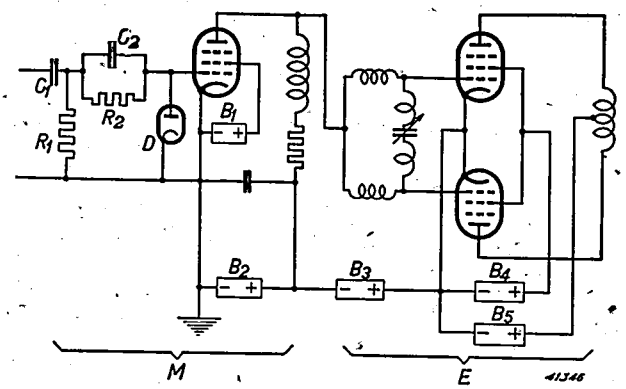


Fig. 13. The modulator stage (M) is constructed as a DC amplifier in order to be able to apply the DC voltage component of the video-frequency signal to the grid of the transmitter output stage (E). As a result all the sources of supply of the output stage are at the positive anode voltage of the modulator with respect to earth.
B_1 screen-grid supply, modulator, B_2 anode supply, modulator, B_3 source of grid bias, output stage, B_4 screen grid supply, output stage, B_5 anode supply, output stage.

satisfactory method is applied. In *fig. 14*, the variation of the grid voltage v_g of the modulator stage, as it is desired is represented for a completely white, a completely black and any given picture. The differences in the DC voltage component with the different types of picture may be seen. If we apply this signal w i t h o u t the DC voltage component to the modulator stage with the help of the normal grid condenser, while the modulator stage is given a fixed control grid bias v_{g_0}, the grid voltage for a completely white and a completely black picture would have the form drawn in *fig. 15*. Instead of this, however, we can also obtain the desired form according to fig. 14 if instead of giving the modulator a f i x e d grid bias, we give it a bias which can be derived from the incoming signal itself with the help of the grid condenser C_1 and a leakage

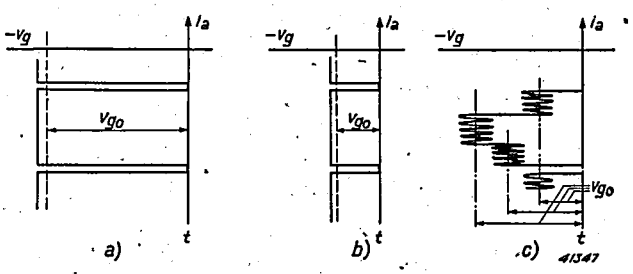


Fig. 14. The desired form of the signal on the grid of the modulator stage, *a*) for a completely white picture, *b*) for a completely black picture, *c*) for any given picture. The differences in the DC voltage component (v_{g0}) may be seen, and in case (*c*) they may also vary within a picture. (The line synchronization signals are again omitted.)

resistance R_1. The bias then adjusts itself automatically so that the voltage peaks reach just to the region of the grid currents ($v_{g0} \approx 0$). Care

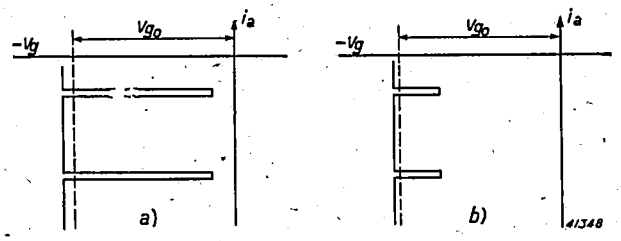


Fig. 15. Signal on the grid of the modulator stage, as it would be obtained upon suppression of the DC voltage ocmponent and with a fixed negative grid bias *a*) for a completely white, *b*) for a completely black, picture.

must be taken however, that the bias thus obtained can be altered rapidly enough to follow any variations of the average brightness in the picture. "Rapidly enough" means that the picture is divided, for instance, into four strips parallel to the direction of the lines and the average brightness is correctly reproduced for each of these strips. Now the possible velocity of variation of the bias is higher, the smaller the product R_1C_1. This product may not, however, become too small, since otherwise the lowest modulation frequencies are not faithfully transmitted, for a lowest modulation frequency of 25 c/sec R_1C_1 must be equal to 0.4, while in order to follow the fluctuations of the average brightness a value at least 20 times as small is required. This difficulty could be solved by introducing a second circuit after C_1R_1 in the manner indicated in fig. 13, with a small C_2 and a large R_2 ($R_2 = 100\ R_1$) of which the product R_2C_2 amounts to only about 0.01. This circuit has practically no effect on the transmission of the video-frequencies; the grid bias, however, can now be considered to consist of the practically constant contribution of the average voltage over R_1 and the contribution of the sufficiently rapidly varying average voltage over R_2. The latter assumes the correct value at every moment due to the fact that at the voltage peaks grid current flows for a moment. Since the adjustment will be more exact, according as the current in the grid circuit grows more rapidly upon the approach of the grid voltage to zero (it already begins to flow at a certain negative grid voltage), a small diode is connected between grid and cathode as indicated in fig. 13, which diode begins to pass current as soon as its anode voltage has nearly become positive.

THE INFLUENCE OF LOSSES ON THE PROPERTIES OF ELECTRICAL NETWORKS

by J. F. SCHOUTEN and J. W. KLÜTE.

621.318.7

Two methods are discussed for determining the influence of losses on the properties of electrical networks such as filters. Especially the method of stretching membranes between certain space curves furnishes a very graphic picture of the attenuation and the variation of the phase rotation of a network as a function of the frequency and the measure of loss.

In calculating electrical networks, and especially electrical filters, it is usually assumed that the self-inductions and capacities used in them are ideal. By ideal is meant that the impedances or admittances of these elements are purely imaginary, so that no losses occur. For these ideal elements the properties of a filter can then be calculated according to known methods [1]).

In practice, however, we are always concerned with elements which are subject to losses, which can be represented for instance by a resistance in series with the ideal coil or a resistance in parallel with the ideal condenser. The properties of a filter which is built up of these "actual" elements can immediately be expressed in mathematical form [2]). The numerical interpretation of this form is ,however, very elaborate and time-consuming.

In this article we shall discuss two methods by which a general and in part very graphic impression is given of the changes undergone by the properties of a filter calculated for i d e a l elements when the influence of the losses is taken into account [3]).

The method of the perpendicular derivative

The transition from the ideal to the actual self-induction amounts to the fact that the impedance Z, instead of being purely imaginary, becomes complex (see *fig. 1*), which may be written as follows:

$$Z = j\omega L \rightarrow R + j\omega L, \quad \ldots\ldots \quad (1a)$$

where R represents the series resistance of the coil.

In the same way, for the admittance Y of the condenser:

$$Y = j\omega C \rightarrow G + j\omega C, \quad \ldots\ldots \quad (1b)$$

where G represents the admittance of the parallel resistance of the condenser. In general the quantities L, R, C and G will depend more or less upon the frequency, but we shall not go into that here.

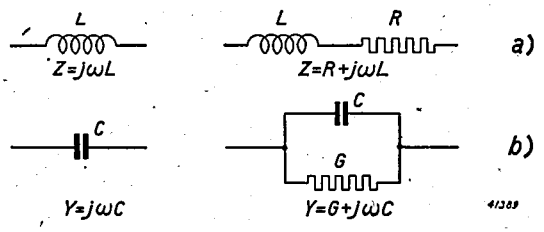


Fig. 1. Equivalent diagram of the ideal (left) and actual (right) elements.
a) The actual self-induction may be considered as an ideal self-induction L in series with a real impedance R.
b) The actual capacity may be considered as an ideal capacity C in parallel with a real admittance G.

We now write formulae (1a) and (1b) in the following form:

$$\left.\begin{aligned} Z &= R + j\omega L = \left(\frac{R}{L} + j\omega\right) L = (\varrho + j\omega)L, \\ Y &= G + j\omega C = \left(\frac{G}{C} + j\omega\right) C = (\gamma + j\omega)C. \end{aligned}\right\} \quad (2)$$

In other words, the transition from the ideal to the actual coil, retaining the self-induction L, amounts only to the substitution of $\varrho + j\omega$ for the quantity $j\omega$. In the former $\varrho = R/L$ represents t h e m e a s u r e o f l o s s o f t h e c o i l. In the same way the transition from the ideal to the actual condenser, retaining the capacity C, amounts only to the substitution of the quantity $\gamma + j\omega$ for $j\omega$. Now $\gamma = G/C$ represents the m e a s u r e o f l o s s o f t h e c o n d e n s e r. Let us for the present assume that the measure of loss of the coils and condensers in the network is the same and equal to k. The following conclusion may then be

---

[1]) B a l t h. v a n d e r P o l and T h. J. W e i j e r s, Electrical filters, Philips Techn. Rev. **1**, 240, 270, 298 and 327, 1936. Referred to below as F.

[2]) Loc. cit. F, p. 302.

[3]) These considerations are generally valid not only for electrical networks, but also for other systems, mechanical for instance. Here, however, we shall confine ourselves in terminology and examples to electrical filters.

drawn. If the properties of an ideal network are known as functions of the frequency and of the values of the various self-inductions and capacities, the properties of the actual network follow from them directly by replacing $j\omega$ everywhere in the formulae by the complex quantity $k + j\omega$. Thus if a property of the ideal network is represented by an analytic function of the argument $j\omega$, which we shall call $f(j\omega)$, the same property for the actual network is represented by the same function of the complex argument $k + j\omega$, thus by $f(k + j\omega)$.

The function $f(k + j\omega)$ is in general complex. It may therefore be written as the sum of a real part α and an imaginary part β:

$$f(k + j\omega) = \alpha + j\beta.$$

For such a complex function the following is valid, and may easily be verified by differentiation with respect to k and ω, respectively:

$$\left.\begin{aligned} \frac{\delta\alpha}{\delta k} &= \frac{\delta\beta}{\delta\omega} \\ \frac{\delta\beta}{\delta k} &= -\frac{\delta\alpha}{\delta\omega} \end{aligned}\right\} \quad \ldots\ldots\ldots \quad (3)$$

These are the well-known differential equations of Cauchy-Riemann from the theory of the functions of a complex variable. These equations may now be interpreted as follows.

The properties of our network, expressed as the sum of a real part α and an imaginary part β, are known as a function of the frequency ω for the case when $k = 0$ (the ideal network). We wish, however, to know the properties for a value of the measure of loss k which differs somewhat from zero (the actual network). The equations (3) show that for a given frequency ω the required change of α with k is equal to the known change of β with ω for the same frequency. In the same way the required change of β with k is equal to the known change of α with ω for the same frequency but with the opposite sign. If we imagine two systems of rectilinear coordinates in which the quantities α and β, respectively, are plotted as functions of k and ω, relation (3) means that at corresponding points (k, ω) the derivative in the direction k on the one surface is equal (with the positive or negative sign) to the derivative in the direction ω perpendicular to the first on the other surface.

If we develop α and β at a certain frequency ω in a Taylor's series in powers of k and apply relations (3), we find as a first approximation [4]:

$$\left.\begin{aligned} \alpha(k) &= \alpha(0) + k\frac{\delta\alpha(0)}{\delta k} = \alpha(0) + k\frac{\delta\beta(0)}{\delta\omega}, \\ \beta(k) &= \beta(0) + k\frac{\delta\beta(0)}{\delta k} = \beta(0) - k\frac{\delta\alpha(0)}{\delta\omega}. \end{aligned}\right\} \quad (4)$$

With this, for not too large values of k, we can determine $\alpha(k)$ as well as $\beta(k)$ from the given dependence of $\alpha(0)$ and $\beta(0)$ on the frequency ω.

The choice of α and β

We have expressly left open the question of which function we wish to choose to characterize our network. We may therefore still make a free choice. In the case of an impedance we may, for example, take for α and β the real and the imaginary parts of the impedance itself, but equally well those of the logarithm of that impedance, or, if it were desirable, the real and the imaginary parts of the square or the e-power of the impedance. For all these cases relations (3) are valid.

It is customary to characterize electrical filters (quadripoles) as follows [5]). An input voltage $e^{j\omega t}$ gives rise to the following output voltage:

$$e^{-T}e^{j\omega t} = e^{-\alpha-j\beta+j\omega t} \quad \ldots\ldots \quad (5)$$

In this α represents the attenuation, *i.e.* the logarithm of the ratio of the absolute values of input and output voltage, while β indicates the phase rotation. For this α and β (whose nomenclature was anticipated in our choice of symbols), which represent the real and the imaginary part of the propagation constant T, equations (3) are thus valid.

Examples

We shall now explain relations (3) and (4) by means of several examples.

a) The low-pass filter of the basic type [6])

The ideal low-pass filter of the basic type (*fig. 2a*) has the following properties. The attenuation α is equal to zero below the limiting frequency ω_1. Below this limit therefore all frequencies are passed without attenuation (transmission region). Above the lim-

---

[4]) The relation (4) was first given by H. F. Mayer E.N.T. 2, 335, 1925. To E. A. Guillemin, Communication Networks II, p. 445 is due the remark that they are ultimately based upon the theory of complex functions.

[5]) Loc. cit. F, p. 270.

[6]) Loc. cit. F, p. 298.

iting frequency (attenuation region) the attenuation rises rapidly (fig. 2b) according to the following relation:

$$\alpha = 2 \cosh^{-1} \frac{\omega}{\omega_1}.$$

The phase β in the transmission region is given by

$$\beta = 2 \sin^{-1} \frac{\omega}{\omega_1}.$$

At $\omega = \omega_1$, β therefore attains the value π; this value is retained for the whole damping region (fig. 2c). From the given behaviour of β as a function of ω we may draw the following conclusions about the influence of the measure of loss k on the attenuation α:

1) In the damping region the attenuation α is in the first approximation independent of the measure of loss k, since $\delta\beta/\delta\omega = 0$.
2) In the low part of the transmission region ($\omega < 2/3\ \omega_1$) the attenuation α is independent of the frequency, directly proportional to the measure of loss k (eq. 4) and inversely proportional to the width ω_1 of the filter, since $\delta\beta/\delta\omega \approx 2/\omega_1$ (fig. 2c).
3) In the high part of the transmission region ($\omega < 2/3\ \omega_1$) the increase in the attenuation is greater than in the low part, considering the increase of $\delta\beta/\delta\omega$ with ω.

For the influence of the measure of loss k on the phase β we find in the same way:

1) In the transmission region the phase β is in a first approximation independent of the measure of loss k, since $\delta\alpha/\delta\omega = 0$.
2) In the damping region the phase β decreases, since $\delta\alpha/\delta\omega$ is positive and directly proportional to the measure of loss k. This decrease becomes smaller with increasing frequency, since $\delta\alpha/\delta\omega$ decreases.

Before passing on to a numerical example it must be pointed out that our original condition, namely that the measures of loss of coils and condensers are equal, is practically never fulfilled in practice. On the contrary the losses of the condensers are usually many times as small as those of the coils. Our formulae (3) and (4) then, nevertheless, remain valid in the first approximation provided the average of ϱ and γ is taken for k, thus:

$$k = \frac{\varrho + \gamma}{2}, \text{ so that } k = \frac{\varrho}{2}, \text{ when } \varrho \gg \gamma.$$

The influence of the inequality of ϱ and γ, which is ignored here, is manifested mainly in a slight shift of the frequency.

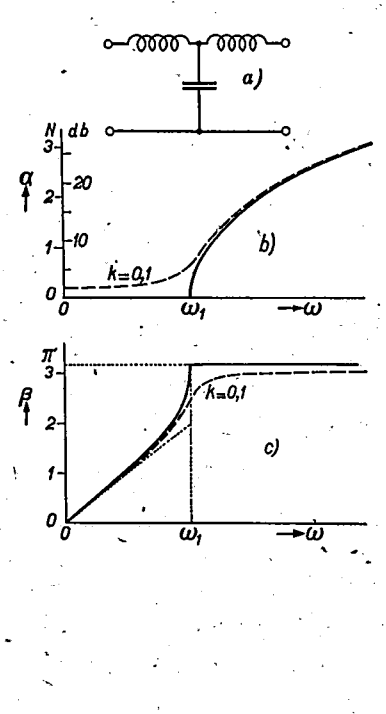


Fig. 2.

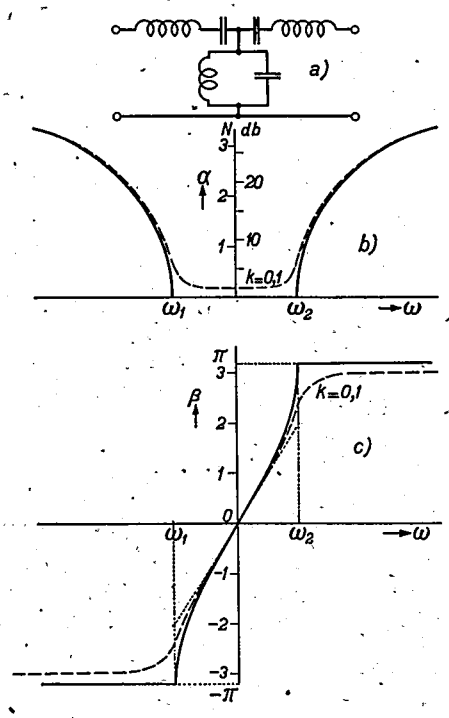


Fig. 3.

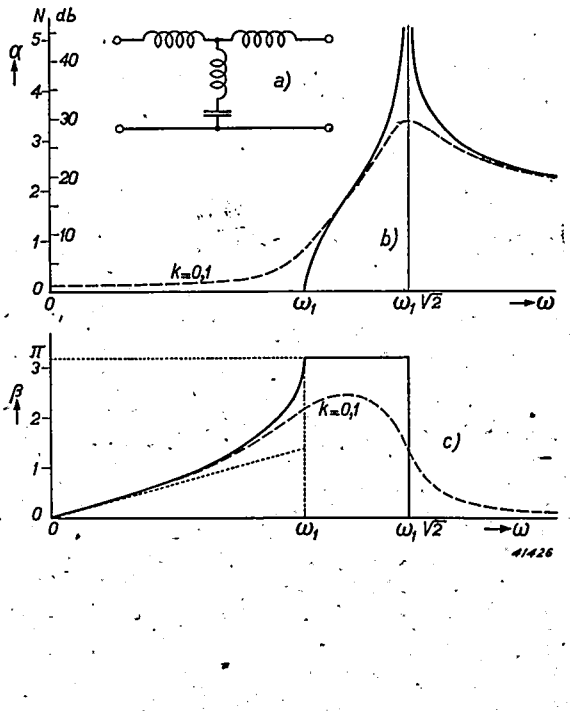


Fig. 4.

Figs. 2, 3 and 4. Low-pass and band-pass filter of the basic type and low-pass filter of the transformation type ($\omega_\infty/\omega_1 = \sqrt{2}$).
a) T-cells of the filter types in question.
b) and *c*) Behaviour of the attenuation α and phase β, respectively, as functions of the frequency for the ideal case when $k = 0$ (full-drawn) and for $k = 0.1\ \omega_1$ (broken line).

We now take as example a low-pass filter consisting of a number of cells. We assume a cut-off frequency of $\nu_1 = 400$ c/sec, thus $\omega_1 = 25\,000$, a measure of loss of the coils of $\varrho = R/L = 125$ ohms per henry and a negligable measure of loss of the condensers. In the low part of the transmission region, independent of the frequency, the damping becomes

$$a = k \frac{\delta\beta}{\delta\omega} = \frac{\varrho}{2} \cdot \frac{2}{\omega_1} = \frac{1}{200} \text{ neper/cell.}$$

The fact that the attenuation here is found in neper per cell is a result of the fact that we began (see (5)) with the natural logarithm. If we wish to express the damping in decibels, since 1 neper corresponds to 8.7 decibels, we find that $a = 0.043$ db/cell. The same formulae are valid for a loaded cable consisting of a number of sections, which may of course also be considered as a low-pass filter.

b) The band-pass filter of the basic type [7])

For this filter practically the same considerations hold as for the low-pass filter. For the damping in the middle section of the transmission region we now find (*fig. 3*):

$$a = 8.7\, k \frac{\delta\beta}{\delta\omega} = 8.7 \frac{4k}{\omega_2 - \omega_1} \text{ db/cell.} \quad (6)$$

From this we draw the conclusion that the attenuation in the transmission region of the band-pass filter is independent of the mean frequency of the transmitting band and that it is determined only by the measure of loss k, to which the attenuation is directly proportional, and by the w i d t h $\omega_2 - \omega_1$ of the band, to which it is inversely proportional.

In connection with this it may be pointed out that the losses of a coil are often expressed not in the terms of the measure of loss ϱ but in those of the quality Q [8]), which is related as follows to the measure of loss:

$$Q = \frac{\omega L}{R} = \frac{\omega}{\varrho}, \text{ or } \varrho = \frac{\omega}{Q}.$$

If one now considers the case often occuring in carrier-wave telephony, namely a number of band-pass filters at different points but with the same band width, these filters undergo the same change in a and β when the values of ϱ and n o t those of Q are the same for the different transmission bands. Considered from the point of view of filter theory, therefore, it is better to use ϱ and not Q to characterize the losses. For the same reason we prefer the measure of loss γ for condensers over the customary tangent [8]) of the phase displacement δ, which is related to γ as follows:

$$\tan\delta = \frac{\omega C}{G} = \frac{\omega}{\gamma}, \text{ o r } \gamma = \frac{\omega}{\tan\delta}.$$

) Loc. cit. F, p. 327.

c) The low-pass filter of the transformation type [9])

Compared with the low-pass filter of the basic type, this filter has the peculiarity that the attenuation already becomes infinite for a finite frequency ω_∞ (in *fig. 4* $\omega_\infty = \omega_1 \sqrt{2}$) and above this frequency gradually falls to a certain limit. It is used when a very steep rise in the damping is required immediately above the cut-off frequency ω_1. In the phase β the frequency of infinite attenuation is manifested in that the phase jumps back abruptly from π to zero. This means, therefore, that in the whole damping region, as for the filter of the basic type, $\delta\beta/\delta\omega$ is zero except at ω_∞, where this quantity becomes negatively infinite. This makes it understandable that here also in the damping region the damping a is in the first approximation everywhere independent of the measure of loss k except exactly at point ω_∞, where the smallest loss already causes the damping to decrease to a finite value.

In the transmission region the phase does indeed vary from zero to π, but more slowly at first, and consequently more rapidly later, than is the case with the basic type. This means that in the low part of the transmission region this type of low-pass filter is l e s s sensitive and in the high part m o r e sensitive to losses than the basic type (fig. 4*b*). This is even truer the closer the point of infinite damping ω_∞ lies to the cut-off frequency ω_1. This influence of the point of infinite damping on the behaviour of the filter in the transmission region must here be accepted as given. In the following, however, on the treatment of the problem with the help of membrane models we shall reach a closer insight into this point.

With the influence of the losses on the phase β we must also note that for frequencies above ω_∞ $\delta a/\delta\omega$ is negative. In this region therefore β must i n c r e a s e with increasing measure of loss k.

These considerations illustrate how with the help of the method of the perpendicular derivative a fairly complete concept may be formed of the behaviour of electrical systems which are subject to slight losses, when the behaviour of the system free of losses is known. We shall now derive another relation between the quantities a, β, k and ω which will lead us to a direct graphic representation of a and β.

[8]) Instead of the quantities Q and $\tan\delta$ the quantities d_L and d_C (loc. cit. F. p. 302), respectively, are also used. These latter quantities are related as follows:

$$d_L = 1/Q \text{ and } d_C = \tan\delta.$$

[9]) Loc. cit. F. Table opposite p. 332.

The method of the membrane model

By differentiation of equations (3) with respect to k and ω we can derive the following relations :

$$\left.\begin{aligned} \frac{\delta^2\alpha}{\delta k^2} + \frac{\delta^2\alpha}{\delta\omega^2} &= 0, \\ \frac{\delta^2\beta}{\delta k^2} + \frac{\delta^2\beta}{\delta\omega^2} &= 0. \end{aligned}\right\} \quad \ldots\ldots (7)$$

These are Laplace differential equations, also well known from the theory of the functions of a complex variable.

These equations do not, like equations (3), represent a relation between α and β, but a condition which not only α but also β, each as a function of k and ω, must fulfil. In spatial representation therefore they represent conditions which are applied to the surfaces $\alpha(k,\omega)$ and $\beta(k,\omega)$. The Laplace equation exhibits great similarity with the differential equation which determines the shape of a soap film, that is to say of the so-called minimum surface. The name minimum surface comes from the fact that such a surface is the smallest which can be laid through a given contour. For very small slopes the differential equation for the minimum surface passes over into that of Laplace. Use has already been made of this in this laboratory, namely for the construction of a model of a two-dimensional potential field by means of a rubber membrane [10]). The behaviour of the rubber membrane is slightly different from that of the soap film, since the tension at every point of the membrane, unlike that of the soap film, is not the same, but depends upon the stretch, which differs from point to point. For small slopes, however, it also satisfies the Laplace equations.

For the surfaces α and β about to be considered the condition that the slopes must be small is by no means satisfied. We shall, however, for the present disregard the inaccuracies thereby caused, and we may then, although roughly, reproduce the form of the surfaces $\alpha(k,\omega)$ and $\beta(k,\omega)$ by representing them as a stretched soap film or rubber membrane.

Just as the solution of the Laplace equation in the interior of a closed contour is fully determined by the values of the function at the edge, in the same way the form of a soap film or of a rubber membrane is completely fixed by its contour. In the practical execution therefore we shall try to determine α and β along a contour, then we shall make a model of it and over it we shall stretch a soap film or a rubber membrane.

[10]) P. H. J. A. Kleynen, Philips Techn. Rev. **2**, 338, 1937.

Examples

The method of the membrane model will now be explained with the aid of several examples.

a) The low-pass filter of the basic type

We have already stated that for the ideal filter (thus in our spatial representation in the coordinates α, k and ω for the plane $k = 0$) the following is valid:

$$\alpha = 2\cosh^{-1}\frac{\omega}{\omega_1} \quad \ldots\ldots (8)$$

Similarly we find for the plane $\omega = 0$

$$\alpha = 2\sinh^{-1}\frac{k}{\omega,} \quad \ldots\ldots (9)$$

and for a plane of constant value a

$$\frac{\omega^2}{\omega_1^2\cosh^2\frac{\alpha}{2}} + \frac{k^2}{\omega_1^2\sinh^2\frac{\alpha}{2}} = 1. \quad \ldots (10)$$

In *fig. 5* a model is shown of the surface $\alpha(k,\omega)$ where the rubber membrane is stretched over the given contours (8), (9) and (10). Here we again encounter the complete behaviour which has already been described in the method of the perpendicular derivative. In the

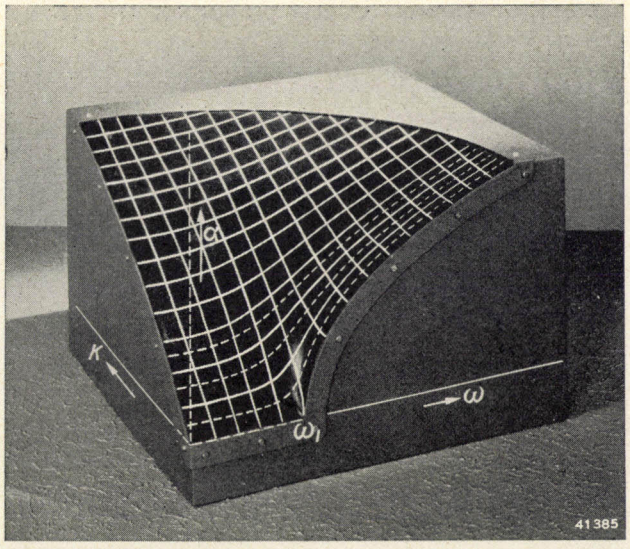


Fig. 5. Rubber model of the damping α (k,ω) for the low-pass filter of the basic type. It should be noted how the attenuation region on the axis $k = 0$ draws the membrane upward for the high part of the transmission region and for $k \neq 0$.

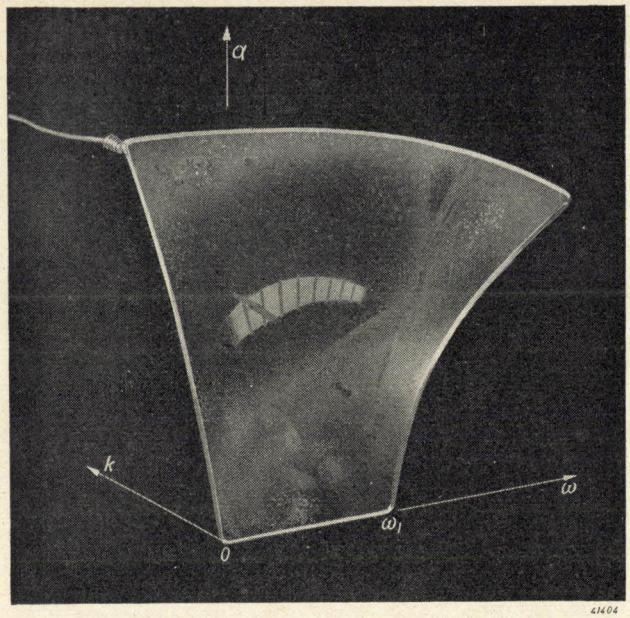
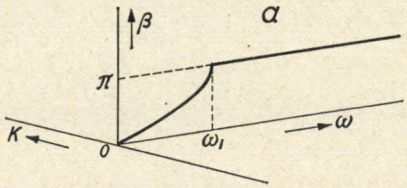


Fig. 6. Soap-film model of the damping α (k, ω) for the low-pass filter of the basic type.

model the values for constant k and constant ω are indicated by white lines. It may clearly be seen how in the low part of the transmission region α increases linearly with k. As, however, we approach the cut-off frequency ω_1, the neighbouring attenuation region already begins to draw the membrane more upwards, so that at $\omega = \omega_1$ the derivative $\delta\alpha/\delta k$ even becomes infinite. Once in the damping region the derivative immediately falls to zero, the membrane is here exactly perpendicular to the plane $k = 0$. In *fig. 6* the same surface is represented in the form of a soap film.

We may also imagine the surface to be continued for negative values of α, k and ω. Lines of constant α are then according to (10) given



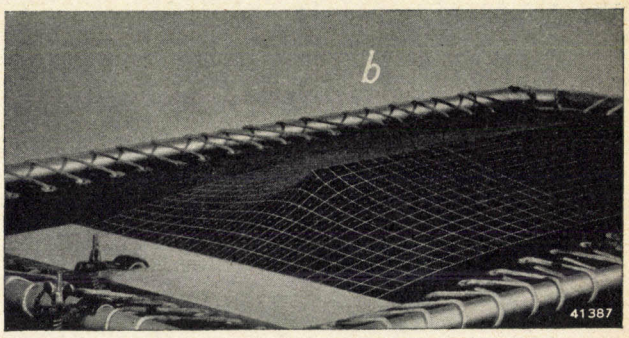


Fig. 8. Rubber model of the phase β (k, ω) for the low-pass filter of the basic type for the values of β between 0 and π.
a) Diagram of the surface β (k, ω).
b) The rubber model of the phase β.
The membrane is fixed by three conditions: the value $\beta = 0$ along the periphery, the value $\beta = 0$ along the axis $\omega = 0$ (fixed by a straight strip above the surface) and the value $\beta = \pi$ along the axis $k = 0$ for $\omega > \omega_1$ (fixed by a straight strip under the surface). The variation according to an arc sine in the transmission region must then be automatically attained.

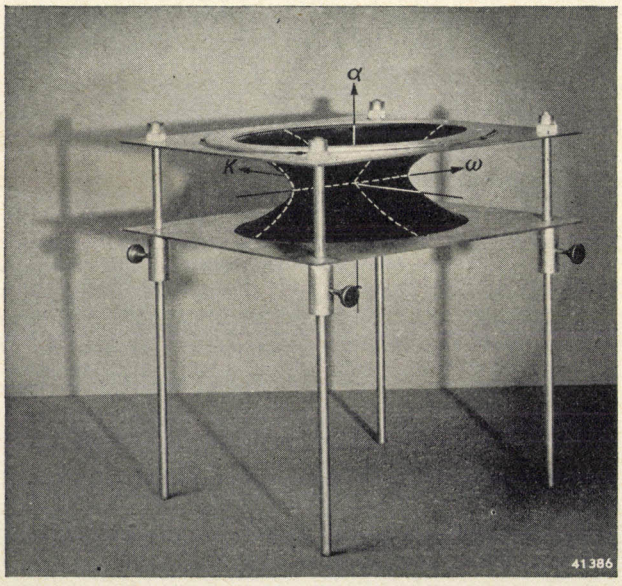


Fig. 7. Rubber model of the complete surface α (k, ω) for the low-pass filter of the basic type (extended to negative valvues of α, k and ω). The model consists of two rubber membranes stretched over two ellipses and fastened to each other along the line which connects the two foci of each ellipse. In the plane $k = 0$ and $\omega = 0$ the surface must automatically assume the form of a hyperbolic cosine and hyperbolic sine respectively. This condition is not completely satisfied by the rubber model, but with a certain approximation.

by a series of confocal ellipses, *i.e.* ellipses with the common foci $-\omega_1$ and $+\omega_1$. We now cover two similar ellipses with a rubber membrane and place the ellipses vertically above each other in a horizontal position (*fig. 7*). If we now fasten the two membranes to each other along the line which joins the foci of each ellipse we shall obtain the desired surface. The variation according to a hyperbolic cosine in the plane $k = 0$ and according to a hyperbolic sine in the plane $\omega = 0$ must then automatically be attained.

For the phase β we can construct a similar model. We may begin by determining that for the frequency 0 the phase must always be zero for each k, since the phase β is always an antisymmetric function of ω (*i.e.* a function which for negative values of the argument takes on a value equal and opposite to that for positive values of the argument). Furthermore it can be proved that, for a large value of k, β approaches

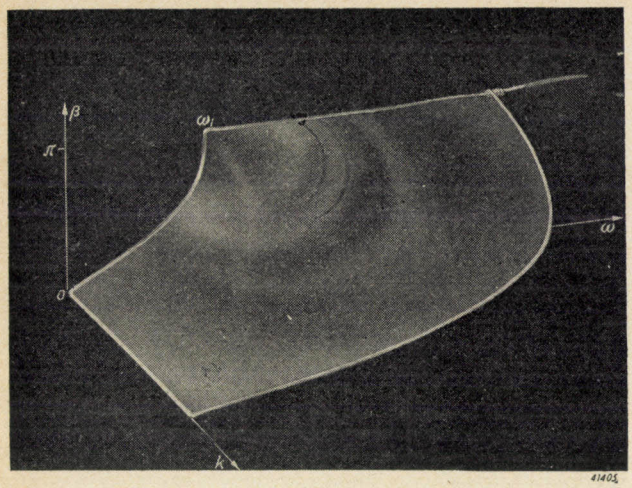


Fig. 9. Soap-film model of the phase β (k, ω) for the low-pass filter of the basic type. The coordinate k is here chosen in the opposite direction for the sake of clearness.

zero independent of ω. Sufficient data are thus available to fix completely the surface β. For this purpose we take a rubber membrane (*fig. 8*) whose fixed periphery approximately represents the phase zero. If now by means of a vertical straight strip u n d e r the membrane we push up the latter along the ω-axis above the cut-off frequency ω_1 to a certain height which represents the height π, and by means of a second strip a b o v e the membrane we keep it down along the k-axis at the height zero, the surface β is entirely fixed. The rectangular network of white lines drawn on the membrane now represents the set of lines of constant k and constant ω respectively. The variation according to an arc sine on the line $k = 0$ for the region between the origin and the cut-off frequency must be automatically assumed.

In *fig. 9* the same function is represented as a soap-film model, but only for positive values of k.

b) The band-pass filter of the basic type

The band-pass filter of the basic type, when the width of the transmission region is small compared with the average frequency of the transmission region, is entirely analogous to the low-pass filter of the same type. It may be considered as a low-pass filter which is reproduced not only for positive but also for negative values of the frequency, and which is shifted in such a way toward high frequencies that the frequency zero is displaced to the mean frequency ω_0 of the transmission region. The limits $-\omega_1$ and $+\omega_1$ of the low-pass filter are then replaced by the limits ω_1 and ω_2 of the band-pass filter.

In *figs. 10* and *11* we show drawings of models of the surfaces α and β. In *fig. 12* the surface α is shown as a soap film.

c) The low and band-pass filter of the transformation type

There is the same relation between these

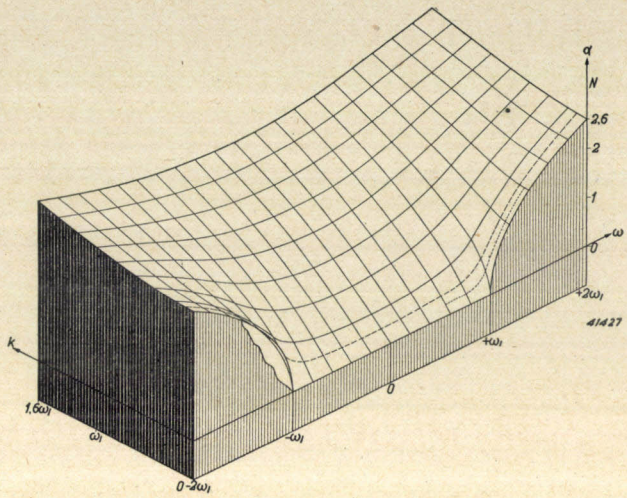


Fig. 10.

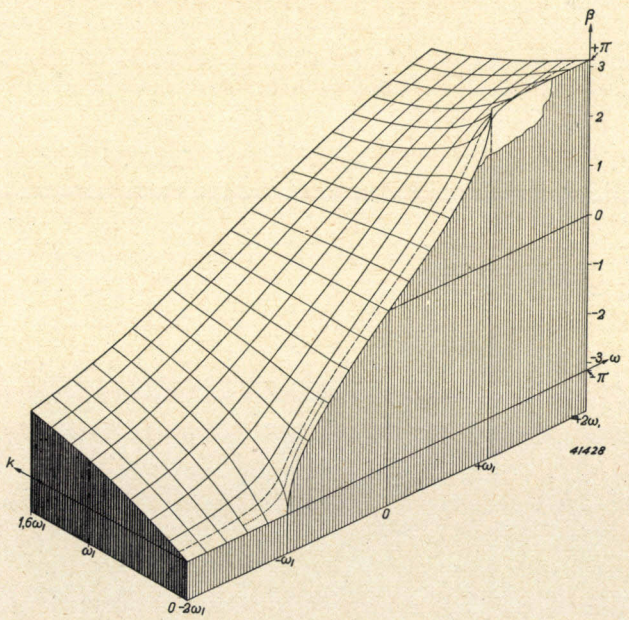


Fig. 11.

Figs. 10 and 11. Drawn models of the attenuation α (k, ω) and the phase β (k, ω) of the low-pass filter of the basic type [11]). By approximation the models also hold for a narrow band-pass filter of the basic type when the mean frequency ω_0 of the transmission region is added to each frequency. The values for $k = 0.1\ \omega_1$ are indicated by a dash line, those for $k = 0.05\ \omega_1$ by a dotted line.

[11]) The models are exact representations of the real and the imaginary part of the function

$$T = \alpha + j\beta = 2 \sinh^{-1}\left(\frac{k + j\omega}{\omega_1}\right).$$

two filters as between the corresponding filters of the basic type. In *figs. 13* and *14* drawings of models of the surfaces α and β are reproduced. In *fig. 15* the surface α is represented as a soap film. Our feeling for the behaviour of the soap film tells us immediately that for $k \neq 0$ the film draws the attenuation α in the transmission region upwards, but in the vicinity of the points of infinite attenuation it draws it downward. A fundamental difference from the band-pass filter of the basic type consists in the fact that for high values of k the attenuation α does not tend towards an infinitely high value, but towards a finite value independent of ω. These low values at the edge of the film will now have a tendency to cause the film to run upwards l e s s strongly in the transmission region than is the case with the basic type. In the vicinity of the cut-off frequencies, however, this influence is overcome by the nearness of the points of infinite damping, which draw the film m o r e s t r o n g l y upward than is the case with the basic type. This results finally in the smaller influence of the losses in the middle and the stronger influence at the edges of the transmission region of the filter. We have already pointed this out when

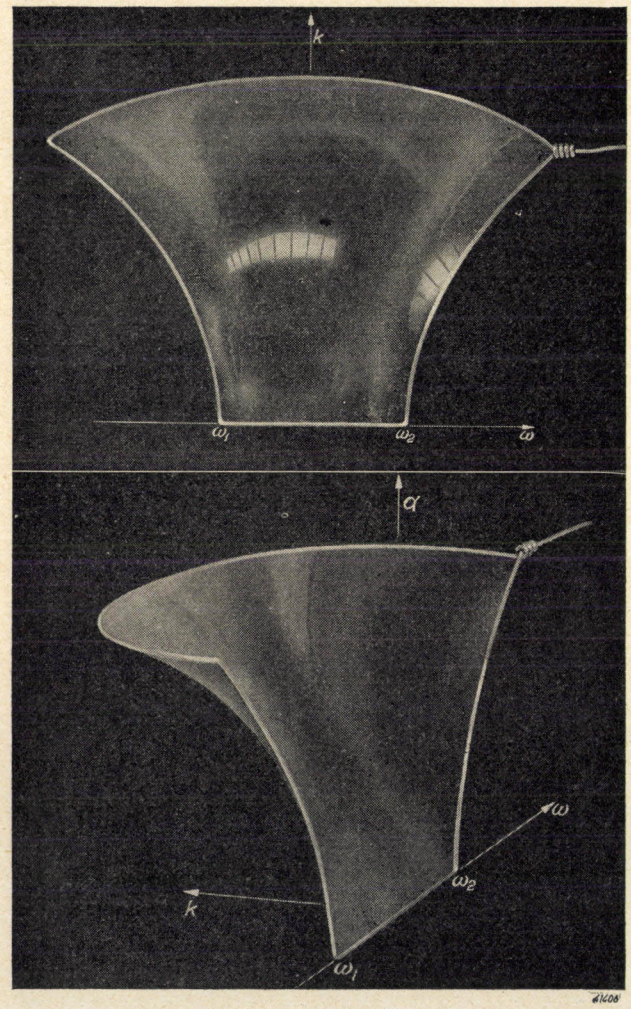


Fig. 12. Soap-film model of the damping α (k, ω) of the band-pass filter of the basic type.

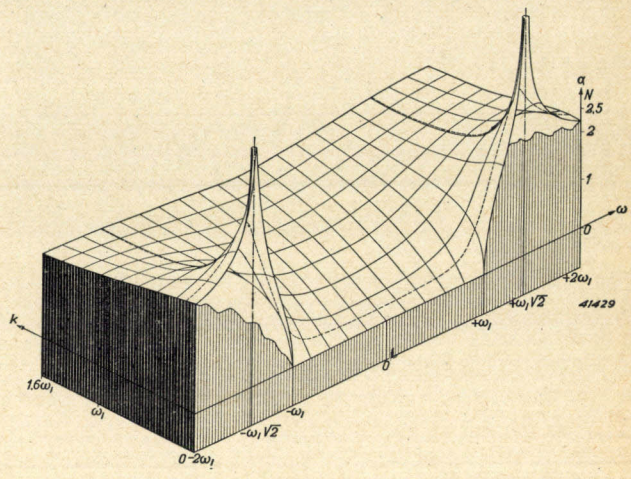


Fig. 13.

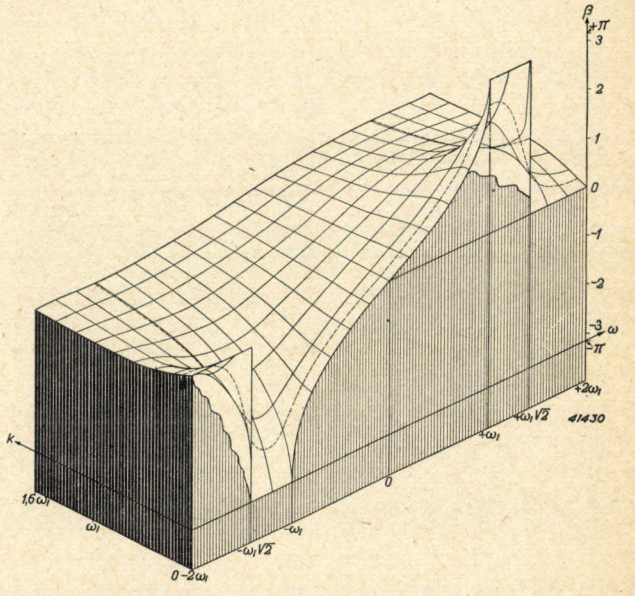


Fig. 14.

Figs. 13 and 14. Drawn models of the damping α (k, ω) and the phase β (k, ω) of the low-pass filter of the transformation type $(\omega_\infty = \omega_1\sqrt{2})$. The models also hold for the band-pass filter of the transformation type in the manner described under figs. 10 and 11.

discussing the low-pass filter of the transformation type according to the method of the perpendicular derivative. In *fig. 16* a soap-film model is shown of the phase of the low-pass filter, while in *fig. 17* a rubber model of the same filter is shown. In the latter model in addition the second part of the surface β is shown which extends from π to 2π. The model is constructed in a way analogous to that of fig. 7. The two membranes are fastened together along the line from ω_1 to ω_∞, and a vertical plate is then inserted between the two surfaces

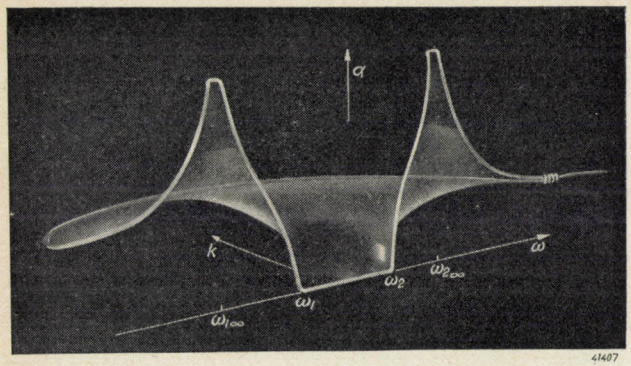


Fig. 15. Soap-film model of the attenuation α (k, ω) for the band-pass filter of the transformation type.

in order to keep the value of β on the ω-axis for frequencies above ω_∞ at zero for the lower and at 2π for the upper surface.

d) A compound band-pass filter for carrier-wave telephony

In carrier-wave telephony use is made of band-pass filters whose chief function is to

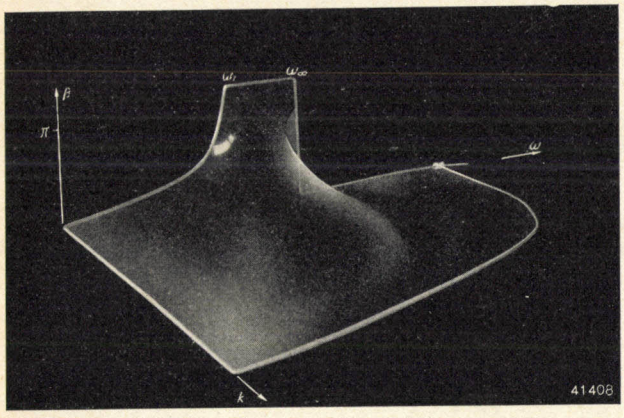
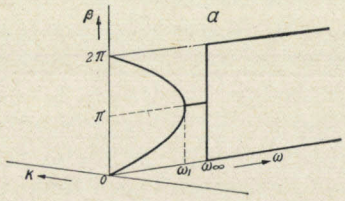


Fig. 16. Soap-film model of the phase β (k, ω) for the low-pass filter of the transformation type. The coordinate k, as in fig. 9, is here chosen in the forward direction.

suppress one of the two sidebands formed during modulation. For this purpose a damping characteristic is chosen which rises steeply on one side and less steeply on the other. For such an asymmetric filter [12]), which has recently been described in this periodical and which consists of three different cells, a soap-film model is shown in *fig. 18*. It is clear that the influence of the losses at the steep edge of the transmission region is greater than at the less steep edge.

The relation between α and β

Upon closer consideration of the results which were produced by the method of the perpendicular derivative and that of the film model, we encounter a remarkable situation. In the first method we deduced for instance that $\delta\alpha/\delta k$ is given by $\delta\beta/\delta\omega$. In the second method,



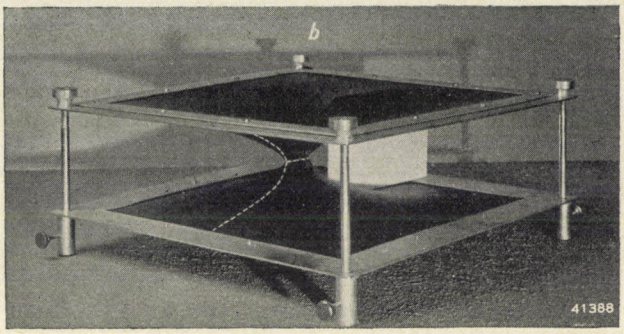


Fig. 17. Rubber model of the phase β of the band-pass filter of the transformation type for values of β between 0 and 2π.

a) Diagram of the surface β.

b) Rubber model of the phase β for the region between 0 and 2π.

The two sheets of the surface are fixed by three known conditions: the values zero and 2π, respectively, along the edges (to which the axis $k = 0$ also belongs), the value π on the ω-axis between ω_1 and ω_∞ (obtained by fastening the two sheets together) and the values zero and 2π, respectively, along the ω-axis for frequencies higher than ω_∞ (obtained by the vertical plate, which, in order to prevent undue damage to the membranes, is not inserted quite as far as the value ω_∞).

however, we gave α along a contour only and from it derived α unambiguously in the whole inner region, thus also $\delta\alpha/\delta k$ along the axis $k = 0$. Thus the given contour in fig. 5 furnished

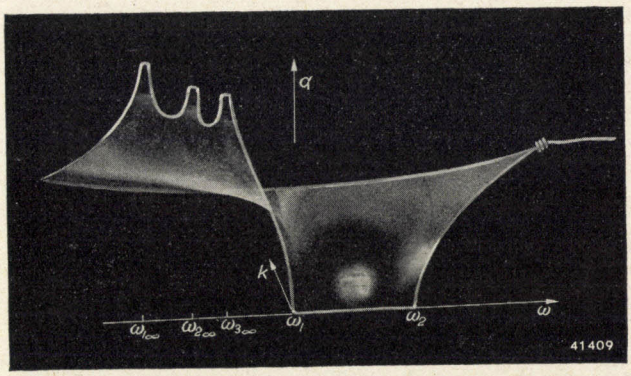


Fig. 18. Soap-film model of the attenuation α (k, ω) of a composite band-pass filter for carrier-wave telephony.

[12]) Th. J. Weijers, Philips Techn. Rev. **7**, 104, 1942.

us with the surface α for the low-pass filter without it being necessary to use our knowledge of the behaviour of β as a function of ω. That this is possible is due to the fact that α and β are not independent quantities. With a given damping as a function of the frequency, the variation of the phase cannot be chosen freely. The explanation of this relation, which we shall not go into here, ultimately comes down to the simple fact that whatever the properties of the filter it is always known with certainty that no output voltage can occur before an input voltage is applied. From this the relation between α and β may be derived.

GENERATORS FOR SHORT-WAVE THERAPY

by J. FRANSEN and J. M. LEDEBOER. 615.846

In the medical treatment with short waves (short-wave therapy) use is made of the fact that by means of high-frequency currents with frequencies above 10^7 c/sec a development of heat can be obtained at any desired spot in the body of the patient. The high-frequency currents are excited by means of short-wave transmitters which in some instances are simpler and in others more complicated than the transmitters used for radio and television. The requirements made of such transmitters and their construction are discussed here on the basis of a description of two generators for short-wave therapy developed by Philips. These generators work on a wavelength of 6 m and can deliver a maximum effective output of 400 and 150 W, respectively, to the patient.

Introduction

Around about 1800, when hardly anything was yet known about the great field of electrical phenomena, attempts were already being made to send electric currents through the human body for the purpose of producing a healing effect upon the sick. Direct current (galvanism) was originally used, later on alternating current of 20 to 200 c/sec (faradism). A specific effect of the electric current was indeed found, namely a prickling sensation which sometimes in the case of nervous disorders was found to have therapeutic effect. At the same time, however, electrolysis occurs in the tissue, and this was the main reason why the current could not be made stronger than 30 or 40 mA.

An increase of the current used only became possible when in 1890 d'Arsonval discovered that with alternating current of a high frequency (above about 100 000 c/sec) the prickle and the harmful electrolysis disappear. In this way currents of 1 A and more could be sent through the body. In this case another phenomenon comes to the fore, namely a considerable development of heat in the body.

Heat development of itself has long been an important aid in medical therapy: a local temperature increase is accompanied by a local increase of circulation and a more intense local metabolism, which especially in the case of certain inflammations, rheumatic disorders, etc. may have a favourable effect. In most methods of heat application, however, such as warm baths and packings, massage, irradiation with light or infra-red rays, it is only the outer layer of the body which is heated in the first instance; the subcutaneous layer of fat forms a sort of heat insulator which prevents a direct heat exchange of the more deeply lying parts with their surroundings. This is of course exactly the function of the fat layer in the human body, although it is to prevent the exchange in the opposite direction. The more deeply lying parts of the body can therefore only receive heat by transportation by way of the blood stream. It is clear that no considerable and certainly no well-localized heating of the organs inside the body can be obtained in this way.

Heating by electric current has the important advantage that the heat is developed all along the path of the current, thus also in the internal organs themselves, so that in this case heating effects of quite a different order of magnitude are possible. In the practical application it is then only a question of pro-

viding that the electric current follows the desired path through the body so that the desired parts are heated. While it was originally held that for this purpose it was only necessary to apply the electrodes which supply the current to the body at the correct spots, more recent knowledge has shown that this is not sufficient. The part of the body to be treated may in the first instance be regarded as consisting of a very large number of heterogeneous elements connected partly in series and partly in parallel, each of which has its own resistance and its own capacity. When a frequency of for instance 10^6 c/sec (wavelength 300 m) is used, as is the case in diathermy, the reactances of the capacities mentioned are still so small that they exert no influence on the distribution of the current through the parallel elements; this distribution then depends only upon the resistances, the current chosing the path of least resistance. Now certain components of the body, such as fat, bone, etc., have such a high specific resistance that they practically act as insulators. It is therefore impossible to send a current through parts which are surrounded by fat, although they lie exactly between the applied electrodes: the current does not pass through those parts but around them, *via* muscle tissue and blood stream. It is for this reason that it is impossible to heat the interior of a joint by diathermy.

The situation changes completely, however, when the frequency of the electric current is still further increased to 10^7 c/sec or more (wavelengths of less than 30 m). The reactances of the body elements then become of the same order of magnitude as their resistances and the current distribution is now also influenced to a decided degree by the dielectric constant ε of the various components. Since the value of the dielectric constants of the different components varies less than their conductivity — for most of the body components ε at such high frequencies is of the order of magnitude 80, corresponding to the value of ε for various aqueous solutions — we obtain a much more homogeneous current distribution than at lower frequencies, see *fig. 1*. Thus it is possible actually to send the current through the desired part of the body, and in this way to cause the development of heat at any desired spot.

This is the great significance of the so-called short-wave therapy which has come into practice since 1926. At the same time the very high frequencies used offered another advantage over the lower frequencies used in diathermy. With direct current and with alternating current of not too high frequencies the current must be applied by direct contact on the body, and to prevent a high transition resistance a good contact between the bare electrodes and the skin has to be provided. This makes the treatment difficult and in many

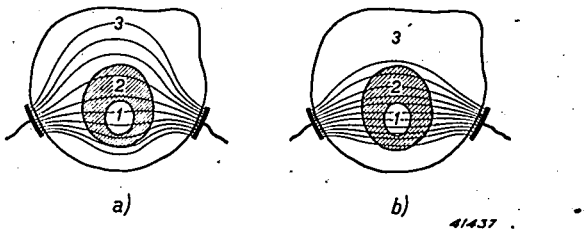


Fig. 1. Current distribution through a part of the body, *a*) at low frequency, *b*) at sufficiently high frequency. *1* is the organ to be heated, *2* an enveloping layer of fat, *3* muscle tissue. At low frequencies the fat has the character of an insulator.

cases, such as those of open skin wounds, impossible. At higher frequencies, however, the electrodes can without objection be placed at some distance from the skin. The current is then transmitted *via* the "coupling capacity" between electrode and skin, so that the equivalent circuit shown in *fig. 2* is obtained. Against the advantages of this treatment from a distance there is no other disadvantage than that for the same current through the patient a higher voltage must be applied to the electrodes. The difference is not very important, since when the distance between electrodes and skin is not too great the coupling capacities C_A are of the same order of magnitude as the capacity C formed by the part of the body being treated.

For a correct understanding it is necessary to point out that the historical development of short-wave therapy by no means followed the lines of the discussion given here. It may rather be assumed that the search for effects of the high-frequency alternating current other than those which result from the local development of heat was the main object in view. Thus, for example, it was thought to have been established that after the elapse of some time the operating personnel of powerful radio transmitters suffered from certain bodily phenomena (such as chronic headache) which were ascribed to the effect of the strong electromagnetic alternating fields. Leaving out the question whether or not these suspicions are based upon truth, it must be noted that while the situation of the patient in the "condenser field" in short-wave therapy is quite analogous to that in which one is situated in the vicinity of a transmitting

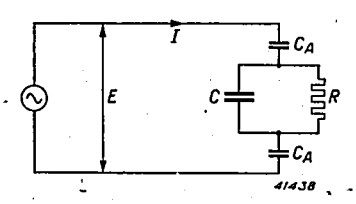


Fig. 2. Equivalent diagram of the circuit which includes the patient in treatment from a distance. The generator furnishes the high-frequency current I which, *via* the coupling capacities C_A existing between the electrodes and the skin, flows through the RC system of the body.

aerial, the specific effects, if they are present, do not play any significant rôle compared with the heat effect. The name "short-wave therapy" is in fact not a very fortunate one, since it emphasizes the wave character, while the fact that waves are also actually emitted by the patient acting as aerial and by the connecting wires is only a subsidiary phenomenon which is undesirable in every respect: it means on the one hand loss of energy and on the other hand it may lead to radio interferences in the neighbourhood. The term "high-frequency diathermy" would therefore be preferable.

The first experiments with short-wave therapy were performed with the help of the oscillations which occur upon the discharge of a condenser battery *via* a spark gap. Such spark generators are still in use at the present time. Short-wave therapy has, however, only been practised on a large scale since it became possible through the development of short-wave transmitting technology, particularly of transmitting valves, to construct entirely noiseless, reliable and easily operated valve generators which deliver a satisfactorily high output. Two such generators which were developed by Philips (types 11 951 and 11 952) will be described in the following.

Choice of the frequency

From the above it follows that the frequency of the generators will be chosen in the region above 10^7 c/sec (wavelength less than 30 m). As to the final choice of the particular frequency, it is more and more customary to work with a wavelength of 6 m. This choice may be explained as follows.

If we consider the diagram of the equivalent circuit in which the patient is included (fig. 2) and calculate the heat Q which is developed per second in the resistance R when an AC voltage of magnitude E and frequency ω is applied to the connections, we find that

$$Q = \frac{E^2}{R\left\{\left(1+\frac{2C}{C_A}\right)^2 + \frac{4}{\omega^2 R^2 C_A{}^2}\right\}} \quad \ldots\ldots (1)$$

It is immediately evident that Q becomes larger, the larger ω is chosen. In order to obtain the greatest possible heat delevopment in the human body, therefore, the frequency must be chosen as high as possible. This conclusion also remains valid when the simple connection in parallel of R and C in fig. 2 is replaced by a complicated circuit of many C's and R's in series and in parallel, which corresponds more to the actual case.

A practical limit is set to the increase of the frequency by two factors. In the first place with all transmitting valves the efficiency decreases when the frequency rises above a certain value. Whereas until a few years ago the limit at which one could work and retain a satisfactory efficiency and output lay at 10 to 30 m, at the present time with newer valves good efficiency can be attained at 6 m and even at 3 m [1]. It is even possible by the employment of magnetron generators [2] to work with wavelengths of 1 m or less. At such short waves, however, another difficulty occurs, namely the emission of radiation by the connections which lead from the generator to the patient. This radiation can only be restricted by constructing the two electrode cables in the form of a Lecher system, *i.e.* by keeping their distance apart small compared with the wavelength. Considering the very different ways in which the electrodes must be able to be placed, this condition can no longer be satisfied at wavelengths of for instance 3 m. For the same reason, when using wavelengths of less than 1 m, the method has been adopted of radiating the energy on to the patient with parabolic reflectors. Because of these complications the advantage in heat development achieved by the increase in the frequency is again endangered, so that for the present it is best not to choose a frequency higher than 5×10^7 c/sec (wavelength 6 m). This wave length is the one most commonly used for short-wave therapy at the present time, and it has been taken as a basis for the Philips generators.

Here again it must be noted that the historical development has not followed this argument as a whole. A group of investigators held the opinion that for heating different parts of the body different frequencies should be chosen, in order to obtain in each case the largest possible heat development [3]. This conception originated from the fact that when equation (1) or one similar to it set up for constant current is partially differentiated with respect to R and dQ/dR is set equal to zero, a relation between R, C, C_A and ω is found for which Q has a maximum. This means, however, — as Daan [4] pointed out — nothing more than that at a given frequency a certain relation between R and C is the most favourable. If on the other hand R and C must be considered as given, as is always the case in practice, the best choice of ω is independent of R and C, as explained above.

The high-frequency output required

The power which must be supplied to the patient depends upon the size of the part of the body to be heated. It must be kept in mind that the energy must be supplied at a certain rate if any appreciable increase in temperature is to take place. The circulatory system inside

[1]) See Philips Techn. Rev. **6**, 253, 1941.
[2]) See for example Philips Techn. Rev. **4**, 189, 1939.
[3]) J. Pätzold, Z. Hochfrequenztechnik **36**, 85, 1930.
[4]) A. Daan, dissertation, Amsterdam 1938.

the body and the perspiration of the skin have a very strong heat-regulating action, *i.e.* they combat local changes in temperature. The heating of a part of the body must overcome these equalizing influences. On the other hand the heating may not take place too quickly, since the body reacts against too rapid disturbances of the heat equilibrium by severe sensations of pain. If for example it is desired to give a temperature increase of 1 to 1.5°C to the chest or pelvis of a patient, a heating time of about 5 min. is found to be necessary, while a power of about 300 W must be supplied to the patient. This is about the highest power which will be needed for the treatment of a single object. Only in the case of the excitation of artificial fever, which was formerly carried out by expressly infecting with malaria, and where the purpose is to heat the whole body several degrees centigrade, are higher powers required, namely 700 to 800 W. Since, however, this treatment is quite rare, it is better to use two separate 400 W apparatus than a heavy apparatus constructed especially for that purpose, also because a 400 W apparatus can more easily be built to make it transportable — a requirement which is always made of apparatus to be used in a hospital. For these reasons the Philips apparatus 11 952 is built for an output of 400 W. This apparatus has the form of a desk on wheels, see *fig. 3*. In order also to be able to treat bedridden patients at home a second type (11 951) has also been developed which with an effective output of 150 W weighs only 28 kg. In this case therefore some of the heaviest treatments are impossible, but the apparatus can easily be transported in a motor car (*fig. 4*).

The above refers only to the maximum output for which the generator must be designed. On the subject of dosage, *i.e.* the regulation and control of the power applied to the patient, some remarks will be made in the following.

Connections of the high-frequency generator

Diagram showing the principle

In principle a generator for short-wave therapy is nothing else but a short-wave radio transmitter. It can, however, be made considerably simpler than a radio transmitter, due to the fact that for therapy it is unnecessary to make strict requirements as to the constancy of the frequency and the amplitude of the oscillation generated. Since an especially constant frequency is not required, driving stages with crystals or the like can be omitted and, even for relatively high outputs, a self-oscillating connection with only one stage can be used. Owing to the fact that a stable amplitude is not required (absence of modulation), the generator can be fed with AC voltage, so that no rectifiers are needed for the supply. It is in fact only these simplifications which make it possible to design a unit which is acceptable economically to the existing circle of users.

The diagram of the fundamental connections of the generators 11 951 and 11 952 is reproduced in *fig. 5a*. The anode circuit, which is tuned to a frequency of 5×10^7 c/sec, is formed by the anode-cathode capacities of the two transmitting valves and the two coils L_a', L_a''. The connections show some similarity to the push-pull connections often used in radio technology (fig. 5*b*). They differ from the latter, however, in that the two valves oscillate in turn, each valve oscillating only in the half period of the low-frequency supply voltage in which its anode is positive, while in the case of true

Fig. 3. Generator for short-wave therapy, type 11 952, in use (treatment of knee joint). This apparatus in the form of a desk on wheels can deliver a maximum high-frequency output (wavelength 5 m) to the patient. Besides the knob for switching on (right) there is no other part to be operated by the medical personnel but the knob for adjusting the desired output (left), which is read off on the meter, and the disc for tuning the circuit with the patient (centre), which is checked on the same meter. Dimensions 46 × 37 × 87 cm, weight 62.5 kg. Two electrodes in the form of plates are fastened to the apparatus by means of two arms which can be turned in all directions.

Fig. 4. Generator for short-wave therapy, type 11 951. The apparatus weighs only 28 kg and can easily be transported in a motor-car. The maximum effective high-frequency output amounts to 150 W.

push-pull connections both valves oscillate continuously, see *fig. 6a*. Nevertheless, even in these halves of the period during which the "other" valve is generating, the full high-frequency voltage acts on both of the anode coils, due to their mutual induction. The high- requency voltage on the complete circ uithas, therefore, at every moment an amplitude which in the most favourable case (full modulation of the voltage) is four times as large as the momentary value of the low-frequency supply voltage, see fig. 6*d*.

The desired power can of itself also be excited without difficulty with a single valve. However, by using two valves each of lower output and connecting them as in fig. 5*a*, several important advantages are obtained. In the first place the supply transformer is s y m m e t r i c a l l y loaded, so that the primary current is about twice as small ($\cos \varphi \approx 1$, while with unsymmetrical loading $\cos \varphi \approx 0.5$), and thus the transformer can be made smaller. Furthermore, the symmetry of the high-frequency oscillation circuit facilitates the oscillation and diminishes the reaction of the circuit in which the patient is included on the generator. Thanks to the symmetrical connections there is also less chance of high-frequency voltages being transmitted to the supply network. Finally it must be noted that when using a single valve intervals occur during which there is no high-frequency voltage present, as may be seen in fig. 6*b* or *c*. On the other hand when using two valves the high-frequency voltage is more nearly continuously present. Some patients seem to find the latter situation more pleasant.

The circuit L_pC_p in which the patient is included is coupled inductively with L_a'—L_a'', so that the low-frequency anode supply voltage, which in the case of valves for several hundred watts must normally amount to several thousand volts, cannot act upon the patient. For absolute safety the point of symmetry of L_pC_p is furthermore connected with earth *via* a high-frequency chok ng coil (not shown in fig. 5*a*). The circuit L_pC_p is tuned to the generator frequency for every case by means of a rotating condenser.

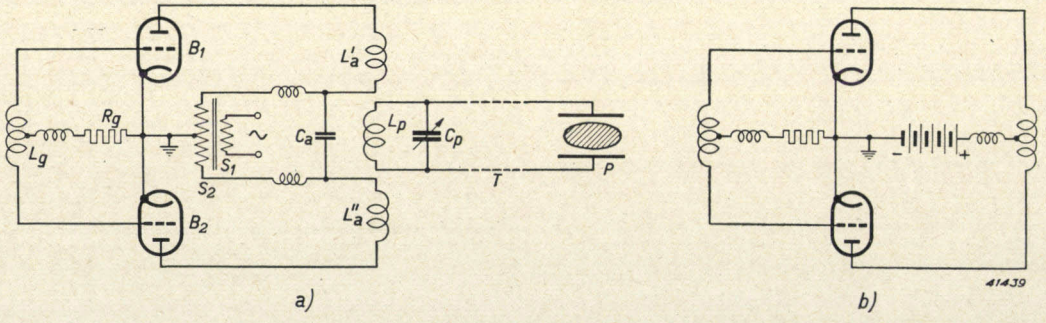


Fig. 5. *a*) Diagram showing the principle of the Philips generators for short-wave therapy. Since the requirements as to constancy of the frequency and the amplitude of the oscillation obtained are not severe, the generator is built up of a single self-generating stage, and as anode voltage simply a 50 c/sec AC voltage is taken which is delivered by the supply transformer S_1, S_2. The s y m m e t r i c a l connection of two valves (B_1, B_2) here employed has various advantages over connections with only one valve. The circuit in which the patient P is included *via* the supply lines T and which is tuned to the wavelength 6 m with the rotating condenser C_p for each case treated, is coupled with the divided anode coil L'_a—L''_a by the coil L_p. C_a is a coupling condenser, L_g grid coil, R_g grid resistance.
b) Normal push-pull connection of two valves.

While the construction of a medical short-wave apparatus is on the one hand simpler than that of a radio transmitter, thanks to the lower requirements of constancy of frequency and amplitude of the oscillations obtained, there are on the other hand complications which are not encountered in a radio transmitter: 1) the loading impedance is different for every case treated, 2) it must be possible to vary the output within wide limits (practically by a factor 10).

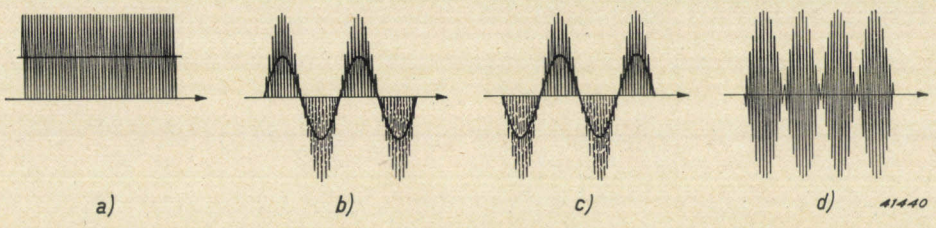


Fig. 6. Form of the resultant anode voltage
a) of each of the two valves in fig. 5*b*,
b) of valve B_1 in fig. 5*a*,
c) of valve B_2 in fig. 5*a*,
d) of the two valves in fig. 5*a* together.
The anode DC voltage and low-frequency anode supply voltage, are indicated by heavy lines. The oscillations indicated by thin broken lines in *b* and *c* are due to induction in the corresponding half of the anode coil during the "wrong" period of the supply voltage.

Matching of the circuit in which the patient is included

As to the loading impedance, the loading of the anode circuit should be matched for each patient by making the coupling between L_a and L_p more or less tight. With the technically unskilled operators of the medical apparatus, however, it is not easily possible to have them make this adjustment, not so much because of the fact that the coupling requires too great precision as far as the magnitude of the output is concerned (with self-generating connections the opposite is true), but because with too tight a coupling an undesired reaction of the circuit with the patient on the generator may occur. Therefore in the apparatus the coupling between L_a and L_p is adjusted permanently in such a way that for the h e a v i e st cases of loading occurring in practice (smallest impedance in parallel with L_pC_p) the correct matching is obtained, while with larger parallel impedance the coupling is actually too loose to make the generator deliver the maximum output. This is no objection since with larger parallel impedances, *i.e.* smaller parts of the body, the maximum output is not applied, in connection with the above-mentioned requirement that the heating up may not take place too quickly.

This brings us to the second complication, the regulation of the power supplied to the patient and the problem of measuring this power.

Regulation of the output

If one begins with the maximum output, the output can be reduced either by decreasing the heating current, increasing the grid resistance or decreasing the anode voltage. The last method is the most economical one, since in that case at lower outputs no unnecessary losses occur in the apparatus. This method has been applied in the larger of the two generators here described, and the regulation is made by varying the primary voltage of the supply transformer. In this way it is of course impossible to take the heating current of the valves simply from an extra secondary winding of the supply transformer, and a separate heating current transformer must be used. Against this disadvantage, however, is the fact that the regulation at the primary side makes possible a simple m e a s u r e m e n t of the effective output. We shall explain this briefly.

For the correct adjustment of the output sometimes the only criterion which is used is that the patient feels a "pleasant warmth". It is obvious that this is not a very reliable method, if only because in many cases (for instance frost bite and various neurological disorders) the sensitivity to heat is disturbed at the spot to be treated. It is impossible to include a direct-reading instrument in the circuit with the patient, since no wattmeter suitable for these high frequencies exists. A useful method of determining the high-frequency output consists in substituting for the patient an electric lamp of about the same resistance, allowing the light flux obtained to fall upon a luxmeter, and afterwards reproducing the indication of the luxmeter by supplying the lamp with direct cur-

rent or low-frequency alternating current with wattmeter connected in series [5]). This method is indeed applied for control in the factory, but it is, too elaborate for use by the doctor.

In this method of regulating the output variation of the primary supply voltage with unvarying grid resistance and heating current, the ratio between the "DC voltage" (actually: low-frequency DC voltage) and the high-frequency AC voltage remains unaltered, *i.e.* the modulation of the generator valve always remains the same. By this means and by means of measures for the limitation of losses, which will be briefly discussed below, it is possible to obtain a practically constant efficiency of the generator for all values of the output chosen. This means that the ratio between the power supplied to the patient and the power taken up from the mains is constant, and we may therefore determine the former simply by measuring the latter, which may be done with a normal wattmeter preceding the supply transformer.

Several particulars of the construction

The apparatus 11 952 contains two triodes type TB 2/200, which type has a permissible dissipation of 130 W, and at 6 m wavelength in the connections described it can develop a high-frequency output of 200 W[6]). In the smaller apparatus 11 951 four small triodes (type no. 29 933) are used and connected two by two in parallel.

With full load the valves of the apparatus 11 952 operate with an efficiency of $200/(200+130) \approx 60$ percent. The total efficiency of the apparatus is of course lower, due to all kinds of losses. These losses are partially proportional to the power supplied to the patient and partially depend upon still other factors. The latter, among which are losses by radiation, were limited as far as possible in order to ensure the proportionality required for the measurement of the output described between the power supplied to the patient and the power taken from the mains. The remaining deviations from this proportionality cause in the most unfavourable case a measuring error of $\pm$ 15 percent, which for medical application already constitutes a very satisfactory accuracy compared with the inaccuracy which had hitherto prevailed.

Moreover, those losses which do not affect these measurements, since they are proportional to the effective output, should also be restricted for the sake of economical operation. Such losses occur for example at the grids in the circuit represented in fig. 5*a*. When the anode of one valve is positive the oscillations excited by that valve act also on the grid of the other valve, due to back-coupling and induction in the grid coil. While of course the anode of the latter valve is negative at that moment, grid current can, nevertheless, flow in the positive peaks of the grid voltage, and this grid current results in an increased grid dis ipation. It has been found possible to decrease this effect considerably by modifying the connections slightly, see *fig. 7*. Each of the two grids is connected with a tap on its "own" half of the supply winding S_2, which tap is so chosen that in the period of negative anode voltage the grid receives a sufficiently large negative bias to prevent grid currents due to the induced high-frequency voltage. Since the grids of the two valves must be at different low-frequency voltages, the grid coil L_g is now interrupted by a condenser, just as was previously done with the anode coil L_a. In the period of positive anode voltage the extra grid bias now becomes positive. In order to prevent the flow of very large grid currents due to this cause, which would again neutralize the desired effect, the grid resistances R'_g, R''_g are increased so much that the grid current takes on a normal relation to the anode current. In this way it has been found possible to obtain an average total efficiency of the apparatus of 45 percent.

The various details mentioned here can be recognized in the photograph of *fig. 8*. The construction of the coupling condenser between the two halves of the anode coil is striking. This condenser is at the full voltage of the supply winding (several thousand volts) and carries the high-frequency current of the anode circuit (order of magnitude 10 A). It is therefore made of a special ceramic material which combines low high-frequency losses with high voltage loadability.

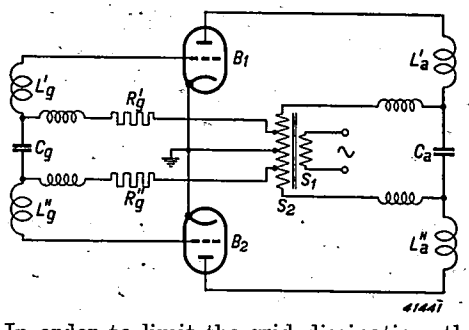


Fig. 7. In order to limit the grid dissipation, the grid resistances R'_g and R''_g, for each of the valves in fig. 5*a* are connected with a certain tap on the corresponding half of the winding of the supply transformer.

[5]) *Cf.* also Philips Techn. Rev. **6**, 215, 1941.
[6]) See the article referred to in footnote 1).

Practical application of the apparatus

Since the apparatus consume only relatively little current — the larger apparatus a maximum of 4.5 A at 220 V —, they may be connected to any ordinary wall contact which is protected by 6 A fuses. The necessary manipulations are limited to the application of the electrodes to the patient, the tuning of the circuit with the patient on the wavelength of the generator and the adjustment of the desired output. The

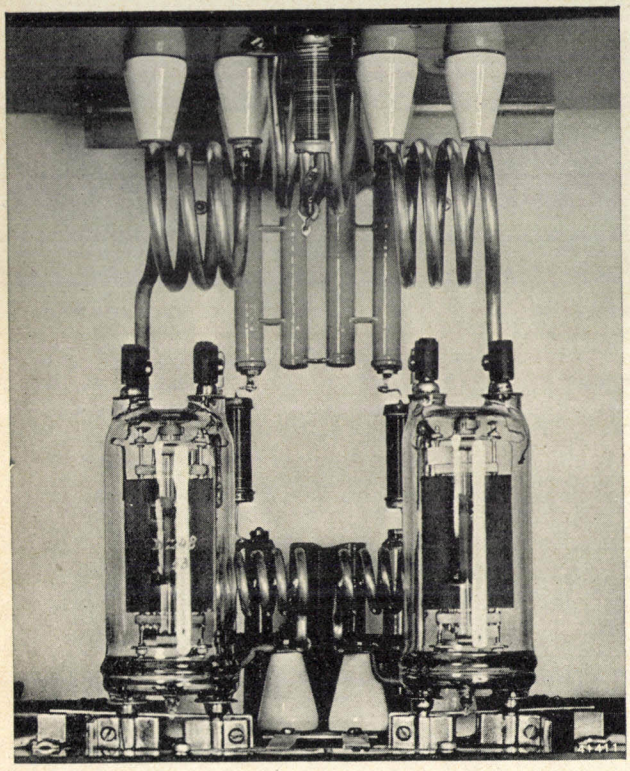


Fig. 8. Part of the inside of the generator 11 952. Above the two valves are placed the two halves of the anode coil; partly between and partly above these coils is the coil of the circuit with the patient. It may be seen that the coupling is only very loose. Behind the coils hangs the coupling condenser of the anode coil, constructed of special ceramic material. Between the valves may be seen the grid coil, which is also divided into two halves, and behind it the corresponding coupling condenser.

tuning is easily performed by means of a rotating condenser with fine regulation (in fig. 3 the tooth-edged disc at the front of the apparatus) and in the case of the apparatus 11 952 it is controlled in a simple way due to the fact that the wattmeter for the measurement of the power delivered gives a maximum deflection at the correct tuning. For control of the tuning a gas-discharge tube can also be used. The tube is placed in the vicinity of the patient or the electrodes and gives the most intense illumination at the correct tuning.

As to the application of the electrodes, it need only be noted that the doctor in charge must have some insight into the expected course of the lines of force between the electrodes.

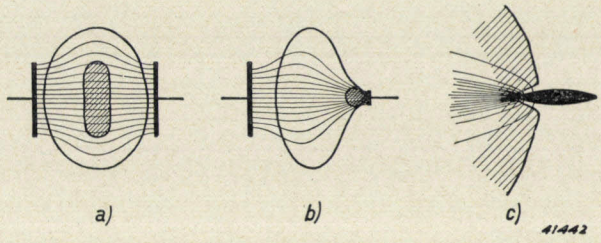


Fig. 9. By giving the electrodes a suitable form and placing them in a suitable position the desired localization of the heat development in the body can be realized. *a*) Uniform heating of an extensive region with two plate electrodes. *b*) Concentration of the heat development in a small region (furuncle) with a small plate electrode. *c*) So-called electro-surgery: by connecting the knife which makes the incision in an operation to one pole of the short-wave apparatus, while the other electrode is placed arbitrarily, such a strong concentration of lines of force occurs at the point where the incision is made that the blood at that point immediately coagulates from the heat (bloodless operation). The cutting itself, *i.e.* the severing of the tissue, also takes place much more easily.

By a correct choice of shape, size, position and distance from the skin he can then cause the heat to be developed as nearly as possible in the desired spots. The diagrams in *fig. 9 a-c* and *fig. 10* give a few examples, while the photograph of fig. 3 shows how the electrodes are fastened to the generator and how in a given case the arrangements can be made. A more detailed discussion of the various forms of electrodes in use and of the methods of treatment, however, falls outside the scope of this article.

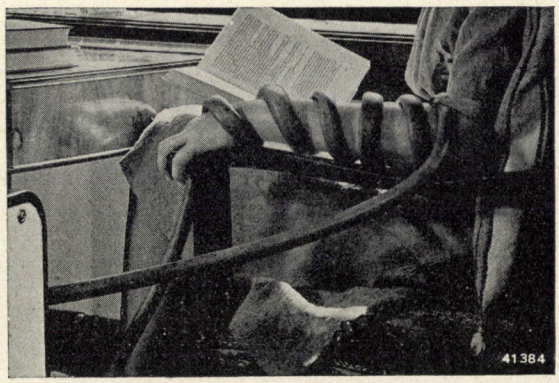


Fig. 10. In order to heat an arm or a leg uniformly, the member in question is not placed between two electrodes, but a coil connected with the two poles is wound around it (method of Kowarschik).

THE VIBRATION OF CONTACT SPRINGS

by J. A. HARINGX.

621.3.066.6

For the satisfactory operation of relay contacts it is found that it is important that the closing of the relay should take place very gradually, because otherwise the contact shows a tendency to rebound repeatedly before it remains definitely closed. The occurrence or absence of an impulse is found to depend upon the resonance frequency of the contact spring. In this article the resonance frequencies are calculated which must be chosen in order to obtain a velocity of closing equal to zero. It is furthermore shown that with only a slight deviation from these resonance frequencies the velocity of closing will increase appreciably.

In different kinds of electrical apparatus, such as for example the electric bell, relay contacts are used which are opened and closed periodically at a fairly high frequency. For the sarisfactory functioning of the mechanism it is very important that the opening and closing should occur exactly at the moments fixed for the structure in question, and, moreover, not only when the apparatus is new, but also after long use.

One of the phenomena which must be considered is the elastic vibration of the contact springs. In particular, it may sometimes be observed that after each mutual contact between the two parts of the contact, the spring continues to vibrate so that the electrical connection is broken several times before it is finally made. Since a spark occurs upon each interruption of the current circuit, it is obvious that such a vibrating contact burns through much more rapidly than a contact which is only broken once per period. Moreover, disturbing high-frequency phenomena occur as a result of these sparks.

The problem of the vibration of contact springs has already been discussed in this periodical in connection with a vibrator for the conversion of direct current into alternating current [1]). In that case the striking together of the two parts of the contact was compared with the collision between two billiard balls, in which case it may also be observed that the contact which occurs is immediately broken again under the influence of the elastic forces. Under special circumstances, however, the contact between the moving ball and the stationary ball is maintained. This case may occur when a third ball is present touching the stationary ball. The impulse of the moving ball can then be transmitted through the stationary ball to the third ball, so that the latter rolls away while the first two balls remain in contact.

In the article referred to above a description is given of the method by which this principle of the third billiard ball can be usefully employed in relay contacts. In this way the two colliding metal parts of the contact are made to remain in contact without vibration, while the impulse is transmitted to the third mass, which is elastically attached. This latter mass can be knocked away without, however, any harmful results, since the broken contact between the second and third parts of the contact does not constitute part of the circuit.

Since this method of combating vibration in contacts is quite complicated, it seemed to us important to find out whether it was not possible to avoid the necessity of such measures. By a suitable choice of the characteristic frequency of the vibrating contact spring it has been found possible to cause the two parts of the contact to come together very gradually, so that no shock at all occurs. Special measures of dissipating the impulse of the collision thus become superfluous.

Mathematical formulation of the problem

The relay contact can be represented by a point of mass *1* at the end of a spring (see *fig. 1*).

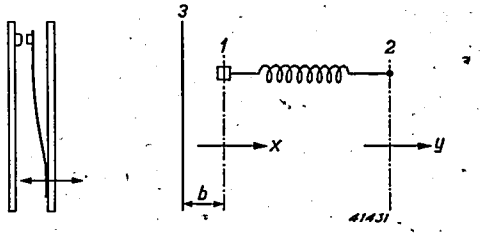


Fig. 1. Principle and mechanical model of a spring contact.

The other end (*2*) of this spring makes a sinusoidal motion according to the formula

$$y = a \cos \omega t \quad \ldots\ldots\ldots\ldots \quad (1)$$

The mass point *1* will then also execute a sinusoidal motion, provided the amplitude A of this motion remains smaller than the distance b between the stop *3* and the position of rest. If m is the mass of the vibrating part of the contact and c the stiffness of the spring, *i.e.* the force per unit of elongation, the following

[1]) Philips techn. Rev. 6, 342, 1941.

is the differential equation of this motion:

$$m\frac{d^2x}{dt^2} = c\,(y-x),$$

or if we substitute for y the value according to (1)

$$\frac{d^2x}{dt^2} + \frac{c}{m}x = \frac{c}{m}a\cos\omega t.$$

The general solution of this is

$$x = A\cos\omega t + B\cos\omega t_0 + C\sin\omega_0 t, \quad \ldots \quad (2)$$

where the characteristic frequency $\omega_0 = \sqrt{c/m}$

and $$A = \frac{a}{1-\omega^2/\omega_0^2} \quad \ldots \quad (3)$$

The term $A\cos\omega t$ represents the forced vibration of the mass point, while the other two terms of equation (2) describe the free vibration of the mass point. Let us now examine what happens when the deflection of the mass point is limited by the stop *3*. In a certain part of the period the mass point will then have the deflection $-b$ and will lie against the stop with a certain pressure. The motion of the mass point will only begin again when the pressure ceases to exist, in other words when the compressed spring has just recovered its original length. Then evidently $y = -b$, so that for the moment t_1 of the beginning of the motion we may write

$$a\cos\omega t_1 = -b \quad \ldots \quad (4)$$

The motion of the mass point must therefore be such that at the moment $t = t_1$, there is a deflection $-b$ and the velocity is zero. These conditions are sufficient to calculate B and C, so that the mechanical problem is hereby solved in principle.

A possible solution is reproduced in *fig. 2*. The motion of point *2* may be seen and the forced vibration of point *1*, as well as the free vibration which must be added to fulfil the initial conditions. At the same time the resultant motion of the mass point is reproduced. t_1 is the moment at which the movement begins; t_2 the moment at which the contact is again closed. It may be seen that the velocity of the motion

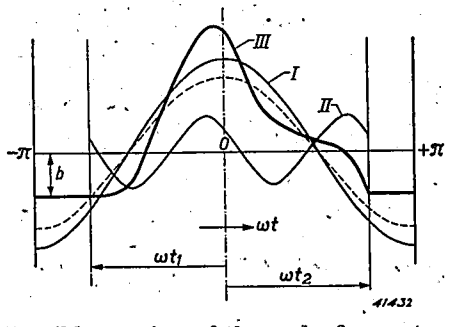


Fig. 2. Possible motion of the end of a contact spring whose other end moves sinusoidally according to the broken-line curve. The curves *I*, *II*, *III* indicate respectively the forced vibration, the free vibration and the resultant motion.

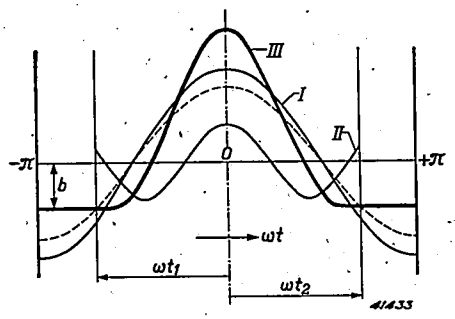


Fig. 3. Motion of the end of a contact spring whose resonance frequency is so chosen that the closing velocity is zero.

(slope of the curve) at the moment of closing has a very discontinuous character, which means that a strong impulse will occur upon closing. Our object of procuring a gradual coming together of the parts of the contact is thus not realized in this case.

In order to obtain improvement it is now sufficient merely to choose the characteristic frequency of the contact spring somewhat lower, so that the diagram of *fig. 3* is obtained. The forced vibration and the free vibration of the mass point are now both symmetrical with respect to the dot-dash line, which represents the moment $t = 0$. The resultant motion of the mass point is thus also symmetrical with respect to this line, and since the motion begins with a velocity zero, it will also end with a velocity zero, so that the desired impulse-free contact is obtained.

The symmetry condition can be formulated mathematically by saying that the free vibration is a pure cosine function; the term $C\sin\omega_0 t$ in equation (2) must therefore dissappear [2]). Since C depends also upon ω_0, by a suitable choice of the resonance frequency of the spring provision can be made to satisfy this condition.

Calculation of the desired resonance frequency

As we have seen, the constants B and C of the motion of the contact piece given in equation (2) are determined by the conditions that for $t = t_1$ there should be a deflection $x = -b$ and a velocity $dx/dt = 0$. Thus in formulae:

$$A\cos\omega t_1 + B\cos\omega_0 t_1 + C\sin\omega_0 t_1 = -b,$$
$$-A\omega\sin\omega t_1 - B\omega_0\sin\omega_0 t_1 + C\omega_0\cos\omega_0 t_1 = 0.$$

[2]) Actually it has only been proved that the disappearance of the term $C\sin\omega_0 t$ is sufficient for securing a closing velocity of zero. To complete the mathematical reasoning it should therefore also be proved that impulse-free closing cannot be secured in any other way, so that the condition $C = 0$ is actually necessary. We shall not, however, go into that here.

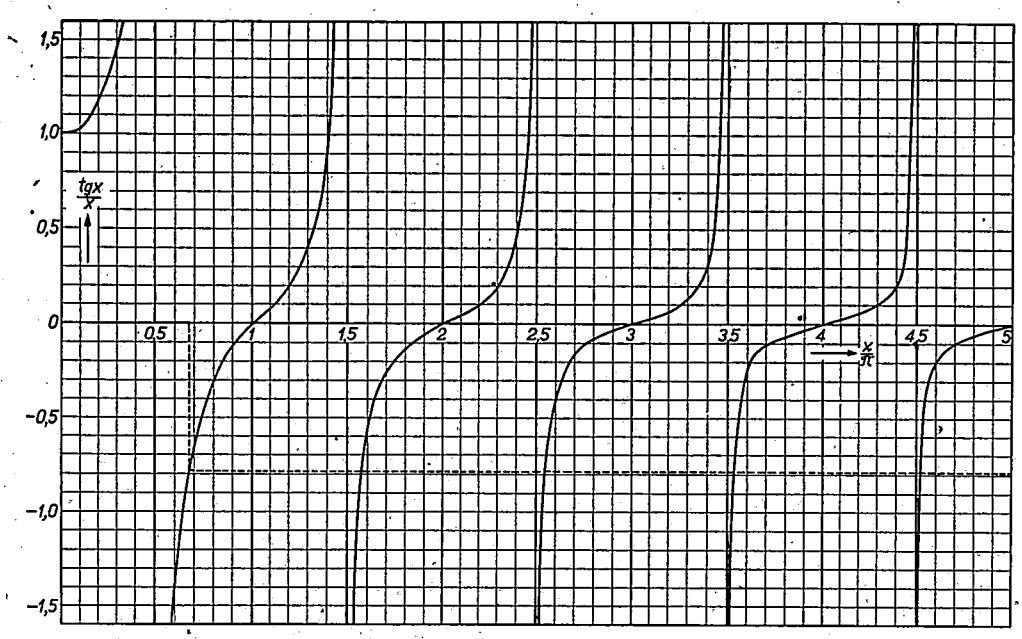


Fig. 4. The function tan x/x which is used to solve the transcendental equation

$$\frac{\tan kx}{kx} = \frac{\tan x}{x}$$

for k. From the figure it may, for example, be deduced that $x/\pi = {}^2/_3$ corresponds to $kx/\pi = 1.57$; 2.59; 3.54; 4.52; etc., and therefore to $k = 2.35$; 3.88; 5.32; 6.78 etc.

By substituting in these formulae the following values from equations (4) and (3):

$$b = -A\,(1-\omega^2/\omega_0{}^2)\cos\omega t_1,$$

one finds

$$\left.\begin{aligned} B\cos\omega_0 t_1 + C\sin\omega_0 t_1 &= -\frac{\omega^2}{\omega^2{}_0} A\cos\omega t_1, \\ -B\sin\omega_0 t_1 + C\cos\omega_0 t_1 &= \frac{\omega}{\omega_0} A\sin\omega t_1. \end{aligned}\right\} \quad (5)$$

If now, in agreement with the condition for impulse-free closing, we make the condition that $C = 0$, we may divide the two equations (5) by each other and in that way find the condition that

$$\tan\omega_0 t_1 = \frac{\omega_0}{\omega}\tan\omega t_1 \quad . \quad . \quad . \quad . \quad . \quad (6)$$

In order to draw practical conclusions from this equation we shall pass over to new variables which are adapted to the form of the technical problem. The function of the mechanism usually prescribes that the contact shall be open during a certain part of the period lying between 50 and 100 percent. If we let this fraction equal α, then according to fig. 3

$$\omega t_1 = -\pi\alpha.$$

If, moreover, we set ω_0/ω equal to k, (6) may be written in the form

$$\frac{\tan k\pi\alpha}{k\pi\alpha} = \frac{\tan\pi\alpha}{\pi\alpha}$$

and we obtain in this way a relation with whose help k can easily be determined for every prescribed value of α. To do this it is only necessary to draw the function tan x/x and to find the values of x for which the ordinate is equal to tan $\pi\alpha/\pi\alpha$ (*see fig. 4*).

If α is chosen equal to 0.5, which means that the relay is open for one half of the period, tan $\pi\alpha/\pi\alpha$ is infinitely large, and the same must be true of tan $k\pi\alpha/k\pi\alpha$. From this for k all the odd numbers follow:

$$k = \frac{\omega_0}{\omega} = 1, 3, 5, 7, 9, \ldots$$

In the limiting case where $\alpha = 1$, thus for a relay which is open during the whole period, tan $\pi\alpha/\pi\alpha = 0$. The tangent of $k\pi\alpha$ (with $\alpha = 1$) must therefore also be zero, and from this it follows that

$$k = 1, 2, 3, \ldots$$

Thus if α is allowed to increase continuously from 0.5 to 1, k must pass from the series of odd numbers to the series of all whole numbers. The transition conditions are easily constructed with the help of fig. 4, as may be seen from the example given. Finally, in *fig. 5* the result of this graphic calculation is given, namely the ratio $\frac{\omega_0}{\omega}$ as a function of α for the case where the contact is closed with no impulse.

Tolerances for the resonance frequency

For the practical application of the method discussed for obtaining a closing velocity of the

contact equal to zero it is not only important to know the suitable resonance frequency of the contact spring, but at the same time one must know how accurately this frequency must be adjusted in order to obtain the desired effect.

If we consider as example the case where $a = 0.5$, we see that the forced vibration $A \cos \omega t$ has the maximum velocity just at the moments t_1 and t_2 at which the contact opens or closes. For the free vibration $B \cos \omega_0 t$, in the case of a closing velocity of zero, the same is true, since ω_0 is an odd multiple of ω. The coefficient B has a value such that the velocities of the forced vibration and of the free vibration compensate each other exactly for t_1 and t_2.

If ω_0 is now allowed to vary, the two velocities will continue to compensate each other for the initial moment t_1 of the motion, but for the moment t_2 this is in general no longer the case.

If for example ω_0 is an even multiple of ω, the two velocities will reinforce each other as much as possible, so that the resultant velocity is twice as great as that of the forced vibration alone. If the resulting closing velocity v is drawn as a function of ω_0/ω, a wavelike curve is therefore obtained which has maxima for all the odd whole numbers. The minima are zero, while the maximum velocity is twice as high as that of the forced vibration alone. The curve derived qualitatively in this way is reproduced in *fig. 6*.

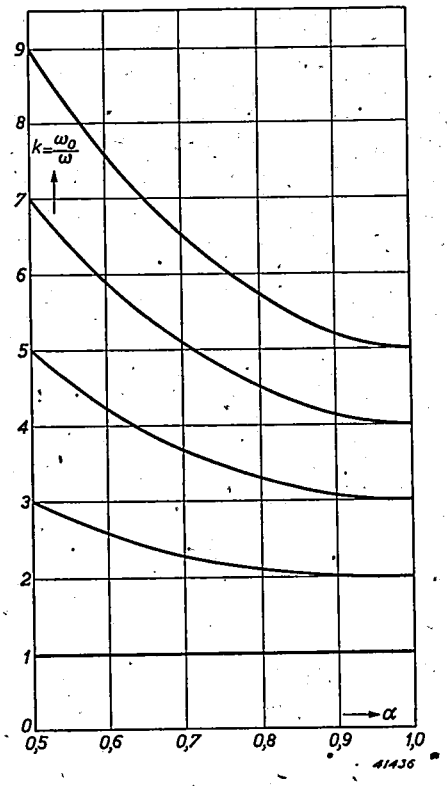


Fig. 5. The behaviour of $\omega_0/\omega = k$ as a function of a for the case of impulse-free closing.

The horizontal line at the height 1 indicates the amplitude of the velocity of the fixed end of the spring. Thus there can only be a question of a decrease in the closing velocity compared

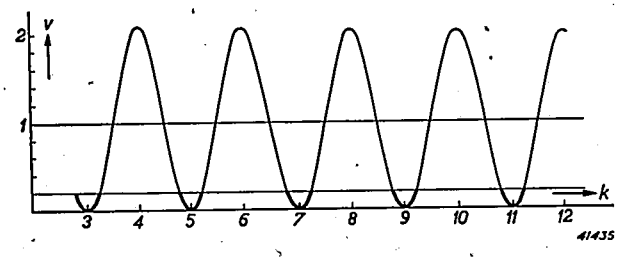


Fig. 6. Closing velocity v as a function of $\frac{\omega_0}{\omega}$ for $a = 1/2$. The ratio of the closing velocity to the velocity which would hold for a rigid connection instead of a spring is plotted as ordinate.

with the velocity of the driving motion at those values of k for which the wave curve lies below this horizontal line. As may be seen, this is the case only in sufficiently close proximity to the odd whole numbers; the maximum difference which may be permitted amounts to about $\Delta k = \pm 0.5$.

If the condition is made that the decrease in the closing velocity should amount at least to a factor 5, the permissible deviation Δk is reduced to 0.2. This means that the adjustment to the correct resonance frequency is quite critical, since with the relay constructions in practical use values of k lying in the neighbourhood of 7 or 9 are quickly reached. The permissible deviation would then amount to only about 3 percent.

In the foregoing we have been assuming that the fixed end of the spring executes a fixed motion. Usually, however, this motion itself is again a free vibration of a mass M attached to some kind of elastic system. A closer consideration of the more complicated case of a coupled vibrating system shows that in general ω must be set equal to the characteristic vibration of the main mass M, while ω_0 can be calculated with the help of the stiffness of the whole elastic system, considering the mass M as stationary.

When the desired resonance frequency of the relay spring has been found there remains the practical problem of the method of adjusting the resonance frequency accurately. We shall not go into that at this point, since in every case it depends upon the special construction. It may only be stated that in the construction of the vibrator for the conversion of direct current into alternating current mentioned at the beginning of this article, the correct setting of the contact spring was actually found to produce the desired result.

SUMMARY OF RECENT SCIENTIFIC PUBLICATIONS ISSUED BY N.V. PHILIPS' GLOEILAMPENFABRIEKEN

Reprints of these publications will be gladly supplied free of charge upon request to the Administration of the Natuurkundig Laboratorium, Kastanjelaan, Eindhoven. Reprints of any publications marked with an asterik, however, are not available in sufficient number for distribution.

1570: E. J. W. Verwey: The electrical double layer of oxidic substances especially in non-aqueous media (Rec. trav. chim. Pays Bas **60**, 625-633, July-August 1941).

The electrical double layer on the surface of particles of oxides is studied : 1) by the examination of suspensions of these substances in water, alcohols, acetone or solutions of electrolytes in these liquids and 2) by electro-endosmosis through membranes in these liquids. The double layer is caused by ions originating from the materials examined. The ions of hydrogen and of the hydroxyl group determine the potential. The zero point of the charge is very closely related to the size of the particles. Just as in the case of silver iodide, large particles have a stronger negative charge than small particles.

1571: J. A. M. van Liempt: Eine einfache Methode zur Bestimmung der Diffusionskonstante von Metallen. (A simple method of determining the diffusion constants of metals) (Rec. trav. chim. Pays Bas **60**, 634—639, July-August 1941).

A simple method is developed for determining the diffusion constant in metals. A thin wire or strip of one metal is embedded in a piece of a second less precious metal by metal-ceramic means. The duration and temperature of the heating are determined at which the diffusion has progressed so far that the limit of resistance is just reached, over the whole section of the wire or strip.

1572: J. L. Snoek: Concerning the internal damping of ferromagnetic substances. (Ned. T. Natuurk. **8**, 177-189. April 1941).

Internal damping phenomena in iron and other ferro-magnetic materials are discussed. The damping of torsional vibrations in iron and nickel wires is almost exclusively due to ferromagnetic hysteresis; external changes in the magnetisation, however, are not observed. The hysteresis damping appears to be proportional to the amplitude of the vibration. In longitudinally vibrating ferro-magnetic rods considerable changes in the induction can be observed when the magnetostriction is great, the material is properly annealed and an external magnetic field of a suitable strength is applied. When, moreover, the frequency is so chosen that the eddy current losses in the rod are the maximum, exceptionally large values are obtained for the logarithmic decrement, as observed experimentally. In iron and iron alloys containing small quantities of carbon or nitrogen damping phenomena of a certain kind are observed. This is purely an after-effect and shows no hysteresis. The great temperature-dependency observed leads one to suppose that diffusion of carbon or nitrogen through the iron lattice plays a part. In cold-treated samples after-effect phenomena are found in another temperature range. This is complementary to what is observed with well annealed specimens.

1573: Balth. van der Pol: Two little known experiments. (Ned. T. Natuurk. **8**, 390-393, September 1941).

At a meeting of the Netherlands Physical Society at Eindhoven some striking, subjective colour phenomena were demonstrated with a Benham optical disc, such phenomena occurring in spite of the fact that the disc consists only of black and white parts. These colour phenomena also occur when monochromatic yellow sodium light is used. Further, a demonstration was given with two constructions of a hot air engine in its most primitive form (Griffiths' experiment).

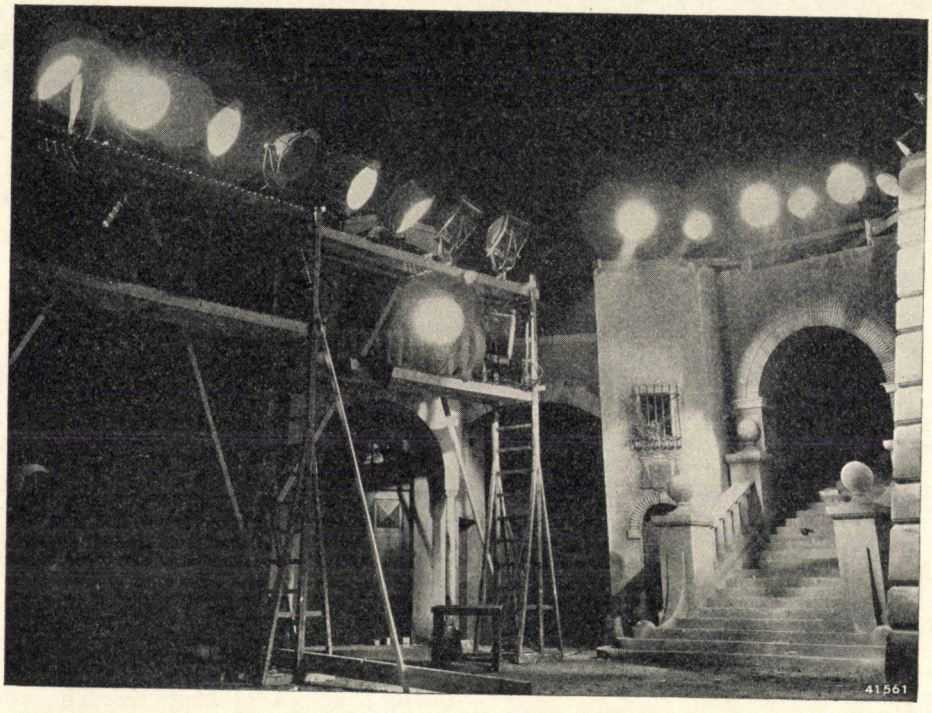
1574: M. J. Druyvesteyn and J. L. Meyering: An approximate calculation of the thermal expansion of solids I (Physica **8**, 851-861, September 1941).

The expansion coefficient for a solid body is practically equal to the third differential quotient of the lattice energy in length. For the alkali metal an expansion coefficient is calculated which agrees very well with the experimental values.

7e No. 6, pp. 161-192 JUNE 1942

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



LIGHT SOURCES FOR CINEMATOGRAPHY

by Th. J. J. A. MANDERS.

621.32 : 778.53

For making cinematographic exposures lamps are needed which have high light intensity and good actinic effect. These requirements led originally to the use of arc lamps, while at present incandescent filament lamps are often used. The latter are more satisfactory, especially for making sound films, since, in contrast with arc lamps, they burn absolutely noiselessly. In this article the construction of incandescent lamps for film making is discussed. These lamps are characterized by a heavily loaded incandescent body with relatively small dimensions of the bulb. The fixtures used in the film studio (banks and batteries of lights and spotlights) are also briefly discussed, while in conclusion several special constructions are dealt with which are used when extremely heavy requirements are made.

When the first cinematographic pictures were made by Edison in 1889 he had already had ten years' experience in making incandescent lamps. At first sight, therefore, it seems strange that Edison made no attempt to use incandencent lamps for illuminating the set, but worked exclusively by daylight. A closer consideration of the properties of the carbon filament lamps and the photographic emulsions of those times, however, shows that the electric lamp as an aid in photography was out of the question at that time. In order to obtain a good photographic effect it is necessary that a sufficiently large part of the radiation of the light source be emitted in those wavelengths for which the film emulsion is chiefly sensitive. Now in those days the emulsions were only sensitive to light of short wavelengths, while the carbon filament lamp emits light mainly of long wavelengths. The actinic effect of the light was thus extremely small; in fact it was so small that it was practically impossible to use it for photography.

In order to form a quantitative idea we shall assume that the emulsion possesses the same sensitivity for all wavelengths below a certain

limiting wavelength λ_0 and is totally insensitive above this wavelength. Furthermore we shall consider the radiation of the incandescent lamp as that of a black body of a given temperature T. The percentage of radiation energy which is actinically effective then depends exclusively upon the product of λ_0 and T; this calculated dependence is reproduced in *fig. 1*.

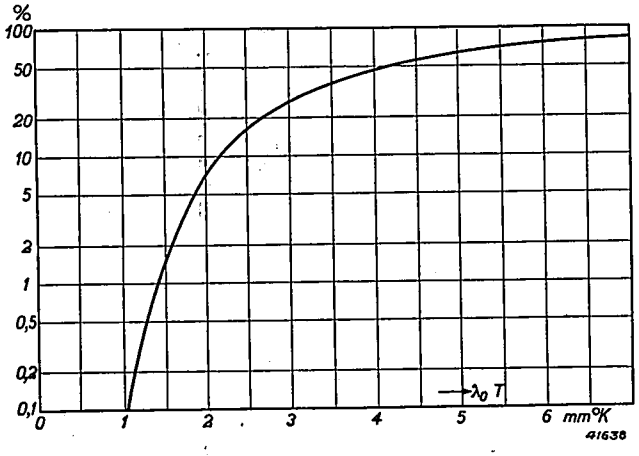


Fig. 1. Percentage of the radiation energy of a black body with a temperature T (°K) in the region of wavelengths shorter than λ_0 (mm) as a function of the product of λ_0 and T.

The unsensitized emulsion has a limiting wavelength of $\lambda_0 \approx 0.5 \times 10^{-3}$ mm, while the carbon filament lamp is used at a temperature of 2000°K. Thus $\lambda_0 T \approx 1$, and for this from fig. 1 it may be deduced that only about 0.8 per thousand of the energy radiated is actinically effective.

In the case of daylight the effective radiation energy amounts to about 25 percent and is thus 300 times as high. Since ordinary daylight possesses an intensity of about 300 W/m², an equivalent illumination with carbon filament lamps would possess an intensity of $300 \times 300 \approx 10^5$ W/m², [1]), which with an efficiency of 0.3 Dlm/W amounts to 300 000 lux! It is easy to understand that such an illumination can never be realized with carbon filament lamps, if only because of the heating of the illuminated objects. If the temperature is calculated which the illuminated objects must possess in order to dissipate the incident energy by radiation, a temperature of more than 1000° C is found.

The circumstances become much more favourable when light sources are used which radiate light of shorter wavelengths. If, for example, an electric arc is used with a colour temperature of 3800° K, the effective radiation energy (wavelength shorter than 500 mμ) amounts to about 6 percent of the total. Indeed already in 1903 it was found possible to take cinematographic pictures of interiors with the help of arc lamps and in this way to become independent of the weather and the position of the sun. Of course the intensity of illumination required, which on the basis of the above estimation would amount to about 10 000 lux, is still very considerable, but thanks to the great intensity of the electric arc this was not found to be a serious objection. In the end, in spite of certain inconveniences connected with its use, the electric arc was generally accepted and formed practically the only source of light used for making films. The incandescent lamp, although it was then already more suitable than the original carbon filament lamp, was for the time being not considered.

If we now consider the progress in photographic technique and in light technique during the early days of the film industry, we can recognize an attempt to bring the actinism curve of the film and the intensity curve of the incandescent lamp closer and closer to each other. The film, originally sensitive only to blue rays, was sensitized for yellow and later for red, whereby the limit of sensitivity was shifted to 600 and 670 mμ, respectively. On the other hand metals with higher and higher melting points came into use as filaments in the lamps, so that the temperature of incandescence could be raised from about 2000° K to about 3000° K. The radiation of the lamp in this way became richer and richer in light of short wavelengths.

In *fig. 2* several stages of this development are given by plotting for various years the intensity curves of the incandescent lamps then in use and the sensitivity curves of the then available types of film. It is immediately evident that the common spectral region of the two curves has become much larger in the course of these years. The actinic part of the radiation energy has increased from 0.8 per thousand to 8 percent, so that the illumination energy required has become a factor 100 smaller. Actually the progress is even greater than would be concluded from this, because simultaneously with the film and incandescent lamp the film camera was also improved. This improvement is difficult to indicate in figures, since the large lens apertures which are now at our disposal cannot be used under all circumstances because of the required depth of focus. It may, however, be said that the reduction in the amount of light required thereby achieved is not estimated too low at a factor 5. According to our calculation, therefore, the intensity of radiation at present still necessary would amount to

[1]) This value can of course only represent a rough estimation of the required illumination intensity, since the actual light requirements are entirely dependent upon the scene to be rendered.

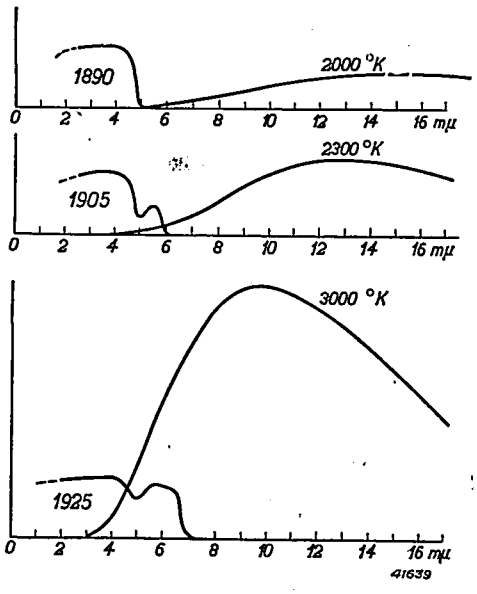


Fig. 2. Spectral distribution of the intensity of various incandescent lamps developed at different successive periods, and spectral sensitivity curve of different photographic materials which were successively put on the market. The effective part of the radiation has increased in the course of time from 0.8 per thousand to 8 percent.

$$\frac{10^5}{5 \cdot 100} \approx 200 \text{ W/m}^2, \text{ or about } 5000 \text{ lux},$$

and this is an illumination intensity which can easily be attained.

The time at which the use of incandescent lamps in the film studio began to be a practical possibility was around 1925. By this time the work in film studios had already reached such an an extent that the much more convenient manipulation of the incandescent lamp (compared with the carbon arc) made its adoption attractive. Another important advantage of the incandescent lamp is its absolutely constant and quiet operation, which is very important for players and operators. The danger of flicker or extinction at critical moments of the scene, which with the carbon arc sometimes results in faulty exposures, is absent with the incandescent lamp. The repetition of scenes due to incorrect exposure therefore need not occur. Even with the usual alternating current supply the light emission of the incandescent lamp is practically constant since the thick filament of the large lamps used for illuminating the film studio possesses a sufficiently large heat capacity to equalize the fluctuations of the energy supply. Because of this, with incandescent lamps there is no danger of stroboscopic effects. With arc lamps, in order to exclude this danger, it was often necessary to use direct current. The latter is obtained by means of motor aggregates, which considerably increase the installation costs.

Further advantages of the incandescent lamp are the following :

1) Simple and cheap operation (no adjustment of the carbons).
2) Ease of transportation of the lighting apparatus.
3) Less danger of fire.
4) Less danger of inflammation of the eyes due to ultra-violet radiation.

The advantages of the incandescent lamp as a source of light for film making were further added to when the sound film made its appearance. Attempts were of course then being made to silence the hissing, sputtering arc, but no complete and reliable success had been achieved, so that the cause of the silent incandescent lamp was quickly won.

The task of the lamp manufacturers now consisted in developing incandescent lamps suitable for the illumination of film studios. In the following we shall give an outline of some of the fundamental problems which were encountered in this development. We shall then discuss the construction and use of the existing film studio lamps and finally the limits of the field of application of the incandescent lamp in the film studio.

Spectral distribution of the light and lifetime

As we have seen, for securing sufficient actinic effect it is desirable to choose the temperature of the filament as high as possible. With every increase of the temperature, however, the life of the filament is shortened and therefore a halt must be called at a certain temperature in order to prevent the use of incandescent lamps for film-making from becoming uneconomical. A life of 100 hours forms a satisfactory compromise. Evaluated according to the actinism of a panchromatic plate, the incandessent lamp with this life delivers about twice as much light as with the life customary for ordinary illumination purposes [2]). The temperature of the filament is about 200° higher than in the case of a life of 1000 hours, and with large lamps amounts to about 3200° K. The content of blue and yellow radiation is high enough at this temperature, so that films can be made not only with panchromatic material but also with orthochromatic material.

When working with colour films, which are at present less sensitive, on the other hand, the actinic effect of the radiation is still on the low side with a colour temperature of 3200°K. Added to this is the fact that the spectral composition of the light at the tempe-

rature mentioned is not entirely suitable for securing optimum colour rendering. Originally the emulsion of the colour film was so tuned that upon the use of arc lamps for projection as well as for photography colours are obtained which approximately resemble those seen by daylight. Later on, in connection with the use of incandescent lamps, the emulsion of the colour film was adapted to a lower colour temperature than that of the arc lamp. It was impossible, however, entirely to reach the colour temperature of the incandescent lamps and a compromise was chosen such that the optimum colour temperature of the lamp for the exposure is higher than 3200° K. This resulted in the acceptance of a shorter life than 100 hours for lamps used for taking colour films.

In practice incandescent lamps with a life of 25 hours and a temperature of about 3300°K are used, which means a colour temperature of 3375°K. If a still higher colour temperature is required it is realized by means of filters, since otherwise the life would probably be much too short. It is clear that the use of filters always involves a loss of light. This device is therefore avoided as much as possible.

Directions for the use of the incandescent lamp in film studios

Just as in the case of the utility illumination of office buildings and factories, a distinction can be made in the illumination of a film set between the general illumination and the local illumination. The general illumination serves to imitate reality as far as light and shadow are concerned; the purpose of the local illumination is usually to produce artistic effects and for instance to focus the attention upon a certain object. The view of a film studio in *fig. 3* provides striking examples of both methods of illumination.

The way in which the camera man works with general or local illumination depends very much upon his personal taste, so that it is very difficult to give generally valid rules. One director prefers a balanced rendering of a whole set, another desires to focus the attention of the audience upon a definite point in every picture. The former will prefer a fairly uniform illumination of the whole, while the latter will keep the general illumination low in order to be able to accentuate the most important details of the picture the more strongly by means of spotlights. He will usually depend entirely upon his own visual impression, since light measurements in the studio take too much time. An exception to this is made in the case of colour films, which, due to the properties of the emulsion, permit much less play in the exposure and with which, moreover, the camera man has not yet obtained as much experience as with black-and-white films.

A result of these very divergent methods of illumination is a great variety in the technical devices used. Nevertheless, some attempt will be made to make some restrictions and to design lamps and lighting fixtures which can as far as possible be used universally. By the above-mentioned division of the lighting installations into general illumination and spotlights this object is found to be best achieved: the general illumination has relatively little to do with the artistic part of the problem of illumination, so that certain general guiding principles can be applied to it. On the other hand the spotlights are so constructed that they are easily transported and adjusted, so that with relatively little variation of the lighting apparatus a great variety of methods of illumination is possible. In the following we shall discuss the most important constructions for the two lighting systems.

General illumination

The required intensity of the general illumination depends upon many factors, such as the nature of the object to be photographed, the lens aperture of the camera used, the transport velocity of the film, the sensitivity of the emulsion and the process of development chosen. The intensity of the accentuation by spotlights which are to be used in addition to the general illumination will also be of importance for the general level of illumination.

Ordinarily with indoor scenes illumination intensities between 1000 and 4000 lux are used. In landscape sets (built up in a film studio) higher intensities of illumination, for instance 10 000 to 14 000 lux, are used. From this it follows that very many lamps will be needed for the general illumination, especially since it is advisable not to use too large units with a view to securing a uniform light distribution.

The lamps are mounted side by side in a diffusely reflecting bank or separately in fixtures which are combined to form batteries. These banks or batteries are placed in front of, beside or above the scene to be photographed, *Fig. 4* shows an example; in this case a number of "Argaphoto" lamps[3]) with the corresponding reflectors (type SC 255) are combined to form a battery illuminating the scene to be filmed

[2]) The power supplied to the lamp increases by about 40%; the further improvement is due to the increase in the actinism of the radiation emitted. For data on this subject see the article: Lamps for use in Photography, Philips Techn. Rev. 6, 259, 1941.

[3]) For the "Argaphoto" lamp and the "Photomirenta" lamp see the article referred to in footnote [2]).

from above. The intensity and the concentration of the beam formed can be varied by a special arrangement. For this purpose the toothwheel E is rotated whereupon the rods Z and D, respectively, for the outer and inner ring of lamps, are so manipulated that the individual reflectors are directed more or less toward the outside.

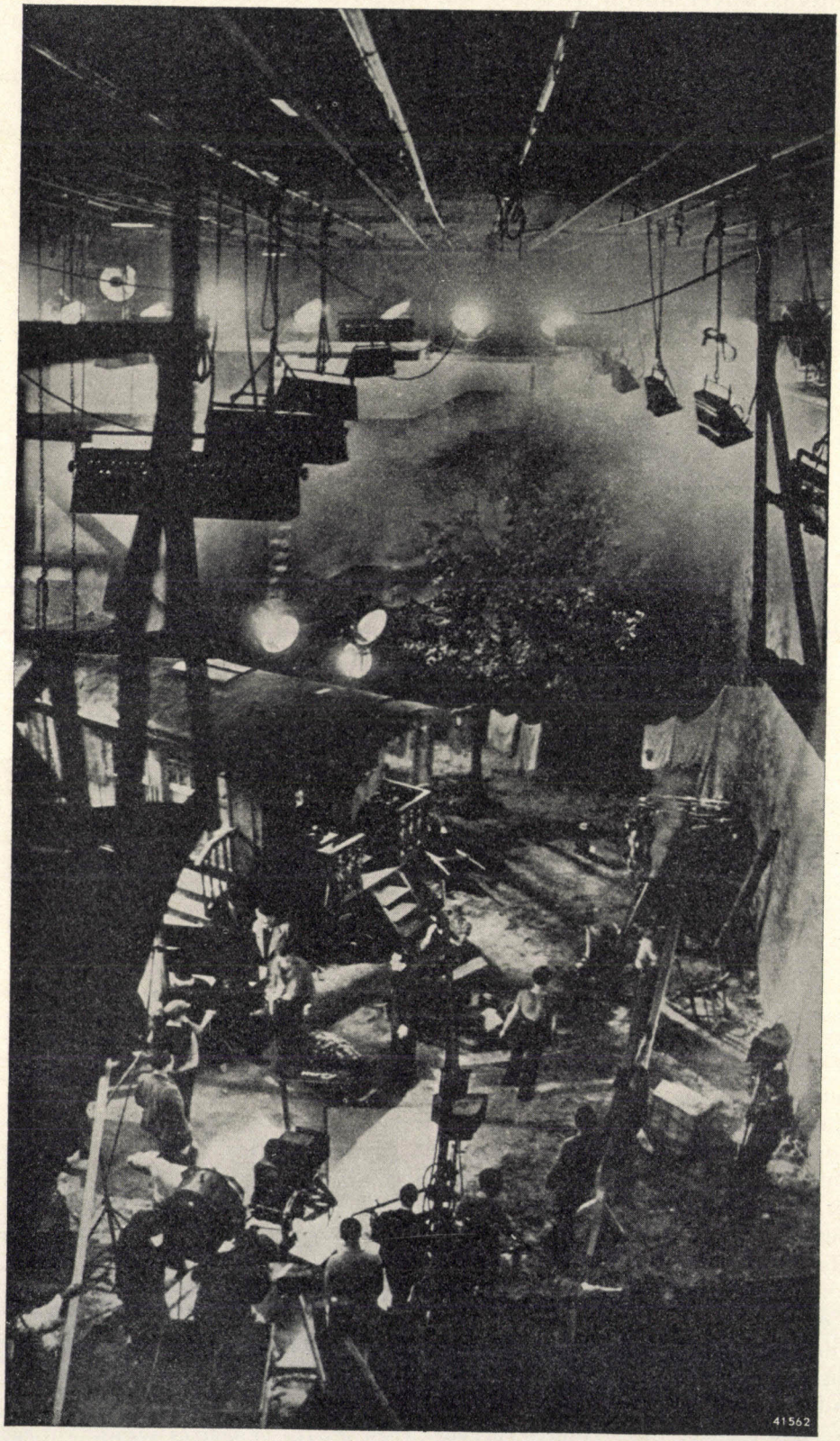


Fig. 3 View of a film studio. Hanging from the ceiling may be seen numerous lights for the general illumination, while a number of spotlights for local illumination are seen on the ground.

Instead of "Argaphoto" lamps, "Photomirenta" lamps, also of 500 *W*, can be used .The reflectors are then superfluous since the "Photomirenta" lamp itself contains a reflector in the form of an internal silver mirror as rear wall. Finally tubular incandescent lamps can also

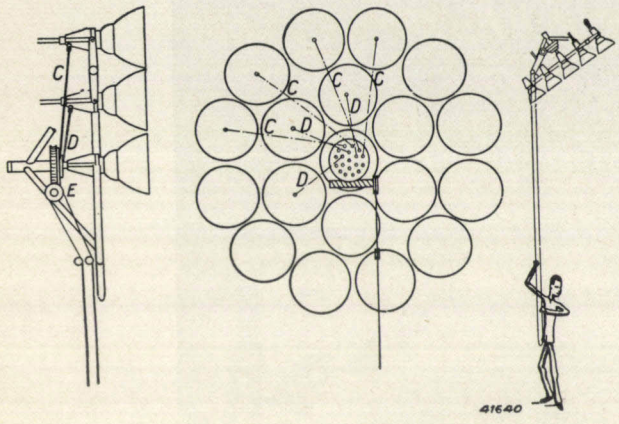


Fig. 4. Battery of "Argaphoto" lamps for general illumination.

very well be used for the general illumination. These lamps with their linear filament are especially suitable for obtaining a uniform illumination. Suitable models for film studio illumination are the "Linea" lamps of 250, 500 and 1000 W (see *fig. 5*).

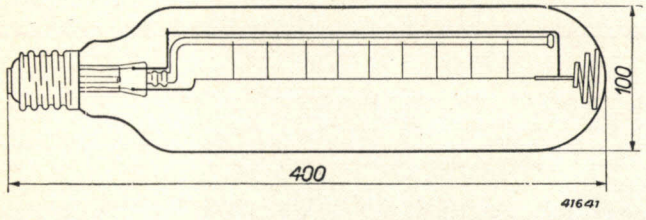


Fig. 5. The "Linea" lamp of 1000 W for general illumination of film studios.

In this connection it may be noted that the oldest type of the low-pressure mercury lamp,the so-called "Cooper-Hewitt" lamp, has a form very much resembling that of the "Linea" lamp. In the early days of film technique when emulsions were much less sensitive, so that it was very difficult to obtain sufficiently strong illumination, use was in fact made of the very good actinic effect of mercury light. This use of the mercury lamp began in 1909. Later, when red-sensitive films were brought on the market, the advantage of the mercury lamp was of much less significance. Since, moreover, the incorrect colour rendering under mercury light was felt more and more as an objection as the requirements as to quality increased, the use of the low-pressure mercury lamp in film studios was given up entirely.

In recent times, however, the use of the low-pressure mercury lamp has experienced a vigorous revival due to the appearance of mercury-discharge lamps with fluorescent tubes. These so-called TL lamps [4]) can be manufactured with a spectral composition which very nearly resembles that of daylight and which is very suitable for cinematographic purposes.

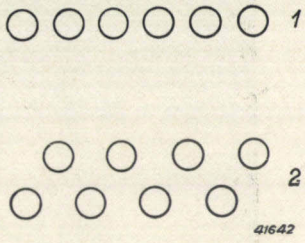


Fig. 6. Arrangement of the parts of the helix in the single-plane filament and in the bi-plane filament.

Spotlights

For spotlights lamps with a small filament are needed. In order to obtain a beam of perfectly parallel rays, these rays would have to start from the focus of some kind of optical system consisting of lenses or parabolic mirrors or a combination of the two. Since, however, exact parallelism of the rays in the beam is not required

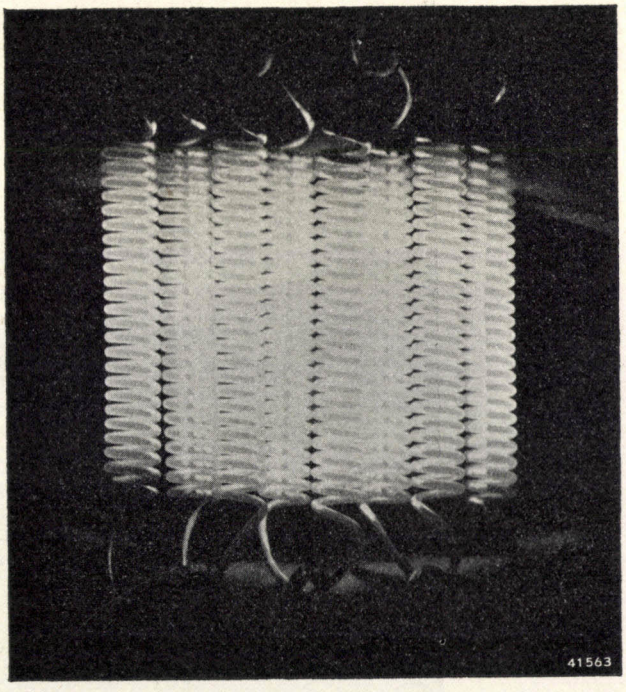


Fig. 7 The bi-plane filament seen from the front. The two groups of helixes supplement each other to give a homogeously luminous plane.

[4]) See Philips Techn. Rev. **3**, 272, 1938; **4**, 337, 1939; **6**, 65, 1941.

and is even undesired in connection with the sharpness of shadows, a certain amount of space around the focus is available for the filament. If the optical system is not to become too large, however, this space remains so limited that it is desirable to concentrate the filament as much as possible.

The way in which attempt are made to realize this concentration of the filament depends upon the desired light distribution over the cross section of the beam. Most film studio lamps are made with single coils, which are placed as shown in *fig. 6* in one plane or in two planes one behind the other. In the latter case a very high average brightness is obtained in the direction perpendicular to the plane of the coils, since the luminous plane can be practically entirely filled with incandescent wires (see *fig. 7*). Upon deviation from this direction, however, the brightness decreases more rapidly than for a single-plane filament, since the coils then partially overlap.

The ordinary powers of the lamps for spotlights are 500, 1000, 2000, 3000, 5000 and 10 000 watts. In order to keep the dimensions of the bulb as small as possible in the high-power lamps, which is of importance for the construction of the fixtures, the bulbs of the types for more than 500 W are made of hard glass. Two examples of the construction of the 5000 W film studio lamp with single-plane filament are given in *fig. 8*. Attention is called especially to the unusual inner construction, the object of which is to make the lamp mechanically as strong as possible. The holder is provided with two pins, which if desired may be replaced by flexible cables.

In the construction reproduced in *fig. 8a* it may be seen that there is quite a long section below the bulb, namely the neck and the holder. By using a modern technique of fusing in, in which the copper pins, which serve simultaneously for fastening and current supply, are fused directly to the glass, this section could be made much shorter (see fig. *8b*). In spite of the fact that the coefficients of expansion of hard glass (40×10^{-7} per degree) and copper (180×10^{-7} per degree) are very different, this seal can easily be made provided the upper edge of the collar of the pin is made sufficiently thin to be deformed before the deforming forces give rise to intolerable stresses in the glass.

A disadvantage of the small dimensions of the bulb is that owing to the deposition of sputtered tungsten the surface of the glass is more strongly blackened in the course of time than in the case of ordinary lamps. At the end of the life of the lamp the absorption means not only an important loss of light but results at the same time in an extra heating of the bulb wall, which

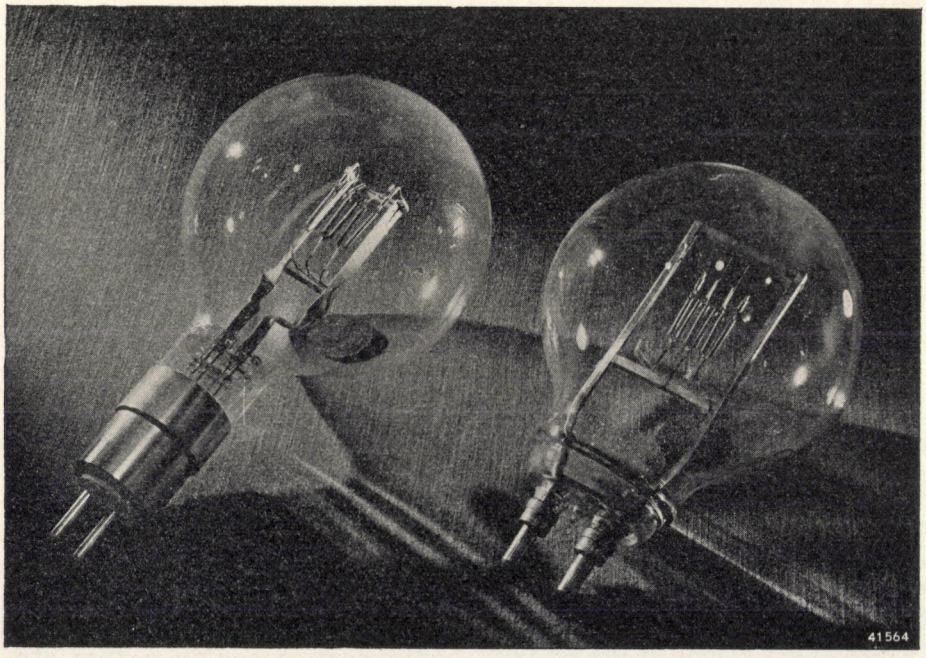


Fig. 8. Two models of the 5000 W film studio lamp with single-plane filament. In type *a*) the bulb has a fairly long neck. This could be very much shortened by fusing the pins directly to the glass (type *b*). The metal powder in the bulb (tungsten) serves to scour the blackness off the inside of the bulb. The diameter of the bulb is about 200 mm.

is greater the smaller the diameter of the bulb. This sets a limit to the reduction of the diameter.

In order to prevent loss of light and to be able to make the lamps smaller, a metal powder is introduced into the lamp to scour the sputtered material from the bulb wall upon shaking. This metal powder can clearly be seen at the bottom of the bulbs of the film studio lamps shown in *figs. 8a* and *b*.

The most commonly used optical systems for concentrating the light of the film studio lamps are the following:

1) Smooth parabolic mirrors. These give a narrow powerful beam. A disadvantage, however, is

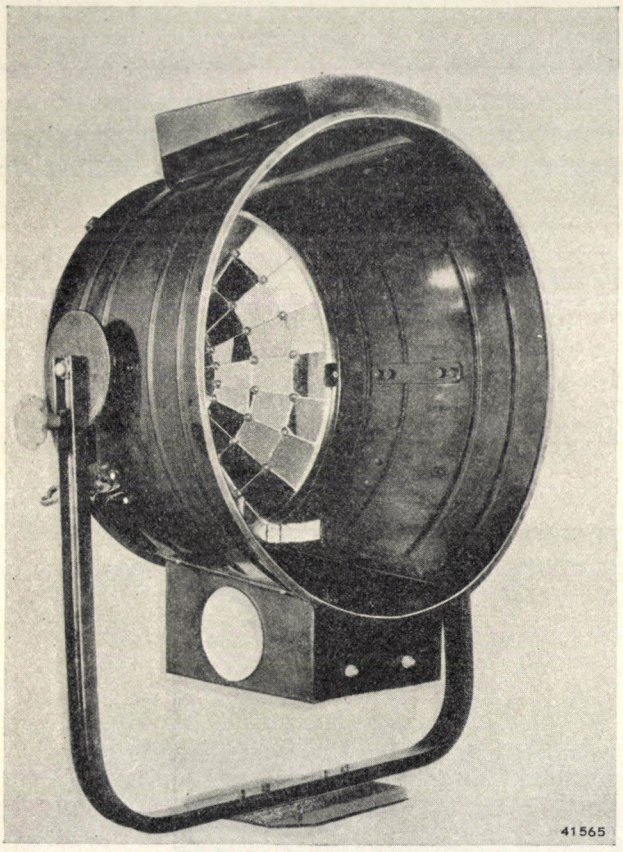


Fig. 9. Parabolic mirrors for film studio lamps. The curved mirror surface is replaced by a number of facet mirrors in order to prevent a sharp projection of the image of the filament.

that it is difficult to prevent the image of certain parts of the coils from being projected somewhere in space. If an actor accidently occupies that position, the contours of the filament become disturbingly visible in the picture.

2) Apparatus in which parabolic facet mirrors are mounted (see *fig. 9*). Small plane mirrors are fastened to the parabola and a broad uniform beam is obtained. *Fig. 10* shows the light distribution of such a spotlight with a film studio lamp of 115V, 5000 W.

3) Apparatus in which a combination of spherical mirrors and Fresnel lenses [5]) is used (see *fig. 11*). The spherical mirror provides that the spaces between the various coils of the single-plane filament are filled by the

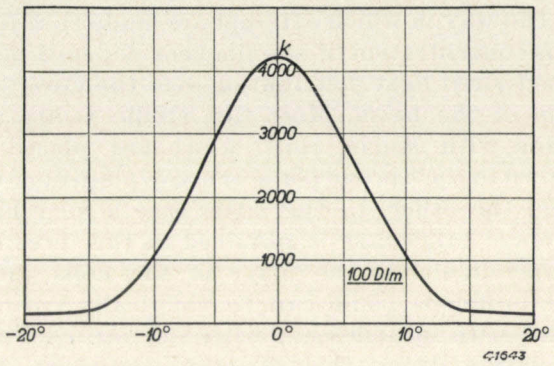


Fig. 10. Light distribution over the cross section of the beam which is obtained with the help of the reflector shown in fig. 9 and a film studio lamp with single-plane filament (115 V, 5 kW).

mirror images of these coils. When a bi-plane filament is used this mirror is unnecessary. *Fig.* 12 shows a model of such an apparatus with a film studio lamp of 3000 W. This construction is the most complicated, but it also has the highest efficiency. About 20 percent of the light emitted is concentrated into a beam by the apparatus shown, while the parabolic mirror with facets shown in fig. 9 has an efficiency of only 9 percent.

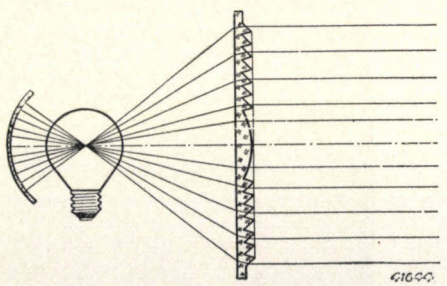


Fig. 11. Beam formation of the light of an incandescent lamp with a spherical mirror and a Fresnel lens.

Several light sources for special cases

Although the lamps described in the foregoing are already reasonably well adapted to the requirements made in a film studio, there are, however, special cases in which these lamps are less suitable.

[5]) A detailed description of the Fresnel lens may found in the article: Lamps for Lighthouses, Philips Techn. Rev. **4**, 33, 1939.

One of the limits set for the use of the film studio lamps described is connected with their heat development. This restricts the intensity of illumination with these lamps to about 20 000 lux, since a more intense radiation begins to hurt the skin. But even when this limit has not been reached, the temperature in the studio may rise intolerably due to the great energy dissipation. In order to combat this a water-cooled studio lamp has been designed (see *fig. 13*). A glass envelope surrounds the lamp and the water flows between this envelope and the bulb. The greater part of the infra red-radiation is absorbed by this water jacket, while hardly any light is intercepted.

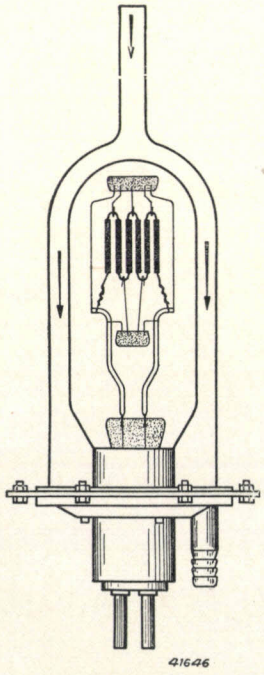


Fig. 13. Film studio lamp with double envelope for water cooling.

Another case in which the film studio lamps are inadequate occurs when an intense beam of relatively small cross section is required. The dimensions of the filament of the ordinary film studio lamps are then still too large for the object in view. A decided improvement is obtained by replacing the single coil filament by a coiled-coil filament. A film studio lamp with coiled-coil filament is shown diagrammatically in *fig. 14*; it may at the same time be seen from this figure how the lamp is placed in the reflector. In this low-voltage lamp the brightness of the filament amounts to 25 cp/mm², while with ordinary film studio lamps (for instance 110 V 3000W) a brightness of only 15 cp/mm² is obtained.

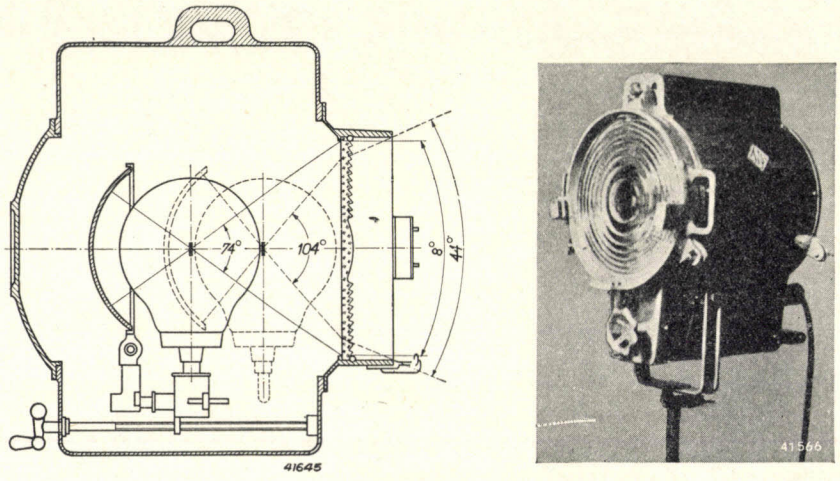


Fig. 12. Reflector with Fresnel lens for film studio lamps. This construction has about twice the efficiency of the spotlight with parabolic mirror.

When the requirements of light intensity or concentration of the beam are very high, it may occur that the object in view cannot be achieved at all with incandescent lamps. In such cases one may return to the carbon arc lamp, which possesses a much smaller luminous surface with the same light flux. If because of the objections

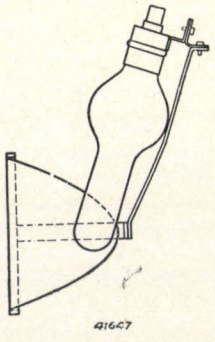


Fig. 14. Film studio lamp with coiled-coil filament (24V, 1000 W). Due to convection in the bulb the sputtered tungsten particles are carried upward and deposited in the wide part of the bulb.

mentioned it is not desirable to use a carbon arc lamp, there is at present the possibility of using water-cooled mercury lamps of super high pressure. In contrast to mercury lamps of low pressure the high-pressure mercury lamps give a practically white light which can be used without hesitation for making black and white films. With this new light source, the brightness of which corresponds approximately to that of the high-intensity arc lamp, experiments have been carried out continuously during recent years. In addition to the uniform and noiseless performance another special advantage is the small proportion of infra-red radiation, so that an intense illumination with mercury lamps has a much smaller heating effect than the same illumination with carbon arc lamps.

STABILITY AND INSTABILITY IN TRIODE OSCILLATORS

by J. VAN SLOOTEN.

621.396 : 615.1

The state of a triode oscillator whose grid voltage is obtained by means of a condenser and a leakage resistance can be characterized for any moment by the value of the bias (V_g) and the value of the oscillator amplitude (W). Graphically, therefore, the state can be characterized by a certain "operating point" in a W-V_g diagram. Each point of the diagram represents a possible operating state. In general, however, this is not stationary, for in the course of time the operating point will shift. This movement is investigated in detail in this article, and the possible stationary states automatically become evident, while at the same time it becomes clear how the oscillator reaches the stationary state. In certain cases the possible trajects of the operating point do not approach the stationary state but converge to a fairly large loop enclosing this state. One then speaks of blocking of the oscillator. The circumstances under which this blocking occurs are studied in detail. It is found that the chance of blocking can be reduced by making the grid current characteristic of the oscillator triode as steep as possible.

In a previous article in this periodical [1]) we devoted a detailed consideration to the way in which the adjusment of a triode oscillator to a stat ionary state of operation is achieved. There we confined ourselves to a discussion of the simplest connections, which are represented by *figs. 1a* and *b*. In the system *a* the control grid bias is taken from a potentiometer, while in the system *b* it is obtained by means of a grid condenser and a leakage resistance.

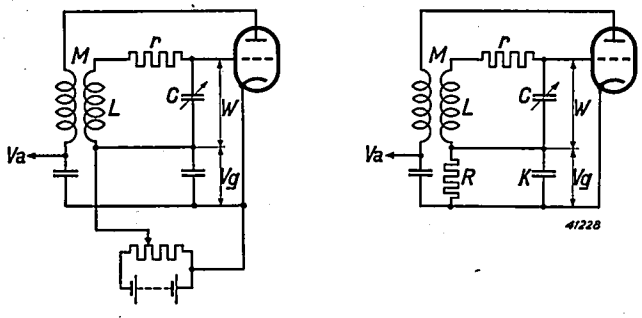


Fig. 1. Diagram of oscillator with triode.
a) The grid bias is excited with the help of a battery and a potentiometer.
b) The grid bias is excited with the help of the grid condenser K and the leakage resistance R.

The latter connections are by far the more important technically, and from the theoretical point of view they also offer the most interesting aspects. One of their remarkable properties, which was already pointed out in the previous article, is that the adjustment to a certain operating point can be stable or unstable according to the magnitude of the grid condenser K. In this article we shall go into the question of the stability or instability of a triode oscillator more thoroughly. We shall begin with a short résumé of what has already been written on this subject, referring for the arguments concerned to the previous article.

The behaviour of the oscillator can best be characterized by plotting in a diagram the amplitude W of the oscillator voltage as a function of the grid bias V_g. In the case of connections *a* this diagram may simply be recorded experimentally by setting the grid bias at different values with the help of the potentiometer and measuring the values of the oscillator voltage which then occur. In the case of connections *b*, on the other hand, the grid bias cannot be chosen arbitrarily, for it adjusts itself automatically to a certain value depending upon the grid resistance R[2]).

Thus in order to record the W-V_g diagram the grid resistance must be varied. When the grid resistance is varied (beginning with small values) the negative grid voltage also rises in the stationary state. Remarkably enough, however, a maximum is then reached for a certain grid resistance and, upon further increase of R, V_g begins to decrease again.

The same holds qualitatively for the oscillator voltage W. With increasing values of R the oscillator voltage first increases, the relation between W and V_g being exactly the same as in the case of direct adjustment of the bias with the help a of potentiometer. When, however, the grid voltage begins to fall again upon further increase of R, the oscillator voltage does n o t follow the same curve in the W-V_g plane but passes through appreciably lower values. This is shown clearly in the diagram reproduced in *fig. 2*.

[1]) Philips Techn. Rev. **7**, 40, 1942

[2]) The grid bias in the stationary state always has a negative value which is slightly less than the amplitude W of the grid AC voltage, so that during a small part of the period the total grid voltage is positive. Only during that part of the period does any appreciable grid current flow. The condenser K is charged by these current impulses, while at the same time it is continuously discharged by the resistance R. The bias adjusts itself automatically to such a value that during a whole period the discharge is equal to the charge in each current impulse. This adjustment is independent of the capacity K, provided the latter is not too small.

An explanation of this remarkable behaviour was given in the article referred to in footnote [1]) by examining for every value of V_b the manner in which the slope S_b necessary for oscillation and the "effective" slope S_{eff} vary

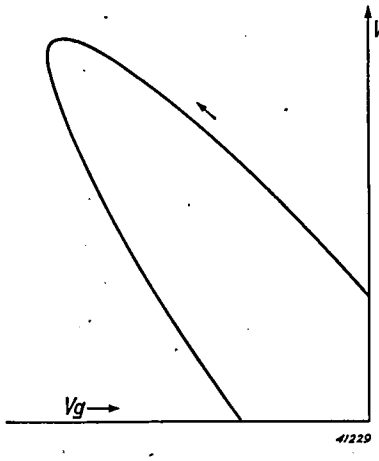


Fig. 2. W-V_g diagram of oscillator connections. Points on the line drawn represent possible stationary states of the oscillator connections according to fig. 1*b*. Upon enlarging the leakage resistance R the curve is traced in the direction of the arrow.

as functions of the oscillator voltage. If, the effective slope is greater than the slope necessary for oscillation the amplitude of the oscillation will increase, while in the reverse case it will decrease. A stationary value of the amplitude must thus correspond to a point of intersection of the curves representing S_b and S_{eff} as functions of W.

Now the curves for S_b and S_{eff} at the ordinary values of the grid bias exhibit the form shown in *fig. 3*, there being two points of intersection B and A. The values of W corresponding to these two points of intersection are found to correspond to the oscillation amplitudes which are observed with a given value of V_g respectively on the ascending and descending branches of the curve for the motion of the stationary operating point in the W-V_g diagram.

In this way the remarkable form of the W-V_g diagram can be explained in principle. Various questions, however, still remain open for discussion. While the stability of a stationary state on the upper branch of the curve in the W-V_g diagram is immediately ensured, this is not the case on the lower branch (point A). It is easily understood that the intersection point A, at a constant value of the bias, *i.e.* in connections according to fig. *1a*, represents a labile adjustment. If, for example, due to an accidental fluctuation the oscillation amplitude is made slightly greater than corresponds to the intersection A in fig. 3, the effective slope S_{eff} will become greater than the required slope S_b, so that the oscillation amplitude will further increase and therefore not return to the stationary value. The same labile behaviour is found for a fluctuation in the opposite direction.

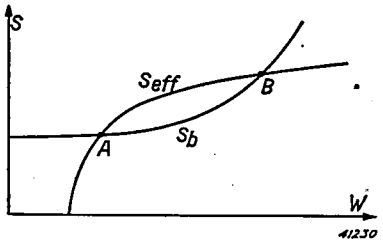


Fig. 3. Required slope S_b and effective slope S_{eff} as functions of the amplitude W of the oscillation. The slopes depend actually not only on W but also on the grid voltage V_g. In drawing this diagram a given constant value of V_g has been assumed.

On the other hand, in the case of the oscillator connections with grid condenser and leakage resistance it is found that the adjustment to the operating point A may be stable under certain circumstances. When, for example, in these connections the amplitude W deviates from the stationary state due to a fluctuation, this results, as far as the further behaviour of the oscillation is concerned, not only in a change in W but also in a change in V_g, and in the previous article we have attempted to prove that this change in V_g has a stabilizing effect. In this article we shall now study in more detail the motion of the state point of a triode oscillator in the W-V_g plane in order to obtain a more accurate insight into the conditions under which this stabilizing effect is sufficient to make a stationary adjustment possible.

The behaviour of the operating point in the W-V_g plane

In order to investigate the character of the motion in the W-V_g plane, we shall first determine the factors which may cause the grid voltage V_g to change. To each set of values of W and V_g there belongs a certain value $\overline{i_g}$ of the average grid current. In order to keep in mind that this value is a function of W and V_g we indicate it by $\overline{i_g}\ (W, V_g)$.

The current through the leakage resistance R in fig. 2 is given by

$$i_R = \frac{V_g}{R}.$$

It is now clear that the condenser K is charged by an average current.

$$i_K = \overline{i_g}\ (W,\ V_g) - i_R.$$

As a result the condenser voltage varies according to the known law:

$$\frac{dV_g}{dt} = \frac{i_K}{K} = \frac{1}{K}\left\{\overline{i_g}\ (W,\ V_g) - \frac{V_g}{R}\right\} \quad \ldots\ (1)$$

This equation, in which the negative grid voltage is considered as a positive quantity, describes the manner in which V_g varies. To this we shall add a second equation describing how W will vary. To do this we make use of the theorem that the AC voltage in a damped L-C circuit with the series resistance r in free oscillation decreases according to

$$\frac{dW}{dt} = -\ W\ \frac{r}{2L},$$

which result is obtained by differentiating the familiar expression

$$W = W_0 e^{-\frac{r}{2L}t}$$

with respect to time.

Now the series resistance present in the oscillation circuit is equal to the resistance of the circuit r decreased by the negative damping which occurs in the oscillation circuit as a result of the back coupling and which can be represented in the diagram by a negative resistance in series with the oscillator coil. Expressed by the usual symbols this negative resistance is given by

$$r' = \frac{MS_{eff}}{C} \quad \ldots\ (2)$$

M is here the mutual induction of back-coupling coil and circuit coil, while S_{eff} represents the effective slope, which is a function of W and V_g.

Besides the resistance r of the circuit itself and the negative resistance r' still a third resistance appears due to the damping resulting from the flow of grid current. Since, however, this grid current is very small at the ordinary adjustment of the oscillator, we may disregard its damping effect[3]). The equation for the variation in W thus becomes:

$$\frac{dW}{dt} = -\frac{r-r'}{2L}\ W \quad \ldots\ (3)$$

We can now eliminate the time variable dt immediately from (1) and (3) and we then obtain:

$$\frac{dV_g}{dW} = -\frac{2L}{KW}\ \frac{i_g\ (W,\ V_g) - \dfrac{V_g}{R}}{r-r'\ (W, V_g)} \quad \ldots\ (4)$$

By $r'\ (W, V_g)$ is meant that the above-defined resultant negative resistance r' (like Ig) is a fuction of W and V_g. From equation (4) it therefore follows that for each set of values of W and V_g, thus for every point in the $W-V_g$ diagram, the direction in which the operating point moves is determined.

When these directions have been determined the solution of the differential equation (4) consists in drawing a smooth line which everywhere follows the local direction.

Although it would be a complicated business to indicate completely the directional field in the W-V_g plane, its qualitative character can be very simply determined. For this purpose we indicate the lines where dW/dt and dV_g/dt respectively are zero. On the first mentioned line W remains constant, while V_g may vary: the operating point is thus displaced horizontally. In the same way it may be seen that the last mentioned line is the locus of all points where the directional field is vertical. Any points of intersection of the curves $dW/dt = 0$ and $V_g/dt = 0$ represent operating points for which both W and V_g are constant: these operating points are thus stationary working states which may or may not be stable.

The line $dW/dt = 0$ is determined by the relation already discussed between W and Vg in fig. 2, which is again reproduced in fig. 4 by curve I. Along this line the required slope is equal to the effective slope and is the condition for the constancy of the amplitude of the oscillation. In the region enclosed by the curve $S_{eff} > S_b$, *i.e.* the oscillation is built up ($dW/dt > 0$), while outside the curve I $S_{eff} < S_b$ and therefore $dW/dt < 0$.

In order to determine the line $dV_g/dt = 0$ we must ask what value V_g finally assumes

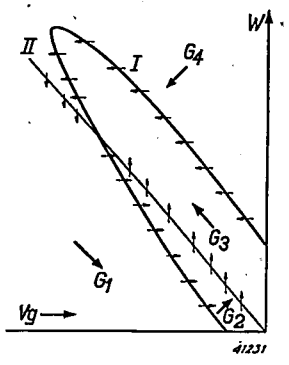


Fig. 4. W-Vg diagram of oscillator connections. Curve I, which is also drawn in fig. 2, represents the locus of all points with stationary oscillator amplitude W; curve II represents the locus of all points with stationary grid bias V_g. The point of intersection indicates the stationary working state. The arrows indicate the direction in which the operating point moves in the different regions.

[3]) In principle this effect can also be taken into account by including the grid current damping in the total (negative) damping resistance r' of the triode. In that case equation (2) would have to be replaced by a more complicated expression. For the less usual adjustments on the upper branch of the curve in the W-Vg plane this correction is important.

with a given value of W. This will depend upon the magnitude of the leakage resistance. If the leakage resistance is infinitely large, the required curve is a straight line at an angle of 45° to the V_g-axis. A stationary adjustment is only reached when the condenser K is charged to such a point that the negative bias V_g is equal to the peak value W of the AC voltage. If the leakage resistance is not infinitely large the line is steeper and has, for instance, the shape indicated by curve II in fig. 4. It will be clear that in the region lying under line II the grid current is smaller than the current through the leakage resistance, so that the condenser is discharged ($\mathrm{d}V_g/\mathrm{d}t \vee 0$). In the same way in the space above line II $\mathrm{d}V_g/\mathrm{d}t < 0$.

On the basis of these properties of curves I and II we can now distinguish four regions in the W-V_g plane, which are indicated in fig. 4 by G_1 to G_4 inclusive and are bounded as follows:

G_1	below	curve II	and	outside	curve I
G_2	,,	,, II	and	inside	,, I
G_3	above	,, II	and	inside	,, I
G_4	,,	,, II	and	outside	,, I

In these four regions the operating point of the oscillator will move in different ways and, as will be clear from the foregoing, it will move in the direction of the large arrows drawn in fig. 4. Each of these four arrows thus represents a region of directions which includes an angle of 90° between the horizontal and the vertical. The small arrows cutting the curves I and II indicate the direction in which the boundaries of the regions are passed.

Examples

With the help of the directions of motion in the W-V_g plane thus found we shall now examine what happens when the oscillator is left to itself with given but otherwise arbitrary values of W and V_g. In practice such an initial state could be realized by placing a battery in parallel with the condenser K and impressing on the L-C circuit a certain value of W by coupling with a generator which produces an oscillation of the same frequency as the oscillator. If the battery and the generator are now removed simultaneously, the oscillation then begins with these given values of W and V_g. A moment later, however, other values of W and V_g will have adjusted themselves, whereby the operating point in the W-V_g plane always follows the arrows of the field of directions.

As a first example we shall consider an oscillator whose grid resistance R and condenser K are very large. It is found experimentally that with such an oscillator no stationary working state is possible, but b l o c k i n g occurs. The oscillation begins to build up, the amplitude W and the negative bias V_g both increasing. The maximum of W, however, is quickly reached and a moment later the oscillation breaks off ($W \to 0$). The condenser K is then discharged *via* the resistance R, so that V_g also decreases again. The operating point in this case moves in the W-V_g diagram along the V_g-axis, as indicated in *fig. 5* by the arrow. At a given moment it will pass curve I and in doing so enter the region in which the oscillation of the oscillator can automatically build itself up. From that moment the value of W will again increase.

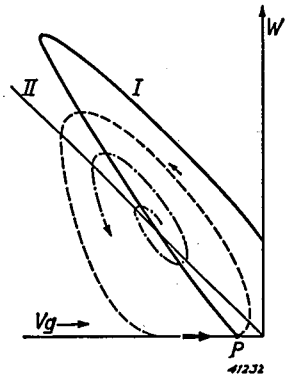


Fig. 5. Motion of the operating point in the W-V_g plane with a large value of the grid resistance R and the grid condenser K. The paths converge to a closed loop; a stationary adjustment is not reached.

The resulting motion of the operating point is indicated by the broken-line curve. As may be seen, the arrow direction dicussed in connection with fig. 4 has everywhere been followed. Typical of blocking, therefore, is the fact that the AC voltage decreases to a very small value before the grid condenser has been sufficiently discharged to cause the oscillation to be built up again.

Under the circumstances considered (large values of R and K) the path of the operating point also passes over to the broken-line curve when it is left to itself at a point of the W-V_g plane not lying on this curve[4]). If for example, the path of the operating point begins somewhere inside the closed loop, it will then describe a spiral towards the outside which finally passes over into the loop. A quantitative formulation of the condition under which this phenomenon takes place will be given later in this article.

---

[4]) The mathematically-minded reader will have noticed that the closed curve which the oscillator point describes upon blocking represents a "cycle limit" of the differential equation for the motion of the operating point.

Fig. 6 shows the shape of the path of the oscillator point when the grid condenser K is so far reduced that the oscillator becomes stable. The operating point in that case describes a spiral toward the inside from a given point in the W-V_g plane, which spiral approaches the stationary adjustment, namely the point of intersection of curves I and II. This point of intersection, since V_g as well as W is stationary, represents the only possible stable adjustment.

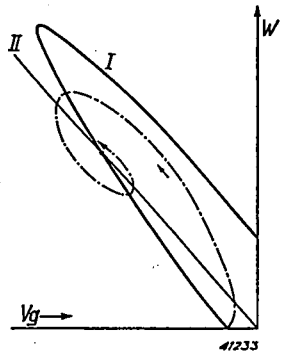


Fig. 6. Motion of the operating point in the W-V_g plane with a value of the grid condenser K smaller than in fig. 5. The operating point describes a spiral which converges towards the stationary state.

The character of the motion of the operating point is again different when the grid resistance is chosen so small that the point of intersection of curves I and II of fig. 4 falls in the upper branch of curve I. It may be recalled from the previous article that an operating point on the upper branch of curve I is always stable and may be realized with the aid of a bias voltage battery instead of with a grid resistance. The diagram in question is now sketched in *fig. 7a*. It may be seen that, retaining the arrow directions indicated in fig 4, the operating point cannot leave region G_3 when it has once entered it. The operating point thus approaches the point of intersection asymptotically without moving around it as in fig. 6.

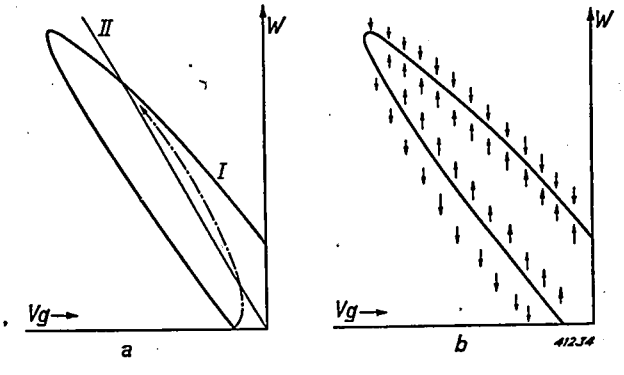


Fig. 7. *a*) Motion of the operating point with a low value of the grid resistance R. The stationary state is situated on the upper branch of curve I. It is reached without the operating point spiraling around it.
b) Motion of the operating point with oscillator connections with a fixed bias. The direction of the motion in the diagram is always vertical. Stable states are only possible on the upper branch of the curve.

If the bias is exited with a battery the field of directions in the W-V_g plane changes in the manner indicated in fig. *7b*. Each arrow of fig. *4* must now be replaced by its vertical component, since movements of the operating point in the horizontal direction do not now occur. It is clear from this diagram that the operating points on the upper branch of the curve are stable and on the lower branch labile.

Summarizing, three cases can be distinguished:

A) *Instability or blocking.* The operating point of the oscillator describes a closed curve in the W-Vg plane or approaches such a curve.

B) *Stability of the first type.* The operating point moves along a spiral towards the inside in the direction of the stationary state.

C) *Stability of the second type*, also to be realized with a fixed grid voltage. The oscillator point moves toward the stationary state from one side only.

For the sake of completeness it should be remarked that an intermediate case is possible between *A*) and *B*) in which in the case of a s m a l l deviation the stationary state is recovered by a spiral motion towards the inside, while a l a r g e deviation from the stationary state leads to a spiral motion towards the outside which passes over into the closed curve mentioned under A). This behaviour, of which *fig. 8* gives an indication, may also occur in practice, although it may be considered more or less a rarity.

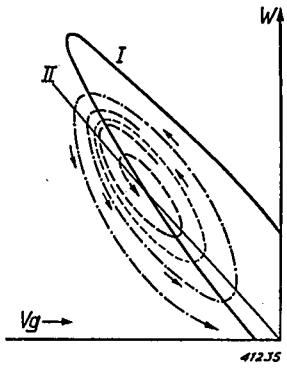


Fig. 8. Motion of the operating point with an oscillator in which, according to the initial point of the motion, blocking as well as stable oscillation may occur. Operating points inside the broken-line curve give a converging spiral, those outside a diverging spiral.

Calculation of the stability

The foregoing considerations were of a somewhat qualitative nature. We shall now study

quantitatively the behaviour of the operating point of the oscillator in the immediate vicinity of the stable or unstable stationary state, in order to furnish evidence that the spiral motions towards the inside or towards the outside, can actually take place.

For this purpose we assume that the coordinates V_g and W of the operating point deviate by small amounts ΔV_g and ΔW from the coordinates of the intersection point of curves I and II representing the stationary adjustment. How will the operating point then move?

The change in V_g is described in general by equation (1). Since no change occurs in the point of intersection itself we obtain from this for small deviations

$$\frac{dV_g}{dt} = \frac{d\Delta V_g}{dt} = \frac{1}{K}\left\{\frac{\delta \bar{i}_g}{\delta V_g}\Delta V_g + \frac{\delta \bar{i}_g}{\delta W}\Delta W - \frac{\Delta V_g}{R}\right\} \quad \ldots\ldots (5)$$

In this equation the negative grid voltage V_g is again considered as a positive quantity. Connected with this is the fact that $\delta i_g / \delta V_g$ will be negative, since with constant oscillation amplitude W the grid current will decrease with increasing negative bias.

The change in the oscillation amplitude $(d\Delta W/dt)$ in the vicinity of the operating point can now be derived in exactly the same way as the change in the bias. From equation (3) it follows immediately that

$$\frac{dW}{dt} = \frac{d\Delta W}{dt} = \frac{W}{2L}\Delta r' \quad \ldots (6)$$

As already mentioned, r' is here a negative substitute resistance which must be imagined to be in series with the coil of the oscillator circuit in order to represent the effect of the back-coupling coil and of the valve on the oscillator circuit. Since this negative resistance is given by equation (2), the following is true:

$$\Delta r' = \frac{M}{C}\left(\frac{\delta S_{\text{eff}}}{\delta W}\Delta W + \frac{\delta S_{\text{eff}}}{\delta V_g}\Delta V_g\right)$$

and by substituting this in equation (6) we obtain

$$\frac{d\Delta W}{dt} = \frac{WM}{2LC}\left(\frac{\delta S_{\text{eff}}}{\delta W}\Delta W + \frac{\delta S_{\text{eff}}}{\delta V_g}\Delta V_g\right) \quad (7)$$

Equations (5) and (7) together describe the behaviour of the oscillator upon the assumed deviation from the state of equilibrium. For the sake of clearness we write these equations of motion of the operating point in the form:

$$\left.\begin{aligned}\frac{d\Delta W}{dt} &= a\Delta W - b\Delta V_g \\ \frac{d\Delta V_g}{dt} &= c\Delta W - d\Delta V_g\end{aligned}\right\} \quad \ldots (8)$$

Here a, b, c and d are quantities defined by (5) and (7) having positive values.

From the two equations (8) we may simply eliminate ΔW or ΔV_g. If we choose the latter we obtain for ΔW a differential equation of the second order:

$$\frac{d^2\Delta W}{dt^2} + (d-a)\frac{d\Delta W}{dt} + (bc - ad)\Delta W = 0,$$

while for ΔV_g exactly the same differential equation may be derived.

In the cases of interest to us $(bc-ad)$ is found to be positive[5]). The solutions of the differential equation then have the character of a sinusoidal oscillation whose amplitude varies with the time according to a power of e, namely proportinal to

$$e^{-(d-a)t}.$$

The spiral motion of the operating point is composed of two such oscillations of W and Vg shifted in phase. A stable adjustment is obtained when the amplitude of the oscillation decreases in the course of time, for which the condition that $d > a$ must be fulfilled, or written out in full :

$$\frac{1}{KR} - \frac{1}{K}\frac{\delta \bar{i}_g}{\delta V_g} > \frac{WM}{2LC}\frac{\delta S_{\text{eff}}}{\delta W} \quad \ldots (9)$$

From this it is immediately clear that the stability can be promoted by a reduction in size of the grid condenser K. Since $\delta \bar{i}_g / \delta V_g$ is negative it further follows from relation (9) that an increase in the slope of the grid current characteristic $(\delta i_g / \delta V_g)$ also has a favourable effect upon the stability [6]). The knowledge of the latter fact has led in practice to new constructions in oscillator valves, for example to the use of a control grid which is closely wound at its ends. The closely wound part draws a grid current at positive grid voltages but has no further effect on the characteristics of the oscillator valve.

---

[5]) The condition that $bc\text{-}ad > 0$ means only that one is concerned with an intersection point on the returning branch of the characteristic. Referring back to fig. 2, we may then say that the slope of curve I is greater than that of curve II. These slopes are easily calculated with the help of equation (8). For curve I $dW/dt = 0$; this curve therefore has a slope $\triangle W/\triangle Vg = b/a$. In the same way one finds for the slope of curve II a value d/c. For a point of intersection on the returning branch therefore $b/a > d/c$ or $bc\text{-}ad > 0$.

[6]) In the graphical method of representation which we have employed in figs. 5 and 6 the stabilizing effect of a steep grid current characteristic can also be detected. If the curve for $dW/dt = 0$ (curve I) is drawn for a steep grid current characteristic, the loop becomes very narrow and the field of directions is so distorted that it becomes almost impossible to draw a spiral motion of the operating point towards the outside.

In connection with these practical consequences we shall go somewhat more deeply into the significance of the grid current characteristic. For that purpose we shall consider fig. 9, in which it is indicated how the negative grid bias and the grid AC voltage together determine the grid current. The relation between grid current and grid voltage is here represented by a broken straight line. The shaded area is a measure of the average grid current $\bar{ig}$.

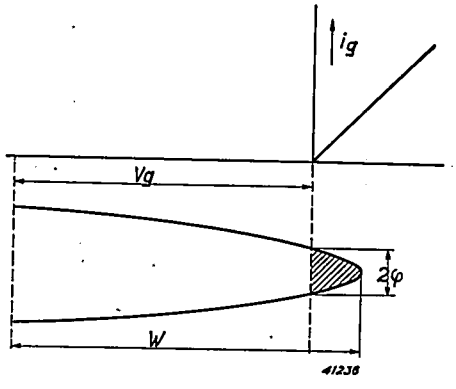


Fig. 9. Diagram for the calculation of the grid current for given values of the grid bias V_g and of the oscillator amplitude W.

For positive grid voltages in the approximation chosen here the grid current is proportional to the grid voltage, so that we may speak of a certain resistance R_i between cathode and control grid. If, further, we indicate by 2φ the phase angle of the part of the period during which grid current flows (see fig. 9) then the average grid current follows from

$$\bar{i}_g = \frac{W}{\pi R_i}\int_0^{\varphi}(\cos x - \cos\varphi)\,dx = \frac{W}{\pi R_i}(\sin\varphi - \varphi\cos\varphi) \quad (10)$$

We may assume that φ is small, so that we may set:

$$\sin\varphi = \varphi - \frac{\varphi^3}{6},$$

$$\cos\varphi = 1 - \frac{\varphi^2}{2}.$$

Equation (10) thus resolves into:

$$\bar{i}_g = \frac{W}{\pi R_i}\frac{\varphi^3}{3} \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

On the other hand between $\bar{i}_g$, V_g, W and R there exists the relation already used in equation (1)

$$\bar{i}_g = \frac{V_g}{R} \approx \frac{W}{R}, \quad . \quad . \quad . \quad . \quad . \quad (12)$$

in which the symbol for "approximately equal to" expresses the fact that V_g and W differ only slightly from each other, which again corresponds to the assumption that φ is a small angle. From (12) and (11) it follows that:

$$\varphi^3 = \frac{3\pi R_i}{R}.$$

With the help of the angle φ thus found we shall now determine the quantity

$$\left(\frac{\delta\bar{i}_g}{\delta V_g}\right)_{(W = \text{constant})}$$

occurring in the criterion of stability, equation (9). For that purpose we make use of the relation:

$$\frac{V_g}{W} = \cos\varphi,$$

from which it follows that

$$\frac{\delta\varphi}{\delta V_g} = -\frac{1}{W\sin\varphi} \approx -\frac{1}{W\varphi}.$$

Bij differentiating (11) we now find that

$$\left(\frac{\delta\bar{i}_g}{\delta V_g}\right) = \frac{W}{\pi R_i}\varphi^2\left(\frac{\delta\varphi}{\delta V_g}\right) = -\frac{\varphi}{\pi R_i} = -\sqrt[3]{\frac{3}{\pi^2 R_i^2 R}} \quad (13($$

In the cases occurring in practice the grid resistance R is large compared with R_i (for instance $R_i = 1000\Omega$, $R = 50\,000\Omega$). From this it follows that the denominator of the term under the radical is much smaller than R^3, so that $\delta\bar{i}_g/\delta V_g$ is many times as large as $1/R$. We may therefore ignore the first term in the stability condition given by equation (9) and then by substituting (13) in (9) we obtain the stability condition in the form

$$\sqrt[3]{\frac{3}{\pi^2 R_i^2 R}} > \frac{1}{2}\frac{M}{L}\frac{K}{C}\,W\frac{\delta S_{\text{eff}}}{\delta W} \quad . \quad . \quad . \quad (14)$$

From this result the same main conclusions may be drawn as from equation (9). The relation between the stability and, respectively, the internal resistance Ri and the external resistance R between cathode and grid is shown more clearly than in equation (9).

THE INVESTIGATION OF TEXTURE WITH ELECTRON RAYS

by J. F. H. CUSTERS.

539.27 : 620.18.

Due to the strong absorption experienced by electron rays in matter, they are more suitable than X-rays for the investigation of the structure of very thin films. As an example of such an investigation the texture is here discussed of aluminium mirrors which are obtained by depositing the metal by evaporation on a cold glass wall or on one heated to 200° C.

It is now known that in some respects moving particles of matter behave like waves of the wavelength

$$\lambda = \frac{h}{mv} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

(m = mass of the particle, v = its velocity, h = Planck's constant). The first experiments from which this could be directly concluded became known several years after de Broglie had set up equation (1) (1924): by allowing electrons to impinge upon a crystal, diffraction and interference phenomena could be observed quite analogous to those which occur upon the diffraction of X-rays, whilst the phenomena were also quantitatively similar when the wave length given by (1) was assigned to the "material waves".

While in the beginning the phenomena in question were mainly of interest for theoretical physics, they soon proved to be extremely useful as an aid in practical research. In the same way as X-rays, upon diffraction at material particles electron rays are able to give information as to the system, orientation, regularity, etc. of the arrangement of these particles, in other words, like X-rays, electron rays can be used for structural analysis.

In this connection there are two properties of the electron rays which are of special importance: 1) there is a much stronger reaction between the electron rays and matter than between X-ray and matter; 2) the electron rays which can be used practically have a much shorter wavelength than the X-rays suitable for structural analysis.

The first difference involves the fact that with electron rays still smaller and more rarefied quantities of matter are sufficient to cause the diffraction phenomena than is the case with X-rays. Thus electron rays are used preferably to investigate the internal structure of gases or for the structural analysis of organic chemical substances of which sometimes only minimum quantities are available. Furthermore, the intense mutual reaction results in the fact that the electron rays are very strongly absorbed. The penetrating power is a factor of the order of magnitude 10^3 smaller than that of X-rays, and after passing through a layer 1000 Å thick, for example, they are already practically entirely absorbed. This fact makes electrons particularly suitable for the investigation of very thin films, such as metal films deposited in some kind of base. With X-rays in such a case not much can usually be done, since the latter, even when they are given a glancing incidence on the surface to be examined, penetrate too deeply, so that the diffraction pattern of the superficial film is drowned out by that of the more deeply lyings parts.

As to the second point of difference, according to (1) the wavelength of the electron rays depends upon the velocity. If the electron ray consists of electrons which are accelerated by an electrical potential difference V then each electron (charge e) has the kinetic energy

$$^1/_2\ mv^2 = eV \quad . \quad . \quad . \quad . \quad . \quad (2)$$

When the velocity v is eliminated from (1) and (2) and the values of the constants e, m, h are filled in one obtains

$$\lambda_{\text{Å}} = \sqrt{151/V_{\text{volt}}} \quad . \quad . \quad . \quad . \quad . \quad (3)$$

If, for example, V = 60 000 volts, λ = 0.05 Å, while the X-rays used for structure analysis have a wavelength of the order of magnitude of 1 Å (for example λ = 1.539 Å for the K_a radiation of copper). The much shorter wave length of the electron rays results in the fact that they are diffracted through much smaller angles than the X-rays. The diffraction of X-rays as well as of electron rays may be considered as a regular reflection at the lattice planes of the crystal lattice of the material investigated, on the understanding that this reflection takes place only when the angle Θ between the incident beam of rays and the lattice plane satisfies Bragg's condition:

$$n\lambda = 2d\sin\Theta \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

(d = lattice plane distance, n = a whole number, the "order" of the reflection). With a given set of lattice planes (d) and a given order n of the reflection, the angle Θ also becomes very small with a small value of λ. If for example d = 1 Å and λ has the value already mentioned of 0. 05 Å, the first order reflection (n = 1) occurs

at $\Theta = 1.5°$. In order to obtain diffraction diagrams of reasonable size which can be measured with sufficient accuracy, therefore, the photographic film on which the diffracted rays are allowed to fall must be placed at quite a large distance from the specimen being investigated.

As an example of the application of electron diffraction we shall discuss here the investigation of aluminium mirrors which are prepared by depositing the evaporated metal on a glass surface. Such mirrors, which are employed because of their good reflection of infrared, visible and ultra-violet radiation, are made in the following way. Two tungsten coils are set up in a vacuum opposite the glass wall to be covered. At regular intervals on one coil short pieces of aluminium wire are fastened. When this coil is heated the aluminium evaporates and is deposited on the glass wall in the form of a mirror whose thickness can be regulated simply by the length of time for which the coil is heated. The mechanical properties of the layer of aluminium, especially the hardness, are, found to depend very much upon the temperature of the glass surface, which can be varied by heating the second tungsten coil to a higher or lower temperature. While a mirror formed by evaporation onto cold glass (18° C) is soft and unless handled very carefully is easily scratched and marred, uton evaporated onto hot glass, for instance 200° C, hard mirrors are obtained which can be used for technical purposes. A necessary condition is that the glass surface must have previously been carefully cleaned of traces of grease and other contaminations. The effect of heating the glass should also be considered as a cleaning of the surface, because when the heated glass wall is first allowed to cool before the metal is deposited upon it a hard mirror is also obtained.

The obvious question is now whether or not the difference in hardness is correlated with a demonstrable difference in structure of the layers deposited by evaporation. Because of the extreme thinness of the films this investigation can be carried out practically only with electron rays.

By allowing a beam of electrons to fall upon the surface of the mirror at a glancing angle as is sketched in *fig. 1*, the diffraction diagrams reproduced in *figs. 2a* and *b* were obtained in the two cases of evaporation on cold and hot glass respectively. In all respects these diagrams resemble diffraction diagrams obtained with X-rays when there are allowed to fall on a (sufficiently thick) layer of a polycrystalline metal : a number of rings, Debije-Scherrer rings, occur, each of which belongs to a definite set of lattice planes of the crystal lattice of the metal concerned. Such a set of lattice planes will lead to reflection of only those crystals which are so

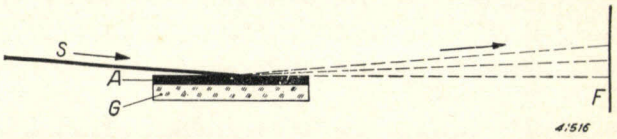
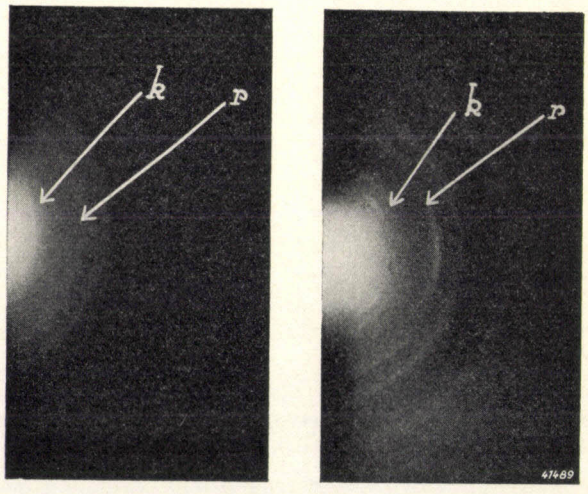


Fig. 1. The electron beam S falls with glancing incidence on the aluminium mirror A deposited on the glass wall G. The rays which are diffracted through very small angles fall on the film F, which is placed at a fairly large distance away (in practice for instance 35 cm).

oriented that the lattice plane makes the angle Θ determined by equation (4) with the electron beam. The rays reflected at this set of lattice planes must therefore always make an angle of $2\,\Theta$ with the direction of the incident beam, *i.e.* they all lie on the surface of a cone around this direction and can only strike the photographic film on the ring where the surface of this cone cuts the film. With glancing incidence of the beam on the specimen being examined there is also the peculiarity that the rays which are reflected toward the inside of the surface are absorbed in the film, so that in the diffraction diagram only half of the diffraction pattern can be seen: it is as if the rings were half covered by the shadow of the plate-shape specimen.

If the crystals in the polycrystalline metal are entirely at random, the lattice plane being considered will occur in all possible positions, the corresponding ring in this case being uniformly black around its whole circumference. This case is obtained in fig. 2*a*. In fig. 2 *b* on the other hand the rings are not uniformly blackened



a *b*

Fig. 2. Electron diffraction diagrams of an aluminium film deposited by evaporation on a cold (*a*) and on a hot (*b*) glass surface. r is the Debije-Scherrer ring for the rhombododecahedron planes, k for the cube planes.

around the whole circumference, from which it follows that in this case the crystals do not occur in all possible positions, but possess a preference for a certain orientation. The film which has been deposited on hot glass and which is much harder thus possesses a texture.

Several months ago it was explained in this periodical [1]) how such a texture can be further investigated and described by means of a so-called pole figure. In the investigation by means of electron rays this process is often simplified thanks to the fact already mentioned that the angles of reflection Θ are so small, namely of the order of magnitude of 1°. The direction of the electron beam itself, due to the slight divergence always present, cannot be defined more exactly than within about 1°, so that it may be said that the reflecting lattice planes must be practically parallel to the incident beam. Furthermore, as is illustrated by *fig. 3*, every lattice plane is perpendicular to the line joining the centre points of the diagram with the interference spot caused by that lattice plane.

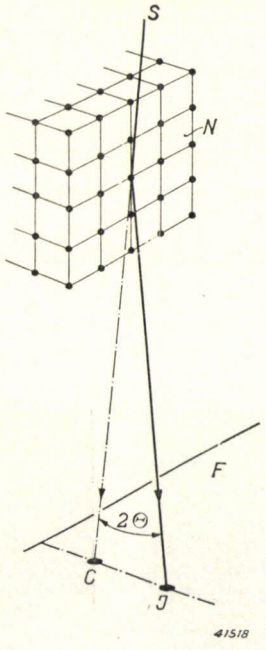


Fig. 3. Reflection of an electron ray at a set of lattice planes. Because of the very small angles of reflection Θ, it may be said that the lattice plane N is practically parallel to the incident beam S. Furthermore the lattice plane is perpendicular to the line joining the spot I which it causes on the film F with the centre interference spot C at which the non-diffracted beam strikes the film.

Let us apply this to the diagram of fig. 2*b* and consider especially the Debije-Scherrer ring (r) which belongs to the rhombododecahedron planes of the cubically crystallizing aluminium. One part in particular of this ring is strongly blackened (white in the reproduction) and the line joining this with the centre is perpendicular to the edge of the shadow of the plate of metal. From this it follows that the *rhombododecahedron planes of the crystals lie preferably parallel to the surface of the mirror.* In *fig. 4* a model is given showing a number of

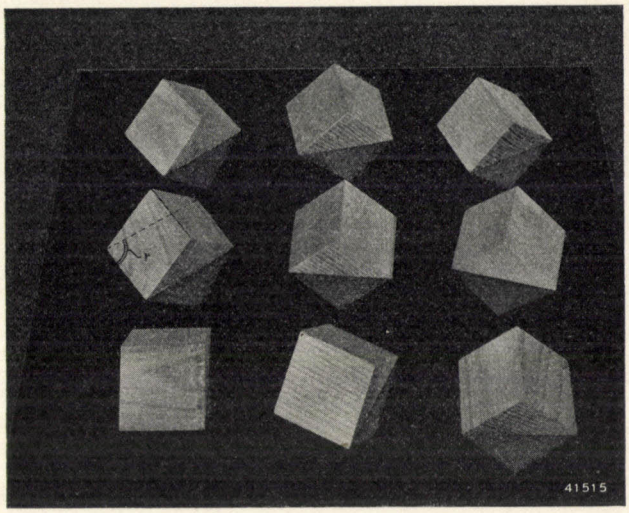


Fig. 4. The aluminium crystals, represented by their elementary cubes, all lie on the glass in such a position that a rhombododecahedron plane r is parallel to the surface.

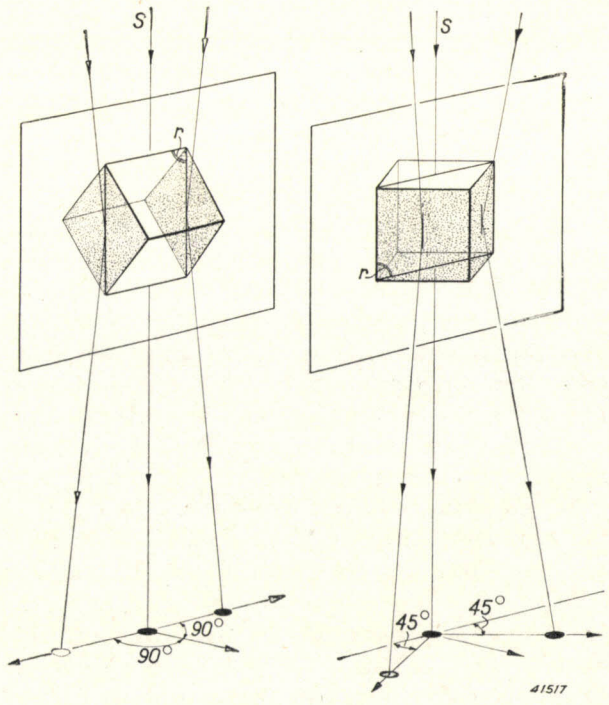


Fig. 5. When a rhombododecahedron plane r is parallel to the surface of the mirror (thus also parallel to the glancing electron beam S), there are only two positions of the crystal in which the cube planes also give rise to reflection. It may be seen that each of two positions produces two interference spots.

[1]) Philips Techn. Rev. **7**, 13, 1942.

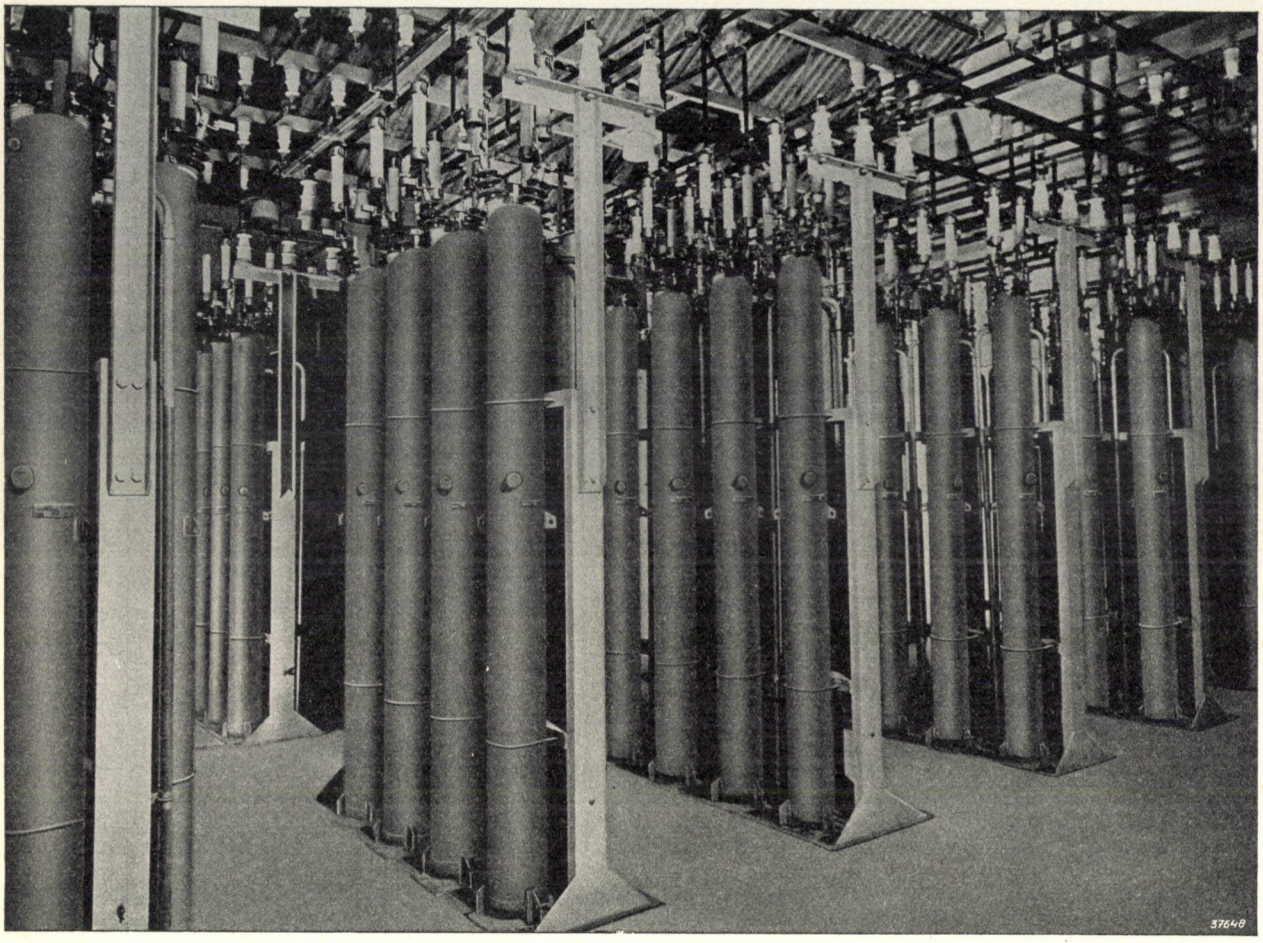
crystals in different positions satisfying this condition.

Now in order to ascertain whether all of these positions actually occur we consider the Debije Scherrer rings of other lattice planes as well, for instance the ring for the cube planes. Out of all the positions of crystals shown in fig. 4 there are only two in which a cube plane lies parallel to the incident beam. These two positions are drawn separately in fig. 5 and at the same time the way in which the reflection will take place is indicated. It is evident that a blackening may be expected on the Debije-Scherrer ring in question at only four points, namely at the intersections with the edge of the shadow and at two points 45° away from that edge. This is actually what is observed in the diffraction diagram. If the mirror is now rotated through any desired angle around the normal to its own surface and a diffraction photograph is again made, exactly the same picture is obtained. From this it follows that then also crystals always occur in the position of fig. 5, in other words all the positions given in fig. 4, in which a rhombododecahedron plane is parallel to the surface of the mirror, actually do occur.

In conclusion it should be mentioned that exactly the same texture as described here is also found in aluminium films deposited by evaporation on molybdenum at 220° C and then heated in a vacuum to 600° C. On the other hand some investigators obtained deviating results in the case of aluminium films on glass, for instance a texture in which a cube edge of the crystals is perpendicular to the surface of the mirror, or where the same is true of a space diagonal of the crystals. Thus in the deposition of the aluminium certain experimental conditions apparently play a part, which conditions must have been different in the different experiments.

A LARGE BATTERY OF CAPACITORS

621.319.4



Battery of capacitors with a reactive power of 10 000 kVA, 6.8 KV/50 c/s, for phase shifting, consisting of 96 Philips pressure capacitors of 105 kVA each.

Attention has already been called in this periodical to the significance of improving the power factor in electric heavy current mains by means of capacitors.

In an AC mains without phase shift the generators at the power station have to supply not only the active power that is transformed in current consumers into mechanical driving power, heat and light, but also the reactive power that maintains the necessary magnetic fields. The production of the active power requires the application of mechanical energy by means of steam engines, water turbines, internal combustion engines. On the other hand, the production of the reactive power does not require the application of any energy and is thus not confined to the location of the power station. It can straightaway be separated from the latter, which may be done by means of capacitors installed in the vicinity of the current-consumer. The reactive current need then no longer follow the long path from the power station over the lines and transformers, which is often more than 100 km (about 65 miles) long. The generators and all the means of transmission lying between the power station and the connecting point of the condensers are relieved of the reactive current. Their capacity increases, and the losses and voltage drops occurring therein are reduced.

Besides the technical advantages with regard to the voltage control, the economic benefits are a saving in expense for the extension of the network of mains on increasing load and a saving in running costs due to lower fuel consumption as a consequence of a reduction in the losses, which may both be very high. At the same time considerable amounts of important raw materials,

[1]) Philips Techn. Review, **1**, 178, 1936.
[2]) Philips Techn. Review, **4**, 254, 1939.

such as copper, aluminium, iron, transformer oil, etc. for enlarging the mains, coal or oil for running, are saved.

Now a few words as to the choice of location for the capacitors. Next to separate installation at the individual works, the concentrated installation of a large condenser capacity at the consumers' end of long transmission lines or at the main point of the reactive load is gradually gaining ground; in this way the power supply company has it in hand to switch capacitors on or off according to the loading conditions.

The above illustration gives an example of capacitors connected at Göteborg (Sweden) to the mains of the Trollhättan power station 140 km (about 90 miles) away. The battery, the reactive power of which is 10 000 kW at 6.8 kV and 50 c/s, comprises 96 Philips pressure capacitors [2]) of 105 RkVA each, which are set up in 12 rows of 8 (diameter 216 mm, height 2735 mm). Each capacitor is protected by high-efficiency fuses. In series with the whole battery are choke coils that consume 4.6% of the mains voltage and detune the frequency of the mains to such a degree that a troublesome amplification of the harmonics of the voltage curve is avoided.

Two further batteries of the same pressure capacitors (5000 RkVA, 6.8 kV each) have been installed at Ornkäldsvik and Brännland on the network of the Norrländska Power Supply Co.

Further to the article mentioned in foodnote 2, we would observe that five years' experience with Philips pressure capacitors for the most varied outputs and voltages is now available and has proved these capacitors to be entirely satisfactory both mechanically and electrically. Almost without exception a refilling of the pressure cylinder with nitrogen was found to be superfluous; in spite of the high field-strength in the dieletric a close inspection did not disclose the slightest traces of ageing phenomena, so that it is perfectly safe to reckon with a practically unlimited life.

THE EQUALIZATION OF TELEPHONE CABLES

by H. van de WEG.

621.392,53

In order to obtain a flat frequency characteristic in a telephone cable a so-called equalization network is connected behind each cable section. In this article the make-up of these networks is discussed and the method is studied which is used to determine the network required for a given behaviour of the cable damping.

When a sinusoidal voltage V_1 is applied to one end of a pair of conductors, a sinusoidal voltage V_2 also occurs at the other end which is in general of smaller amplitude and more or less shifted in phase. This is expressed by writing

$$V_2 = V_1 e^{-g} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

and the transmission is in this way characterized by a "transmission factor" g which is in general complex:

$$g = \ln \frac{V_1}{V_2} = a + jb \quad . \quad . \quad . \quad (2)$$

The real part a of the transmission factor indicates the attenuation (damping), the imaginary part b the phase rotation experienced by the sinusoidal voltage upon transmission through the cable. The transmission factor depends not only upon the properties of the cable, but also upon the terminal impedances at the beginning and the end of the cable [1]). More generally, the transmission over any given network terminated in any manner (quadrupole) can also be characterized by such a transmission factor.

In the transmission of speech, music, telegraphy signals, etc. through the cable AC voltages of different frequencies must be transmitted. For each of these frequencies the factor g may have a different value: the cable (with its termination) has a certain "frequency characteristic". For an *undistorted* transmission of all signals this frequency characteristic must now satisfy two requirements: the damping factor a must be independent of the frequency: $a = a_0$, and the phase factor b must increase proportionally with the frequency ω so $b = \omega\tau$. In that case with any given variation with time $F(t)$ of the signal transmitted, the signal received always has the form $e^{-a_0} F(t - \tau)$, *i.e.* the signal received, apart from an attenuation by a factor e^{-a_0} and a retardation by a time τ, is the same as the signal transmitted at every moment.

[1]) In this connection one also speaks of transmission factor in use, attenuation in use, etc., when one means the transmission upon termination of the cable by the impedance occurring during normal use. If the cable is terminated at both ends by its characteristic impedance, which case practically is the most favourable and theoretically the simplest to deal with, g, then calculated per km length of the cable, is indicated as the propagation constant $\gamma = \alpha + j\beta$.

If we now consider an ordinary telephone cable we find approximately the desired variation with the frequency as far as the phase b is concerned, see *fig. 1*. As to the attenuation a,

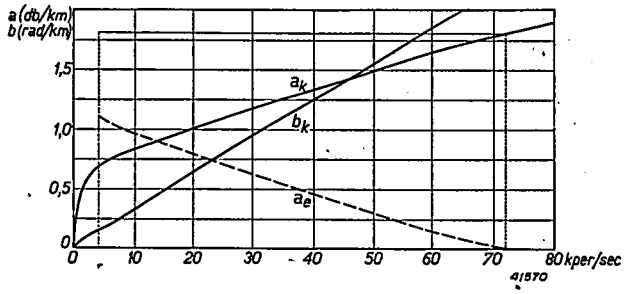


Fig. 1. Attenuation a_k (db/km) and phase rotation b_k (radians/km) of a given cable as a function of the frequency. The broken-line curve indicates the damping a_e which a network should have for the equalization of a section of 1 km of cable in the frequency region from 4 to 72 kc/sec.

however, the requirement of a flat frequency characteristic is by no means satisfied: the damping increases rapidly with the frequency. If the cable is to be used for telephony or telegraphy measures must therefore be taken to make the attenuation independent of the frequency. This can be accomplished by placing between the cable and the receiver a network whose transmission factor g depends upon the frequency in such a way that the cable and network t o g e t h e r have exactly the desired frequency characteristic. Since according to the definitions (1) and (2) the damping a of two transmission elements connected one behind the other is at every frequency simply the sum of the dampings of the separate elements, the network to be added must obviously possess the variation in damping indicated in fig. 1 by the broken line. This process, in which the damping of the cable is added to at each frequency to give the same fixed value, is called e q u a l i z a t i o n of the cable and the network in question is called the e q u a l i z e r.

In most cases no attention need be paid to the phase rotation, since, as mentioned in connection with fig. 1, the deviations of the phase from the desired dependence on the frequency are small; the distortion thereby caused can only be disturbing in the case of very long

cables, for instance for transatlantic communication, or of instance for the transmission of television signals, since the picture is affected by phase rotations to a much larger degree than sound perception [2]). Only in these cases therefore must an "equalization" also be applied for the phase, and certain networks are again used for that purpose.

It may here be noted that equalization is only necessary for the frequency region that is of importance for the signals to be transmitted. For telegraphy, for example, the characteristic need only be made flat to about 75 c/sec.; for the transmission of music to about 10 000 c/sec, etc.

According to the above, equalization always means the artificial increasing of the damping, which from the point of view of economy is undesirable. There is also the possibility of obtaining the desired flat characteristic to a certain extent without extra damping, by changing the cable itself, for instance by coil loading which has previously been discussed in detail in this periodical [3]). In this case the damping in a certain frequency region is not only rendered flat but in addition is considerably lowered. The disadvantage of this method, however, is that in the vicinity of a certain frequency, the limiting frequency, the damping rises so suddenly that the transmission of higher frequencies is quite impossible [4]). In the case of cables for carrier-wave telephony, where a large number of calls are transmitted simultaneously over each pair of conductors by shifting [5]) the speech frequencies to different frequency regions (channels), and where therefore fairly high frequencies must also be transmitted — in a 17-channel system designed by Philips up to 72 000 c/sec, for example —, loading cannot in general be considered. In this case, therefore, in order to obtain an attenuation independent of frequency, recourse must be had entirely to equalization.

It might be asked whether in the case of carrier-wave telephony an equalization over the whole frequency region is actually necessary. The requirement of freedom from distortion of the speech to be transmitted, for which a flat frequency characteristic is necessary, is only made of each separate channel, which for instance in the case of the 17-channel system mentioned includes speech frequencies from 300 to 3400 c/sec. If we consider these successive frequency regions (*fig. 2*) the attenuation is found to vary only very little within each channel. When we consider a cable section 30

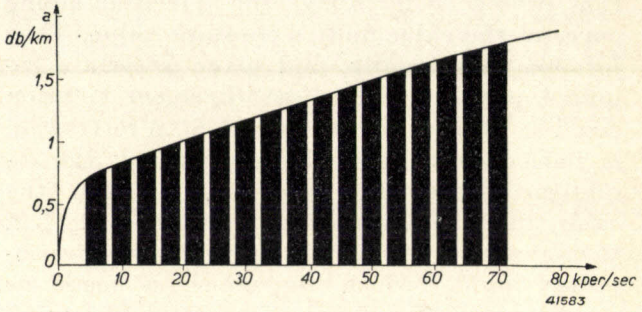


Fig. 2. Cable-damping in the frequency regions of the successive speech channels in a carrier-wave system with 17 channels.

km long the variations for the highest channels still lie practically within the permissible limit of about 1 db, and for the lowest channels they are not greater than 2.5 db, so that as far as the distortion is concerned an equalization in each channel, which can be realized in a simple way, would be sufficient. The attenuation curve would in this way be supplemented forming a kind of step-like curve. For another reason, however, equalization over the whole frequency region from 4 000 to 72 000 c/sec. is necessary, namely in order to avoid overloading of the line repeaters. These repeaters must be of such dimensions that they provide an amplification which is sufficient to bring the highest channel, which is the weakest at the receiving end, up to the desired level. Since with a cable length of 30 km the lowest channels are already about 30 db stronger than the highest,the repeaters would be very heavily overloaded by the lowest channels. The result of overloading is in the first instance a non-linear distortion in which higher harmonics of each speech frequency as well as combination tones of different frequencies occur which may fall in higher or lower channels and thus cause cross-talk between the various channels. Thus in carrier-wave telephony also an equalization network must be placed behind every cable section in front of the input of the following line repeater (*fig. 3*).

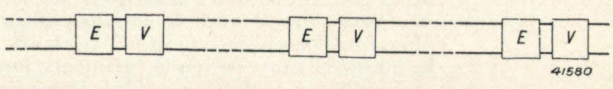


Fig. 3. Cable sections with equalization networks *E* and line repeaters *V*.

[2]) On the influence of the phase on the form of a complex vibration and on the sound impression see: J. F. Schouten, Philips Techn. Rev. **4**, 167, 1939.

[3]) See especially: F. de Fremery and G. J Levenbach, Carrier-wave telephony on coil-loaded cables, Philips Techn. Rev. 4, 20, 1939.

[4]) The damping in the region below the limiting frequency is, moreover, never absolutely constant, so that in addition to loading an equalization must also be applied.

[5]) See: Philips Techn. Rev. **6**, 325, 1941 **7**, 83, 1942; **7**, 104, 1942.

Equalization networks

We shall now examine more closely the networks which can be used for equalization.

Given a cable with a certain damping curve, it is a question of finding a network whose damping factor has a prescribed variation as a function of the frequency, *i. e.* variation which will bring the given damping curve of the cable up to a constant value. Thus formulated generally, and given a certain required precision, the most divergent solution could be proposed for the problem. A restriction is immediately imposed, however, by the requirement that the matching between the cable and the terminating resistance R_0 (*i.e.* the impedance of the final apparatus or the amplifier input, which impedance is made as nearly as possible equal to the so-called wave resistance of the cable) must not be disturbed by the insertion of the equalization network between them. This means that when it is terminated by the ohmic resistance R_0 the equalization network, must also exhibit the resistance R_0 as input impedance. At the same time there is then the advantage that if necessary a second equalization network (satisfying the same conditions) can be placed between the first equalization network and the terminating resistance R_0, and in this way any desired number of equalization networks can be connected one behind the other, each with a different damping curve if necessary, without any change in the transmission properties of each network. Thus it is possible to divide, as it were, the total required damping curve in to different parts and assign each part to a network, which is of great advantage in the case of a fairly complicated shape of the required damping curve.

In *fig. 4* several forms of networks are now given which satisfy the condition mentioned. The impedance indicated by Z_1 may in each case still be composed in any desired way of resistances, self-inductions and capacities. The reciprocal impedance R^2_0/Z_1 occurring in other branches is then also composed of such elements, and this impedance is found by replacing the elements R_1, C_1, L_1 of the network Z_1 by elements R_2, L_2, C_2 respectively, which satisfy the equations:

$$R_1 R_2 = L_1/C_2 = L_2/C_1 = R^2_0,$$

and with which every connection in series in the network Z_1 is replaced by a connection in parallel and *vice versa*. If, for instance, Z_1 is a parallel connection of a capacity C_1 and a resistance R_1, then, as can easily be verified, R^2_0/Z_1 becomes a connection in series of a self-induction $L_2 = R_0{}^2 C_1$ and a resistance $R_2 = R_0{}^2/R_1$.

The following equation is valid for the damping of a network according to fig. 4*a* :

$$a_{\mathrm{db}} = 20 \log \left| \frac{1 + Z_1/R_0}{1 - Z_1/R_0} \right| \quad . \quad . \quad (3)$$

for the networks according to figs. 4*b*, *c*, *d*:

$$a_{\mathrm{db}} = 20 \log |1 + Z_1/R_0| \quad . \quad . \quad (4)$$

It is clear that by a varied choice of Z_1 a very varied behaviour of a with the frequency can be realized. In *fig. 5* the qualitative behaviour of a is reproduced for the network of fig. 4*b* with several types of impedances Z_1.

The function of an equalization network is to a certain extent analogous to that of a filter. It is therefore perhaps advisable to point out their points of difference. In the case of a filter it is a question of obtaining a damping in a certain frequency region and in another frequency region preferably an entirely unattenuated transmission (zero damping). Therefore for the composition of filters in principle only reactive elements (coils and condensers) are considered which have the lowest possible energy dissipation. "Damping" (without dissipation) is in this case obtained in certain frequency regions only due to the fact that the matching between the filter impedance and the terminating

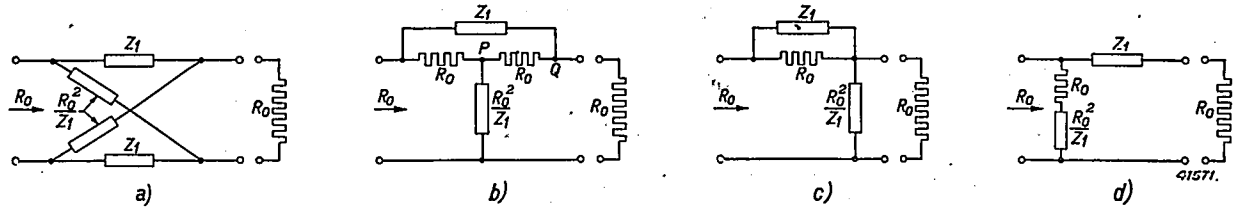


Fig. 4 *a-d*). Four types of equalization networks. All have the property that upon being terminated by a resistance R_0 the input impedance also becomes equal to R_0. With the same value of Z_1 the three types of networks *b*), *c*) and *d*) have the same damping curve (equation 4). This can be explained in the following way: If the network *b*) is terminated by a resistance R_0 it forms a Wheatstone bridge, one diagonal of which lies between the input terminals of the network, the other between points P and Q. It is easily verified that with the impedances indicated the bridge is balanced, so that the resistance R_0 between P and Q carries no current. This resistance can for example: therefore be short-circuited, network *c*) then being obtained. Or the resistance in question may be made infinitely large, network *d*) then being obtained. The transmission is the same in all three cases, provided the left-hand terminals are used as input. If, however, the direction of the transmission is reversed, networks c) and *d*) no longer satisfy the initial condition, so that these two networks cannot be used for transmission in both directions.

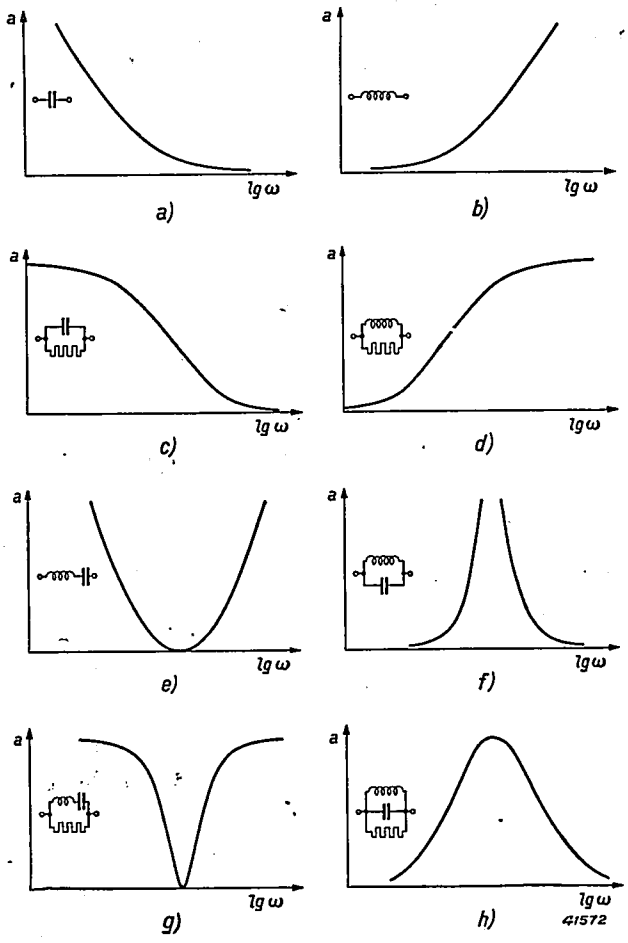


Fig. 5. By building up the impedance Z_1 of the networks in fig. 4 in different ways from coils, condensers and resistances, damping curves of different types can be obtained. For the network of fig. 4*b* the approximate form of the damping curve is drawn for the most important types of impedance Z_1.

impedance is poor in these frequency regions. Because of this the energy of the incoming signal is partly reflected and does not reach the termination behind the filter. Once the transmission region has been determined, however, the behaviour of the damping of each type of filter in the damping region is entirely fixed, so that owing to the limited number of types of filter there are only limited possibilities of variation. It is clear that by also including resistances in the network much more extensive possibilities are obtained for realizing a given damping behaviour.

It must further be pointed out that equalization is not only used in cables, but that it is sometimes also necessary in the case of other transmission elements (quadrupoles). In the latter case it is sometimes not actually a question of "equalization", but of obtaining a definite (thus not flat) frequency characteristic. Thus, for example, the previously described psophometer characteristic [6]) can also be realized by means of "equalization" networks.

Determination of the network required

How does one now find the impedance Z_1 that has to be taken in the networks of fig. 4 in order to obtain a prescribed form of damping curve? This can be done in different ways. In the first place a certain type of impedance Z_1 may be chosen for the approximate shape of the desired damping curve, and then the necessary equations may be set up for the n elements in this impedance — in the case above these were the two elements C_1 and R_1 — by setting the attenuation a given by (3) or (4) exactly equal to the damping desired at n different frequencies for those frequencies [7]). The calculations are generally quite elaborate and it cannot be seen in advance how much the attenuation for other frequencies will deviate from the desired values. Moreover, it may happen that from the equations values are found for the impedance elements which are not realizable with the available means, for instance negative capacities and the like.

A different method is therefore often preferred, namely the accurate drawing of the damping curves for a large number of networks with different kinds of impedances Z_1, consisting of elements of different values. Given such curves, it is possible by interpolation to find a curve which approaches the desired damping curve as closely as possible. It may seem a hopeless task to draw such curves and to use them, considering the large number of parameters. If, for instance, we take for Z_1 the connection in parallel of C and R already used as an example, the damping curve $a\ (\omega)$ of the equalization network has the three parameters R_0, R and C. Thus for this single type of Z_1 we obtain a threefold family of curves, or ∞^3 different curves. Fortunately, however, by a suitable choice of the variables the number of curves to be drawn is found to be very much reduced. If, for example we substitute in equation (4) for a network of the form of fig. 4*b* the function of Z_1 which corresponds to the connection in parallel of R and C:

$$Z_1 = \frac{1}{\frac{1}{R} + j\omega C},$$

we find for the damping:

$$a = 20 \lg \left| 1 + \frac{R/R_0}{1 + j\omega RC} \right| . \quad . \quad . \quad . \quad (5)$$

By now choosing the quantity ωRC as independent variable instead of ω we can represent equation (5) as a single family of curves in which only R/R_0 still appears as parameter. In *fig. 6* a number of curves of this family are drawn.

The fact that, with R/R_0 given, the damping a

[6]) Philips Techn. Rev. **7**, 108, 1942.

[7]) O. J. Zobel, Bell System Techn. J. **7**, 438, 1928

depends only on the expression ωRC means that two different prescribed damping curves

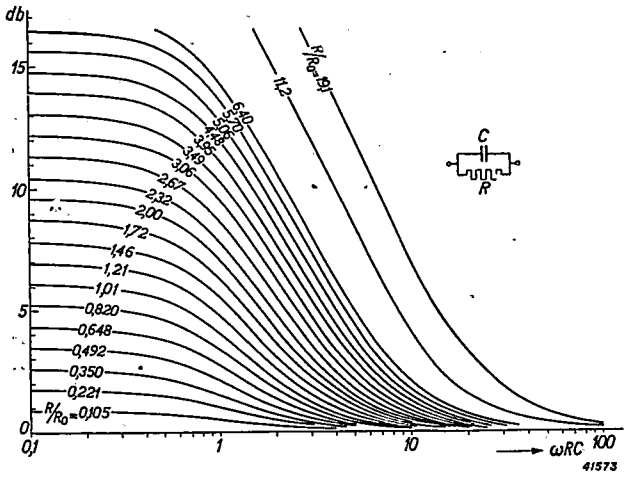


Fig. 6. Damping curves (a as a function of ωRC) of networks according to fig. 4*b*, with, a capacity C and a resistance R, connecd in parallel for the impedance Z_1, for a series of different values of the parameter R/R_0. (The same curves indicate the damping for the case where Z_1 consists of a self-induction L and a resistance R; conneted in parallel; the figures along the abscissa then, indicate the quantity $R/\omega L$.)

$a_1(\omega)$ and $a_2(\omega)$ can be realized with the *same* curve $a(\omega RC)$ of our family (thus with the same value of R/R_0), provided a_1 and a_2 satisfy the condition that

$$a_1(\omega) = a_2(k\omega) , \quad , \quad . \quad . \quad . \quad . \quad . \quad (6)$$

where k represents any arbitrary constant. The network securing the damping a_1 (ω) must then be made with $RC = 1$, the network for $a_2(\omega)$ with $RC = k$. With a linear scale for the frequency ω equation (6) means that the curve a_2 (ω) is obtained from the curve $a_1(\omega)$ by expansion or contraction in the direction of the abscissa; with a logarithmic frequency scale on the other hand $a_2(\omega)$ is obtained from $a_1(\omega)$ simply by translation of the curve in the direction of the abscissa, because $\log k\omega = \log \omega + \log k$. For this reason the family of curves in fig. 6 is drawn with a logarithmic abscissa: the prescribed damping curve $a(\omega)$, which has been drawn on transparent paper with the same logarithmic scale for the abscissa, can then simply be laid on fig. 6 and by shifting it in the direction of the abscissa an attempt may be made to make it coincide with one of the curves. If this attempt is successful, the value of the abscissa ωRC can be read off at a definite frequency ω and the required value of RC is then found as the quotient of the two abscissae. An example is given in *fig. 7*, where it may at the same time be seen that in order to make two curves coincide a mutual shift of the two graphs in the direction of the ordinate may also be used. Such a shift means only an increase in the damping by an

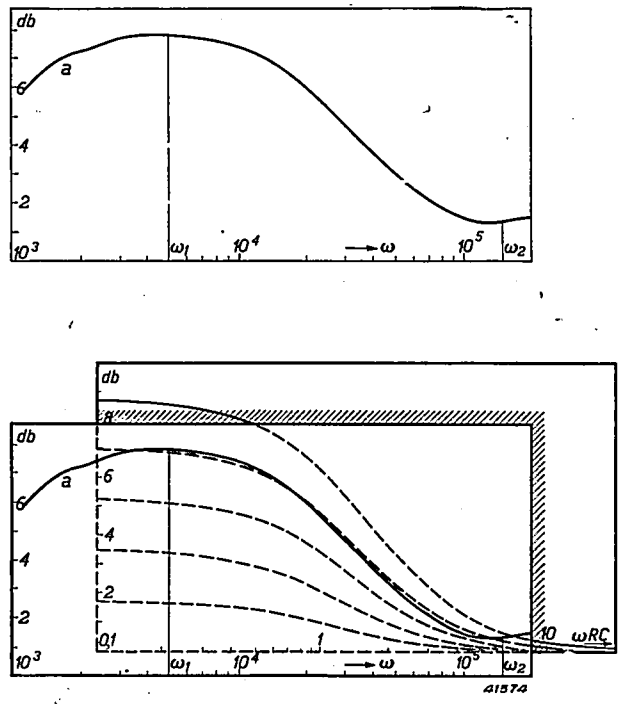


Fig. 7. Example of the manipulation of the families of curves. a is the prescribed damping curve which must be realized as well as possible in the frequency region indicated from ω_1 to ω_2. The curve a drawn on transparent paper is laid on the family of curves of fig.6 and by moving it about it can be made to coincide approximately with the curve for which $R/R_0 = 1.21$. With this position of the two pieces of paper the abcissa ω_1 lies at $\omega RC = 0.22$, the abscissa ω_2 at $\omega RC = 6.6$. From this it follows that $RC = 0.22/\omega_1$ or $6.6/\omega_2$, which must of course give the same result.

amount which is the same for all frequencies, and is therefore - although not desirable in practice — permissible for the equalization.

Fig. 8 gives the family of curves for an equalization network according to fig. 4*b* with a different impedance Z_1, namely a connection in parallel of a resistance R with a self-induction L and a capacity C in series. In this case the four parameters R, L, C, R_0 can be reduced to two, *viz.* $\sqrt{L/C}/R_0$ and R/R_0, when $\omega\sqrt{LC}$ is plotted as abscissa. A twofold family of curves is thus retained. Here also by the use of a logarithmic abscissa scale an attempt may be made to make a given damping curve coincide with one of the curves simply by shifting.

The construction of the family of curves need only be done once, and the number of curves whose shape must be calculated is very much reduced in the manner described. Nevertheless, the preparation of the graphs always requires extensive calculations. This can be entirely avoided by a simple method of

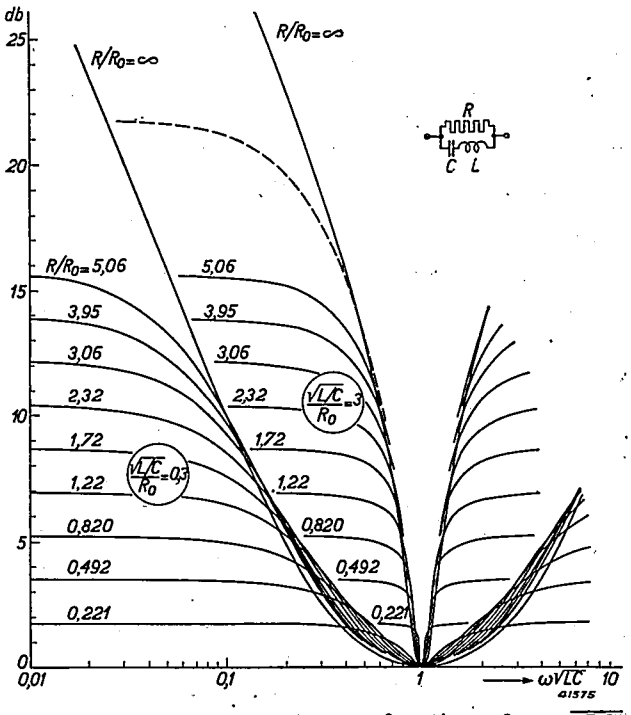


Fig. 8. Damping curves (a as a function of $\omega\sqrt{LC}$) of networks according to fig. 4*b* with, for the impedance Z_1, a resistance R in parallel with a self-induction L in series with a capacity C. In this case a family of curves is obtained, each family being characterized by a certain value of the parameter $\sqrt{L/C}/R_0$ and each curve by a certain value of the parameter R/R_0.

measuring. The arrangement sketched in *fig.* 9 is used. Between the output voltage V_2, which can be measured with the voltmeter V, and the input voltage E, which is furnished by a tone generator, there is the following relation:

$$V_2 = \frac{E \cdot R_0/2}{R_0 + Z_1}.$$

Furthermore the voltage V_1 is always equal to E_2, since the adjustable attenuator T has at all positions an input resistance R_0. For the ratio V_1/V_2 the following thus holds:

$$\left|\frac{V_1}{V_2}\right| = \left|\frac{R_0 + Z_1}{R_0}\right| = \left|1 + \frac{Z_1}{R_0}\right|.$$

If now for every frequency of the tone generator the attenuator is so adjusted that $V_2 = V_2'$, which is checked by comparing the voltmeter indications at the two positions of the switch S, and if the attenuator is calibrated in db, the position the attenuator immediately gives the quantity

$$20 \log \left|1 + \frac{Z_1}{R_0}\right|.$$

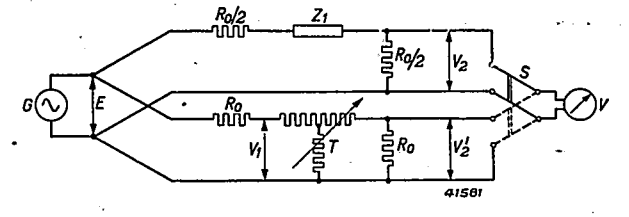


Fig. 9. Connections for recording the damping curves of equalization networks according to fig. 4*b*. G tone generator, T variable attenuator calibrated in db, S switch, V voltmeter.

According to equation (4) this is exactly the expression for the damping of a network according to figs. 4 *b*, *c*, *d* with the impedance in question Z_1. The variation of damping with the frequency can in this way be very rapidly and easily determined for every set of values of the parameters of Z_1.

Actually the method foils down to the same thing as if the damping of the network of fig. 4*b* were measured directly. Since, however, in this measurement the condition concerning the terminating impedances does not need to be satified, the network could be replaced by the simpler one shown in the upper half of fig. 9. The advantage in this case is that for each set of parameters it is not also necessary to realize the reciprocal network, which is particularly convenient when one can be saved the trouble of regulating coils, which is always much more difficult than setting condensers.

We shall now explain in more detail the practical process of equalization on the basis of an example.

Practical process of equalization

First of all in a practical case it has to be determined what damping curve the equalization network must possess. The data required can most easily be obtained by measuring the damping of the section of cable to be equalized as a function of the frequency. If this is impossible, as for instance in the case of installations which have to be fuelly worked ont in advance, the damping curve must be calculated. When doing so account must be taken of the fact that the amplifiers, transformers etc. may also all have a definite frequency characteristic which is not flat. All these characteristics should be combined. Furthermore, the termination of the cable plays an important part. If the cable is terminated by its characteristic impedance, only travelling waves occur in the cable and its damping curve can be derived in a simple way from the cable characteristics (L, C, R and G per km). Usually, however, the impedance of the termination of the cable will not be exactly equal to the lowest frequencies. The energy arriving at the ends is then partially reflected, the degree of reflection (*i.e* the magnitude of the energy loss) depending upon the frequency. In the case of not too long cables the situation becomes even more complicated, due to the fact that the energy reflected at the end of the cable is partly reflected again and then once more furnishes an appreciable contribution at the end, which this may even be repeated several times. Formulae have been developed for making extact allowances for all these corrections to the variation of the damping of any given quadrupole as a function of the frequency. We shall not go into them here, however, and shall assume that the damping curve to be equalized is already given, either by calculation or by measurement.

Let us now suppose that curve *1* in *fig. 10* is the desired variation of the damping of the equalizer. It is here a question of equalizing a carrier-wave cable for the frequency region from 4 to 72 kc/sec with a tolerance of ± 1 db. Upon

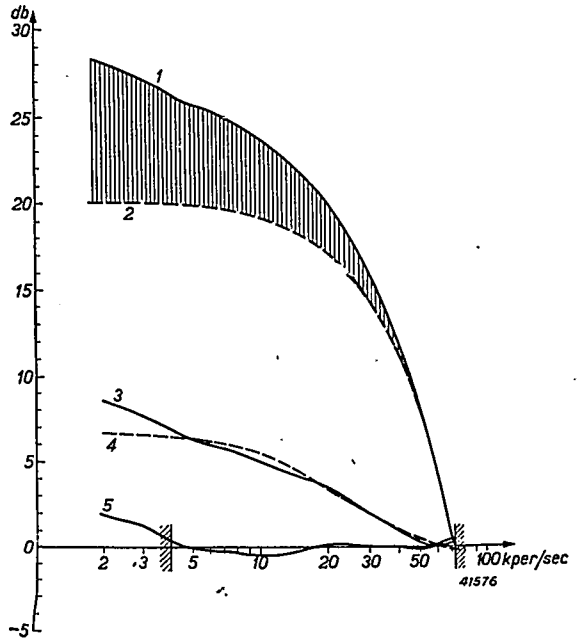


Fig. 10. Example of the manner in which equalization is accomplished. Curve *1* is the desired damping curve; curve *2* is furnished by a first equalization section (fig. 11*a*). The difference (curve *3*) is approximated by the damping curve *4* of a second equalization section (fig. 11*b*). The discrepancies finally remaining are shown in curve *5*: at all frequencies within the frequency band to be used they are smaller than 0.5 db.

comparison of this curve with the curves of the different graphs it is evident that the desired shape cannot be realized with a single equalization network of the types considered. A damping curve must therefore first be separated which can be obtained with the first equalization section, and then the rest of the damping must be contributed by a second or if necessary several more sections. This separating can be done more or less arbitrarily. For curves with the approximate shape of fig. 10 — a rapid drop in the damping towards high frequencies — it will always be necessary, however, for practical reasons to begin by matching the the damping to the requirement at the h i g h e s t frequencies. Furthermore, the damping for the first equalization network to be chosen should lie preferably entirely b e l o w the desired curve, since otherwise, in order to avoid having to provide negative dampings in the following sections, a constant damping would have to be added over the whole frequency region. In this way we arrive at the choice of curve *2* in fig. 10, which, apart from a constant damping difference for all frequencies (displacement in the direction of the ordinate), is exactly realized by a curve of the family drawn in fig. 8. The curve in question is indicated by a dotted line in fig. 8.

For the parameters of the impedance Z_1 of the corresponding network the following relation may be read off from the graph:

$$\sqrt{L_1/C_1}/R_0 = 3,$$
$$R_1/R_0 = 11.2,$$
$$1/\sqrt{L_1C_1} = 2\pi \cdot 80\,000.$$

Since the wave resistance with which we choose to terminate the cable lies in the vicinity of 150 ohms in the case of carrier-wave cables, we choose $R_0 = 150$ ohms and then have three equations for R_1, C_1, L_1. In *fig. 11a* the network is drawn and the values of R_1, C_1, L_1 thus calculated, as well as of the elements R_2, C_2, L_2 of the reciprocal impedance (R_0^2/Z_1) are indicated.

The damping curve *2* in fig. 10 subtracted from the prescribed damping *1*, gives the damping curve *3* which must now be realized with additional sections. This curve is satisfactorily approximated by one of the curves of fig. 6, namely the curve with parameters

$$R/R_0 = 1.458,$$
$$1/RC = 2\pi \cdot 15\,000.$$

R_0 must of course again be chosen equal to 150 ohms. In this way one obtains the second section shown in fig. 11*b* with the values there indicated for the elements R_3, and C_3 of the impedance Z_1, and R_4, L_4 of the reciprocal impedance R_0^2/Z_1. The damping curve of this equalization network, again except for a constant damping difference, is given by curve *4* in fig. 10. The difference between this curve and curve *3*, which finally gives the remaining errors in the equalization, is plotted as curve *5*. It may be seen that the deviations in the whole frequency region from 4 to 72 kc/sec amount to a maximum of ± 0.5 db.

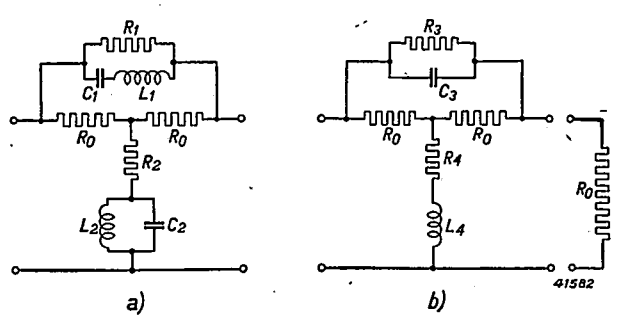


Fig. 11. Equalization network consisting of two sections *a*) and *b*) for the equalization of a carrier-wave cable (see fig. 10).

a)	$R_1 = 1680$	ohms	$R_2 = 13.4$	ohms
	$C_1 = 0.00442$	μF	$L_2 = 99.5$	μH
	$L_1 = 895$	μH	$C_2 = 0.0398$	μF
b)	$R_3 = 219$	ohms	$R_4 = 103$	ohms
	$C_3 = 0.0484$	μF	$L_4 = 1090$	μH

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

1575: M. J. Druyvesteyn: An approximate calculation of the thermal expansion of solids II (Physica **8**, 862 — 867, September 1941).

The elasticity theory of an isotropic is extended by taking into account the anharmonic terms of the mutual elastic coupling of the particles of the material. Terms which would give rise to anisotropic properties are ignored. An equation is found for a relation between the constant γ of Grüneisen and the lateral contraction . For most metals crystallising with cubic symmetry the results agree very well with the experimental findings, but in the case of alkali metals there are large deviations.

1576: A. Claassen and J. Visser: Die Trennung und Bestimmung von Titan und Aluminium mit Ortho- Oxychinolin (Rec. Trav. chim. Pays Bas **60**, 715 - 727, September - October 1941). The separation and determination of titanium and aluminium with ortho-oxychinnoline.

Berg's method for determining titanium with ortho-oxychinnoline yields doubtful results owing to the incomplete precipitation of the titanium. The titanium can be completely precipitated when the reaction takes place in a tartaric acid solution in the presence of an excess of ammonium sulphate and this solution after boiling is left on the water bath for one to two hours. A further condition is that the hydrogen ions must have a concentration of at most $10^{-5.2}$ grams equivalent ($P_H = 5.2$). If a citric acid solution is used no ammonium sulphate need be added. Quantities of tartaric acid or citric acid larger than 1 gram increase the solubility of the precipitate. Specifications are given according to which quantities of titanium up to 40 mg can be determined analytically to an accuracy of 0.1 — 0.2 mg. Gravimetrical determinations always yield too high results owing to the difficulty in washing out the precipitate. Titanium can be very sharply separated from aluminium with the aid of ortho-oxychinnoline in oxalic acid solution provided the hydrogen ion concentration lies between $P_H = 5.6$ and 6.5. Separation in malonic acid solutions yields inaccurate results.

1577: J. van Slooten: The transformer properties of a four-polar system. (T. Ned. Radio-Genootsch. **8**, 217-234, November 1941).

An electric four-polar system can be regarded as a transformer converting an impedance between the output terminals into another impedance between the input terminals. The latter is conceived to be a function of the former. It is known that generally speaking two output impedances (so-called iterative impedances) can be indicated and in this transformation they do not change in value. With the aid of this property the four-pole equations can be cast in a form which offers advantages for various applications. In this article two simple diagrams are given for cases where the four-polar system consists of a non-ideal transformer (with finite self-inductances and spread) or of a piece of no-loss cable.

1578: E. J. W. Verwey and P. W. Haayman: Electronic conductivity and transition point of magnetite ("Fe_3O_4") (Physica **8**, 979-987, November 1941).

Experiments with some sintered rods of magnetite (Fe_3O_4) have shown that when passing a conversion temperature in the neighbourhood of 120°K the electrical resistance changes by leaps and bounds. When cooling the resistance increases by a factor of 100. This article shows that this phenomenon depends to a large degree upon the stoichiometrical composition of the homogeneous spinel phase, which can be expressed by the ratio of iron to oxygen or by that of ferro-oxide to ferri-oxide. With a small excess of oxygen the above mentioned sudden increase in the resistance is gradually reduced and the conversion temperature lowered several tens of degrees until ultimately the phenomena of conversion entirely disappear. These phenomena are believed to be related to a theory previously put forward in connection with the crystal structure of magnetite (in particular the distribution of the bivalent and trivalent ferro ions in the lattice) and the related mechanism of electro-conductance.

1579: M. J. Druyvesteyn and J. L. Meyering: Elastic constants in the system Cu-Zn (Phy sica **8**, 1059-1074, November 9141)

A method is described for determining the elasticity and torsion moduli of rods from their longitudinal vibrations (*cf.* Philips Techn. Rev. **6**, 372, 1941). The elasticity and torsion moduli are measured for poly-crystalline material in the system Cu-Zn, from which the characteristic temperature is then calculated. It appears that in the system mentioned these moduli have a minimum value (for β brass) and a maximum value (for γ brass).

1580: E. J. W. Verwey: The charge distribution in the water molecule and the calculation of the intermolecular forces (Rec. Trav. chim. Pays Bas **60**, 887-896, November 1941).

From calculations of the electrical alternating potential between monovalent ions and water molecules in aqueous solutions it has been found that the model according to Bernal and Fowler (and also similar models) is inadequate for the charge distribution in the water molecule. A model is therefore proposed which in essential points means an improvement upon that of Bernal and Fowler. The centre of the negative charge of the molecule is then only a small distance away from the oxygen nucleus. The model supplies the sublimation heat of ice with sufficient accuracy. In particular the influence is investigated of various hypotheses concerning the degree of screening of the proton charges in the molecule.