JOURNAL OF

The British Institution of Radio Engineers

(FOUNDED IN 1925 - INCORPORATED IN 1932)

"To promote the general advancement of and to facilitate the exchange of information and ideas on Radio Science."

Vol. 6 (New Series) No. 3 JUNE, 1946

VOLUNTARY SERVICE

During the past few years the nation has become so accustomed to voluntary service in one form or another that compulsory duty is more likely to receive widespread comment than the public or civic service which is freely given. It would, however, be wrong to imagine that there is any lack of gratefulness for work freely done in any cause. Achievement is usually a just reward, but this does not justify the omission of appreciation by those who have benefited by the labour of others.

Custom dictates that thanks to voluntary workers be expressed at annual general meetings; in those necessarily brief expressions, however, it is easy to overlook the sacrifices which may have been made by those who serve on the Councils and Committees of various Institutions, including our own.

It is usually the busy man who finds the most time to contribute to the common good; the man who is successful and efficient in his acknowledged profession is frequently the man best qualified to assist the work of Committees. Whatever the reason, the fact remains that such voluntary service is a commitment honourably undertaken and kept without tangible personal reward and, as such, deserving of great appreciation.

The Institution could not properly fulfil its aims and objects were it not for the members who comprise the Council, the local Section Committees, the Standing Committees, or the Special Committees which are appointed from time to time. An examination of the records of those who have so served the Institution during the difficult days of war, and particularly during the conversion period from war to peace, shows that invariably such men have had many other responsibilities to the industrial organisations of which they usually form an essential part. Their attendance at committee or other meetings has, in the main, been most gratifying; they have done voluntarily everything necessary for framing the policy and administration of the Institution's work.

The principal officers have necessarily devoted more time than other colleagues on Committees, but all have been animated by the desire to push forward the work of the Institution and to meet fully the declared aims and objects of the organisation.

In every community there are many who, for various reasons, are unable or do not wish to give such voluntary service; all members, however, have responsibility in directing the affairs of the Institution when elections are held which appoint the Council and the various Committees.

Another year has passed, bringing us again to the time when such elections must be held. The Council feel it is appropriate to tender thanks, on behalf of the members, to those who have served the Institution in the past and to give welcome and help to those who will serve the Institution in the future.

Membership of the General Council, the Standing and Local Committees of the Institution and representation on such bodies as the Parliamentary and Scientific Committee, British Standards Institution, City and Guilds of London Institute, etc., involves some one hundred and twenty-one members voluntarily serving the Institution. The personnel changes from year to year and, in general, there is ample opportunity for every member of the Institution, in one way or another, to further the declared objects of the Institution. Some sections of the membership have, indeed, advocated ways of extending opportunity for members to take an active part in the Institution's work and some of these proposals are discussed in the annual report of the Council to be published in the next issue of the Journal.

Not the least of these proposals is the important one that, with the development of the Sections, the Chairmen of Section Committees should become ex-officio members of the General Council, as well as, perhaps, the Chairman of certain Standing Committees, including the very important Charter Committee.

In such ways, the national character of the Institution will be emphasised and the important work which lies immediately before the Institution will benefit by the services which are voluntarily but sincerely given.

It is in this spirit, therefore, that all members of the Institution will be interested in the elections and appointments to the Council and various Committees now due to take place and marking the first twenty-one years of the Institution.

THE INFLUENCE OF AN ECLIPSE OF THE SUN ON THE IONOSPHERE

by

R. L. Smith-Rose, D.Sc., Ph.D., M.I.E.E.*

(*A lecture delivered before the London Section of the Institution on March 20th*, 1946.)

1. Early Observations on the Effect of the Sun on Radio Transmissions

The early years of the present century were conspicuous for the rapid progress made in the practical demonstration of the possibilities of the use of radio waves for communication over distances of more than a few miles. Conspicuous among the experimental workers of those days was Guglielmo (later Marchese) Marconi, who with his collaborators first transmitted wireless signals across the Atlantic between Cornwall, England, and St. John's, Newfoundland, in December, 1901. This achievement, and other work which rapidly followed, soon made it clear that if sufficient radio-frequency power could be generated and transmitted, and a sufficiently sensitive receiving system could be devised, then it was probable that signals could be exchanged between any two points on the earth's surface under suitably favourable conditions. At the time these conditions were not understood, and for the next twenty years or so there was a widespread and often vigorous controversy among both the theoretical and experimental workers in the subject as to what exactly was the precise mechanism by which electric waves in the radio frequency part of the spectrum could be caused to travel around the curvature of the earth, when it was quite clear that the corresponding waves in the visible part of the spectrum were subject to only very slight bending round the earth's surface.

Among the significant experimental evidence which ultimately led to the true explanation of the phenomena involved was the fact that over distances exceeding a hundred miles or so large differences in signal strength were obtained for day and night transmission, and very marked fading conditions were frequently experienced when sunset or sunrise traversed the path between the sending and receiving stations. In the earliest experiments conducted across the Atlantic it was found that signals which could be received at night up to distances of 2,000 miles, were not detectable in daylight beyond 800 miles. Later, when communication was established across the Atlantic between stations operating on wavelengths in the region of 4,000 m. (frequency 75 kc/s), variations in signal strength of the order of ten to one were a common experience; frequently the signals were reasonably steady all day, and much more variable at night. The precise effects experienced over this and other paths, however, varied very markedly with the wavelength and path of transmission, and with the time and season.

From the point of view of this lecture, the year 1912 is a notable landmark in the history of the subject, and for more than one reason. In the first place, one of the outstanding features of the meeting of the British Association for the Advancement of Science held in that year was a discussion on some of the unsolved problems of wireless telegraphy.[1] This discussion was opened by Dr. (later Sir Ambrose) Fleming, and among the contributors were Dr. W. H. Eccles, Prof. G. W. O. Howe, Lord Rayleigh and Prof. A. E. Kennelly. Secondly, in June, 1912, W. H. Eccles read a paper before the Royal Society,[2] putting forward the hypothesis that at night electric waves are propagated round the world by successive reflections from the permanently ionised regions of the atmosphere, the existence of which had been previously suggested independently by Heaviside and Kennelly in 1902. To account for the different propagation phenomena experienced during the daytime, it was supposed that the effect of sunlight was to bring into action another ionised layer in the atmosphere underneath the more permanent one.

In the third place, what must have been the earliest systematic study of the effect that an eclipse of the sun had upon the transmission of radio waves was conducted by Eccles[3] during the eclipse of April 17th, 1912. On that occasion, as there was no conveniently placed radio transmitting station in operation, observations were made in London on atmospherics originating from sources in Spain or North-West Africa. These observations indicated that with the onset of the eclipse the radio propagation conditions first improved rapidly, then deteriorated to a maximum absorption effect as the umbra crossed France intersecting the path of transmission, and later improved to the normal conditions for the day.

In connection with the solar eclipse of May 29th, 1919, a committee of the British Association arranged for experiments to be made on the effect of the eclipse on signals transmitted continuously across the central line during the transit of the umbra across the Atlantic Ocean.[4] Many of the observations on both atmospherics and signals showed no unusual effects; but in some cases the records showed that there was a marked improvement in signal strength which reached its highest value long before the umbra intervened between the stations, and this value persisted after the umbra had passed.

* National Physical Laboratory.

The eclipse of the sun of April 8th, 1921, was such that the shadow of the moon just passed across the northernmost part of Scotland between 0730 and 1000 G.M.T. On that occasion the British Post Office arranged for special observations to be made at a number of radio receiving stations in this country.[5] At most of the stations operating on a frequency of 500 kc/s (wavelength 600 m.), no change in conditions was observed, but one station reported a marked increase in the strength of signals, more particularly on those emanating from distant transmitters. At the Devizes station the strength of signals received from various European stations working on a frequency in the region of 150 kc/s (2,000 m.) increased markedly during the eclipse to about normal night-time value, returning afterwards to the normal daylight strength. An increase in the strength and frequency of atmospherics was also observed at several stations during the eclipse.

2. Investigations during Solar Eclipses, 1925-1929

The first occasion on which members of the staff of the Department of Scientific and Industrial Research participated in an investigation of radio conditions during a solar eclipse was on January 24th, 1925, when the path of the total eclipse crossed the North Atlantic ocean between 1400 and 1550 G.M.T. (See Fig. 1.) On this occasion the staff of the Radio Research Board worked in co-operation with engineers of the Post Office and the International Western Electric Company, to make measurements on the signals received in this country from transmitters situated on Long Island, U.S.A.[6] As shown in Fig. 2, transatlantic signals sent on a frequency of 57 kc/s increased in strength as the shadow of the eclipse left the neighbourhood of New York. After passing through a maximum of about twice the normal strength, the signal decreased to a well-defined and very low minimum value at a time which coincided to within a few minutes with the intersection of the path of totality and the great circle path of the radio waves. At one of the receiving stations in this country a range of signal intensity of nearly 20 to 1 in amplitude was recorded, whereas measurements made on the days before and after the eclipse showed no corresponding variations in field strength. (See Figs. 2 and 3.) Simultaneous observations made with a radio direction finder at the Radio Research Station,

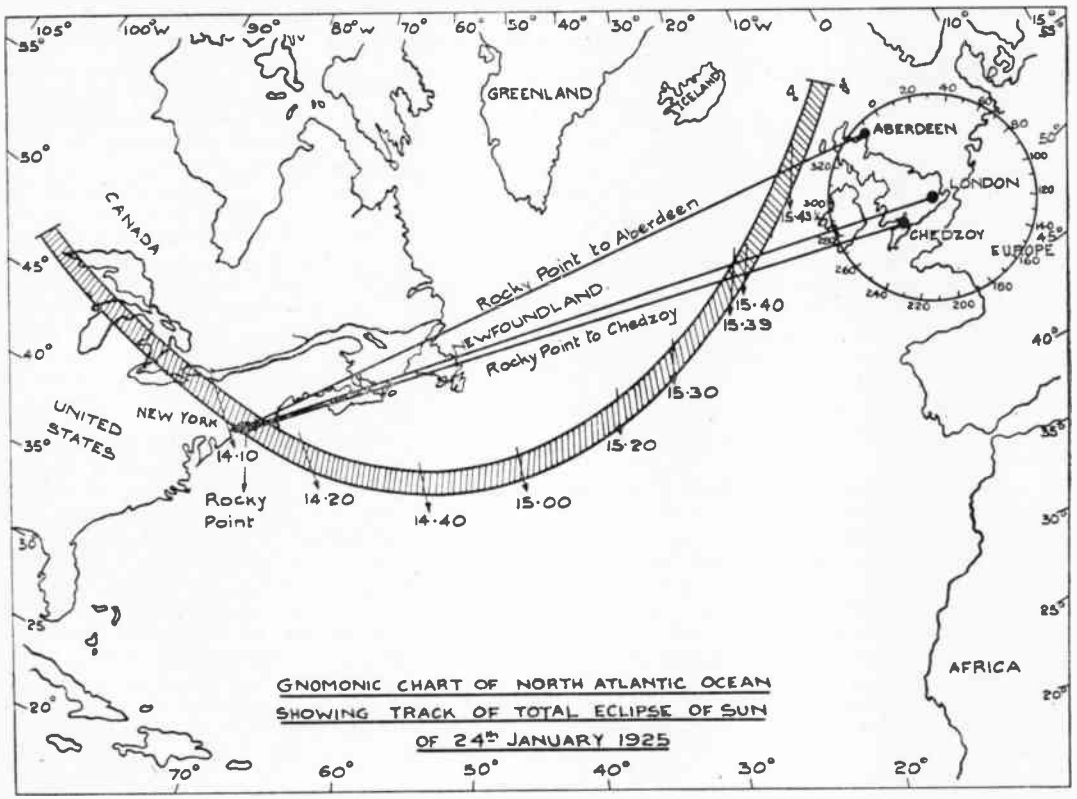


Fig. 1

Slough, showed that the direction of arrival of these transatlantic signals was very steady, and that there was no detectable effect due to the eclipse.

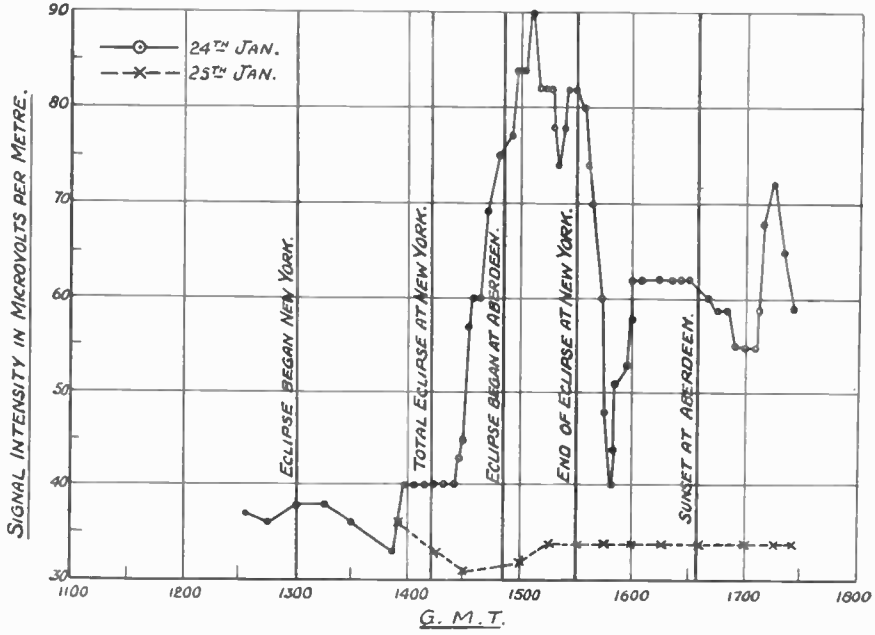


Fig. 2.—Signal Intensity of Rocky Point, U.S.A., observed at Aberdeen during the Solar Eclipse of 24th January, 1925. Frequency—57 Kc/s.

During the same eclipse of January 24th, 1925, measurements were made by Professor (now Sir Edward) Appleton and Dr. M. A. F. Barnett,[7] of the signal intensity received at Cambridge from the London broadcasting station. The variations in the received intensity during the eclipse were similar to those experienced during the normal sunset, which on the day in question occurred nearly an hour after the eclipse.

Other observers have since noted similar effects on the transmission of radio waves during an eclipse, with the general conclusion that the interception of the solar radiation proceeding towards the earth produces a result similar to that of a return to night conditions. It was realised, however, that the phenomena involved were by no means simple, and it was not until a more satisfactory understanding was obtained of the part which the upper atmosphere plays in the propagation of radio waves that more useful experiments could be conducted during an eclipse of the sun. Before dealing with the results of later investigations, therefore, a diversion will be made to introduce a brief survey of our knowledge of the ionosphere and of the transmission of radio waves through it.

During the next two years, the now classical technique of examining the propagation of radio waves through the ionised regions of the upper atmosphere had been developed by E. V. Appleton and M. A. F. Barnett, and the application of this experimental method enabled much wider and more definite information to be obtained on the occasion of the eclipse of June 29th, 1927, than had hitherto been possible. An official report of the Radio Research Board[8] describes the results obtained on medium wave broadcasting transmissions, with the addition of some long-wave observations by J. Hollingworth and R. Naismith, and corresponding measurements with radio direction finders by R. L. Smith-Rose and R. H. Barfield.

The centre line of this eclipse swept across Northern England, and the North Sea to Southern Norway, while as shown in the accompanying map (Fig. 4), the plan position of this central path at a height of about 100 km. traversed the Midlands. A variety of observations of signal intensity and direction of transmission were conducted on waves crossing the path of totality. The time of occurrence of the eclipse (about 0500 G.M.T.) was rather soon after sunrise to obtain a satisfactory discrimination of the effects in some cases.

The results of measurement of the strength of signals received in the North of England from the London broadcasting station on a wavelength of 365 m. are shown in Fig. 5, from which there is seen to be a very marked increase in intensity at the time of total eclipse. For transmissions in the reverse direction from Manchester to Slough, Fig. 6 shows the variations observed to be similar, although there was more fluctuation in the received signal strength during the actual eclipse. A similar, but less marked effect was observed in the West of England on signals from a station in Norway transmitting on a wavelength of 12,100 m. In all these cases observations made at the same times on the days before and after the eclipse showed no corresponding variations, there being a fairly smooth transition from night to day conditions from the occurrence of sunrise shortly before 0400 G.M.T.

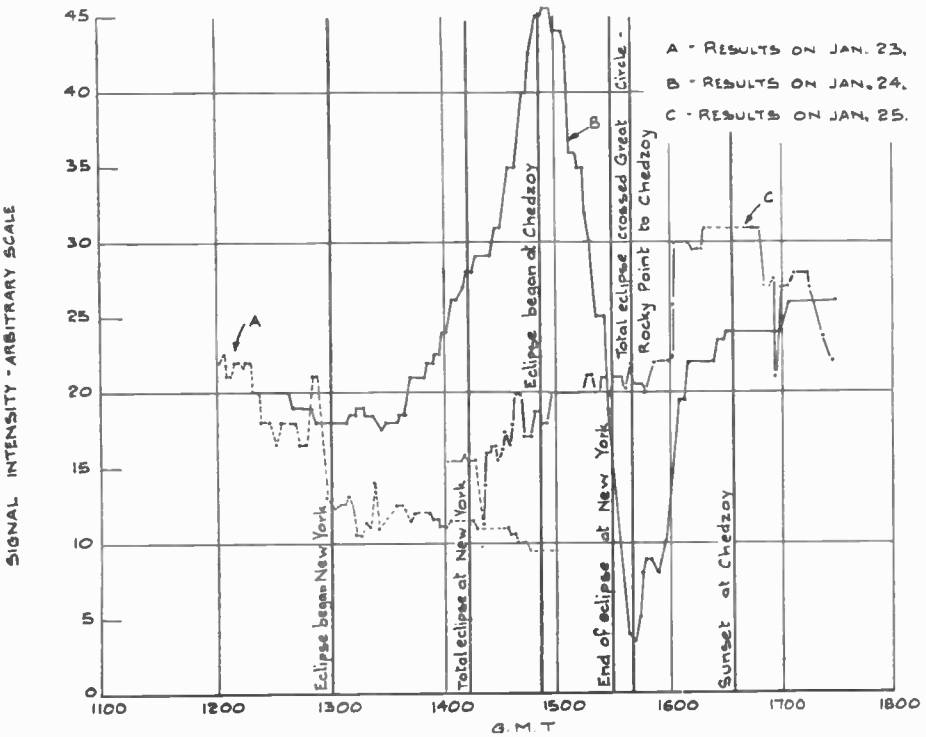


Fig. 3.—Signal Intensity of Rocky Point, U.S.A., observed at Chedzoy, Somerset, during Solar Eclipse of 24th January, 1925. Frequency—57 Kc/s.

In addition to the straightforward observations of signal strength, the methods of measurement used on medium waves permitted the investigation of the characteristics of the waves deviated from the upper atmosphere by using special transmissions in which the wavelength of an unmodulated carrier wave was changed continuously through a small range. The resulting signal maxima and minima due to interference between ground and atmospheric waves, were measured at receiving stations equipped with galvanometers and photographic recording apparatus. From the records so obtained it was possible to deduce the relative amplitudes of the atmospheric and ground waves and the equivalent height of the ionised layer responsible for reflecting the atmospheric waves back to the receiving station. The required "frequency-change" transmissions were available from the Newcastle and Birmingham stations of the B.B.C.

The results showed that there as a large increase in the intensity of the downcoming ray, which was observed at both near and distant receiving stations during the eclipse period only. A striking feature of the observation was the short time the eclipse effect lasted, this period varying from 20 to 50 minutes at the different stations, while the total time taken for the moon's shadow to pass across the earth was nearly two hours. This means that quite an appreciable fraction of the sun's radiation was cut off before the effect could be detected by radio methods.

On the occasion of the solar eclipse of May 9th, 1929, a French expedition to Indo-China carried out a variety of observations of signal strength and directional bearings of short-wave stations (wavelengths 17 to 30 m.).[9] A remarkable coincidence was observed between the occurrence of totality and the diminution in strength of signal from stations in Japan and Java becoming too weak to measure. There were also erratic variations in the bearings on a direction finder observed during the eclipse on signals from a station 175 km. distant, transmitting on a wavelength of 670m. Half-an-hour after the cessation of the eclipse all these radio conditions returned to normal.

From all these observations made prior to 1930, it was evident that an eclipse of the sun produced a marked effect on the propagation of radio waves through the ionised regions of the atmosphere, and there was a very marked coincidence in time between the transit of the moon's shadow across the earth and the changes in radio transmission conditions resulting therefrom.

3. Propagation of Radio Waves through the Ionosphere

(a) Early Measurements on Characteristics of Ionosphere

The first direct evidence of the existence of an ionised layer in the atmosphere capable of deflecting radio waves incident upon it, was provided by the experiments of E. V. Appleton and M. A. F. Barnett[10] in 1925. These experiments were conducted on transmissions over a short distance in which the question of curvature of the earth's surface was not involved. Besides yielding experimental proof of the existence of the Kennelley-Heaviside layer, a measurement of the

height of the layer above the earth's surface was also obtained. In this early work a frequency-change method was used, in which the frequency or wavelength of the sending station was varied over a small range. This gradual change of frequency produced at the receiving station regular maxima and minima of signal intensity as a result of interference between the waves transmitted along the ground, and those which had been reflected from the ionised layer. From a knowledge of the change in frequency and the number of interference fringes resulting therefrom, the height of the reflecting layer can be obtained. In a second series of experiments conducted at the same time, Appleton and Barnett compared the signal variations obtained

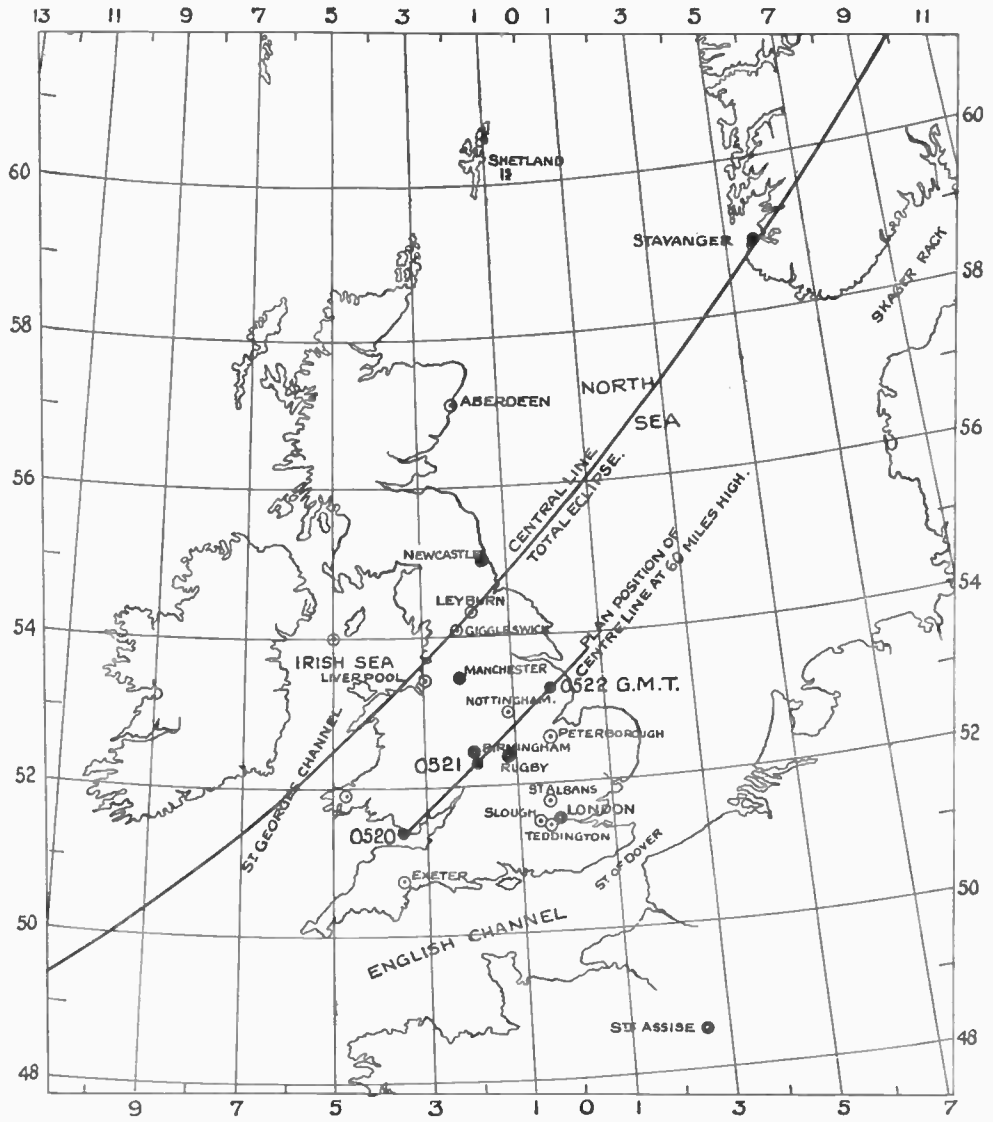


Fig. 4.—Gnomonic Chart of British Isles showing track of total eclipse of sun at ground level and at 100 km. (60 miles) above the earth. 29th June, 1927.

on a vertical aerial and on a loop the plane of which coincided with the plane of transmission. In this way the angle of incidence of the downcoming waves received at the ground was measured, and again the

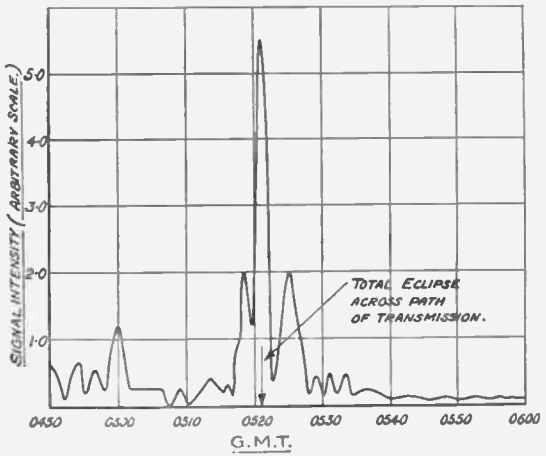


Fig. 5.—Graph showing the variation of received signal intensity at Giggleswick from the London Transmitting Station. Total Eclipse of sun, 29th June, 1927.

height of reflection of the waves was obtained. The height of the reflecting region in each experiment was found to be about 90 km. for the range of frequencies 750 to 1,000 kc/s (wavelengths 300 to 400 m.). This experimental proof of the existence of the ionised region was followed by confirmatory independent experiments carried out by R. L. Smith-Rose and R. H. Barfield,[11] who also obtained equivalent heights of the order of 100 km. for a similar range of frequencies.

Shortly after the above work, G. Breit and M. A. Tuve proposed and used, an amplitude-change or pulse-modulation technique to measure the difference in time of transit of the ground waves and the waves travelling up to the ionised region and back to the receiver. This method was rapidly adopted, and has been used by investigators in all parts of the world for studying the reflection of waves from the ionised regions of the atmosphere over the past twenty years or so. Pulse-modulated waves are emitted from a sending station and both the ground wave and ionospheric or sky wave are received a short distance away, and the difference in time of arrival of the two signals is measured accurately by a cathode ray tube technique, which is to-day well known. By varying the frequency of the sending station and keeping the receiving station in tune therewith, the reflecting properties of the ionosphere can be studied over a wide range of frequencies. In this manner Appleton[13] showed, as long ago as 1927, that while low-frequency waves are reflected from a layer, now termed region E, at a height of about 100 km., waves of higher frequency are reflected from a higher layer, now known as region F, the height of which is of the order of 250 km.; and while these actual heights vary slightly with the frequency in use, there is a marked discontinuity between the two regions. If the frequency in use is still further increased, the height of the F region

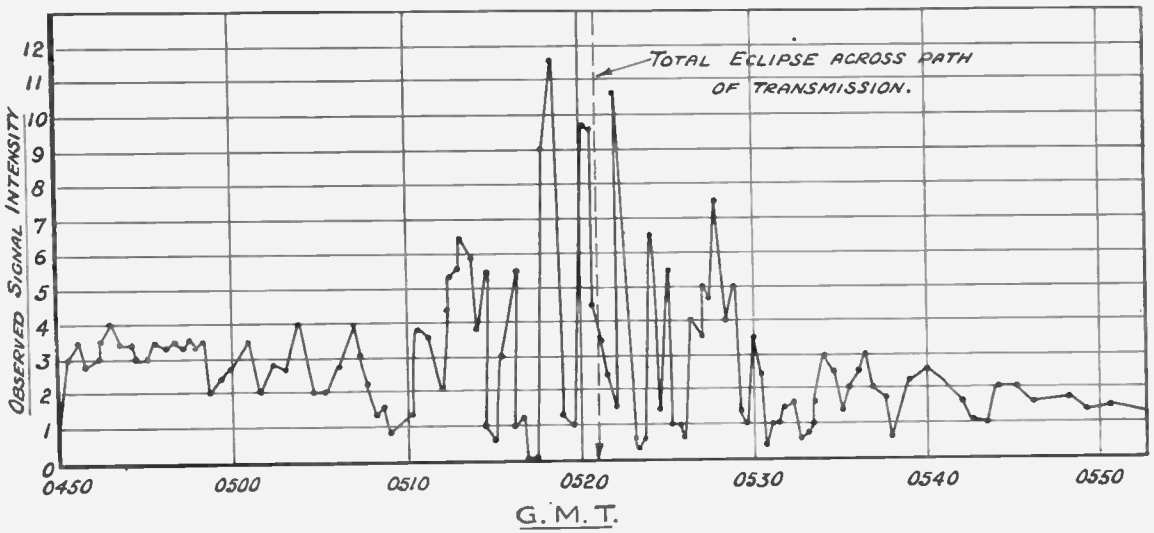


Fig. 6.—Graphs of signal intensity of the transmissions from Manchester observed at Slough. Total eclipse of sun, 29th June, 1927.

reflection increases somewhat, but ultimately the waves penetrate the whole region of ionised atmosphere and pass out into space. The actual frequencies at which the waves penetrate first the E and then the F regions are termed the critical frequencies of the regions; and it is the measurement and study of these critical frequencies, as well as of the heights of the regions, that have yielded a great deal of important information on the characteristics of the ionosphere and of the propagation of waves through it. Later work has shown that at times the F region is divided into two portions, termed F_1 and F_2 respectively, while very low frequency waves are reflected at a lower D region, the height of which is usually about 60 km. above the earth.

(b) Ionisation Density and Frequencies for Radio Communications

Now it can be shown that the maximum frequency, f in kc/s., at which radio waves incident on an ionised layer are reflected from it, depends first upon the density of ionisation N, and secondly upon the angle of incidence φ, these quanitites being related by the expression

$$N = \frac{f^2}{81} \cos^2 \varphi \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

For the purpose of studying the conditions in the ionosphere at any position on the earth's surface, it is customary to take soundings at vertical incidence, by setting up a suitable sending and recording equipment at the place in question, and projecting the radio waves vertically upwards. Under these conditions the angle $\varphi = 0$, and thus we can determine the density of ionisation of any particular region in the ionosphere from its proportionality to the square of the highest frequency which is returned from that region, i.e., from a measure of the critical frequency f^0 for the region in question, as given by the relation

$$f^0 = 9\sqrt{N} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

Since the critical frequency is greater for the F region than for E region, we know that the density of ionisation is correspondingly greater for the upper than for the lower region. Moreover, if the value of the critical frequency is studied at intervals of the day and night and also at different seasons, it will be found to vary very markedly, the frequency and so the ionisation density being a maximum during the daytime and a minimum at night. Typical variations of the critical frequencies for the E and F_2 regions are shown by the graphs in Fig. 7, from which it will be appreciated that there is a close correlation between the density of ionisation in the two regions and the occurrence of sunrise and sunset.

Returning now to the simple equation (1) given above, it will be realised that as we separate the sending and receiving stations, the value of the angle φ will increase and so, for a given density of ionisation N, the critical or penetrating frequency will increase in accordance with the relation

$$f = \frac{9\sqrt{N}}{\cos} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

This equation gives the upper limit of the frequency which may be used for radio communication between two points on the earth's surface. As the distance of transmission is increased the maximum frequency which may be used under a given condition of ionisation increases in the manner defined by equation (3). If higher frequencies are used, the radio waves will penetrate the ionised region and no signals will reach the receiving point. If a frequency less than this critical value is used, signals will at first still arrive at the receiver, but of a reduced intensity; and then as the frequency is very much reduced the waves

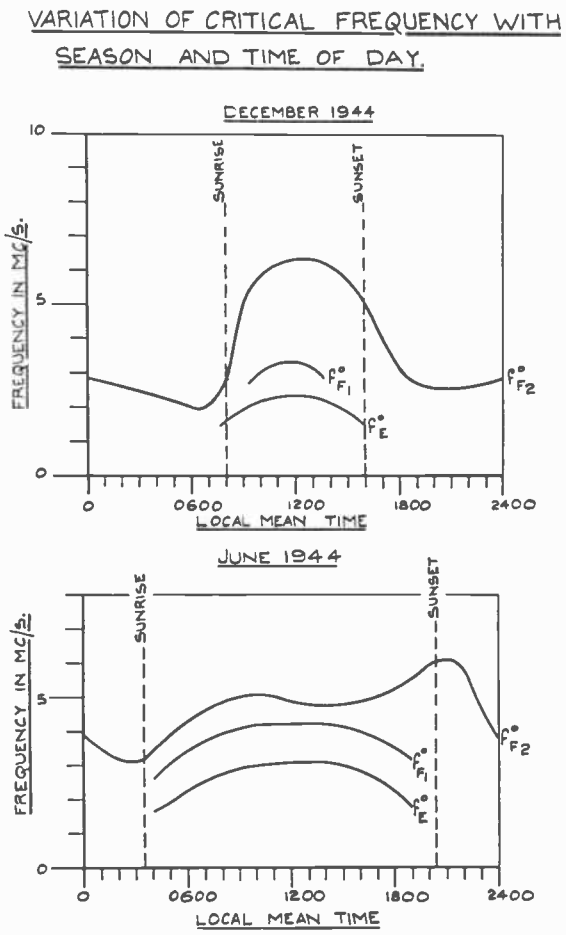


Fig. 7.—Variation of critical frequency with season and time of day. Critical frequency values latitude 50° N.

may be reflected from an ionised region below the one it is intended to use for the particular range over which the transmission is required.

In order to translate the principles outlined above into the requirements of the communication engineer, a series of what are termed maximum usable frequency curves are drawn up, for the required conditions of time and place, and a typical set of these for a latitude in the South of England is shown in Fig. 8. The lowest curve in the diagram shows the variation of the critical frequency (i.e., for zero distance transmission or normal incidence condition) over the 24 hours; while the other curves show the maximum permissible values and their diurnal variations for the different distances of transmission indicated. The actual magnitudes of the frequencies and to some extent the shapes of these curves vary considerably from summer to winter conditions and for various localities on the earth's surface; while for a fixed place on the earth and for corresponding times and seasons there is a continual change in conditions from year to year during the sunspot cycle which has a recurrence period of about eleven years.

(c) Cause of Ionisation in Upper Atmosphere

From this brief review of the subject it will be clear that radio communication over any but the shortest distances depends upon the transmission of radio waves through the ionosphere, and that the effectiveness of such transmission depends upon the height of the reflecting region and the density of ionisation therein. From the known characteristics of the ionosphere as illustrated above, it seems clear that the cause of the ionisation is of solar origin, and the close correlation of the phenomena involved with the occurrence of sunrise and sunset shows clearly that, in general, the greater part of the ionising radiation travels in straight lines. The suggestion that the atmosphere is ionised was first made by Balfour Stewart, who in 1878 put forward the theory that the daily changes in the earth's magnetic field were due to electric currents in a conducting medium above the surface of the earth.

Ever since it was suspected that radio transmission took place through this upper conducting region of the atmosphere, however, speculation has been made as to the cause of this conductivity. For example, J. A. Fleming in 1921[14] suggested that the ionisation was caused by dust particles projected by light radiation pressure from the sun according to a hypothesis previously put forward by Arrhenius and others. An estimate was made that such dust particles might, under suitable conditions, reach the earth's orbit in from 22 to 76 hours after leaving the sun's outer sphere, and arrive with velocities varying from 350 to 1,700 km. per second. It was further estimated that due to the viscosity of the earth's atmosphere, these dust particles would not penetrate below a certain level, probably 60 to 80 km in height, and thus any ionisation caused

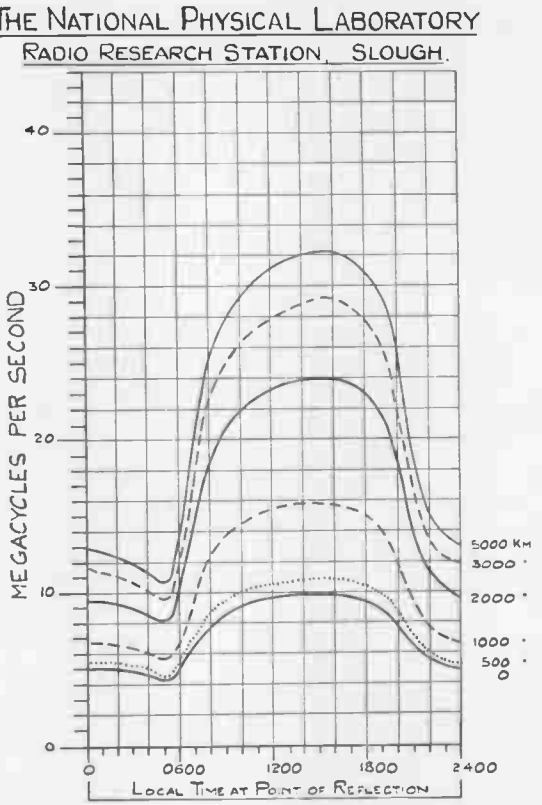


Fig. 8.—The National Physical Laboratory Radio Research Station, Slough. Predicted curves of maximum usable frequencies. Latitude 51½°. *March*, 1946.

by their arrival would have a fairly well-defined under surface determined by the rapid rise in viscosity of the air with increasing density.

On the basis of the results obtained during the 1927 eclipse, already described, Appleton concluded that ultra-violet light from the sun was the main cause of the ionisation in the E region, which was responsible for the reflection of the waves of frequencies in the region of 1,000 kc/s used on this occasion. In 1931, however, S. Chapman[15] suggested that while the ionisation in the upper or F region might be due to ultra-violet radiation, there was reason to suppose that the lower E region was ionised by the arrival of neutral corpuscles emitted by the sun.

This matter was discussed in some detail at a meeting of the Royal Astronomical Society[16] in 1932, and attention was there drawn to an important consequence of the difference in the speed with which ultra-violet light and the neutral particles travel. Due to the fact that these particles or corpuscles travel at a speed

much less than that of light, there are important differences in time and place for what may be termed the "optical eclipse" and the "corpuscular eclipse" respectively when the moon cuts off the stream of radiation and particles emanating from the sun. The difference in times of the two eclipse effects is illustrated diagrammatically in Fig. 9, in which it is seen that the motion of the moon from left to right across the stream of radiation from the sun, causes the corpuscular shadow to lag behind the optical shadow. It can be shown that if the speed of the particles is about 1,600 km. (1,000 miles) per second, the parallel shadow beam carved out by the moon will have a backward slope of about 1° relative to the light shadow giving rise to the optical eclipse. On the occasion of the solar eclipse of 1927 in Great Britain, the results obtained were interpreted as supporting the view that ultra-violet radiation is at least partly operative in the ionisation of the lower region, since a partial return to night-time conditions was experienced at about the period of total eclipse. But it follows from the above reasoning that the corpuscular eclipse occurred nearly two hours earlier, and no radio observations were made at this time. Accordingly, in the eclipses from 1932 onwards, particular attention has been devoted to making observations at times appropriate to both the corpuscular and optical eclipses. Apart from the evidence obtained from such investigations, however, it is to be noted that during intense magnetic storms the ionisation density in the E region sometimes increases so much that radio waves emitted from ground stations, at frequencies normally above the critical value for the E region are unable to penetrate this region and get through to the F layer. Correlation between such storms and solar activity supports the view that the ionising influence in this instance proceeds from the sun with a speed of the order of 1,000 km./sec. and thus is corpuscular in character. These particles may thus account for the excess ionisation experienced in the lower layers on abnormal days.

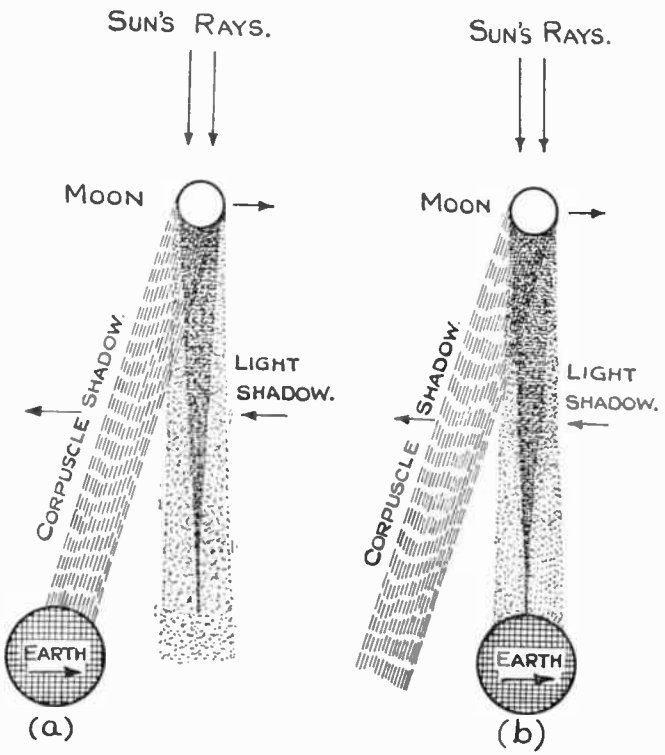


Fig. 9.—Difference in incidence of corpuscular and optical shadows.
(a) *Corpuscular eclipse in progress before optical eclipse begins.*
(b) *Optical eclipse in progress after end of corpuscular eclipse.*

4. Eclipse Observations made between 1932 and 1940

During the various solar eclipses which have occurred since 1930 many radio research institutions in various parts of the world have taken the opportunity of observing the effect of the actual changes in the ionised layers caused by the eclipse, and the resulting effects on the propagation of radio waves through the ionosphere. A summary of the results obtained by various investigators between 1932 and 1934 has been given by E. V. Appleton and S. Chapman[17] in a report presented to the International Scientific Radio Union.

Fig 10 shows the ground track of the optical eclipse of August 31st, 1932, which passed over North America between about 2000 and 2100 G.M.T.; and on the same chart is shown the track of the corresponding corpuscular eclipse, which for an assumed speed of the corpuscles of 1,600 km. (1,000 miles) a second, would have occurred about two hours earlier and over 1,000 miles further east. Observations made in Canada and America showed that the ionisation in the three main regions (E, F_1 and F_2) of the ionosphere decreased by amounts varying from 58 to 30 per cent., the variations taking place approximately in phase with the optical eclipse. The experimental evidence thus obtained indicated very definitely that ultra-violet light is the principal ionising agency for both the E and F_1 regions. On the other hand, some observers recorded effects which occurred before the optical eclipse, in one case two hours before, leading to a conclusion that in the case of the F_2 region at least part of the ionisation changes may have been due to the corpuscular eclipse. The marked effect produced on the field strength of transatlantic signals received in Scotland on the low frequency of 57 kc/s is indicated in Fig. 11; but it is to be noted that nothing abnormal was observed on the corresponding signals received in the South of England.

The whole subject was studied further by international co-operation during the next few years, and a report presented to the above mentioned International Union in 1938 by E. V. Appleton and R. Naismith,[18] the National Physical Laboratory working under the auspices of the Radio Research Board of the Department of Scientific and Industrial Research to make a brief series of measurements of the conditions affecting

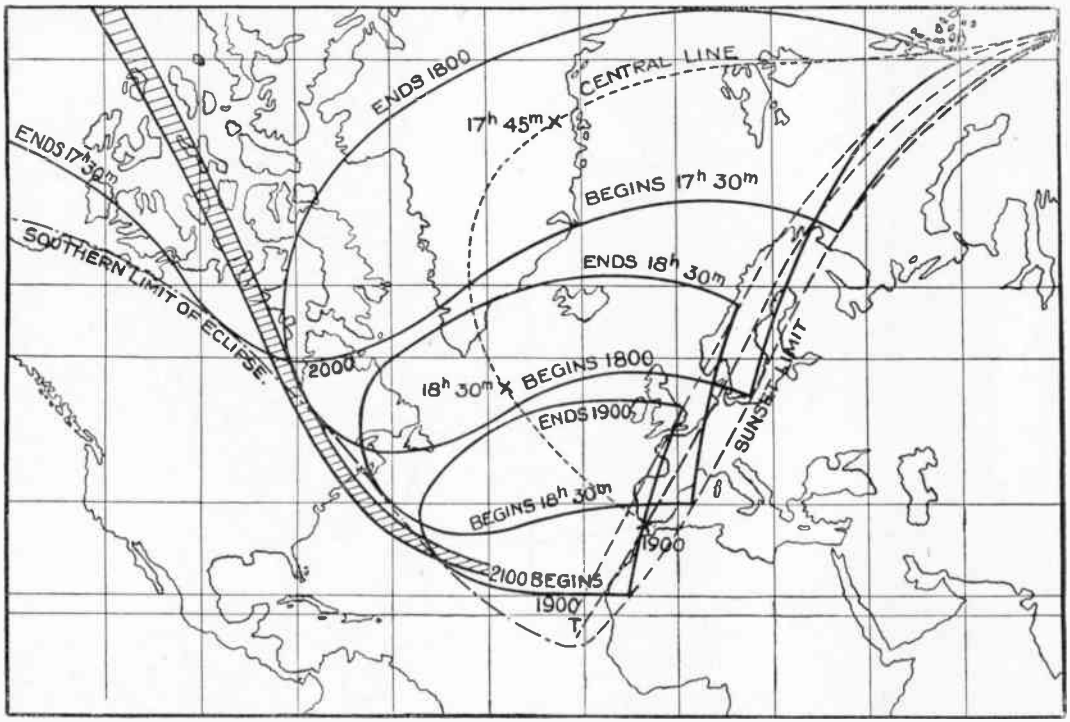


Fig. 10.—*Solar Eclipse,* 31*st August,* 1932. *Chart showing optical shadow (shaded) and centre line of corpuscular shadow.*

substantially confirms the above conclusions for the E and F_1 regions. The results obtained for the F_2 region by different observers still presented some conflict of evidence which is perhaps partially explained by the known abnormal behaviour of this region under different conditions of the sun's altitude.

During the past war it has obviously been difficult to pursue very actively this type of scientific investigation but during the total solar eclipse of October 1st, 1940, observations were made at three places in South Africa by representatives of three different institutions.[19] The conditions were particularly favourable, and the results first confirmed the previously observed behaviour of the E and F_1 regions, and secondly, revealed a marked effect of ultra-violet light in the F_2 region. Here the maximum ionisation density decreased by about 25 per cent., the minimum being reached about 30 minutes after totality.

5. Ionospheric Conditions during the Solar Eclipse of July 9th, 1945

With the termination of the war in Europe, it became possible for the staff of the Radio Division of the transmission of radio waves through the ionosphere on the occasion of the partial eclipse of the sun on July 9th, 1945. These measurements were supple-

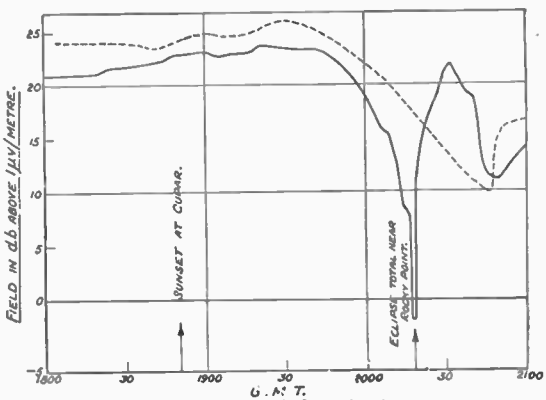


Fig. 11.—*Signal intensity of Rocky Point, U.S.A., observed at Cupar, Scotland, during the solar eclipse of* 31*st August,* 1942. *Frequency—*57 *kc/s.,* 30*th August;* 31*st August,* 1932.

mented by a variety of other observations made by British Service establishments and other organisations working to a general co-operative programme suggested and arranged by Sir Edward Appleton, Secretary of the above-mentioned Department. The detailed analysis and study of the results obtained are still in progress, but a preliminary account of the major effects observed can be given at this stage to supplement the brief statement already published.[20] The path of totality of the optical eclipse at the earth's surface and at a height of 100 km., corresponding to the E region, is shown in Fig. 12.

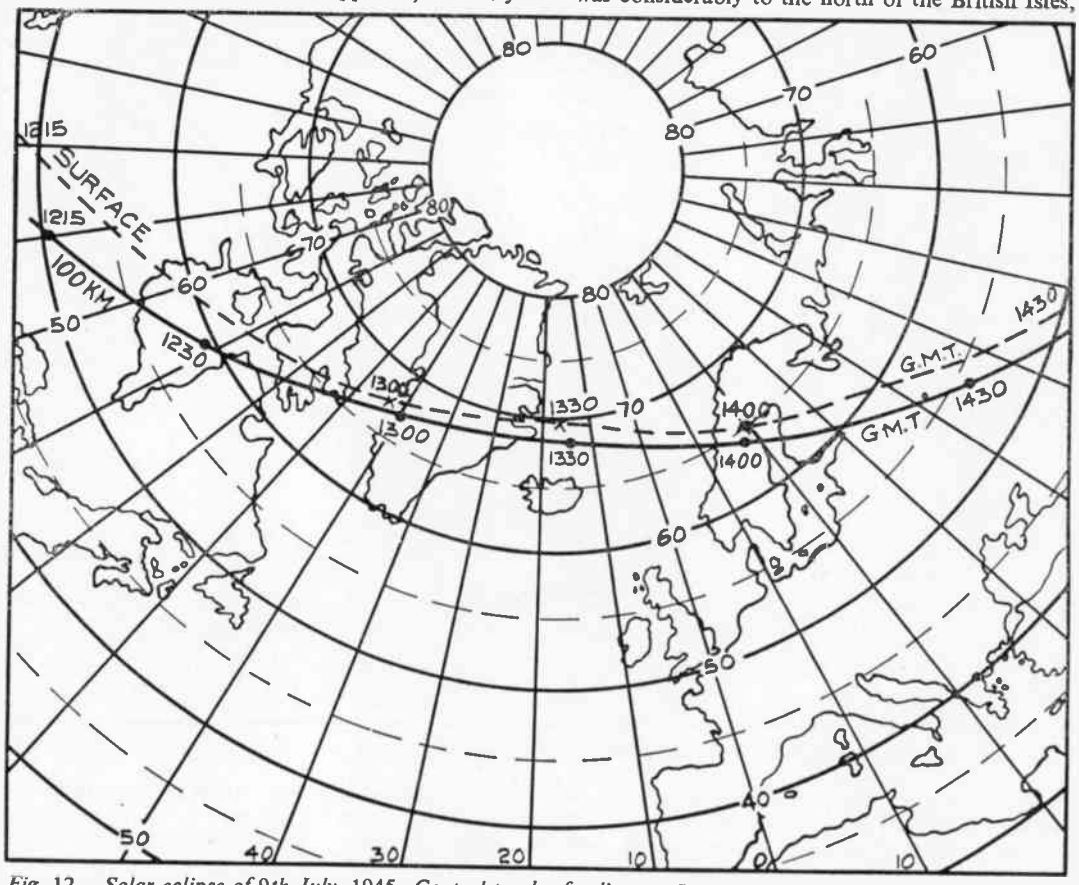


Fig. 12.—Solar eclipse of 9th July, 1945. Central track of eclipse at Earth's surface and at height of 100 km.

The National Physical Laboratory scientific team on this occasion comprised Mr. R. Naismith and Dr. W. G. Beynon, who carried out vertical incidence ionospheric soundings at Loth in Scotland (latitude 58° 4′ N., longitude 3° 47′ W.); Mr. W. R. Piggott, who made similar measurements at Slough, England (latitude 51° 30′ N., longitude 0° 33′ W.); at the Slough station also some radio direction finding observations were made by Mr. W. Ross, while Mr. L. H. Ford and Dr. B. G. Pressey conducted some ultra-short wave field measurements at Teddington and Slough respectively. As will be seen from Fig. 12, the path of totality of the solar eclipse on this occasion was considerably to the north of the British Isles, and at the time of maximum phase, which was about 1400 G.M.T., the magnitude of the partial eclipse was 76 per cent. at Loth and 61 per cent. at Slough. The observations at each station may be conveniently divided into two groups. In the first set of measurements the critical frequencies and equivalent heights of the various ionospheric regions were measured with the aid of a completely automatic equipment generally known as the vertical incidence (h', f) or (P', f) equipment. In the second group of measurements, a close watch was maintained on the amplitude of the signals returned by the ionospheric layers. From these amplitude measurements the variation in the reflection coefficient, or absorption characteristics, of the ionosphere could be studied. Both sets of measurements were made almost continuously on the day of the solar eclipse, and at short regular intervals of half-an-hour or less on each of three days before and after

July 9th. It was hoped that measurements on these six "control" days would furnish mean data for an undisturbed day with which the observations on the eclipse day might be compared.

(a) Vertical Incidence (h′, f) results

It may be convenient to recall here that earlier ionospheric work, including long period observations and measurements during previous solar eclipses, had established the fact that at least two of the ionospheric regions (normal region E and region F_1) were produced largely, if not entirely, by solar ultra-violet radiation. As we have seen, however, in the case of region F_2 the evidence for solar radiation was by no means decisive. During the winter period and also during the summer night period, regions F_1 and F_2 merge into one another to form a common F region. On the other hand, during the summer day the F_1 and F_2 regions can always be recorded as two distinct and separate regions. Hence it was expected that the eclipse last year, occurring as it did near mid-day and mid-summer, might yield some interesting information on the ionising agency responsible for region F_2.

Before considering the actual results obtained, it may be useful to estimate approximately the change in critical frequency or ionisation density to be expected during a partial solar eclipse. It can be shown that if the sun is producing ion pairs at the rate of q per second, and there is a state of quasi-equilibrium between the rate of production and its rate of dissipation, due to the recombination of ions and electrons, then

$$q = \alpha N^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (4)$$

where α is a constant and N is the density of ionisation. Furthermore, it can be shown that if $f°$ is the normal incidence critical frequency then N is proportional to $(f°)^2$. Hence we can write

$$q = \alpha N^2 = K(f°)^4 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (5)$$

where K is a constant. Now in a partial solar eclipse we can assume that the rate of production of ionisation q is reduced in proportion to the degree of totality. From the above simple relation we can thus estimate the change in $f°$ or N to be expected for a given change in q. If the solar disc is 61 per cent. covered then we find that N should be reduced to 63 per cent. of its normal value, and $f°$ reduced to 79 per cent. of its undisturbed value. These figures would thus be expected to apply to the Slough results. For Loth, with 76 per cent. of the disc covered, at maximum totality we should expect N to be reduced to about 50 per cent. of its normal value and $f°$ reduced to 70 per cent. of its undisturbed value. It will be interesting to compare these approximate theoretical estimates with the reductions actually observed.

The results obtained during the eclipse for the three principal ionospheric regions are shown in the accompanying table and in Fig. 13, and may be summarised as follows :—

(i) Normal Region E

At Slough during the optical eclipse period $f°_E$ was reduced from a normal value of 3·3 Mc/s to about 2·9 Mc/s (reduction of 12 per cent. in critical frequency corresponding to a 23 per cent. reduction in ionisation density). At Loth the corresponding reduction in $f°_E$ was from 3·3 Mc/s to 2·6 Mc/s, representing a 21 per cent. fall in frequency or 38 per cent. in ionisation density.

(ii) Region F_1

At Slough the normal incidence critical frequency dropped from a normal value of 4·5 Mc/s to 3·8 Mc/s (15 per cent. in frequency and 29 per cent. in ionisation density). At Loth the corresponding figures were 4·5 Mc/s to 3·4 Mc/s (25 per cent. in frequency or 43 per cent. in ionisation density).

(iii) Region F_2

At Slough $f°_{F_2}$ dropped from a normal value of about 5·0 Mc/s to 4·2 Mc/s, this being a decrease of 16 per cent. in frequency and 30 per cent. in ionisation density. At Loth the corresponding figures were 4·8 Mc/s to 3·9 Mc/s, a fall of 19 per cent. in frequency and 34 per cent. in ionisation density.

SOLAR ECLIPSE OF JULY 9th, 1945

Estimated and Observed Reduction in Ionisation Density in Upper Atmosphere

Place	Magnitude of Optical Eclipse	Approximate Reduction in Ionisation expected from Optical Eclipse	Observed Reduction in Ionisation for Regions		
			E	F_1	F_2
	(per cent.)	(per cent.)	(per cent.)		
Loth..	76	50	38	43	34
Slough	61	37	23	29	30

It will be noticed from the results that in general the observed reduction in critical frequency was always less than the estimate given by our simple calculation. Furthermore, the results for all regions show clearly that solar radiation is certainly an important, if not the sole agent responsible for ionisation of the region.

A theoretical method has now been developed for converting these normal incidence critical frequencies into maximum usable frequencies (M.U.F.) valid for oblique incidence transmission problems. From the reduction in critical frequencies observed during the eclipse it is thus possible to predict that over certain circuits it would not be possible to maintain short wave communication during the eclipse period since the frequency in use might exceed the reduced M.U.F. This would account for the complete fade-out of signals observed during an eclipse over certain circuits

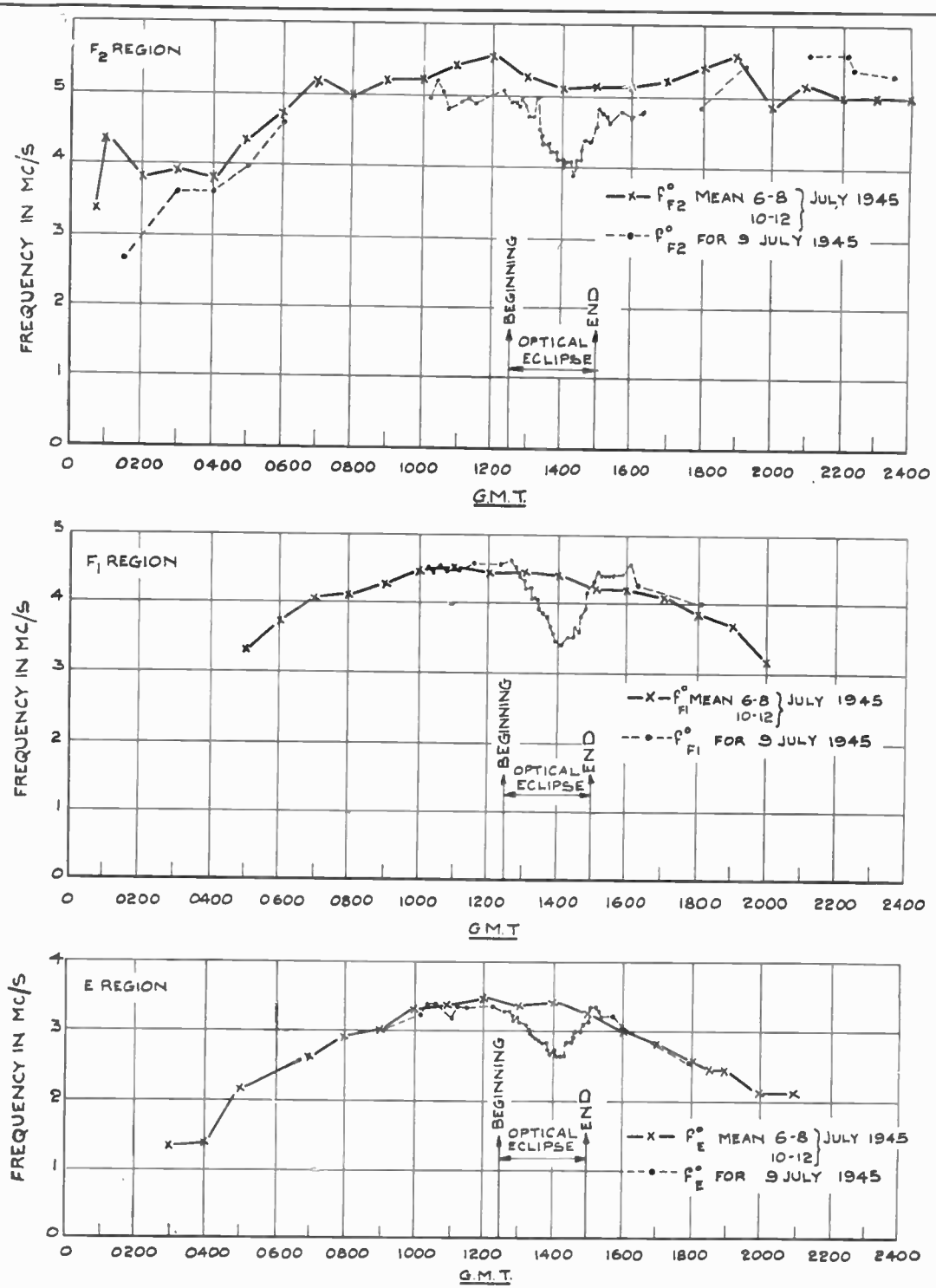


Fig. 13.—*Measurements of critical frequency (f°) made at Loth, Scotland, during the solar eclipse of 9th July*, 1945.

(b) Absorption Measurements during the Eclipse

We may now consider the second group of measurements made during the eclipse period, viz., measurements on the absorption of radio waves incident normally on the ionosphere.

If $A_1, A_2, \ldots A_n$ be the amplitudes of successive echo signals reflected normally from the ionosphere then the reflection coefficient of the ionosphere (ρ) can be written

$$\rho = \frac{2A_2}{A_1} = \frac{3A_3}{2A_2} = \ldots . \frac{(n+1)A_n + 1}{nA_n} \quad \ldots .(6)$$

Hence measurements of the amplitudes of successive reflections from the ionosphere give us the reflection coefficient ρ and the logarithm of ρ will express the ionospheric absorption. It must be remembered that the amplitudes of signals reflected from the ionosphere are usually varying quite quickly, and such changes do not always represent true ionospheric absorption changes. It is to be expected therefore that curves showing ionospheric absorption variations during the eclipse period are rather more irregular than was the case for the critical frequency measurements.

very frequent collisions with neutral atoms or molecules ; this type of absorption is to be expected in that part of the ionosphere where the atmospheric density is large, i.e. in the lowest part of the ionosphere. Although the gas density is large, the ionisation density at this low level is small, with the result that the radio wave is not deviated appreciably—hence the term "non-deviation" or D layer absorption which is applied to this type of absorption. For most practical problems particularly for long distance short wave communications during the daylight period, this D layer absorption is believed to be of considerable importance.

During the eclipse absorption measurements, it was decided to avoid critical frequency complications and to concentrate the bulk of the measurements on frequencies which could be expected to remain well removed from penetration conditions. In this way it was hoped to minimise the effects of deviation absorption, and study closely the solar control of D layer absorption. Most of the normal incidence absorption measurements were thus made on frequencies in the range 1 to 3 Mc/s.

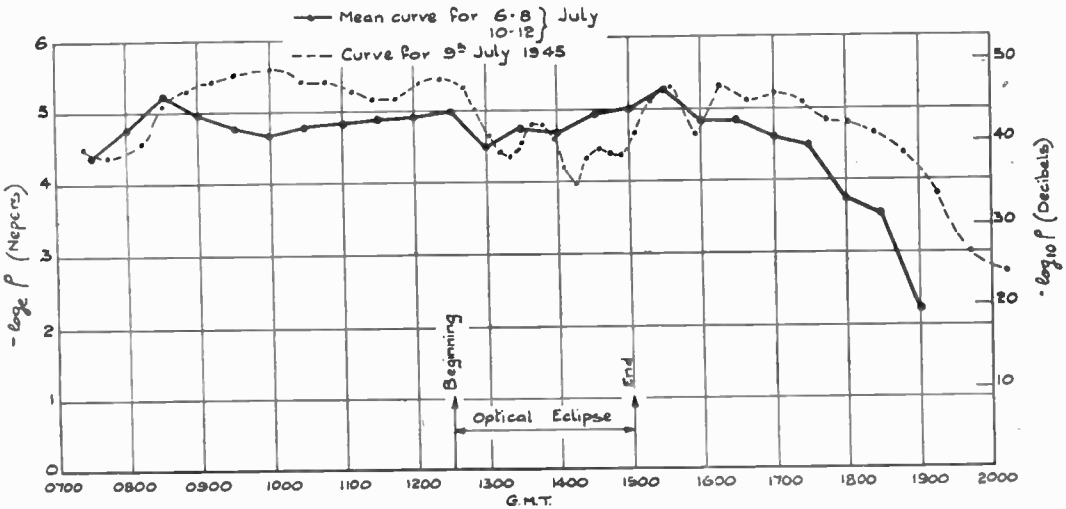


Fig 14.—*Normal incidence Ionospheric absorption measurements made at Loth, Scotland, during the solar eclipse of 9th July*, 1945. *Frequency*—1.4 *Mc/s.*

It is now well known that the net absorption by the ionosphere of an incident radio wave is made up of two main types. In any practical case one or other of these types may predominate, but in general both are present. Pronounced attenuation of the incident signal may occur if the frequency of the signal is near the normal or vertical incidence critical frequency of one or other of the ionospheric regions. This type of ionospheric absorption is known as "deviation absorption," since it is an absorption process associated with marked deviation or bending back of the wave by the ionosphere. In the second place, absorption may be large when the electrons in the ionised layer make

A typical sample curve of the absorption results at vertical incidence on 1·4 Mc/s made at Loth is shown in Fig. 14. The ordinate scale expresses the ionospheric absorption in nepers and decibels and it will be noted that with the commencement of the eclipse at 1230 G.M.T. the ionospheric absorption decreases quite quickly. A minimum value is reached near 1415 G.M.T. after which a recovery sets in, but this recovery is not completed until about 1530 G.M.T., some half-hour after the end of the optical eclipse. Thus during the solar eclipse the ionospheric absorption for a frequency 1·4 Mc/s incident normally decreased by about 1·2 nepers, as a result of which the ampli-

tude of the measured signal increased by about 10 db. In general the eclipse absorption measurements amply confirm the solar control of the ionisation in the lowest part of the ionosphere (below 100 km.).

From these normal incidence measurements it can thus be stated that during a solar eclipse we should expect an increase in signal intensity on all radio frequencies which undergo major attenuation in the D layer (this would include the medium wave and parts of the short wave bands). Such an increase in signal intensity was in fact observed during this eclipse.

(c) Effect of Eclipse on Communication

The above measurements made under vertical incidence conditions were supplemented by observations of the received signal strength of ordinary communication and broadcasting transmissions carried out by the staff of the Post Office and British Broadcasting Corporation. Full reports of these will be available later, but a few typical results may be quoted here. During the eclipse period, observations on transatlantic transmissions on a frequency of 18 kc/s and 58·5 kc/s showed no appreciable change; but on a frequency of 216 kc/s transmitted over a 1,260 km. path, a slight increase in field strength was noted. The absence of a change in field strength on the frequency of 58·5 kc/s confirms the experience of the eclipses of 1932 and 1936 for reception in the South of England of transatlantic signals on a frequency of 57 kc/s, although these are in marked contrast to the effects observed in Scotland in 1932 (Fig 11) and in both England and Scotland in 1925. There is clearly room for further study of the phenomena associated with the propagation of these relatively low frequency signals over the North Atlantic path.

(d) Direction Finding

An extensive series of radio direction finding observations was carried out at the Radio Research Station, Slough, during the eclipse on transmissions from stations in America, U.S.S.R., India, China and Spain using frequencies between 10 and 15 Mc/s. A rotating spaced-loop direction finder of high accuracy was used but the results obtained on the day of the eclipse did not differ significantly from those observed on other days before and afterwards. In most cases the deviations in bearings were under 1°, the highest being 1·6°, and the results in general illustrate the very high accuracy which is obtainable in radio direction finding over long distances—1,300-8,500 km—in favourable circumstances.

(e) Transmissions on Ultra-Short-Waves

In addition to the above work, the opportunity was also taken to examine any effects produced by the eclipse on the transmission of waves of much higher frequencies than those referred to above. By arrangement with the British Broadcasting Corporation, a programme of transmissions was provided from Alexandra Palace station on a frequency of 46·3 Mc/s, and the field strength received at both Teddington and Slough, some 25 miles distant, was recorded continuously for over 30 hours, including the eclipse period. No significant effects due to the obscuration of the sun were observed, and the received signal was sensibly constant to better than 1 db. for the whole period of the measurements. Similar results were obtained on a frequency of 92 Mc/s during measurements of the field strength received at Slough from a transmitter at Teddington, 11 miles away.

6. Conclusions

It has been the object of this lecture to review the results obtained in the many and various types of radio observations made during the successive total or partial eclipses of the sun which have occurred during the past thirty years or so. While in the early work the simple study of the strength of signals and atmospherics yielded useful results, much more has been learnt in the more recent and systematic study of the main characteristics, height, critical frequency and absorption or reflection coefficient, of the ionised regions of the upper atmosphere. The measurements made in the high frequency band, between 3 and 30 Mc/s, in which the ionosphere determines the effectiveness of long-distance transmission, have shown that ultra-violet radiation from the sun is the main cause of atmospheric ionisation. When this radiation is intercepted by the passage of the moon across the path, the density of ionisation decreases and the critical frequencies of the ionised layers fall in a corresponding manner, in synchronism with the occurrence of the optical eclipse. The results obtained in recent years still leave open to doubt, however, the question as to whether some part of the ionisation of the highest or F_2 region is due to the arrival of neutral particles which are emitted from the sun. The effect of the interception of such particles by the moon, should cause a corresponding corpuscular eclipse effect some two hours earlier than the optical eclipse, but a satisfactory and definite conclusion on this point still appears to be lacking.

The effect produced by an eclipse on the signal strength received over any radio communication channel depends very much upon the actual conditions in the ionosphere in relation to the frequency and distance of transmission. If the frequency in use is well below the optimum value for the path or circuit, the reduction in critical frequency and absorption of the ionosphere may well improve the strength of the received signal. If, however, the frequency is at or very near the optimum value prior to the eclipse, then it may well be greater than the maximum usable frequency under conditions of total or partial obscuration of the sun's radiation, and a complete fade-out of the signal will be observed. Thus simple measurements of field strength will only yield useful results if they can be related to the known conditions of the

ionosphere at one or more points along the path of transmission during the solar eclipse under observation.

When the frequency in use is so low that it is always very different from the maximum usable value, such as is the case for the transatlantic transmissions on a frequency of about 57 kc/s, the effects observed will clearly depend upon the characteristics of the lower regions of the ionosphere ; and the results obtained on the above frequency in several eclipses suggest that we still lack a complete understanding of the transmission of radio waves through the ionosphere over paths of the order of 3,000 miles or more in length. This gap in our knowledge is at least partly due to the difficulty of determining the ionospheric conditions (height, critical frequency and absorption), simultaneously at various points along the path of transmission.

After nearly half a century's active and intensive research, therefore, we are still left with a number of problems to be solved before we can say that we fully understand all the features associated with the transmission of radio waves around appreciable portions of the earth's surface.

References

1. J. A. Fleming and others. " Discussion on the Scientific Theory and Outstanding Problems of Wireless Telegraphy." Report of British Association for Advancement of Science, 1912, 401-5, *Marconigraph*, 1912.
2. W. H. Eccles. " On the Diurnal Variations of the Electric Waves Occurring in Nature and on the Propagation of Electric Waves Round the Bend of the Earth," *Proc. Roy. Soc.*, A 1912, 87, 79-99.
3. W. H. Eccles. " Effect of the Eclipse on Wireless Telegraphic Signals," *Electrician*, 1912, 69, 109, *Electrician*, 1925, 94, 208.
4. W. H. Eccles. " Radiotelegraphy during the Solar Eclipse of May 29th, 1919." Report of British Association for Advancement of Science, 1919, 40-42, *Nature*, 1919, 104, 323-4.
5. F. Addey. " Eclipse of the Sun, April 8th, 1921 : Effect Produced at Wireless Stations," *The Radio Review*, 1921, 2, 226-227.
6. F. Addey. " Wireless and the Eclipse." *Electrician*, 1925, 94, 152-153.
7. E. V. Appleton and M. A. F. Barnett. " A Note on Wireless Signal Strength Measurements made during the Solar Eclipse of January 24th, 1925," *Proc. Camb. Phil. Soc.*, 1925, 22, 672-675.
8. E. V. Appleton and M. A. F. Barnett. "Wireless Observations during the Eclipse of the Sun, June 29th, 1927," *D.S.I.R.—Radio Research Special Report No.* 7 (H.M.S.O.).
9. (*a*) G. B. Galle and G. Talon. " Researches on the Propagation of Radio-Electric Waves during the Eclipse of May 9th, 1929," *Comptes Rendus*, 1930, 190, 48-52.
   (*b*) G. B. Galle. " Eclipse Observations at Paulo Condore (Indo-China)," *Onde Electrique*, 1930, 9, 312-348.
10. (*a*) E. V. Appleton and M. A. F. Barnett. " Local Reflection of Wireless Waves from the Upper Atmosphere," *Nature*, 1925, 115, 333.
    (*b*) E. V. Appleton and M. A. F. Barnett. " On Some Direct Evidence for Downward Atmospheric Reflection of Electric Rays," *Proc. Roy. Soc.*, A 1925, 109, 621-641.
11. R. L. Smith-Rose and R. H. Barfield. " An Investigation of Wireless Waves arriving from the Upper Atmosphere," *Proc. Roy. Soc.*, A 1926, 110, 580-614.
12. G. Breit and M. A. Tuve. " A Radio Method of Estimating the Height of the Conducting Layer," *Nature*, 1925, 116, 327.
13. (*a*) E. V. Appleton. " The Existence of more than One Ionised Layer in the Upper Atmosphere," *Nature*, 1927, 120, 330.
    (*b*) E. V. Appleton. " On Some Measurements of the Equivalent Height of the Atmospheric Ionised Layer," *Proc. Roy. Soc.*, A 1930, 126, 542-569.
    (*c*) E. V. Appleton and J. A. Ratcliffe. " Some Simultaneous Observations on Downcoming Wireless Waves," *Proc. Roy. Soc.*, A 1930, 128, 133-158.
    (*d*) E. V. Appleton and A. L. Green. " On Some Short Wave Equivalent Height Measurements of the Ionised Regions of the Upper Atmosphere," *Proc. Roy. Soc.*, A 1930, 128, 159-172.
14. J. A. Fleming. " The Coming of Age of Long Distance Radiotelegraphy and Some of its Scientific Problems," *Journal Roy. Soc. Arts*, 1921, 70, 66-76 and 82-96.
15. S. Chapman. " Phenomena of the Upper Atmosphere," *Proc. Roy. Soc.*, A 1931, 132, 353-374.
16. (*a*) S. Chapman. " Geophysical Discussion on Upper Air Ionisation," *The Observatory*, 1932, 55, 75-81.
    (*b*) S. Chapman. " The Influence of a Solar Eclipse upon Upper Atmospheric Ionisation," *Roy. Astron. Soc.*, M.N., 1932, 92, 413-420.
17. E. V. Appleton and S. Chapman. "Report on Ionisation Changes during a Solar Eclipse," *Proc. I.R.E.*, 1932, 23, 658-669.
18. E. V. Appleton and R. Naismith. " Reports on Ionospheric Changes during a Solar Eclipse," *Proceedings International Scientific Radio Union*, 1938, 5 Fascicule 1, 314-322.
19. J. A. Pierce, A. J. Higgs and E. C. Halliday. "Ionospheric Observations during the Solar Eclipse of October 1st, 1940." (*a*) *Nature*, 1940, 146, 747 ; (*b*) *Nature*, 1942, 149, 701.
20. R. L. Smith-Rose. "The Solar Eclipse of 1945 and Radio Wave Propagation," *Nature*, Jan. 1946, 157, 40.

NOTICES

Honours

The Council have tendered congratulations to :—

Admiral Lord Louis Mountbatten, C.B., G.C.V.O., D.S.O., Personal Naval Aide-de-Camp to The King (Member and President-elect of the Institution), who, as announced in His Majesty's Birthday Honours Lists, is to receive a Peerage of the United Kingdom.

Air Commodore William Charles Cooper, M.A. (Member and Chairman of the Technical Committee), on his appointment as a Commander of the Most Excellent Order of the British Empire.

Major Ronald William Fenemore (Associate) on his appointment as a Member of the Most Excellent Order of the British Empire.

Obituary

Council record with deep regret the death, after a short illness, of **Jerrold Burns**, of London, N.W.11 (Registered Student).

Mr. Burns was registered as a student of the Institution in February, 1942.

Annual General Meeting

Corporate Members of the Institution are requested to note that the twenty-first Annual General Meeting of the Institution will be held on Wednesday, September 25th, 1946, at 6 p.m. The meeting will be held at the London School of Hygiene and Tropical Medicine, Keppel Street (Gower Street), London, W.1. Tea will be served from 5.15 to 5.45 p.m.

In accordance with Article 32, the Council will shortly be circulating a list of members whom they nominate for the vacancies about to occur on the Council.

" After the issue of the Council's lists and not later than twenty-one days after the date of such issue, any ten corporate members may nominate any other duly qualified person to fill any such vacancy by delivering such nomination in writing to the Secretary, together with the written consent of such person to accept office, if elected."

Dates of Graduateship Examination

In future, the Graduateship Examination will be held in the usual principal centres throughout the British Empire on the third Thursday and Friday in May and the third Thursday and Friday in November of each year.

The next Graduateship Examination will, therefore, take place on November 21st and 22nd, 1946. Application to write the November examination must be lodged not later than October 1st.

The results of the May 1946 Examination will be published in the next issue of the Journal, dated July-August.

Anniversary Dinner, October 31st, 1946

Members who have already advised the Institution of their definite intention to be present at the 21st Anniversary Dinner should now apply for their tickets of reservation by forwarding the appropriate remittance.

Available accommodation at the dinner has already been over applied for and in order to give fair opportunity to all members who have indicated their intention or wish to be present, such members are requested to apply for their reservation tickets as quickly as possible and such applications will be treated strictly in order of receipt.

Papers for Reading at the Convention

Council confirm that the 1947 Convention will commence with a reception in Bournemouth, on Monday, May 19th, 1947. The Convention arrangements will extend over the following Tuesday, Wednesday and Thursday, members dispersing on Friday, May 23rd, 1947.

Papers offered for reading and discussion at the Convention must be lodged with the Papers Committee not later than December 1st, 1946.

GRADUATESHIP EXAMINATION—NOVEMBER 1945.

Final Supplementary Pass List

The following overseas candidates have satisfied the Examiners, as indicated :

Passed Whole Examination

EPSTEIN, Jan Leo — Haifa.

Passed Parts III and IV

PARTRIDGE, Jack Edward — Wellingborough

Passed Parts I, II and IV

ROTTENBERG, Robert Ruben — Cairo

Passed Parts I, II and III

MARTIN, Ronald Stewart — Sydney

Passed Parts I and II

THOMPSON, Arthur Lawrence — Melbourne

MOURAD, Abraham — Cairo

Passed Part I only

BURTON, George — Tarporley

THE ELECTRON GUN OF THE CATHODE RAY TUBE—PART II.*

by

Hilary Moss, Ph.D., B.Sc., (Eng.). (*Associate Member*).

A Paper read before the London Section of the Institution on Wednesday, April 17th, 1946, and the North-Western Section on Tuesday, April 30th, 1946.

LIST OF SYMBOLS USED (in addition to those defined in Part I)

E_{a_1}, E_{a_2}, etc. = Potentials on first, second, etc., anodes in volts.

I_a = Total electron current emerging from grid hole in μA.

$I_a|_o$ = Total electron current emerging from grid hole in uA. at $E_g = O$.

E_c = Negative grid voltage required to cut off cathode current I_a.

E_g = Grid voltage. Negative unless otherwise stated.

E_d = Grid drive, i.e. $|E_c \sim E_g|$ where $E_g > E_c$.

D = Diameter of grid hole in mm.

t = Thickness of grid material in mm.

f = Anode to grid spacing in mm.

b = Cathode to grid spacing in mm.

k = Constant of scale. Note that k in Part I is also defined as the Boltzmann constant, but no confusion is likely.

l, ρ, s, q, k_1, k_2, . . . , K_1, K_2, . . . , , , = Various constants.

Notes

Experiments are described on two types of triode, designated A and B, shown in Figs. 12A and 12B respectively. Curves referring to triode A have a reference number followed by A—curves referring to triode B have a reference number followed by B.

Section and diagram numbering follow on from Part I.

Ref. p1 means "Electron Gun of the Cathode Ray Tube"—Part I*, and ref. p2 means "Engineering Methods in the Design of the Cathode Ray Tube."†

CONTENTS

2.1. Introduction

In Part I of the paper, certain inherent limitations in the performance of any electron gun were discussed. Although, ultimately, the design is circumscribed by the two laws there treated, considerable freedom exists for effecting compromises in operational performance.

* Part 1. *Journal Brit. I.R.E.*, January, 1945.

† *Journal Brit. I.R.E.*, December, 1945.

It may be necessary to produce a wide range of compromises between the performance with regard to focus, brilliance and deflection defocusing. Some of this work can be adequately treated by reference to quite simple laws, ref. (p2). However, it is often necessary to employ methods based on a deeper knowledge of the electron gun than can be obtained by similitude and scaling theory. An accurate knowledge of the changes effected by alterations in electrode potentials and gun geometry is generally essential in all but the simplest problems. It is with these questions that we now concern ourselves.

The detailed theory of the operation of the electron gun is exceedingly intricate and it is doubtful that much progress can be made without introducing many approximations. The theory to be presented seems to the author to be the simplest possible, consistent with a reasonable regard to what happens in practice, but it is certainly not suggested that it is more than an approximation. The most important thing about approximations is some understanding of the magnitude of the errors which they involve and an adequate appreciation of the postulates on which they are based. An attempt has been made in this paper to treat both these features, but as far as possible without obscuring the development of the theory.

2.1.1. Gun Analysis—the Final Lens

We saw in the introduction that it is convenient to divide the whole electron gun into two fairly distinct portions—the triode and the final lens. The latter has been the subject of extensive treatment in the literature.[6] This work has been largely directed to developing the analogies between Gaussian " thick " lens light optics and the equivalent electron optical cases, comprising co-axial cylinders, discs, etc. From a designer's point of view such investigations have only a limited importance. Whatever type of lens is employed, it can be considered merely as a device producing a certain magnification between the crossover and the spot, and (usually) intercepting some of the current in the process. In the absence of all lens aberrations the conditions at the screen are similar to those at the crossover only on a different scale of size.

In the case of the electrostatically focused tube, an alteration in the geometry of the final lens will entail an alteration in the focusing voltage ratio. However, in practice, for almost any normal type of gun, it will be found that the focusing anode voltage lies between about $\frac{1}{3}$ and $\frac{1}{8}$ of the voltage on the final anode. In a well-designed gun this focusing anode takes a negligible current, so that it can be supplied from a very high resistance chain across the main H.T. supply. It is so easy to vary the focusing anode voltage by tapping in at a suitable point on this resistance chain that there is little point in worrying much during the design of the gun as to what the required focusing anode voltage will be.

Again the question of the spherical aberration in the final lens is generally of little moment, at least with electrostatically deflected tubes. We saw in section 1.1.3 that for this class of tube we are normally obliged to limit the beam width to some 3 mm. in order to avoid deflection defocusing. It has been shown in these laboratories by means of Hartmann diaphragm tests that the simplest form of lens, consisting of a pair of equal diameter coaxial cylinders, will give negligible aberration for the focal lengths in customary use, provided only that the diameter of the cylinders is not less than approximately eight times the stop diameter. The use of such a cylinder diameter usually presents no problem, so that the spherical aberration is easily avoided. With the larger beam widths employed in magnetically deflected tubes, rather more care is necessary if the focusing is electrostatic, but the problem is not serious. In general, therefore, the final lens is of limited interest to the designer. It is sufficient to treat it in terms of the equivalent "thin" lens obeying the simple optical laws which define the magnification it produces between the crossover and the spot.

2.2. The Triode

This is the portion of the electron gun around which most of the interest is centred.

In order to discuss the problems on a quantitative basis, we shall discuss two specific forms of triode geometry. These are shown in Figs. 12A and 12B respectively. These diagrams give a letter code defining the significance of the variables concerned. It will be noted that the dimension f in Fig. 12A has not quite the same significance as in 12B.

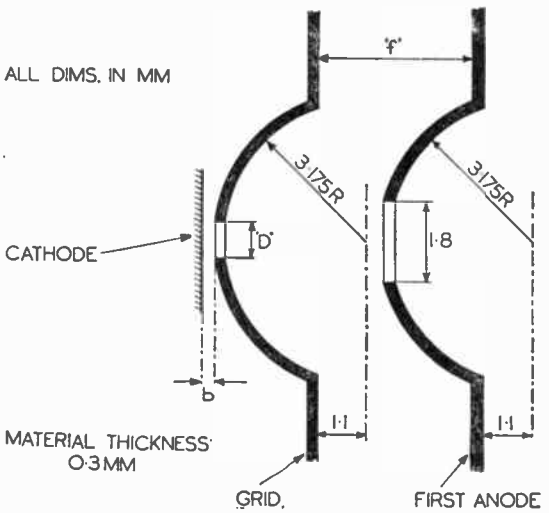


Fig. 12A.—Dimensions and Letter Code for Triode A.

In the latter diagram f denotes the difference between the tangent planes resting across the adjacent sides of the anode and grid apertures. In Fig. 12A the dimension f is not quite equal to the distance between such tangent planes on account of the difference in sizes of the holes in the grid and anode, but the discrepancy is slight.

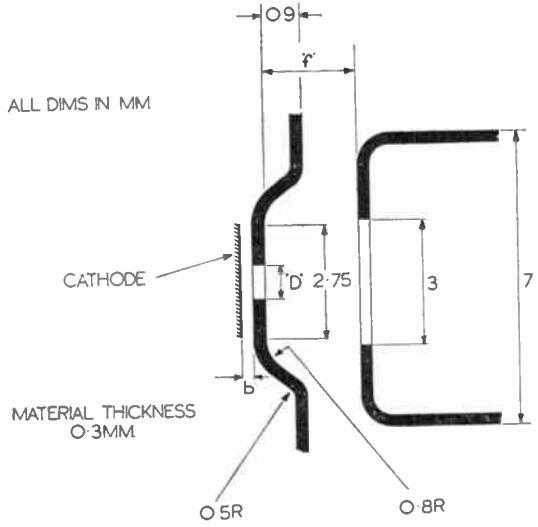


Fig. 12B.—*Dimensions and Letter Code for Triode B.*

2.2.1. Significance and Determination of the Cut-off Voltage E_c

It will be seen later that the cut-off voltage is a most vital factor in defining the whole performance of the electron gun. In many cases quite large geometrical changes can be made, but providing the cut-off voltage is maintained constant, the effect of these changes is negligible. Several important formulæ involve the cut-off voltage. Its determination is thus quite vital to triode design.

We commence with a dimensional investigation. It is postulated that the cut-off voltage must depend on the grid hole diameter D, the grid material thickness t, the cathode to grid spacing b and the anode to grid spacing f. Experimental evidence shows that the anode hole diameter has a negligible effect and it will be ignored. A further approximation is neglect of the radius of curvature of the grid and anode dishes.

Thus $E_c = k.\Phi[D,t,b,f]$

whence we assume the usual

$$E_c = k.D^n.t^l.b^s.f^q \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (27)$$

The requirements of dimensional homogeneity then give

$$n + l + s + q = 0 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (28)$$

since we assume that the term k in (27) has the dimensions of a voltage.

Dimensional analysis yields no further information than that implicit in equation (28). A theoretical analysis is exceedingly difficult and the determination of the best fitting values of the indices n, l, s and q was done experimentally.

At first sight a satisfactory method of such experimental investigation would appear to be to construct a series of triodes each of fixed geometry and measure the cut-off voltages. In practice, however, it proves virtually impossible to avoid appreciable changes in other parameters while varying the one under investigation, and the resulting analysis of the data is very involved. The correct solution is to construct triodes in which one parameter is continuously variable during operation, so that no other changes can occur.

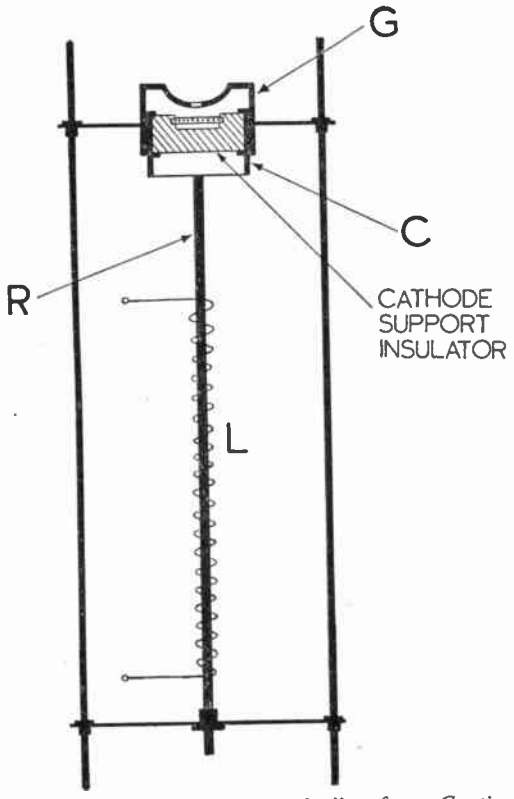


Fig. 13.—"*Thermal Expander*" *for Continuous Variation of Cathode/Grid Spacings.*

Fig. 13 illustrates the method employed for continuous variation of the cathode to grid spacing. The cathode assembly is mounted in a cylindrical "cradle" C which can slide freely inside the grid cylinder G. The movement is controlled by the thermal expansion of

the rod R which is heated by the coil L. Satisfactory results were obtained by using a rod of "Staybrite" stainless steel about 6 in. long. A heating power of some 10 watts sufficed to reduce the cathode to grid spacing from about 0·8 mm. to zero in about five minutes. All intermediate values were obtained by reduction of heater wattage, a waiting period of about five minutes being generally sufficient to attain equilibrium.

Figs. 14A and 14B show some results obtained using this form of continuously variable cathode to grid spacing device. As always, 14A refers to triode A and 14B to triode B. For comparison purposes, a "best fit" rectangular hyperbola has been drawn over the curve of Fig. 14B. It is apparent that the "fit" is reasonable, and the same will be found to apply to the curves of Fig. 14A, so we conclude that the cut-off voltage is inversely proportional to the cathode to grid spacing. Hence in equation (28) $s = -1$.

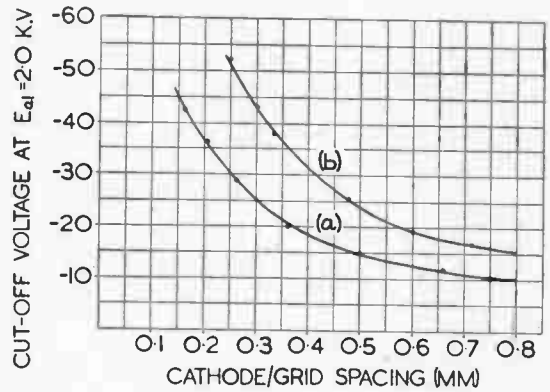


Fig. 14A.—*Cut-off/Cathode to Grid Spacing Relationship for Triode A.*
In Curve (a), D = 0·8 *mm.* f = 2 *mm.*
(b), D = 0·8 *mm.* f = 1·3 *mm.*

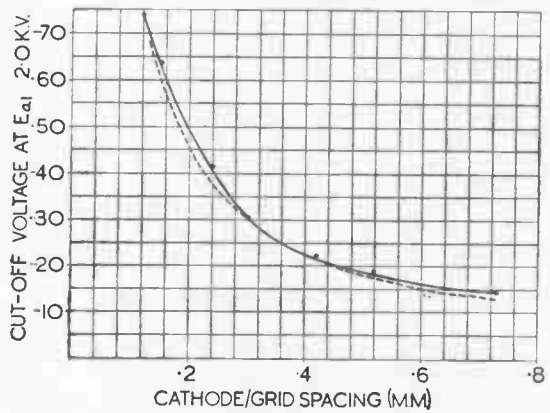


Fig. 14B.—*As* 14A *but for Triode B.* D = 0·8 *mm.* f = 1·77 *mm.*

Similarly, triodes were constructed in which the anode to grid spacing could be continuously varied during operation. In this case, on account of the greater range of movement involved, a less sensitive mechanism than thermal expansion was adequate to effect spacing variations. The anode was arranged to slide smoothly along guide rods when the tube was tapped.

Typical cut-off voltage against anode to grid spacing curves are displayed in Figs. 15A and 15B. Rectangular hyperbolæ have been drawn over these curves to indicate the degree of "fit." It will be seen that the cut-off voltage is approximately inversely proportional to grid spacing, so that in equation (28) $q = -1$.

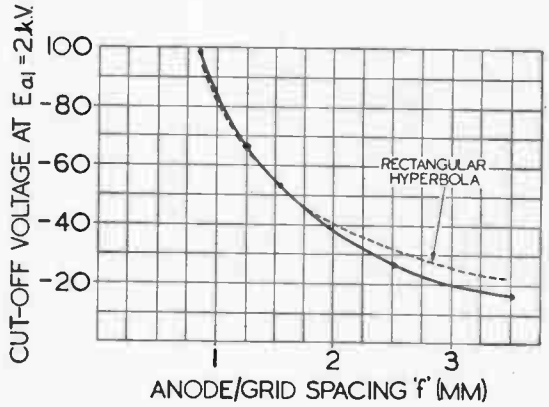


Fig. 15A.—*Cut-off/Anode to Grid Spacing Relationship for Triode A.* D = 0·8 *mm.* b = 0·15.

The determination of the effect of grid hole diameter change on cut-off voltage is more difficult, since it is scarcely possible to construct triodes with continuously variable grid hole sizes. The method adopted, therefore, was to construct a series of separate triodes with different grid hole sizes, all other parameters being held as closely constant as possible. Correction for the residual minor differences in the other parameters was made on the basis of $s = q = -1$, as already established. A typical result is shown in Fig. 15B, which indicates that the cut-off voltage varies approximately as the cube of the grid hole diameter. Hence in (28) $n = 3$.

Putting in the values of s, q and n, already established, into (28) then leads to the conclusion that $l = -1$, so that the cut-off voltage appears to be inversely proportional to the grid material thickness.

Finally, therefore,

$$E_c = \frac{k.D^3}{t.b.f} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (29)$$

It must be emphasised that these results are only approximations, and will not yield accurate results for *wide* changes in the parameters of the triode.

This fact, however, does not prevent (29) from being of considerable practical utility.

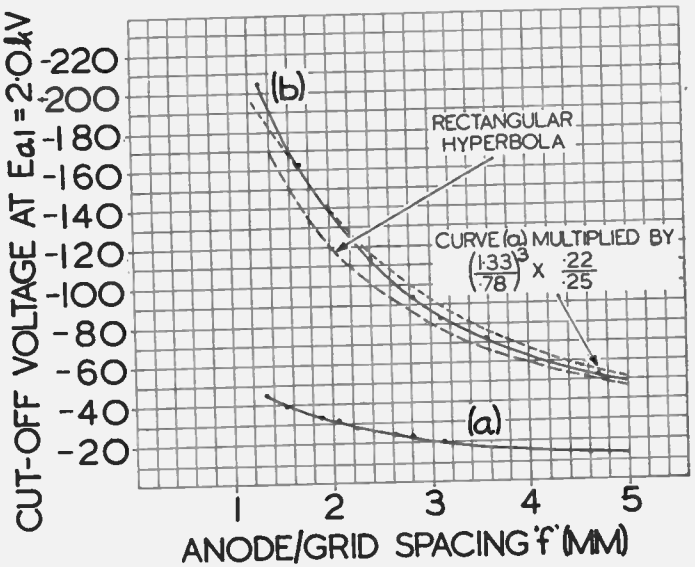


Fig. 15B.—*As for* 15A *but relates to Triode* B.
In Curve (*a*) D = 0·78 *mm.* *b* = 0·22 *mm.*
In Curve (*b*) D = 1·33 *mm.* *b* = 0·25 *mm.*

Where it is desired, for example, merely to estimate the change in cut-off voltage due to some small change in geometry, such as might be caused by inaccuracy in tooling or assembly, then (29) is entirely adequate and may be applied with confidence.

The relationship between the cut-off voltage and first anode voltage is shown in Fig. 16. See ref. p.2, section 2.2, Fig. 1. It will be seen that the cut-off voltage is very closely a linear function of the anode voltage, but is not quite proportional to it, as the curves do not pass through the origin. This displacement at the origin is due to the effects of emission velocity and contact potential. However, for most practical purposes, provided the cut-off voltage is not too small, we may write

$$E_c = k_1 E_{a_1} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (30)$$

where E_{a_1} is the first anode voltage.

2.2.2. Effect of Cut-off Voltage on Total Current at $E_g = O$

Consider any triode of constant geometry. For anode voltage E_{a_1}, suppose the corresponding cut-off voltage is E_c. Let $I_a|_o$ be the total cathode current (i.e the current emerging through the grid hole) at zero grid ($E_g = O$). It is found experimentally that if E_c is varied by variation of E_{a_1}, in accordance with equation (30), that

$$I_a|_o = k_2 . E_c^{3/2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (31)$$

For any one triode of fixed geometry, this relation is easily justified theoretically. For under space charge limited conditions

$$I_a|_o = \gamma . E_{a_1}^{3/2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (32)$$

This is merely the familiar three halves power law which applies to any space charge limited device (ref. 2, section 2.3). Substituting for E_{a_1} in (32) from (30) then yields (31).

The important point about equation (31) is that it appears to hold quite well when the cut-off voltage E_c is changed by certain geometrical alterations, and not merely by change of anode voltage E_{a_1}.* Careful measurements made by the author with special triodes, in which the anode to grid spacing could be varied during operation, have confirmed that (31) holds to within about 10 per cent. for wide changes in anode to grid spacing or grid hole diameter, provided that the ratio of grid hole diameter to cathode to grid spacing lies between about 3 and 6. Table 1A, for example, shows some results obtained on a triode belonging to type A (Fig. 12A), in which the anode was mounted on a sliding rod system so as to permit adjustment of the anode to grid spacing *f* during operation. It will be seen that the cathode current at zero grid ($I_a|_o$) is defined by the cut-off voltage E_c, but is substantially independent of the *individual* values of anode to grid spacing or anode voltage which together determine the cut-off voltage. This means that to a first approximation the constant k_2 in equation (31) is independent of anode voltage or anode to grid spacing.

Table 1B is similar and refers to triode B. Again it is seen that the cathode current is dependent on the cut-off voltage only.

TABLE 1A.

Anode to Grid Spacing (*f*)	Anode Voltage (E_{a_1})	Cut-off Voltage (E_c)	Current at Zero Grid ($I_a|_o$)
1·32	1,000	− 32·6	727
2·08	2,000	− 34	766
3·12	4,000	− 34	770
0·9	1,000	− 47	1,200
1·56	2,000	− 49	1,310
2·08	3,000	− 48	1,290

Cathode to grid spacing *b* = 0·15 mm. Grid hole Diameter D = 0·8 mm. } Triode A.

* The existence of a relationship of this type was first pointed out to me by members of the C.R.T. Research staff at G.E.C., Wembley.

Furthermore, it is apparent that approximately the same currents are obtained as for triode A for equal cut-off voltages. It can be concluded that the shape of the grid and anode has little effect on the constant k_2 in equation (31).

TABLE 1B.

Anode to Grid Spacing (f)	Anode Voltage (E_{a1})	Cut-off Voltage (E_c)	Current at Zero Grid ($I_a|_o$)
1·26	1,190	− 33	693
2·23	2,020	− 33	715
3·4	3,240	− 33	710
1·26	1,810	− 48	1,180
2·23	3,000	− 48	1,200
3·5	5,150	− 48	1,210

Cathode to grid spacing b = 0·22 mm.
Grid hole diameter D = 0·78 mm. } Triode B.

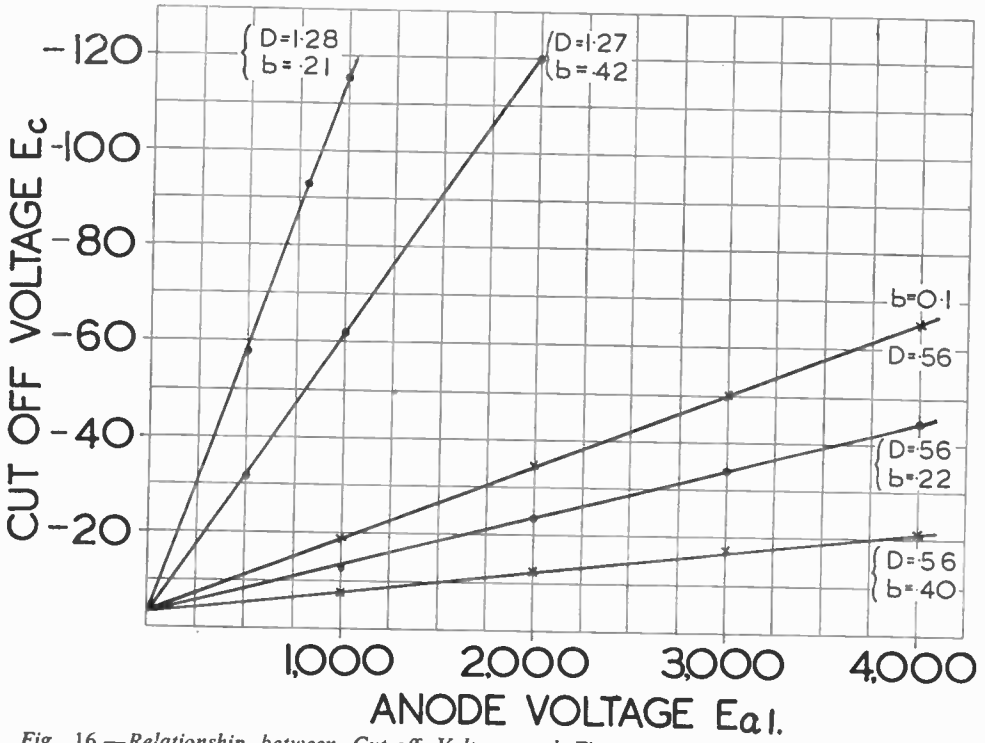


Fig. 16.—Relationship between Cut-off Voltage and First Anode Voltage. Note that Grid Voltage necessary to suppress Emission at zero Anode Voltage was − 3 Volts. See ref. p2, section 2.2.

However, equation (31) can be justified only for a fixed triode geometry, and it is clearly impossible for it to hold generally for any type of triode. Further investigations by the author, which can only be summarised here, have shown that in general for *any* triode equation (31) can be more generally expressed in the form

$$I_a|_o = k_2.E_c^p \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (31A)$$

For *any* fixed triode (31A) holds quite accurately, but the terms k_2 and p are not constants. They are functions of the grid hole diameter, grid material thickness and cathode to grid spacing, A complete investigation of the nature of these functions has not been made, as the matter is rather academic. Fig. 17A, relating to triode A, shows some values of k_2 and p with variations of cathode to grid spacing only. The individual variations in these quantities are quite large, but they tend to partially compensate one another for cut-off voltages in the region 30-60 volts, so that the net changes in $I_a|_o$ are not too marked.

The best values of k_2 and p appear to be 3 and 3/2 respectively for the range of cathode to grid spacings in general use, so that the empirical formula for cathode current at zero grid becomes

$$I_a|_o = 3.E_c^{3/2} \ldots\ldots\ldots\ldots (31B)$$

where $I_a|_o$ is in μA. for E_c in volts.

2.2.3. Effect of Grid Modulation on Total Current

An extended series of measurements made in these laboratories has confirmed earlier findings of the G.E.C. at Wembley that the total cathode current varies as the 7/2 power of the grid drive. By grid drive is meant the modulus of the difference between the actual grid voltage and the cut-off voltage. It is, of course, assumed here that the actual grid voltage is more positive than the cut-off voltage, since otherwise the total current is zero. Thus we have

$$I_a = k_3.E_d^{7/2} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (33)$$

where E_d is the grid drive. Now when the grid drive is numerically equal to the cut-off voltage, the actual

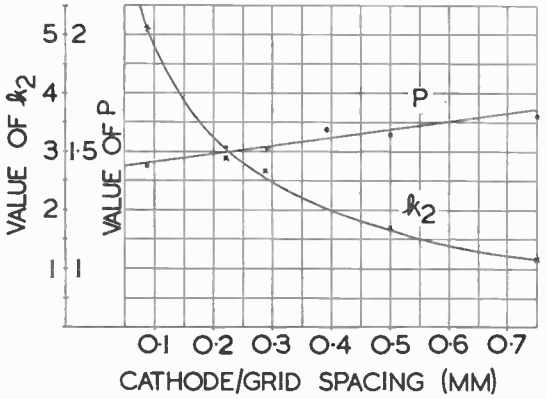


Fig. 17A.—*Variations in* k_3 *and* ρ *with Cathode to Grid Spacing.* (*Formula* 31*A*).

grid voltage is zero, and, therefore, the current given by (33) must equal that given by (31B). Equating these values and setting $E_d = E_c$ in (33) gives

$$3.E_c^{3/2} = k_3.E_c^{7/2}$$

whence

$$k_3 = 3.E_c^{-2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (34)$$

Substituting from (34) into (33) immediately gives

$$I_a = 3.E_d^{7/2} \,.\, E_c^{-2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (35)$$

where I_a is in μA. for E_d and E_c in volts.

Equation (35) is to useful accuracy a "universal" modulation formula which enables us to calculate the total current from a triode of customary geometry when the cut-off voltage is known.

With the restrictions on scale noted in the previous section and provided that $|E_d| > |E_c|/4$, its use will permit the grid drive for a specified total current to be calculated to within about $\pm$ 5 per cent.

Figs. 18A and 18B show typical modulation curves for triodes A and B respectively. It will be seen that to useful engineering accuracy the curves are dependent only on the cut-off voltage. The change in triode geometry from system A to system B has made only a second order difference to the curves.

We note from (35) that the total cathode current for a given grid drive varies inversely as the square of the cut-off voltage.

2.2.4. The Emitting Area of the Cathode Surface

It is clearly of primary importance to know how much of the cathode surface is contributing to the current as given by (35). In order to investigate this point, triodes were made up in which the cathode surface was covered with a very fine mesh nickel gauze, the latter serving as a calibrating scale. An optical photomicrograph of the cathode system thus assembled was then taken through the grid hole. This is shown at the right hand top corner of sheet No. P.1. The triodes were then mounted in a suitable tube and an electron-optical picture of the cathode image (refer to Fig. 1) was then projected on to a fluorescent screen and also photographed. By careful comparison of the optical and electron-optical pictures, it will be seem that at zero grid voltage the emitting area of the cathode surface is that portion of the cathode bounded by the projection of the grid hole on the cathode face. All the triodes in question were of the sliding anode type, and it was conclusively shown that the above result is quite independent of either anode voltage or anode to grid spacing. The truth of these conclu-

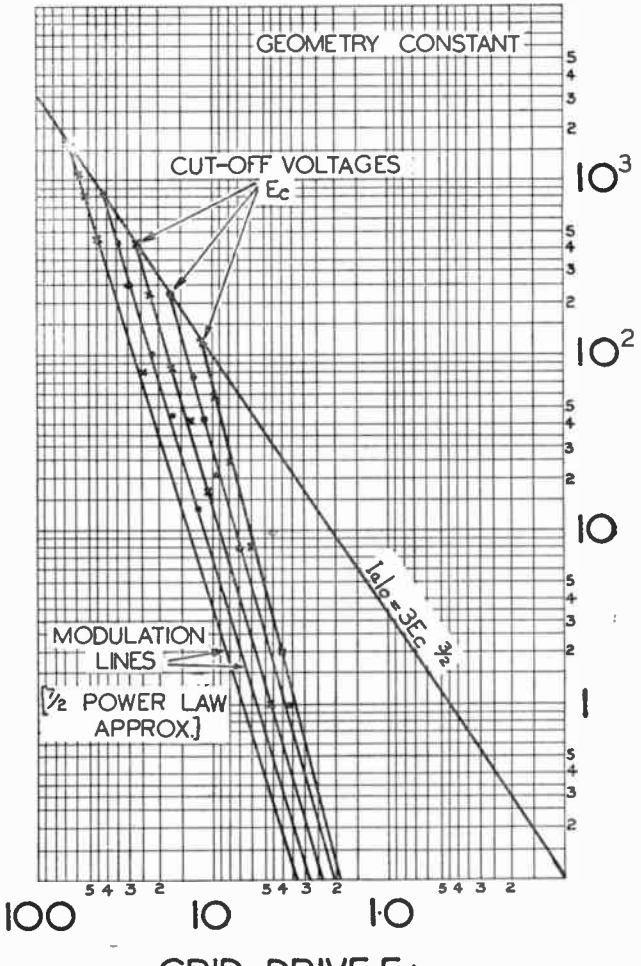


Fig. 18A.—*Modulation Curves for Triode A. Cut-off Voltage Varied by alteration of First Anode Voltage.* D = 0·8 *mm.* b = 0·21 *mm.*

sions can be verified by reference to the right hand photos of sheet P.1. (Photos p. 3, p. 4, and p. 5, facing page 22.)

Further triodes were constructed with continuously variable cathode to grid spacings (refer to Fig. 13), by means of which it was established that the above result is also independent of the cathode to grid spacing, at least up to the point where the ratio—grid hole diameter/cathode to grid spacing is not less than about 3. For values less than 3, the conclusions are uncertain, since the cathode image becomes very blurred, presumably due to severe aberrations in the immersion lens. This conclusion can be verified by reference to the photos on sheet P.2.

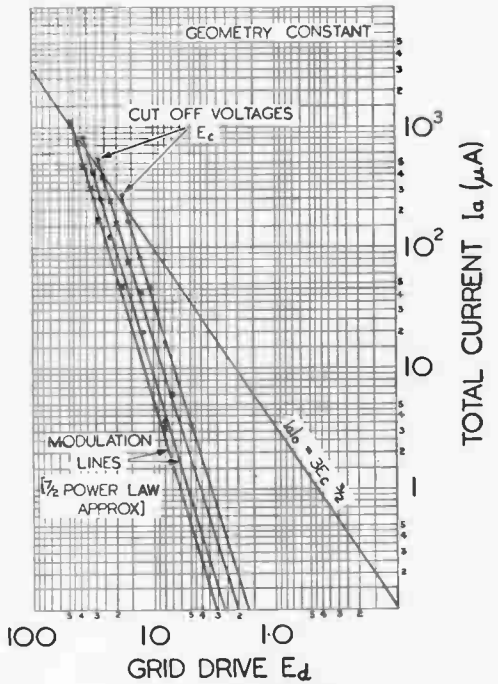


Fig. 18*B*—*As* 18*A but for Triode* **B**
D = 0·8 *mm.* *b* = 0·22 *mm.*

Inspection of the photographs on sheet P.1 will show that the magnification between the cathode surface and its image is almost entirely independent of the grid bias. The effect of making the grid more negative is merely to reduce the cathode emitting area. Application of the principle of voltage similitude (ref. p2) leads us to the important conclusion that the *magnification between the cathode surface and its image is almost independent of the voltages applied to the gun.* Complete independence will be assumed in the theory herein developed.

For a constant anode voltage and geometry the emitting diameter of the cathode surface appears to fall almost linearly with increase in grid bias, becoming zero at the cut-off point.

Hence, we may write

$$2y = \frac{E_d}{E_c} . D \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (36)$$

where $2y$ is the diameter of the emitting portion and D, as always, is the grid hole diameter.

2.2.5. The Distribution of Electron Current Density across the Cathode Surface

Although at $E_g = 0$ all the cathode surface under the grid hole is emitting, it is clear that the emission density must be much higher at the centre of the cathode under the centre of the grid hole than at that portion nearly under the rim of the hole. Furthermore, the system has radial symmetry, so that the density distribution on any diametral line through the cathode centre must be the same.

Fig. 19A shows an experimentally obtained density plot across a cathode surface. In order to ensure space charge limited conditions, the cathode was operated at nearly 1,000° C. Four curves are shown corresponding to four different values of grid bias as marked. The figure is drawn to scale and shows the essential dimensions.

It will be seen that as the grid bias is increased so the peak to mean ratio of the distribution decreases. This is well shown by the dotted line curve, which is actually the curve for $E_g = 0$ on a reduced scale for comparison purposes.

Further experiments have shown that, for any specified value of grid bias, the shape of the distribution is independent of the anode to grid spacing provided that the cut-off voltage is maintained constant by change of anode voltage.

It will be recollected by reference to Fig. 1 that the cathode density distributions in Fig. 19A are those which obtain also in the plane of the cathode image $a'b'c'$. Indeed, this assumption is inherent in Fig. 19A, for the curves shown were obtained by scanning the emergent rays from the triode by means of a small hole. Furthermore, owing to the very large depth of focus of the cathode image, similarly shaped distributions are present over a wide range of distances from the triode. They are thus those which occur in the plane of the final focusing lens, where it is customary to place the beam stop. This aperture stop, which is usually part of the third anode, is bombarded by a very non-uniform current distribution, and for the high voltage guns may become overheated. The diaphragm should, therefore, be thick enough to provide adequate lateral conductivity.

2.2.6. The Peak Cathode Loading

We shall now proceed to establish that the peak

cathode loading must vary as the 3/2 power of the grid drive, and must be inversely proportional to the square of the grid hole diameter.

Consider any one triode of fixed geometry and suppose that the anode voltage is such that the cut-off voltage is E_c. For any applied grid drive E_d, the total cathode current will be given by (35) and the area of the cathode from which this current is drawn will be defined by (36). Eliminating E_c between (35) and (36) gives

$$I_a = \frac{12.y^2.E_d^{3/2}}{D^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (37)$$

This current I_a in (37) comes from the whole of the emitting area which has a diameter of $2y$. Thus the mean loading over this area will be

$$I_a/\pi.y^2 = \frac{12.E_d^{3/2}}{\pi.D^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (38)$$

On the assumption that the peak/mean ratio of loading is a constant and independent of the grid bias, it follows from (38) that the peak loading is given by

$$\rho_c = \frac{k_4.E_d^{3/2}}{D^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (39)$$

We should not expect (39) to hold very accurately since several approximations have been made in its derivation. Fig. 19A, for instance, shows that the transition from (38) and (39) involves an assumption which can only be approximately true. Furthermore, (36) does not hold very accurately, either. The relation between the grid bias and the emitting cathode diameter is a flat " S " shaped curve and is not truly linear. (Fig. 24.) In Fig. 20A we have an experimental plot showing the departures from equation (39). The ordinate scale is proportional to the currents falling into a small hole lying centrally across the cone of rays from the triode. These currents have been converged into peak cathode loadings from a knowledge of the size of the cathode image projected across the hole. It will be seen that the peak cathode loading is not simply a 3/2 power function of the grid drive as in (39), but it is possible to draw a line through the points which is a fair average and conforms with (39). Fig. 21A is similar, but refers to a 0·56 mm. diameter grid hole. Again an average straight line has been drawn through the points. From Figs. 20A and 21A we find that k_4 is approximately ·0145 for D in mm.

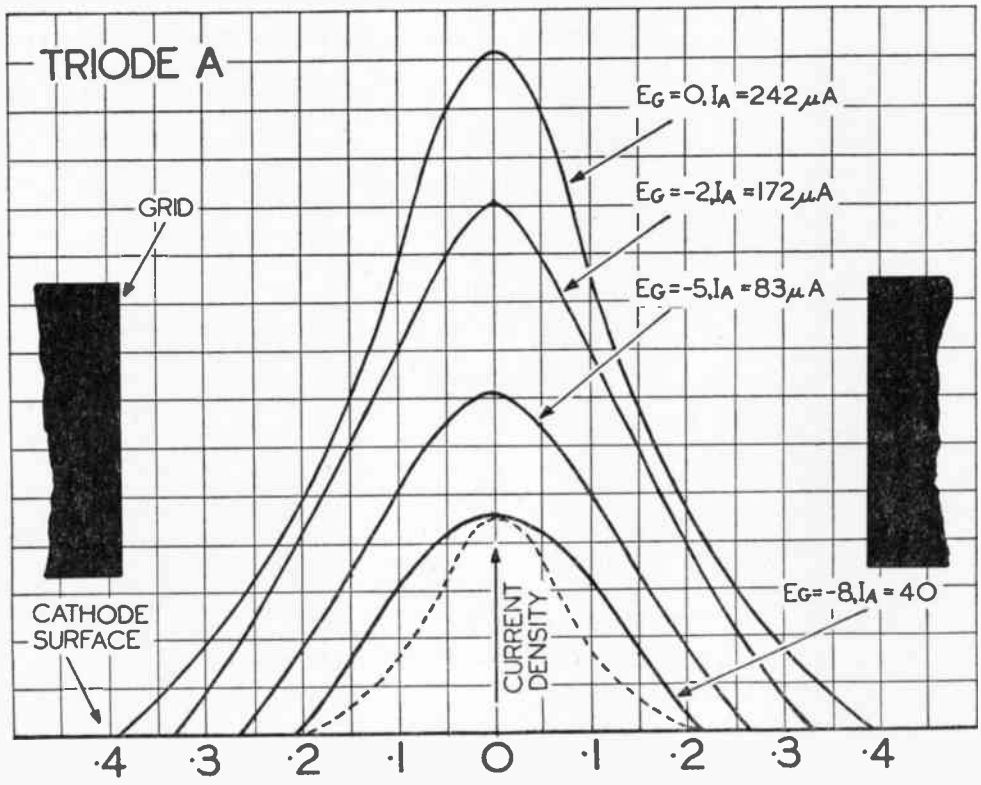


Fig. 19A.—Electron Current Distribution across Cathode Surface for Various Grid Voltages.

and ρ_c in ma./sq. mm., so we have the semi-empirical formula for the peak cathode loading

$$\rho_c = \frac{\cdot 0145 E_d^{3/2}}{D^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (39A)$$

Fig. 21A shows an additional feature of interest, namely, that the peak cathode loading curve shows no discontinuity through zero grid. This would imply that a gun with a (say) — 20 volt cut-off, operating at a grid voltage of + 10 volts, has the same peak loading as a tube with a cut-off of — 50 volts operating at $E_g = -20$, provided, of course, that the grid hole diameters are the same. The total current in the first case would be much higher due to that component which flows straight to the grid. This latter portion of the current would come mainly from that portion of the cathode under the rim of the grid hole, and would thus not contribute appreciably to the peak loading at the centre.

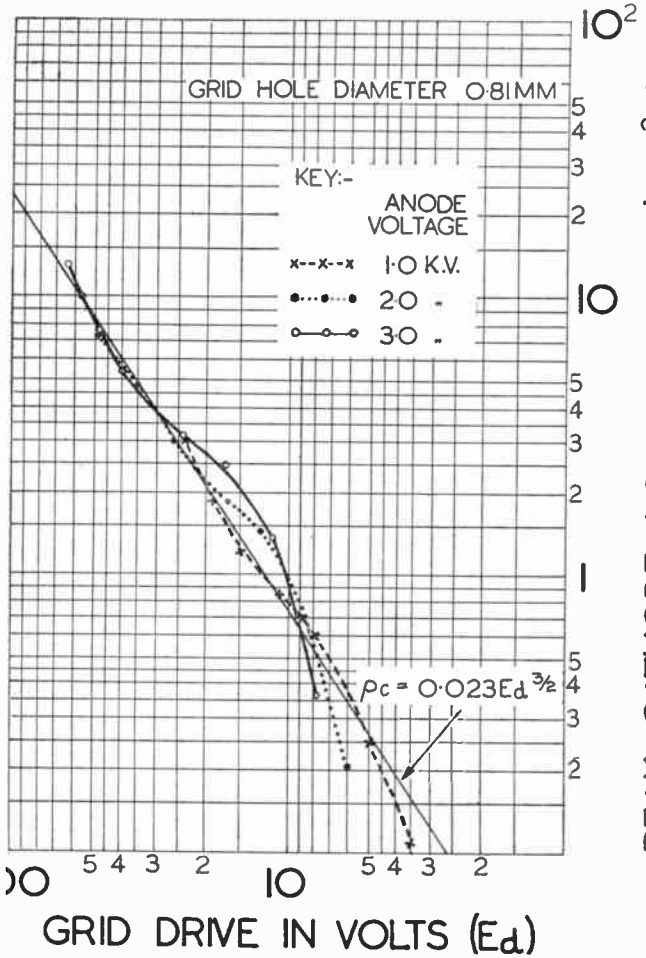


Fig. 20A.—Peak Cathode Loading/Grid Drive Curves for 0·81 mm. Grid Hole.

These statements must not be construed, however, as meaning that positive grid operation of a cathode ray tube is necessarily permissible. Damage may easily be caused, since frequently the grid is not designed to take current.

Finally, it is of interest to consider physically what happens when the grid is made less negative. The total cathode current increases because

(*a*) the emitting area of the cathode increases

and

(*b*) the emission density increases.

Effect (*a*) occurs according to the square of the grid drive as in (36). Effect (*b*) occurs according to the 3/2 power of the grid drive as in (39). The net increase in cathode current therefore occurs according to the 7/2 power of the grid drive as in (35). Again we must emphasise that these results are only approximations, but they do serve as a basis for useful conceptions and calculations.

2.2.7. The Angular Divergence of the Cone of Rays from the Triode

It has already been shown in section 2.2.4 that at zero grid voltage the emitting portion of the cathode is that area bounded by the projection of the grid hole on the cathode surface. This result is almost entirely independent of the triode geometry. However, the angle into which this cone of rays is projected depends markedly on certain geometrical parameters of the triode. The right hand photos of sheet P1 show, for example, that the beam angle varies with anode to grid spacing. This is equivalent to saying that the magnification between the cathode surface and its image is a function of the anode/grid spacing. In this section we investigate the general question of the dependence of beam angle on triode geometry.

Experiments made by the author with sliding screen tubes have shown that the electron cone from the triode is not exactly straight-sided, but tends to be slightly torpedo-shaped with the apex at the crossover. This fact has been ignored in the working which follows, and the semi-angle of the cone of rays, termed α, is considered equal to that of straight-sided cone, apex at the grid plane and with the base formed by the intersection of the divergent cone of rays on a plane 250 mm. from the grid.

Figs. 22A and 22B are experimentally obtained curves which illustrate how the beam angle varies with change of anode to grid spacing for various grid hole diameters. $E_g = 0$ throughout. Since the semi-angle of the cone is fairly small, and in view of subsequent approximations, we may write $\tan \alpha = \sin \alpha$.

From the curves on Fig. 22A it will be seen that an approximate analytical relationship between the anode to grid spacing f and the angle is

$$\sin^{4/3}\alpha = K_1/f \quad \ldots\ldots\ldots\ldots (40)\ (D \text{ constant})$$

where K_1 is a constant.

By drawing lines parallel to the axis of tan α it will be found that the intercepts of the curves for the differing grid hole sizes are approximately proportional to the grid hole diameters, so that

$$\sin\alpha = K_2.D \quad \ldots\ldots\ldots\ldots (41)\ (f \text{ constant})$$

By means of triodes embodying the thermal expanding device (Fig. 13) for continuous variation of the cathode to grid spacing, it was established that changes in this quantity do not greatly affect the beam angle at zero grid voltage. This can be seen by reference to photos on sheet P.2. There is a slight dependence of beam angle on cathode to grid spacing, the nature of which is indicated in curve 23B. In the development of the theory which follows, complete independence is assumed.

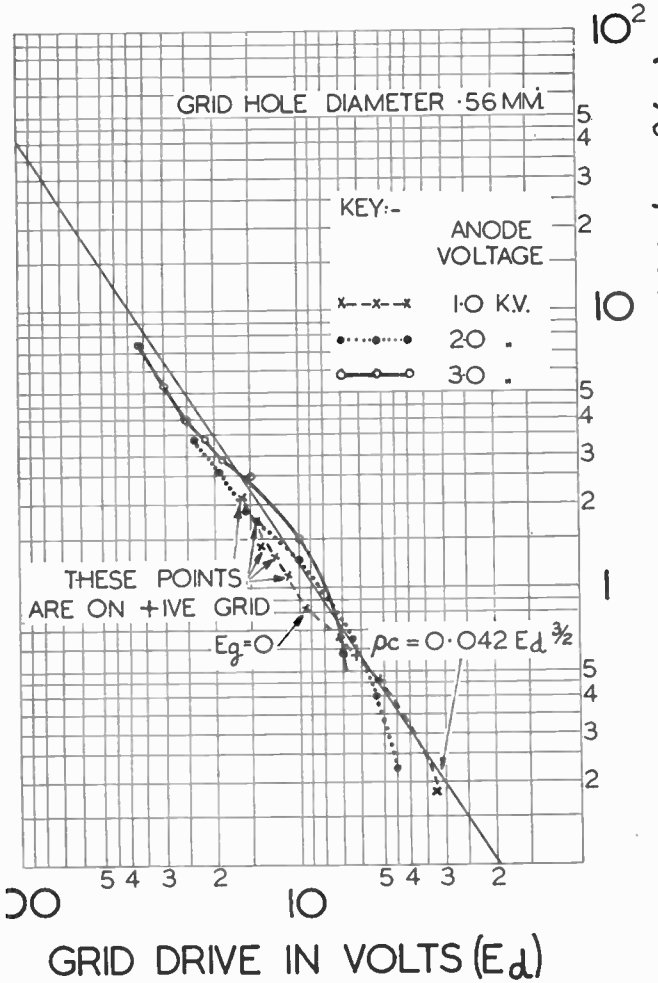


Fig. 21A.—Peak Cathode Loading/Grid Drive Curves for 0·56 Grid Hole.

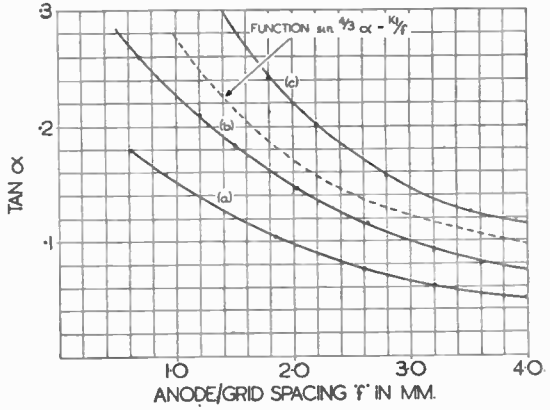


Fig. 22A.—Effect of Variation in Anode/Grid Spacing on Beam Angle from Triode A, at $E_g = 0$.

In Curve (a) D = 0·55 *mm.* b = 0·17 *mm.*
In Curve (b) D = 0·8 *mm.* b = 0·15 *mm.*
In Curve (c) D = 1·27 *mm.* b = 0·33 *mm.*

Figs. 24A and 24B show how the beam angle varies with grid bias for constant geometry and anode voltage. It will be seen that the curves there presented depart considerably from a linear relationship, a fact which is somewhat at variance with the work of L. Jacob.[7] However, to a first approximation and to make the theory reasonably simple, a linear law has been assumed, so that we may write

$$\sin\alpha = K_3.E_d/E_c \ldots (42)\ (\text{constant geometry})$$
$$|E_d| \leqslant |E_c|$$

In this formula K_3 is the value of sin α when the grid bias is zero. By the reasoning presented in this section and section 2.2.4, K_3 is a function of the geometry of the triode, but is independent of the anode voltage.

2.2.8. Effect of Grid Modulation on Crossover Size

We now reach a critical stage in the development of the theory. It will be shown that the nature of the dependence of crossover size on grid modulation is already determined if the truth and accuracy of the relationships so far established be accepted.

Consider a triode system of constant geometry operating at a fixed anode voltage E_{a1} and grid drive of E_d. Let I_a be the total cathode current and A_1 the crossover area. The term "crossover area" is defined more carefully presently. The semi-angle of the cone

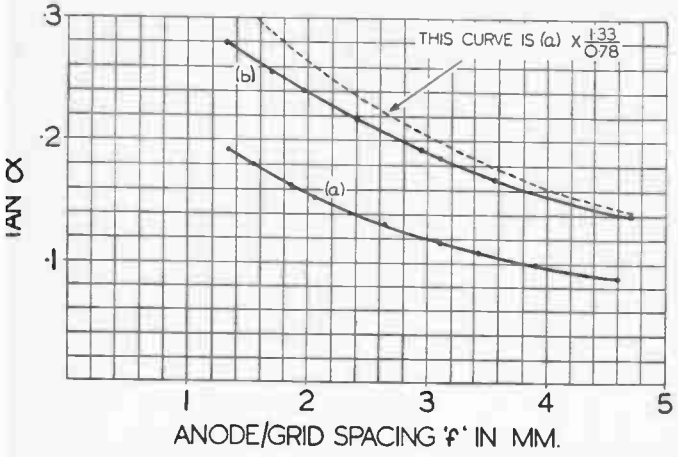


Fig. 22B.—*Effect of Variation in Anode/Grid Spacing on Beam Angle from Triode B, at* $E_g = 0$.

In Curve (*a*) D = 0·78 *mm.* b = 0·22 *mm.*
In Curve (*b*) D = 1·33 *mm.* b = 0·25 *mm.*

of rays from the triode operating under these conditions is α. Let B be the brightness of the crossover so formed, i.e., the current per unit area per unit solid angle. Hence, by definition

$$B = \frac{k_5.I_a}{A_1.\sin^2\alpha} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (43)$$

Now let the grid drive be altered to E_d'. This is the sole alteration. The only restriction imposed is $|E_d| \leqslant |E_c|$ throughout. The new total current becomes I_a', the new crossover area A_1' and the new beam angle is α'. The new crossover brightness of B' and by definition

$$B' = \frac{k_5.I_a'}{A_1'.\sin^2\alpha'} \quad \ldots\ldots\ldots\ldots\ldots\ldots (44)$$

Now consider the Langmuir formula (1). Since $E.e/k.T >> 1$, it may be expressed

$$\rho_0 = \rho_c.K.E.\sin^2\alpha \quad \ldots\ldots\ldots\ldots\ldots\ldots (45)$$

Furthermore, ρ_0 is proportional to $\frac{I_a}{A_1}$ so that (45) shows $\frac{I_a}{A_1.\sin^2\alpha}$ is proportional to $\rho_c.K.E.$ Now by (39), ρ_c is proportional to $E_d^{3/2}$, so that finally $\frac{I_a}{A_1.\sin^2\alpha}$ (i.e. the crossover brightness) is proportional to the 3/2 power of the grid drive for constant geometry and anode voltage.

Hence, if we set $E_d'/E_d = k$, it follows that

$$k^{3/2}.\frac{I_a}{A_1.\sin^2\alpha} = \frac{I_a'}{A_1'.\sin^2\alpha'} \quad \ldots\ldots\ldots\ldots (46)$$

But by equation (35)

$$I_a' = k^{7/2}.I_a$$

and from (42)

$$\sin^2\alpha' = k^2.\sin^2\alpha.$$

Putting these values in the right hand side of (46) then yields

$$k^{3/2}.\frac{I_a}{A_1.\sin^2\alpha} = \frac{k^{7/2}.I_a}{A_1'.k^2.\sin^2\alpha}$$

whence it follows that $A_1' = A_1$.

Hence the important conclusion that the crossover size must be independent of the grid drive if the truth of the preceding relationships is accepted. Any experimentally detected deviations from this simple result are presumably due to approximations implied in the formulæ (1), (35), (39) and (42). It has been carefully pointed out that these formulæ do involve considerable approximations, and this fact is apparent from examination of the experimental curves on which they are based. Hence it would not be surprising to find that in practice the crossover size proves to be somewhat dependent on the grid drive. In particular, and this point is stressed, the reasoning just given is based on the Langmuir equation (1), and this, in turn, is based on two important postulates :

(*a*) neglect of space charge, and

(*b*) assumption *that the triode is aberrationless.*

There is much evidence to show that the latter assumption is far from true, unless the portion of the rays from the triode which are under examination is severely restricted by a beam stop. Hence the independence of crossover size and grid drive in the idealised case might be experimentally verified by using a triode system in which marginal rays were severely limited by a diaphragm.

Before discussing experiments on these lines, it is necessary to consider briefly what is meant by the term " crossover size."

Two difficulties arise here. In section 1.1.1 it was shown that the current density distribution in the crossover has the form $\rho = \rho_0 \exp - K.r^2$ where r is

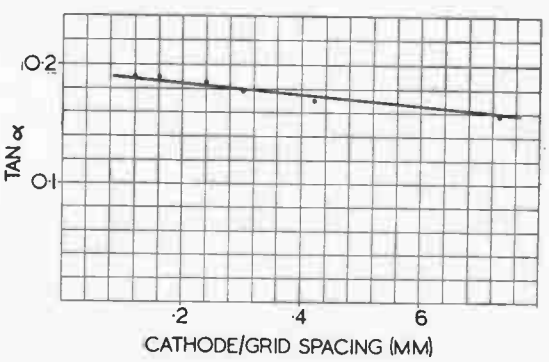


Fig. 23B.—*Effect of Variation of Cathode to Grid Spacing on Beam Angle.* D = 0·8 *mm.* f = 1·77 *mm.*

the distance from the beam axis. Thus ρ is finite for all finite values of r and the crossover has no definite size since it has no definite boundary. We shall here adopt the usual solution to this difficulty and shall define the crossover (or spot) size as that value of r which satisfies the relation

$$exp - K.r^2 = \rho/\rho_0 = 1/5 \quad \ldots\ldots\ldots\ldots\ldots\ldots (47)$$

Thus physically the edge of the crossover or spot is defined as that circle on which the beam density has fallen to one-fifth of its value on the beam axis.

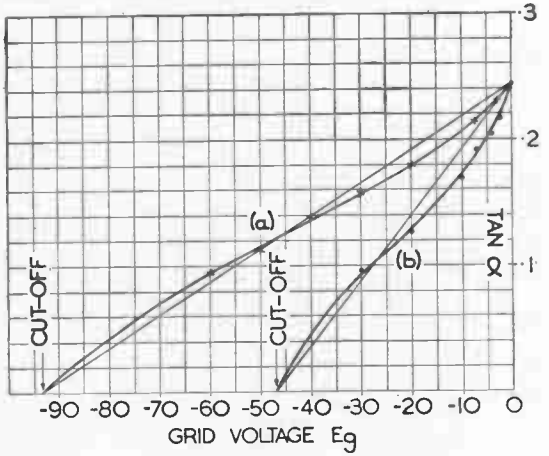


Fig. 24A.—Relationship between Grid Voltage and Beam Angle. Geometry Constant. Cut-off Voltage Varied by Alteration of First Anode Potential from 2 kV. in (a) to 1 kV. in (b).

Note.—D = 0·8 *mm.* b = 0·15 *mm.* f = 0·8 *mm.*

The other difficulty arises because it is difficult to see how it is practicable ever to measure directly the crossover distribution. All that can be done is to make measurements on its image. If it is now assumed that the lens producing this image is free from aberrations, satisfies the Abbé sine law, and if, in addition, the effects of space charge are neglected, then

$$\sqrt{E_1}.y_1.\sin\alpha = \sqrt{E_2}.y_2.\sin\theta \quad \ldots\ldots\ldots\ldots (2A)$$

In the equation the suffix 1 refers to the crossover and the suffix 2 to the image. In practice the value of E_2 is easily obtained by direct measurement, but the determination of E_1 is a difficult problem, since the crossover is formed in a region of high axial potential gradient and it is difficult to see how the position of the crossover could be determined with sufficient accuracy to define E_1. This, however, need not worry us unduly, because in practice it is not the actual crossover size which is of primary interest, but only the spot size when this crossover is focused by a lens of known magnification. It is clear from (2A) that this image size is completely defined for a given lens geometry as soon as $\sqrt{E_1}.y_1$ is known, and it is, therefore, of somewhat academic interest to argue about the individual values of E_1 and y_1. When discussing triode performance, however, it is convenient to be able to introduce the crossover size concept, and in all that follows this will be *defined* as that value of y_1 which satisfies (2A) when E_1 has the value equal to the potential on the first anode.

The effect of grid modulation on spot size for a triode of type A is shown in Fig. 25A. These curves were experimentally obtained by an elaboration of the "slit scanning" method first described by Jacob.[5] Briefly, the image of the crossover under investigation is made to traverse slowly a long slit, the direction of scan being at right angles to the slit.* This scanning

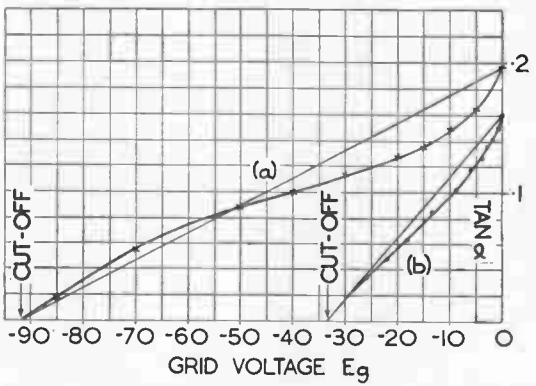


Fig. 24B.—Relationship between Grid Voltage and Beam Angle. First Anode Voltage 2 kV.

In Curve (a) D = 1·33. b = 0·25 *mm.*
In Curve (b) D = 0·78. b = 0·22 *mm.*

is done in exact synchronism with a beam moving along the "X" axis of an auxiliary oscillograph, the two scanning voltages being derived from a common source. Behind the slit is a Faraday cage, which picks up the electrons falling through the slit.

This current develops a proportional potential across a resistance, and this potential, after D.C. amplifica-

* It is not immediately obvious that the value of the distance of the centre of the slit from the centre of the electron beam, for any fraction of the current picked up when the slit lies diametrically across the beam, is equal to the value of r satisfying (47) for the same fraction ρ/ρ_0. This, however, is only one of the many peculiar properties of the solid Gaussian distribution. For, consider a pair of rectangular Cartesian axes drawn in the plane normal to the beam direction, and such that the origin coincides with the beam centre. Suppose the slit lies parallel to the Y axis and distant x_1 from it. Then at any point on the slit, distant y from the X axis, the current density will be $\rho = \rho_0 . \exp - K.r^2$, *i.e.* $\rho = \rho_0 \exp. - K (x_1^2 + y^2)$, since $r^2 = x_1^2 + y^2$. This may be written $\rho = \rho_0 . \exp - K.x_1^2 . \exp - K.y^2$ so that for all values of y on the slit displaced x_1 from the beam centre, the density is $\exp - K.x_1^2$ of the value which would obtain if the slit were lying diametrically across the beam. By integrating for all values of y from $+\infty$ to $-\infty$, we find that the ratio of current in the slit when displaced x_1 from the beam centre, to current in slit when lying diametrically across the beam is also $\exp - K.x_1^2$. But this is precisely the ratio of current density at radius x_1 to current density ρ_0 at the centre, which proves the proposition.

tion, is applied to the "Y" axis of the auxiliary oscillograph. Fig. 25A shows the type of curve thus traced. They are of Gaussian form. The author has shown by a rigorous mathematical analysis that this is a necessary condition for the actual radial distribution to be Gaussian.

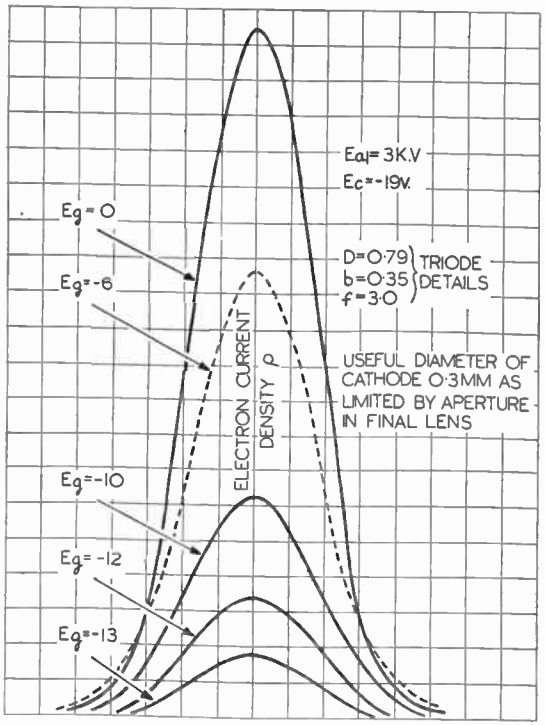


Fig. 25A.—Effect of Variation of Grid Drive on Spot Size.

In the curves of Fig. 25A the only independent variable was the grid potential. The curves show clearly that the size of the image of the crossover, defined by (47), is independent of the grid voltage, i.e., the curves are all members of the family $\rho = \rho_0 \, exp. - K.r^2$, where K is a constant for the family and ρ_0 is the parameter. So far, then, this experiment supports the theoretical conclusions.

This result applies to the spot, i.e., to the image of the crossover. It is necessary to be perfectly clear as to the deduction which can be drawn about the conditions at the crossover itself. This requires some general understanding of the possible aberrations in the triode, a very difficult question which will now be briefly discussed.

In the idealised case shown in Fig. 1, the crossover diameter is independent of the emitting area of the cathode face. This is a necessary (but not a sufficient) condition satisfied by an aberration less triode. In practice it is a condition which is not satisfied except in the region where the emitting area is small (i.e., near cut-off). Therefore electrons from the edges of the cathode at *a* and *c* will not have extreme trajectories passing through A and B, but rather through points more widely separated. This reveals itself as an increase in the crossover size (and in general a departure from the Gaussian density distribution) as the emitting area is increased. If the lens focusing the crossover accepts the whole of the beam from the latter, the spot, i.e., the crossover image, will also change its size in precisely the same manner. Frequently, however, the final focusing lens contains a severe beam stop, so that only the centre of the electron cone from the triode contributes current to the spot. The outer portions of the beam are trapped on the diaphragm, and it is just these electrons which come from the edges of the cathode, as shown in Fig. 1, and increase the crossover diameter. Therefore the effect of the diaphragm in the final lens may be sufficient to prevent the actual spot size from increasing even though the crossover size increases. Hence the important conclusion—if the final lens is aberration free and space charge is neglected the electron density distributions in crossover and spot are always similar (in general on a different scale of size) over and only over that region where the whole of the current in the crossover passes to the spot. As a corollary, we note that in the special case where the crossover diameter is independent of the cathode emitting area, the electron density distributions in crossover and spot are similar regardless of how much current is intercepted by the final lens. A sufficient (although not a necessary) condition for the latter is that the triode be aberration free.

In Fig. 25A the diaphragm in the final lens was such as to limit sin α to 0.0345. This corresponds to an emitting cathode diameter of almost exactly 0.3 mm. Only electrons emitted from within this circle could pass to the image. The cut-off voltage was − 19 and the grid hole diameter 0.79 mm. Equation (42) then shows that the final anode hole was "filled" at about − 12 volts on the grid. It is therefore clear that it is not permissible to deduce from the curves of Fig. 25A that the actual crossover size was constant for values of grid bias between − 12 and zero.

We can, nevertheless, draw certain conclusions about the actual crossover aberrations from Fig. 25A. It is obvious that for any given triode geometry the shape of the equipotential lines in the grid/cathode region will vary with grid voltage and this would be expected to have some effect on the aberrations. Therefore, although the final anode hole was filled at − 12 volts, and further increases in grid drive served merely to produce electrons from a wider circle on the cathode—these electrons all being trapped on the diaphragm—the change in the shape of the equipotentials due to the increase in grid drive could effect the spot size by affecting the aberrations in the useful cathode region. Fig. 25A shows that this has not occurred to any appre-

ciable extent, but there is some evidence of a decrease in spot size between −6 and zero grid volts, presumably due to this cause. This effect has been confirmed on several other occasions.

To sum up. It was shown at the beginning of this section that if the truth of the relationships so far treated be accepted than the crossover size must be independent of grid bias. But one of these relations, Langmuir's equation, is itself based on the assumption, inter alia, that aberrations are negligible. Hence we conclude that independence of crossover size and grid drive is possible only in an aberrationless system, and then only if the approximations implied in equations (35), (39) and (42) are ignored.

Experiments made on *narrow beam angle systems* confirm that independence of apparent crossover size and grid bias is substantially realised. No conclusions can be drawn about the real crossover/grid drive relationship in the practical case where aberrations are present.

However, in many types of cathode ray tube, the beam is heavily trimmed and in effect the triode aberrations are thereby so reduced that the approximation of independence of *apparent* crossover size and grid bias is permissible. It is, indeed, necessary for the sake of internal consistency if the foregoing laws are accepted.

2.2.9. Effect of First Anode Voltage Change on Crossover Size

Suppose firstly that the triode is aberration free. It will be here assumed as usual that space charge effects can be ignored. It will be postulated that the first anode voltage is the only independent variable. To simplify the writing of the equations and where no confusion can arise, the first anode voltage, previously denoted by E_{a1}, will be written simply as E. Then

$$\rho_0 = .\rho_c\left(\frac{E.e}{k.T}+1\right)\sin^2\alpha \quad \ldots\ldots\ldots\ldots(1A)$$

In practice $E.e/k.T >> 1$ so that we may rewrite (1A).

$$\rho_0 = \rho_c.K.E.\sin^2\alpha \quad \ldots\ldots\ldots\ldots\ldots\ldots(48)$$

Now suppose that A_1 is the area of the crossover, where $A_1 = \pi.y_1^2$ and y_1 is defined by (47). Then we may write

$$\rho_0 = \lambda.I_a/A_1 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(49)$$

where ρ_0 is the peak crossover density and I_a, as always, is the total current in the crossover. λ is a constant to allow for the peak/mean ratio of the Gaussian crossover distribution. Eliminating ρ_0 between (48) and (49) yields

$$A_1 = \lambda.I_a/\rho_c.K.E.\sin^2\alpha \quad \ldots\ldots\ldots\ldots\ldots(50)$$

It will now be postulated that the grid voltage is maintained at zero. Hence the grid drive E_d will be proportional to the anode voltage E. From (39) this implies that the cathode leading ρ_c will be proportional to $E^{3/2}$. But from (32) I_a is also proportional to $E^{3/2}$. Therefore I_a/ρ_c in (50) is constant. Also, sin α is constant since $E_g = O$ (section 2.2.7). Hence we may express (50) so far as changes in E only are concerned as

$$A_1 = K_5/E$$

or

$$y_1 = K_5/\sqrt{E} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(51)$$

Equation (51) has been deduced for $E_g = O$ throughout, but it clearly applies generally for any value of grid voltage, since we saw in the last section that ideally the crossover size is independent of E_g. Hence the important result that for constant geometry the crossover diameter is inversely proportional to the square root of the first anode voltage. For a constant gun geometry, the spot size at the screen will follow a similar law.

Fig. 26A shows the results of an experimental investigation of this law. A constant gun geometry was maintained and the anode voltages were varied over the range 900 to 3,000 volts.

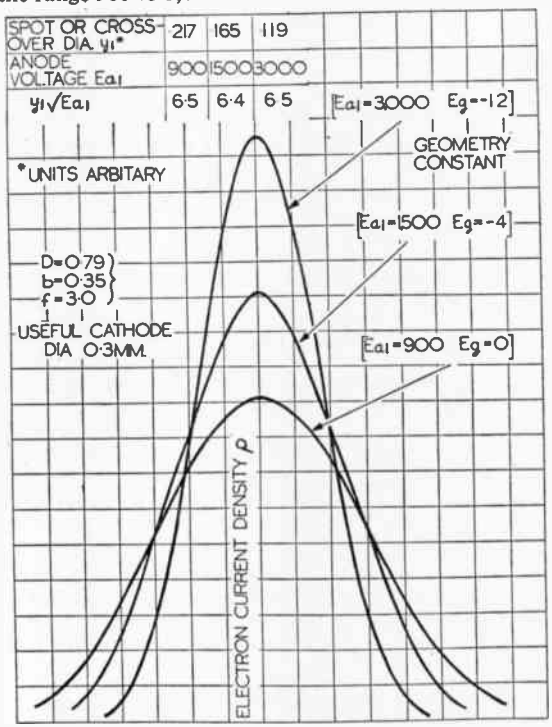


Fig. 26A.—Effect of First Anode Voltage on Crossover Size.

The aperture in the final lens was such as to limit the half beam angle α from the triode to arc sin 0.0345. As usual the curves given are those of the image of the crossover. It is permissible to deduce that the

effective crossover size follows the same law as the spot for the aperture stated. The grid voltage was varied so as to maintain a constant beam current of 10 μA. The table on Fig. 26 indicates the constancy of $y_1\sqrt{E}$ as required by (51).

2.2.10. Effect of Changes in Anode to Grid Spacing on Crossover Size

Let us first investigate this problem on the assumption that the triode is aberration free. The space charge will also be ignored, and it will be assumed during the proof that the grid voltage is zero. The anode to grid spacing f is the only independent variable. A dependent variable is sin α, which is related to f as shown in Figs. 22A and 22B. To a first approximation the analytic relation is given by (40)—

$$\sin^{4/3}\alpha = K_1/f \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (40)$$

On these premises, in equation (50) ρ_c is always proportional to I_a as in section 2.2.9. Hence (50) may be expressed

$$A_1 = K_8/E.\sin^2\alpha \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (52)$$

Using (40) in (52) immediately gives

$$A_1 = K_7.f^{3/2}/E \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (53)$$

so that for constant anode voltage E the crossover area is proportional to 3/2 power of the anode to grid spacing. As an extension, using (29) in (53) shows immediately that when f is the only independent variable

$$A_1 = K_8/E_c^{3/2}.E \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (54)$$

Note, incidentally, that the general form of (52) is immediately obvious from (4).

It must be very clearly understood that (53) and (54) are based on an assumption that the triode is aberration free, and also on the accuracy of (40) which is empirically derived from experiments described in section 2.2.7. (See Figs. 22A and 22B.) Their experimental application has been investigated by means of a special tube embodying a device for continuous variation of the anode to grid spacing. Details are shown on Fig. 27A. Examination of them shows that the effective crossover size does appear to follow (53) and (54) for the range investigated. So far as the actual crossover size is concerned it is permissible to deduce only that the results are not inconsistent with what would be observed if the triode were aberration free.

2.2.11. Effect of Changes in Grid Hole Diameter on Effective Crossover Size

Again we commence by assuming that the triode is free from aberration and that space charge is negligible. The grid hole diameter is the only independent variable. Equation (50) is the starting point so that

$$A_1 = \lambda.I_a/\rho_c.K.E.\sin^2\alpha \quad \ldots\ldots\ldots\ldots\ldots\ldots (50)$$

Suppose $E_g = 0$ throughout. The cut-off voltage E_c is a dependent variable, being related to the grid hole diameter D by means of (29). From (31) I_a is proportional to $E_c^{3/2}$ and from (39) ρ_c is proportional to $E_c^{3/2}/D^2$. Hence I_a/ρ_c in (50) is proportional to D^2, so we may write (50) in the form

$$A_1 = K_9.D^2/\sin^2\alpha \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (55)$$

since E is constant. Using (41) in (55) then immediately shows that the crossover size is independent of the grid hole diameter, *all* other variables being constant.

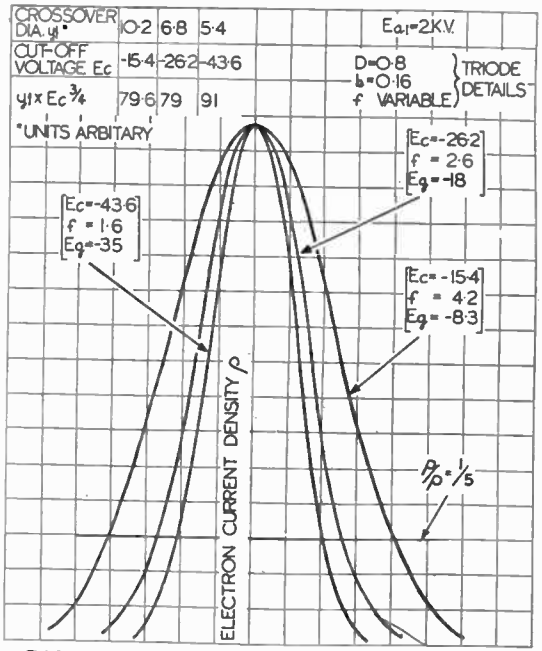


Fig. 27A.—Effect of Variation in Anode/Grid Spacing on Spot Size. Peak Cathode Loading Constant.

An experimental check on this result is shown in Fig. 28A. The limiting aperture in the final anode restricted the useful cathode diameter to 0.15 mm., and it is not, therefore, permissible to deduce from this experiment more than that the result is consistent with what would be observed if the triode were aberration free.

This result has since been confirmed by constructing two tubes with grid hole diameters of 0·55 mm. and 1.25 mm. respectively. All other factors were kept constant. In both cases the spot size was found to be substantially independent of the grid voltage over the majority of the modulation range. The spot sizes did not differ at all so far as could be measured. A 10 per cent. difference would have been detected.

2.2.12. Effect of Change in Cathode to Grid Spacing on Effective Crossover Size

As in the previous section, the starting point is

equation (50). Again we ignore space charge and postulate that the grid voltage is zero. The cathode to grid spacing is the only independent variable. Again, therefore, I_s/ρ_c is constant so that

$$A_1 = K_s/E.\sin^2\alpha \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (52)$$

as before. In so far as it is justifiable to regard sin α as independent of the cathode to grid spacing (52) shows that the crossover size is also independent of this factor.

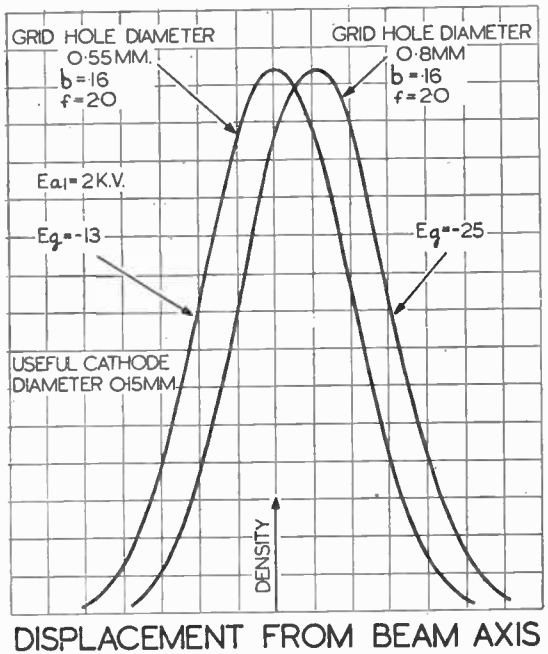


Fig. 28A.—Effect of Change of Grid Hole Diameter on Spot Size. Beam Current Maintained Constant.

In practice sin α increases somewhat as the cathode to grid spacing is reduced, as is indicated in Fig. 23B, so that from (52) it follows that the crossover area must correspondingly fall. This fall is increased owing to the square law dependence of A_1 on sin α indicated in (52).

Fig. 29A confirms these variations of crossover size. For the changes in cathode to grid spacing normally traversed the change in crossover size is not very marked, and it is justifiable to assume that the latter is independent of the cathode to grid spacing in this simplified theory.

2.3. The Complete Electron Gun

This comprises the triode followed by some type of focusing system and, usually, an aperture stop. The focusing may be either electrostatic or magnetic, and speaking generally, from an electron-optical point of view, the choice is of little moment. This is because it would seem there is little difficulty in constructing either type of lens to give negligible spherical aberration. with the apertures permitted by the deflecting system, for the focal lengths in common use.

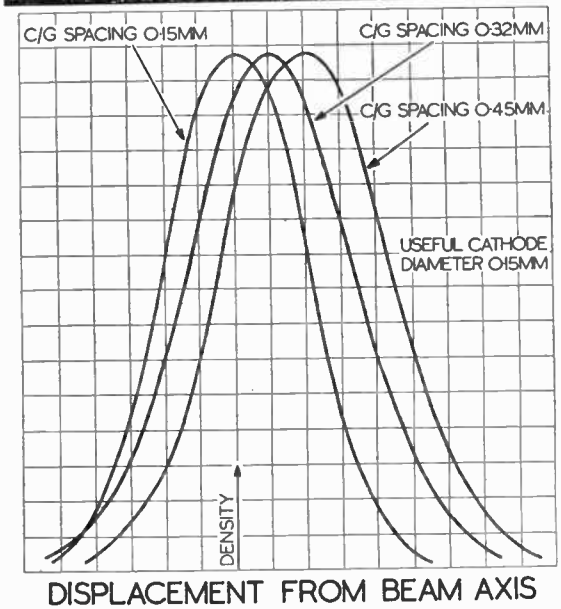


Fig. 29A.—Effect of Variation of Cathode to Grid Spacing on Spot Size (Peak Cathode Loading Constant). Cathode to Grid Spacing continuously Variable by Thermal Expander, as shown in Fig. 13.

However complex the focusing system may be, it can be considered, if aberrationless, as a device which reproduces the crossover conditions at the screen on a different scale of size. It has already been discussed in section 2.2.8 how this statement must be modified if the triode producing the crossover has aberrations *and* if the focusing lens contains a beam stop.

Confusion can easily arise regarding the effect of changes in the refractive indices of image and object spaces. Suppose the potential at the crossover is E_1 and the potential at the screen is E_2. Consider the special case when $E_1 = E_2$. Following the reasoning in section 2.2.8 we may suppose that this condition is obtained merely by operating the first anode potential at the same value as the screen potential. In these conditions a certain magnification, termed M, exists between the crossover and spot.

Applying equation (2A) shows that when the voltage distribution is changed so that the crossover and spot potentials are no longer equal, then the effective magnification is $M.\sqrt{\frac{E_1}{E_2}}$. However, this result is true only in so far as the voltage changes can be effected without alteration of the value of M itself. It is not, for example, permissible to deduce generally that the

spot size, for constant first anode voltage, would vary as $\frac{1}{\sqrt{E_2}}$. In fact, such a relationship is often found in practice, but it arises because in many types of cathode ray tube variations of the final anode and screen potential E_2 in the range considered are *not* accompanied by marked changes in the value of M.

For further information on these questions, the reader is referred to an interesting paper by Dr. G. Liebmann (ref. 8) and especially to the discussion by the author to appear in *Proc. I.R.E.* probably in late spring, 1946.

It is not proposed to enter into detailed discussions in this paper regarding the extent to which the accuracy of the Langmuir formula (1) applies quantitatively in practice. It is generally found that the measured electron current density is about 1/10th, and never appears to exceed about 1/5th of that which the formula predicts. The reasons for this hiatus are presumably the presence of aberration and/or space charge. It would appear that these effects are arising in the triode section and not in the final lens. The performance of the latter is much easier to investigate theoretically, especially in regard to space charge (refs. 3 and 9). The situation is, however, mysterious in that the general *form* of the Langmuir equation is strongly supported by the observed variations in triode performance with geometry and operating conditions. It fits, for example, the observed dependence of spot size with anode voltage as discussed in section 2.2.9 and other sections of the present work as treated. It would seem, therefore, that the effect of the aberrations is to introduce a multiplying factor in front of the whole equation, which would leave its form unchanged and yet would bring the observed and computed spot densities more into line. Further work on these matters is being pursued, and it is hoped that it will form the subject of a later paper.

2.3.1. Some General Design Notes

Fig. 30 shows a three anode type of electron gun which is typical of modern practice. The first and third anodes are maintained at fixed potentials while the second anode potential is varied to provide the appropriate lens strength. This fact explains the inherent advantages of the three anode construction as compared with the two anode type. For many reasons such as variation of deflection sensitivity, spot characteristics, etc., it is not admissible to vary the final anode potential for focusing. Therefore in the two anode gun we should be obliged to vary the first anode potential for this purpose, and this procedure would, therefore, also vary the triode cut-off voltage and hence the whole triode characteristics. This cross coupling between focus and brilliance is very undesirable, so that the three anode construction is much to be preferred.

In magnetically focused tubes this does not arise, so that a single anode is quite adequate.

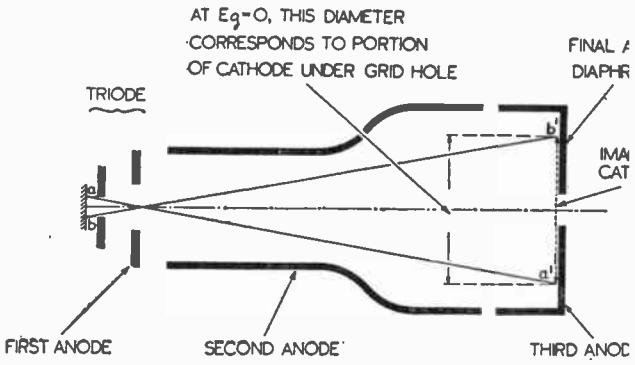


Fig. 30

The aperture stop is best placed in the final anode as shown. As a general principle aperture stops must not be placed in any electrode which is at a lower potential with reference to the fluorescent screen than *all* the subsequent electrodes. This is on account of secondary electron emission from the edges of the stops. Except in the case defined above, such secondaries are liable to reach the screen in a diffuse unfocused beam, thereby reducing contrast. Thus the only alternative position for the beam stops in Fig. 30 would be the first anode, provided that then $E_{a_1} > E_{a_2}$. This position is perfectly satisfactory from an electron optical standpoint, but undesirable mechanically, since the beam dimensions are small in this region and the positional tolerance on the stop is therefore very close. Occasionally beam stops have to be used on the first anode, for the purpose of preventing the marginal electrons from the triode from hitting the second focusing anode and so producing secondaries. In general a preferable solution of this problem is to employ a second anode of larger diameter, or to flare out the end as in Fig. 30. Generally this has the further advantage of reducing the aberrations of the final focusing lens for a given hole diameter.

2.3.2. Useful Cathode Area—the Geometric Loading

Fig. 30 shows diagrammatically a typical three anode electrostatically focused electron gun. As compared with Fig. 1 it will be seen that it has two "final" lenses, i.e., that between the first and second anodes, and that between the second and third anodes. These can be replaced by an equivalent thin lens, located between the two. It should be noted that as the first lens between the first and second anodes is very near to the crossover its effect is small, so that the equivalent thin lens is principally defined by the A2/A3 system.

It was seen in section 2.2.4 that the magnification of the cathode image is almost entirely independent of the grid or first anode voltage. In Fig. 30 a small portion of this image lies across the third anode hole. The extent of this portion depends on the geometrical characteristics of the triode, as dealt with in section

2.2.7, and it also depends on the geometry of the remainder of the gun and on the voltage ratios between the first and second anodes, and between the second and third anodes. Frequently the first and third anodes are internally tied together, so that the voltage ratios E_{a1}/E_{a2} and E_{a2}/E_{a3} are fixed for any cathode ray tube. In this case, then, the area of the cathode surface of which the image falls over the third anode hole is completely fixed and independent of the applied voltages, provided that the tube is focused. All the electrons starting from this portion of the cathode pass into the useful beam, and, conversely, all the electrons from the remainder of the emitting surface strike the third anode diaphragm. This former portion of the cathode will be termed the " useful cathode area."*

The " geometric loading "* will be defined as the reciprocal of the useful cathode area expressed in square millimetres. These expressions give a clear mental picture of the demand imposed on the cathode by the geometry of the gun as distinct from that imposed by the applied voltages. They are also useful because in many types of electron gun the product of beam current and geometric loading is closely equal to the peak cathode loading. This statement is justified because frequently the electron current density in the third anode hole is fairly uniform, since the latter subtends only a small portion of the cathode surface.

The geometric loading is best determined experimentally for any tube by measuring the width of the electron cone in the final anode plane, at zero grid voltage and with the tube focused. We have seen in section 2.2.4 that this image corresponds to a cathode area equal to the grid hole area, so that it is unnecessary to scribe the cathode surface artificially.

* These expressions are not in general use.

Table 2 shows typical data regarding the useful cathode area for a variety of wartime cathode ray tubes. The figures relate to tubes made by the author's company.

The most striking feature of the table is the smallness of useful cathode areas, which demonstrates how vital is the fine structure of the emitting cathode surface

2.3.3. The Beam Current/Grid Voltage Relationship

In an all-magnetic tube there is frequently no final anode diaphragm or other form of aperture stop, so that the whole of the cathode current emerging from the grid hole passes to the screen. The current reaching the screen is defined as the beam current. In this case, therefore, the beam and cathode currents are the same, so that we may use (35) directly to estimate the beam current for a given grid drive. In the electrostatic type of tube, however, a considerable proportion of the total cathode current will usually be wasted by impact on the third anode diaphragm. This proportion will increase towards zero grid voltage and may there reach 80 per cent. or more of the total. A little consideration of this type of tube will show that the beam current/grid voltage curve will have two distinct regions. Firstly, in the region of cut-off, the diameter of the cone of rays from the triode in the plane of the stopping diaphragm will be insufficient to fill the hole, and here the beam and cathode currents will be the same, so that (35) holds and the beam current rises as the 7/2 power of the grid drive. When the hole is filled further increases in beam current must be due only to increases in emission density, since the useful cathode area is constant. In this region, therefore, the beam current will rise as the 3/2 power of the grid drive. This is shown in the curves of Figs. 20A and

TABLE 2

Tube type	Grid hole diameter (mm.)	E_{a3}/E_{a2} for focus $E_{a1} = E_{a3}$	Diameter of image of cathode in final anode plane at $E_g = O$ and E_{a3} set for focus (mm.)	Final anode hole diameter (mm.)	Diameter of cathode imaged over final anode hole (mm.)	Area of cathode imaged over the final anode hole (mm.²)
ACR. 1	0·8	5·26	8	3·0	0·3	0·071
VCR. 87	0·8	4·9	9·5	3·0	0·25	0·05
VCR. 85	0·55	5·0	13	2·3	0·097	0·0074
VCR. 511	0·55	5·7	10	2·3	0·126	0·0126
VCR. 131 / ACR. 12	0·8	4·75	10·5	3·0	0·228	0·041
ACR. 13 / VCR. 97	0·8	5·9	11	1·8	0·131	0·0135
09	0·55	3·34	12	2·3	0·105	0·0088
26	1·25	4·17	10	3·0	0·375	0·11
VCR. 139	0·8	6·15	3	1·25	0·333	0·087
VCRX. 22	0·55	6·15	3·2	0·8	0·137	0·0148

21A which illustrate the case where the aperture is very small.

With most types of electrostatic tube the aperture is small and is filled only a few volts from cut-off. The great majority of the beam current/grid voltage curve therefore follows a 3/2 power law. As zero grid is approached, it is common to find that the curve flattens out somewhat, due to onset of temperature limited conditions in the cathode material. This is a danger sign so far as steady grid potentials are concerned and tubes should be operated so that the mean beam current corresponds to a value lying well inside the 3/2 power part of the characteristic.

TABLE 3.

Effect of multiplying 1st anode voltage by k.
(Final anode voltage constant.)

Parameter under examination (Condition : Constant grid drive)	No beam stop in third anode	Heavy beam stop in third anode
Total cathode current ..	$\times \frac{1}{k^{3/2}}$	$\times \frac{1}{k^{3/2}}$
Beam current	$\times \frac{1}{k^{3/2}}$	$\times 1$*
Beam angle from triode (α)*	$\times \frac{1}{k}$	$\times \frac{1}{k}$
Emitting diameter of cathode surface	$\times \frac{1}{k}$	$\times \frac{1}{k}$
Spot size*	$\times 1$	$\times 1$

2.3.4. Distribution of applied Potentials in Multi-anode Guns

A certain amount of confusion is sometimes present regarding the effect of changing the distribution of applied potentials in a multi-anode gun. A common case is illustrated by reference to the three-anode gun of Fig. 30. Suppose that the final (3rd) anode potential is held constant, while the first anode potential is varied. What are the principal changes effected ?

The analysis is difficult unless it is postulated that aberrations and the space charge are negligible throughout. This assumption will be made. It is supposed that the first anode voltage E_{a_1} is changed to $k.E_{a_1}$. The grid *drive* is maintained constant. As usual, we impose the restriction that $|E_d| < |E_c|$ throughout, so that the grid is never driven positive. Table 3 summarises the principal effects produced. The first line follows immediately from equation (35), since the cut-off voltage is multiplied by k. The effect on the actual beam current, as distinct from the total cathode current, will depend on the degree to which the beam is trimmed in the final anode and the table shows two columns which represent each of the two possible extreme cases : firstly, no beam stop at all, and secondly, a beam stop which masks the great majority of the cone of rays from the triode. Clearly, when no beam stop is present the beam current follows the same law as the cathode current, as indicated in the first column.

When a severe beam stop is present in the third anode, the variations in beam current with first anode voltage for constant grid drive will depend upon the influence of the A_1/A_2 lens on the angular divergence of the cone of rays from the triode passing up the second anode tube. Provided $1/4 < \frac{E_{a_1}}{E_{a_2}} < 4$, which permits E_{a_1} to vary over quite a wide range, it is usually permissible to assume that the effect of the A_1/A_2 lens is negligible.* The reasoning presented in sections 2.2.5 and 2.2.6, illustrated in Figs. 20A and 21A, then shows that the beam current in the presence of a severe stop is unchanged.

The third and fourth lines in Table 3 follow from equations (36) and (42). The constancy of spot size indicated in line 5 again requires the postulate of negligible action by the A_1/A_2 lens, but is then otherwise obvious.

In actual practice there are some advantages to be gained by operating the first anode at the highest voltage applied to the tube. The triode clearances for a given cut-off voltage can thereby be increased so that the mechanical assembly is easier and, in addition, it is likely that the effects of space charge in the cross-over region are reduced. It is not suggested, however, that the overall advantages are very great.

2.3.5. The Similitude Principles

Finally, no account of any electron gun theory would be complete without a reminder on the two similitude principles. These have been treated in detail in ref. p2. The first has been employed in section 2.2.7, and may be stated :

(*a*) In any electron optical system of constant geometry into which the electrons are injected with zero initial velocity, the multiplication of all electrode potentials by the same factor does not alter the electron trajectory. The postulate here is that the space charge is negligible.

The second principle may be stated :

(*b*) In any electron optical system to which constant potentials are applied, multiplication of all electrode dimensions by a constant factor multiplies the scale of the electron trajectory by this factor without altering its shape. In reference p2 this was referred to as the principle of geometrical similitude, and it was shown in the

* Postulates negligible action by A_1/A_2 lens.

* Reference 8 and discussion previously referred to in section 2.3

appendix that it applies even in the presence of space charge, provided the total current in the ray remains constant. This latter condition is automatically satisfied by the scaling operation so that the original and transformed systems are geometrically similar.

This second principle (*b*) is often of great value in checking the consistency of theory* or in providing additional information from known data. This was illustrated in equation 28 of section 2.2.1.

It has been noted that the similitude principle (*a*) holds generally only in the absence of space charge. However, examination of the peculiar form of the space charge equation 22 (ref. p1) shows that provided the beam current varies as $E^{3/2}$, the electron trajectory is unaltered. It has been shown in section 2.2.2 of this paper that if the grid voltage of the cathode ray tube is held at zero the current will vary in this manner. From this fact we conclude that deductions can be made about the behaviour of the gun at high voltages from experiments performed at low values of anode potential provided the observations are not being made on quantities which depend on the magnitude of the effect of initial emission velocity. For example, it would not be permissible to make observations on spot size, which does depend on the initial emission velocity.

2.3.6. Conclusions and Acknowledgments

It must be especially stressed that some of the statements made in the foregoing account must on no account be read out of their context. This is particularly true of sections 2.2.8 and 2.2.11. The results given here apply to the actual crossover only if the aberrations and space charge are assumed absent. The results have a practical value because in many cases the beam divergence is so limited that the effective crossover size does substantially follow the laws stated. It is admittedly unfortunate that the theory cannot be more closely related to practice, but any attempt to achieve this seems likely to lead to such involved conclusions that no clear mental conceptions can be formed.

Further work is clearly needed on the aberrations in the triode. It is particularly of interest to establish how it comes about that the Langmuir formula, based on the assumption of no triode aberration, appears of the form necessary to explain the general observed variations of the effective crossover size, while at the same time it yields values of electron current density many times higher than it would appear possible to obtain with existing guns.

It is a pleasure to record my appreciation of the long and stimulating contact which I have had with the cathode ray tube research department at the G.E.C., Wembley, on many aspects of electron gun design. Specific acknowledgments have already been made in the text, but I am indebted to them in addition for many interesting discussions from which I am very conscious of profit. The author is not entirely clear as to how far he differs from the G.E.C. as regards his theories on these matters. The official G.E.C. accounts have not yet been published* and the acknowledgment is not intended to commit the G.E.C. to any of the views here expressed. I am also much indebted to members of my own staff, especially to Mr. L. A. Woodbridge and Mr. D. C. Phipps, who assisted with the experimental work, and to Miss Webb for her help with the manuscript. Finally, I wish to thank A. C. Cossor, Ltd., for the facilities provided and for permission to publish this work.

Bibliography

6. "Electron-optical System of Two Cylinders as applied to Cathode Ray Tubes." D. W. Epstein, *Proc. I.R.E.*, Vol. 24, p. 1095, August, 1936.

7. "A New Type of Electron-optical Voltmeter." L. Jacob, *J.I.E.E.*, Vol. 91, Part 2, No. 24, December, 1944.

8. "Image Formation in Cathode Ray Tubes and the Relation of Fluorescent Spot Size and Final Anode Voltage." G. Liebmann, *Proc. I.R.E.*, June, 1945.

9. "A Space Charge Problem." Hilary Moss, *Wireless Engineer*, July, 1945.

Note

p1. "Electron Gun of the Cathode Ray Tube, Part I." Hilary Moss, *Journal Brit. I.R.E.*, January, 1945.

p2. "Engineering Methods in the Design of the Cathode Ray Tube." Hilary Moss, *Journal Brit. I.R.E.*, December, 1945.

Books

"Electron-optics in Television." Maloff and Epstein. McGraw-Hill, 1938.

* In order to forestall being "hoist by my own petard," the following note is necessary. In our simplified theory it has been assumed that the beam angle from the triode at zero grid is independent of either grid material thickness or cathode to grid spacing. If equations (40) and (41) are strictly accurate, then the similitude principle (*b*) shows that these assumptions cannot be more than approximations. For otherwise (40) and (41) reveal that sin α varies as $k^{1/4}$, where k is the linear size of the triode. To eliminate this dependence on size it is merely necessary to assign to sin α a dependence on grid material thickness t and on cathode to grid spacing b, such that sin α varies as $k^{1/4}$. If this dependence is equally divided, then sin α is proportional to $1/t^{1/8}$ and sin α is proportional to $1/b^{1/8}$. Therefore sin α changes only very slowly with either grid material thickness t or cathode to grid spacing b, this justifying the assumption of complete independence to simplify the theory. More accurately it would seem that

$$\sin \alpha = K.D/f^{3/4}.b^{\lambda}.t^{\gamma}$$

where $\lambda + \gamma = 1/4$. Experiment has indicated this type of relationship and also the fact that $\lambda > \gamma$ inside the range of values of b and t normally of interest.

* G.E.C. investigations will be given under the following titles and authors. The places of publication are tentative only.

"The Effect of Cathode Ray Tube Triode Parameters on Emission Density." J. Sharpe, *Wireless Engineer.*

"Cathode Ray Tube Gun Design." L. C. Jesty and J. Sharpe, *J.I.E.E.*

"A Study of the Immersion Objective of a Cathode Ray Tube Gun." L. Jacob, *J.I.E.E.*

"Variation of Beam Angle with Modulation, in Cathode Ray Tube Guns." L. Jacob, *Philosophical Magazine.*

" Television." Zworykin and Morton. Wiley, 1940. (Especially chapter 13 on " The Electron Gun." Compare their equation (13.8) with our (11) and their (13.13) with our (40) and (41).

APPENDIX 1

The Constant Brightness Theorem and Related Topics

The results of sections 2.2.10, 2.2.11 and 2.2.12 are special cases of what will be termed the " constant brightness theorem." This can be formally stated :

" For a constant first anode voltage and a constant cathode loading, the crossover brightness is wholly independent of the triode geometry." (Postulates : lens aberrations and space charge neglected.)

To prove this theorem it is merely necessary to transpose equation (1A) and set $\rho_0 = \gamma.I_a/A_1$, so that

$$\frac{\gamma.I_a}{A_1.\sin^2\alpha} = \rho_c\left(\frac{E.e}{k.T} + 1\right)$$

For α small, $\sin^2\alpha$ is proportional to ω the solid angle of the cone from the crossover, so that the left hand side is proportional to the crossover brightness, i.e. current/unit area/unit solid angle. The right hand side does not explicitly involve the triode geometry, which establishes the proposition.

For the purpose of gaining confidence in the internal consistency of some of the foregoing theory, it is helpful to consider again the effects of changing the three fundamental parameters of the triode geometry.

The working will be made more general by writing equation (29) in the form

$E_c = \frac{k.D^n}{t.b.f}$, and equation (40) in the form

$\sin\alpha = \frac{K_1}{f^m}$, so that the values of n and m are left open.

Consider first the case in which the anode to grid spacing is changed. *This is the only independent variable.* Fig. X is the relevant diagram. It shows the linear $\sin\alpha/E_g$ law implied in equation (42). Let symbols without dashes refer to the initial conditions, those with dashes to the final ones. Initially the grid bias is OM, so that U is the working point. The anode to grid spacing is f, so that P is the cut-off point and OP the cut-off voltage E_c.

Now suppose that the anode to grid spacing is changed to f', where $f' = f/k$ and $k > 1$. In consequence the new cut-off voltage rises to OQ. For the constant cathode loading postulated in the constant brightness theorem, equation (39) shows that the grid drive E_d must be kept constant. This is a sufficient condition since the anode to grid spacing does not appear in (39). Thus the new grid bias must be adjusted to be equal to ON, where the point N is defined by QN = PM = E_d. T is the new operating point.

Equation (29) shows that the new cut-off voltage E_c' is equal to $k.E_c$, whence equation (35) shows that the new total crossover current is I_a', where

$$I_a' = I_a/k^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (a)$$

We now proceed to investigate the change in beam angle. From Fig. X, $\tan\psi = OR/OP$ and $\tan\phi = OS/OQ$. Dividing gives $\tan\psi/\tan\phi = \frac{OR.OQ}{OP.OS}$. From equation (29) $OQ/OP = k$ and also from (40), $OS/OR = k^m$. Substituting these values gives $\tan\psi/\tan\phi = k^{1-m}$. Hence $UM/TN = k^{1-m}$, so that

$$\sin\alpha' = k^{m-1}.\sin\alpha \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (b)$$

Finally, equation (53) shows that the new crossover area is A_1' where

$$A_1' = \frac{A_1}{k^{2m}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (c)$$

The original brightness is proportional to $I_a/A_1.\sin^2\alpha$. The final brightness is proportional to $I_a'/A_1'.\sin^2\alpha'$,

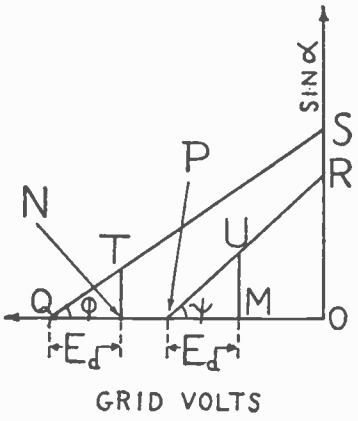


Fig. X

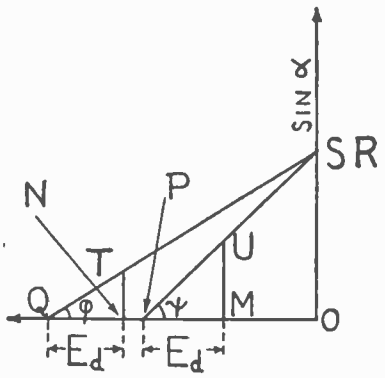
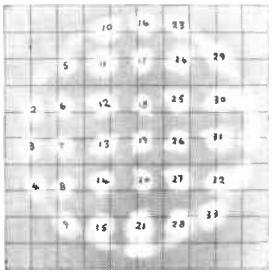
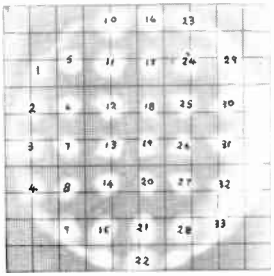


Fig. Y



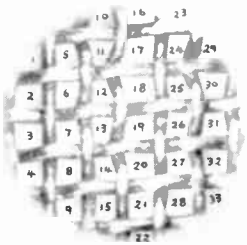
p.6 A/G Spacing 1.3 mm

Eg = -6

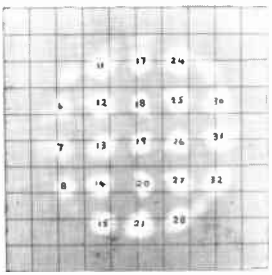


p.2 A/G Spacing 1.3 mm

Eg = 0

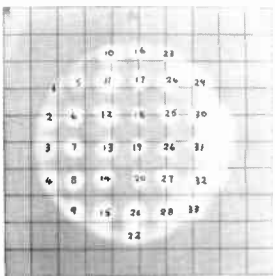


p.1 Optical Image as seen through grid hole.



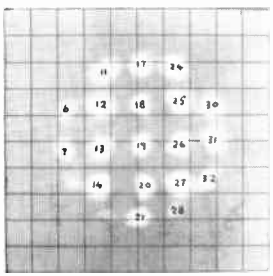
p.7 A/G Spacing 1.3 mm

Eg = -16



p.3 A/G Spacing 2.0 mm

Eg = 0



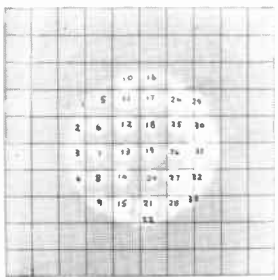
p.8 A/G Spacing 1.3 mm

Eg = -30

SHEET No. P.1.

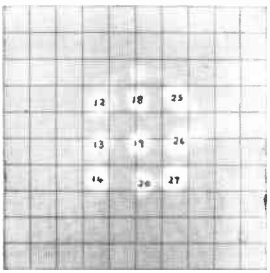
V_A = 4,000

GRID HOLE 0·8 mm



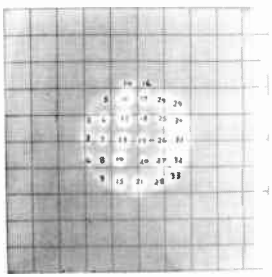
p.4 A/G Spacing 2.6 mm

Eg = 0



p.9 A/G Spacing 1.3 mm

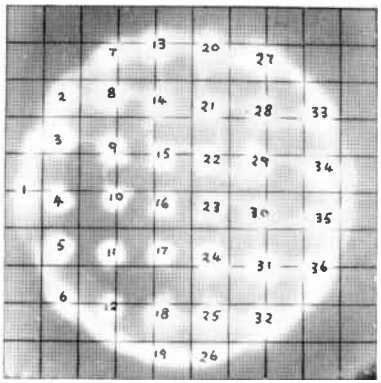
Eg = -50



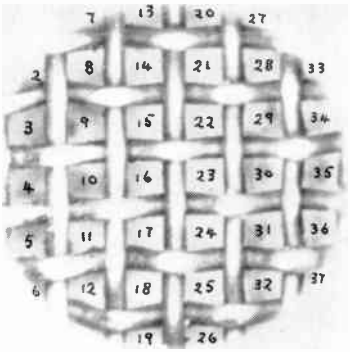
p.5 A/G Spacing 3.3 mm

Eg = 0

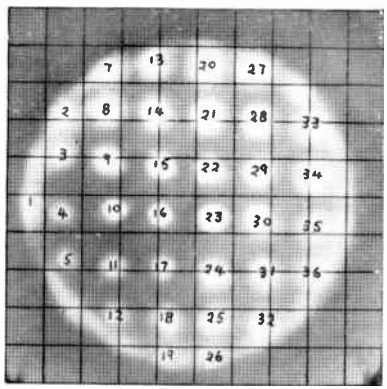
[*To face page* 120]



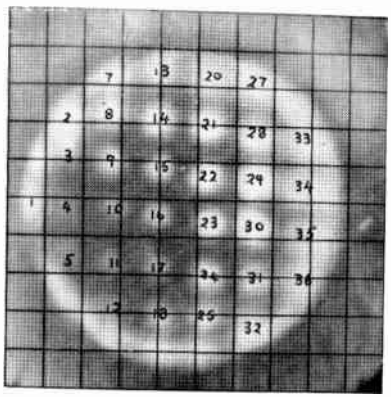
p.18 A/G Spacing 1.53 mm.
C/G Spacing 0.12 mm.
Eg = 0



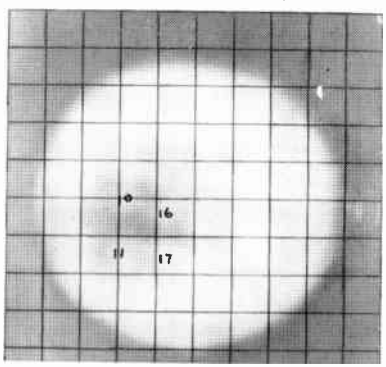
p.17 Optical Image as seen through grid hole.



p.19 A/G Spacing 1.53 mm.
C/G Spacing 0.24 mm.
Eg = 0



p.20 A/G Spacing 1.53 mm.
C/G Spacing 0.30 mm.
Eg = 0



p.21 A/G Spacing 1.53 mm.
C/G Spacing 0.45 mm.
Eg = 0

SHEET No. P.2.

EFFECT OF VARYING
C/G SPACING.

Va = 4,000

GRID HOLE 0.81 mm

i.e. using (*a*), (*b*) and (*c*) is unchanged. The experimental findings are therefore consistent with the constant brightness theorem, in the case where the anode to grid spacing is the variable. Note that the self consistency of the working is not dependent on the value of m which is used.

Next, consider the case in which the cathode to grid spacing is the only independent variable. The same notation and reasoning is used. Fig. Y is the relevant diagram. From section 2.2.7 the angular divergence of the rays from the triode at $E_g = 0$ is substantially independent of the cathode to grid spacing, so that the points R and S coincide. Clearly

$$I_a' = I_a/k^2 \quad \ldots\ldots (a)$$

as previously. But as R and S coincide, UM/TN = OQ/OP = k, so that

$$\sin \alpha' = k^{-1}.\sin \alpha \quad \ldots\ldots (d)$$

Finally, the crossover size (on our simplified theory) is independent of the cathode to grid spacing, so that

$$A_1' = A_1 \quad \ldots\ldots (e)$$

Obviously (*a*), (*d*) and (*e*)result in the constant brightness theorem in the case where the cathode to grid spacing is the variable.

Finally, consider the case where the grid hole diameter is the only independent variable. Refer to Fig. Z.

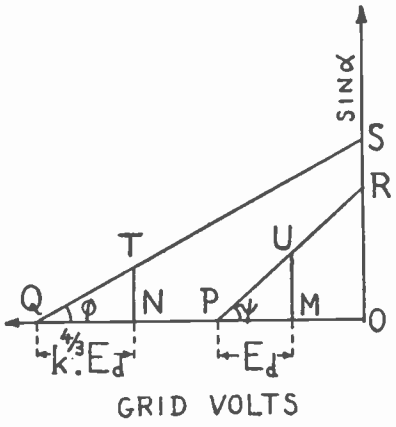


Fig. Z.

The cut-off point is changed from P to Q by alteration of the grid hole diameter from D to D′, where D′ = k.D and $k > 1$. As before, the working point changes from U to T. This time, however, the new grid drive at T will not be equal to that at U, since a constant cathode loading is postulated and equation (39) shows that this depends on the grid hole diameter which has been changed. For a constant cathode loading ρ_c we have

$$\rho_c = \frac{k_4.E_d^{3/2}}{D^2} = \frac{k_4.E_d'^{3/2}}{k^2.D^2}$$

so that $E_d' = k^{4/3}.E_d$.

By application of (29) in more general form, the new cut-off voltage E_c' will be $k^n.E_c$, whence (35) shows that the new total crossover current will be I_a' where

$$I_a' = \frac{I_a.k^{14/3}}{k^{2n}} = \frac{I_a}{k^{2n-14/3}} \quad \ldots\ldots (f)$$

Now consider the change in beam angle. From Fig. Z, $\tan \psi = \frac{UM}{E_d}$ and $\tan \phi = \frac{TN}{k^{4/3}.E_d}$, so that

$$\frac{TN}{UM} = \frac{k^{4/3}.\tan \phi}{\tan \psi}.$$

As previously $\frac{\tan \phi}{\tan \psi} = \frac{OS.OP}{OQ.OR}$. From section 2.2.7, equation (41) OS/OR = k, so that

$$\frac{\tan \phi}{\tan \psi} = \frac{1}{k^{n-1}}.$$

Substituting gives

$$\frac{TN}{UM} = k^{\frac{7}{3}-n}$$

so that

$$\sin \alpha' = k^{\frac{7}{3}-n}.\sin \alpha \quad \ldots\ldots (g)$$

Finally, the crossover size is independent of the grid hole diameter (ideally) so that

$$A_1' = A_1 \quad \ldots\ldots (e)$$

Substituting the values of I_a', etc., from (*f*), (*g*) and (*e*) in $\frac{I_a'}{A_1'.\sin^2 \alpha'}$ we again find that the crossover brightness is unaltered by the change in grid hole diameter.

APPENDIX 2

In the three tables below are summarised the results of changes in the three fundamental geometrical parameters of the triode. To make the working more general, we leave the values of the indices in equs. (29) and (40) open by writing them as m and n, as in Appendix 1. The results of substituting the suggested "best fit values," i.e. $n = 3$ and $m = 3/4$ are inserted in brackets. It should be noted that most of these results are approximate only and accuracy on wide extrapolations must not be expected.

APPENDIX 2(a)

OPERATION		CONSEQUENCES					
Independent variable and operation	*Sole* other variable and extent of adjustment to achieve new condition specified	Cut-off voltage	Crossover diameter	Beam angle (α)	Total current in beam	Current/unit solid angle	Cathode loading
Cathode/grid spacing multiplied by k	Grid material thickness, grid hole diameter and anode/grid spacing all multiplied by k. (Geometrical similitude)	$\times 1$	$\times k$	$\times 1$	$\times 1$	$\times 1$	$\times \frac{1}{k^2}$
	Grid drive multiplied by $\frac{1}{k}$ to maintain constant beam angle	$\times \frac{1}{k}$	$\times 1$	$\times 1$ (postulate)	$\times \frac{1}{k^{3/2}}$	$\times \frac{1}{k^{3/2}}$	$\times \frac{1}{k^{3/2}}$
	Anode/grid spacing multiplied by $\frac{1}{k}$ to maintain constant cut-off voltage	$\times 1$ (postulate)	$\times \frac{1}{k^m}$ $\left(\times \frac{1}{k^{3/4}}\right)$	$\times k^m$ $\left(\times k^{3/4}\right)$	$\times 1$	$\times \frac{1}{k^{2m}}$ $\left(\times \frac{1}{k^{3/2}}\right)$	$\times 1$
	Grid hole diameter multiplied by $\sqrt[n]{k}$ to maintain constant cut-off voltage	$\times 1$ (postulate)	$\times 1$	$\times k^{1/n}$ $(\times k^{1/3})$	$\times 1$	$\times \frac{1}{k^{2/n}}$ $\left(\times \frac{1}{k^{2/3}}\right)$	$\times \frac{1}{k^{2/n}}$ $\left(\times \frac{1}{k^{2/3}}\right)$

APPENDIX 2(*b*)

OPERATION		CONSEQUENCES					
Independent variable and operation	*Sole* other variable and extent of adjustment to achieve new condition specified	Cut-off voltage	Crossover diameter	Beam angle (α)	Total current in beam	Current/unit solid angle	Cathode loading
Grid hole diameter multiplied by k	Anode/grid spacing, cathode/grid spacing and grid material thickness all multiplied by k to maintain constant cut-off voltage. (Geometrical similitude)	$\times 1$ (postulate)	$\times k$	$\times 1$	$\times 1$	$\times 1$	$\times \frac{1}{k^2}$
	Grid drive multiplied by k^{n-1} to maintain constant beam angle	$\times k^n$ ($\times k^3$)	$\times 1$	$\times 1$ (postulate)	$\times k^{\frac{3n-7}{2}}$ ($\times k$)	$\times k^{\frac{3n-7}{2}}$ ($\times k$)	$\times k^{\frac{3n-7}{2}}$ ($\times k$)
	Cathode/grid spacing multiplied by k^n to maintain constant cut-off voltage	$\times 1$ (postulate)	$\times 1$	$\times k$	$\times 1$	$\times \frac{1}{k^2}$	$\times \frac{1}{k^2}$
	Anode/grid spacing multiplied by k^n to maintain constant cut-off voltage	$\times 1$ (postulate)	$\times k^{m.n}$ ($\times k^{9/4}$)	$\times k^{1-m.n}$ $\left(\times \frac{1}{k^{5/4}}\right)$	$\times 1$	$\times \frac{1}{k^{2(1-m.n)}}$ ($\times k^{5/2}$)	$\times \frac{1}{k^2}$

APPENDIX 2 (c)

OPERATION		CONSEQUENCES					
Independent variable and operation	Sole other variable and extent of adjustment to achieve new condition specified	Cut-off voltage	Crossover diameter	Beam angle (α)	Total current in beam	Current/unit solid angle	Cathode loading
Anode/grid spacing multiplied by k	Grid hole diameter, cathode/grid spacing, grid material thickness all multiplied by k. (Geometrical similitude)	$\times 1$	$\times k$	$\times 1$	$\times 1$	$\times 1$	$\times \frac{1}{k^2}$
	Cathode/grid spacing, multiplied by $1/k$ to maintain constant cut-off voltage	$\times 1$ (postulate)	$\times k^m$ ($\times k^{3/4}$)	$\times \frac{1}{k^m}$ $\left(\times \frac{1}{k^{3/4}}\right)$	$\times 1$	$\times k^{2m}$ ($\times k^{3/2}$)	$\times 1$
	Grid hole diameter multiplied by $k^{1/n}$ to maintain constant cut-off voltage	$\times 1$ (postulate)	$\times k^m$ ($\times k^{3/4}$)	$\times k^{\frac{1}{n}-m}$ $\left(\times \frac{1}{k^{5/12}}\right)$	$\times 1$	$\times \frac{1}{k^{\frac{2}{n}-2/m}}$ ($\times k^{5/6}$)	$\left(\times \frac{1}{k^{2/n}}\right)$ $\times \frac{1}{k^{2/3}}$
	Grid drive multiplied by k^{m-1} to maintain constant beam angle	$\times \frac{1}{k}$	$\times k^m$ ($\times k^{3/4}$)	$\times 1$ (postulate)	$\times k^{\frac{7m-3}{2}}$ ($\times k^{9/8}$)	$\times k^{\frac{7m-3}{2}}$ ($\times k^{9/8}$)	$\times k^{\frac{3(m-1)}{2}}$ $\left(\times \frac{1}{k^{3/8}}\right)$

APPENDIX 3.

Illustrative Examples

In this section we illustrate briefly a few typical applications of the theories advanced.

Problem No. 1

The gun of a television type cathode ray tube consists of a triode only, the focusing and deflection both being magnetic. No beam limiting stop is incorporated. When tested it is found that the central focus quality is satisfactory up to full working brilliance, but that the deflection defocusing is then excessive. Measurement shows that full brilliance requires 300 μA in the beam, and that the limit of tolerable deflection defocusing is reached when the beam current is only 200 μA. . How may the triode geometry be altered to meet the requirements indicated for satisfactory operation?

The problem is to reduce the beam angle from the triode, so that the beam width when the beam current is 300 μA is the same as that obtaining in the prototype when the beam current is only 200 μA.

Appendix 2 (*a*) shows that multiplication of the cathode/grid clearance by k, and the grid drive by $1/k$, multiplies the current/unit solid angle from the triode by $1/k^{3/2}$. The current into the maximum solid angle permissible on account of deflection defocusing has to be increased from 200 to 300 μA, so that $200/k^{3/2} = 300$, whence $k = 0{\cdot}76$. Hence a solution is effected by multiplication of the cathode grid clearance by 0·76, and the grid drive by $1/0{\cdot}76 = 1{\cdot}31$. The cut-off voltage is also multiplied by the latter figure.

Appendix 2 (*a*) shows that the cathode loading is then multiplied by 1·5, as is otherwise apparent since the current/unit area/unit solid angle has been multiplied by 1·5.

Problem No. 2

Suppose that, for reasons of mechanical design, it is not practicable to use the solution to problem 1, and to reduce the cathode/grid clearance. Are any alternative triode changes possible?

Looking through Appendix 2 (*b*) we notice that an increase in grid hole diameter does not affect the spot size, and that multiplication of the grid hole diameter by k, and the grid drive by k^2 also multiplies the current/unit solid angle by k. The current has to be increased from 200 to 300 μA into the same solid angle, whence $k = 3/2$. The grid drive, therefore, has to be multiplied by approximately 9/4 and the cathode loading is multiplied by 3/2.

It will be noted that this solution is electrically less satisfactory than the previous one, since the grid drive has to be more than doubled, against a previous increase of only about 30 per cent. Note that the cut-off voltage has been more than trebled.

Problem No. 3

An oscillograph type cathode ray tube using a narrow beam is found when tested at a fixed grid bias and anode voltage to yield a spot size of approximately 1 mm. diameter with a beam current of 30 μA. It is felt that the performance compromise of this tube is wrong in that it has inadequate resolution, but unnecessarily high beam current. A spot size of 0·75 mm. is required. What is the simplest change in gun geometry to effect this alteration?

From Appendix 2(*c*) it is seen that multiplication of the anode/grid spacing by k multiplies the crossover diameter by k^m, and the current/unit solid angle by k^{2m}, provided that the cathode to grid spacing is also multiplied by $1/k$ to maintain constant cut-off voltage. The grid drive is assumed constant. Taking the value of m as 3/4 as suggested in the table, gives $k^{3/4} \times 1 = 0{\cdot}75$ whence $k = 0{\cdot}68$. Hence the anode to grid spacing is multiplied by 0·68, and the cathode to grid spacing by $1/0{\cdot}68 = 1{\cdot}47$. The beam current is then multiplied by $0{\cdot}68^{3/2} = 0{\cdot}56$, and so becomes about 17 μA. The peak density in the spot and the cathode loading are both unchanged.

LONDON SECTION DISCUSSION

Dr. G. Liebmann: I should like to congratulate Dr. Moss on the excellent manner in which he has presented a very complex subject. It is very difficult to make a theory at all, but I think he has succeeded very well.

With regard to Dr. Moss's equations (27), (28) and (29), a theoretical analysis of the triode problem, based on electrostatic field theory, shows that the cut-off voltage E_0 should, in first approximation, be an inversely linear function of grid-cathode spacing b and grid-anode spacing f, as Dr. Moss has found by experiment. The same analysis indicates that the relationship between D and t cannot be described very simply; but equation (29) appears to be a good approximation to the theoretical function over the limited range of dimensions referred to in the paper.

The main point on which I differ from Dr. Moss, and probably from most of his listeners, is that I do not agree with the cross-over theory, which plays a considerable part in his paper. I put forward my views last year in a Proc.I.R.E. paper, and I merely want to show you the experimental results which I obtained when investigating the spot size of several types of television cathode ray tubes as a function of final anode voltage.

In this experiment the spot size for constant beam current (10 ua.) is practically independent of the accelerating voltage for quite different tube designs. (Tube No. 1 is a tetrode, Nos. 2 and 3 are pentodes, No. 4 is a triode and No. 5 a hexode). The spot size decreases with higher anode voltage (only) for a triode tube; this decrease can be fully explained by the

reduction in effective cathode area, because the grid bias has to be increased in this type of tube to keep the beam current constant.

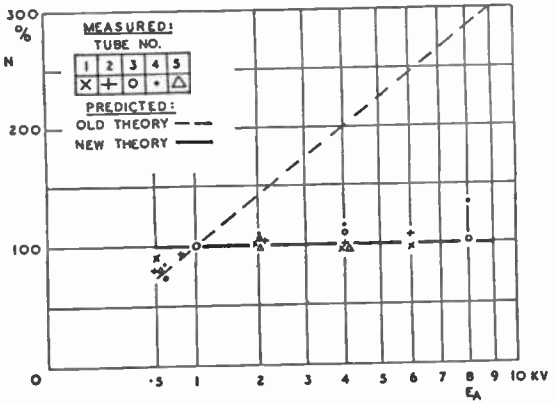


Fig. 1.

Even here, the change in spot size is considerably smaller than is predicted by the cross-over theory in its classical form, as proposed by Zworykin and Morton. This result differs from that shown in Fig. 26A of Dr. Moss's paper; but Fig. 26A is the *only* experimental result reported by Dr. Moss which cannot be equally well explained by the "cathode image" theory. In this theory the spot on the tube screen is supposed to be an image of a minified image of the cathode, situated between grid and first anode. No mystery attaches to the Langmuir condition in this theory, as the Langmuir condition appears only as an *upper limit* set to the current density in the focused spot, whereas the *actual* current density in the spot is given by the cathode current density, the magnification and the amount of superimposed lens aberrations, and current lost in beam stops.

Dr. Moss considers that, in first approximation, the spot size should not depend on the grid modulation; but this seems to apply only to tubes in which the diameter of the beam is severely limited by a small beam stop. The spot size depends quite considerably on the modulation in tubes without stops—for instance, in television cathode ray tubes.

Referring to the fact that the beam angle does not appear to change much with a change in grid-cathode spacing, as shown in Fig. 23B of the paper, I should like to give another presentation of the same problem which indicates that the grid-cathode spacing is an important design parameter so far as the beam angle is concerned.

In this experimental graph the beam angle is plotted as a function of the beam current for four different grid-cathode spacings. We see that for a given current the beam angle becomes smaller as the clearance decreases. In high current guns a small beam angle is of great importance for several reasons, and a small grid-cathode spacing is therefore essential in this type of tube. For example, we designed a television tube in pre-war days which employed a grid-cathode spacing of 0·3 mm.

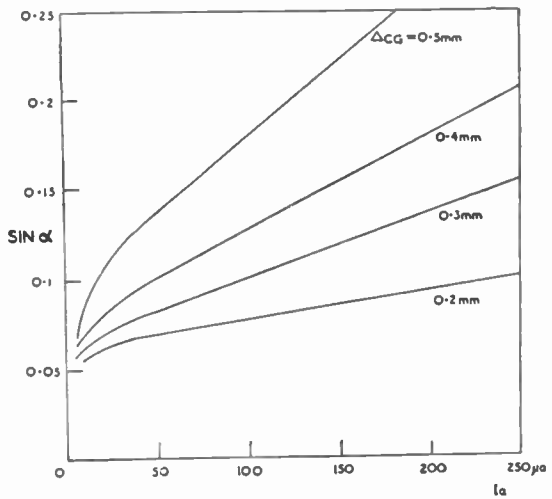


Fig. 2.

By certain reasoning, partly based on our measurements shown in slide 2, we concluded that we could get better performance if we reduced the beam angle by reducing the grid-cathode spacing. This was borne out by our results, as we could maintain in the re-designed tube a good focus up to beam currents of 400 to 500 ua., whereas before the defocusing point lay near 200 ua., a figure usual for this type of tube.

In Fig. 19A in the paper a radial cathode current distribution is shown which is more peaked at high beam currents than at low currents. We have found that the distribution curve is little influenced by the grid modulation, and is of the type shown by the lowest curve in Fig. 19A. This is not in agreement with the findings of Dr. Moss, and I should like to ask if the curves of Fig. 19A have been confirmed.

Mr. E. E. Shelton: Dr. Moss has indicated that there is some mystery about the relationship between $I_a|_0$ and the cut-off voltage, but I don't think there is anything very peculiar about it. Some years ago I. Langmuir proposed a relation for the triode which was an empirical relation, and which holds for fairly large changes of geometry. The Langmuir equation is

$$I_a = k(E_a + \mu E_g)^{3/2}.$$

The usual curve showing the relation between I_a and E_g for one particular value of E_a is of the form shown in Fig. 3.

The Langmuir equation re-written for point B becomes

$$O = k(E_a + \mu E_0)_g^{3/2}.$$

Therefore $E_a = -\mu E_c$

and if rewritten for point A it becomes

$$I_a|_o = kE_a^{3/2}$$

which if substituted gives

$$I_a|_o = k(-\mu E_c)^{3/2} = k\mu^{3/2}(-E_c)^{3/2}$$

μ includes the geometrical variations. This last result brings out the important fact that all the variations in $I_a|_o$ resulting from geometrical changes occur in the multiplying constant $k\mu^{3/2}$ and do not occur in the index of E_c. This fact is in substantial agreement with Fig. 17A in the paper.

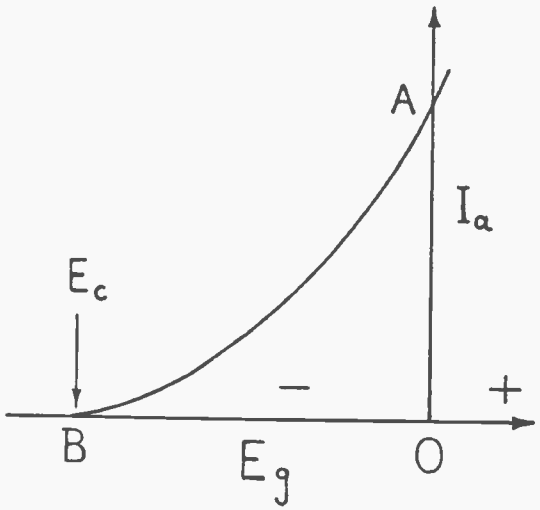


Fig. 3

In the dimensional equations (27) and (29) it is assumed that k has the dimensions of a voltage which implies that the equations incorporate all parameters with the dimensions of length. In the final result (29) t, b and f are multiplied in the denominator. There is, however, some evidence taken over not so wide a range as covered in the paper that t and b should be added and not multiplied, in which case the dimensional equation is rather upset.

Mr. G. L. Hamburger : Formula (35) presumably refers to the total cathode current and its relation to grid drive. I should like to ask what is the actual relation between the beam current and the grid drive.

Mr. W. A. Beatty : I should like to ask to what extent it has been found possible to obtain practical improvements in the performance of electron guns as a result of the theoretical analysis indicated in this paper.

REPLY TO THE LONDON DISCUSSION

Dr. Moss : My thanks are due to Dr. Liebmann for the long journey he has made to attend this meeting and for his interesting contribution. It is pleasing to know that his theoretical analysis supports my findings that the indices of b and f are both -1. It is a little surprising that the theoretical analysis should give so simple a result and it would be interesting to see Dr. Liebmann's reasoning. Of course, as Dr. Liebmann mentioned, I did not for a moment suggest that the terms D^3 and t in equation (29) were really precise : they are engineering approximations only, and it is consoling that his theoretical analysis indicates that they are quite reasonable approximations over the range which we are considering.

It would take too long to do justice to Dr. Liebmann's views on the cross-over theory on this occasion. Mutual discussions on this interesting matter are being published in the American Proc.I.R.E., probably in early summer, 1946.

I hope my remarks about the effect of cathode-grid spacing changes on beam angle and Fig. 23B will not be misinterpreted. So far as I can see, Dr. Liebmann and I are not in disagreement about the effect of this parameter, only he has chosen to view the matter from another angle, as indicated in his Slide 2. His later comment about getting more current into a given solid angle by reduction of the cathode-grid spacing is entirely consistent with my Appendix 2(*a*), line 2. The operation here is cathode-grid spacing multiplied by k and grid drive multiplied by $\frac{1}{k}$ to maintain constant beam angle. It will be noted that the total current as a result of this operation is multiplied by $\frac{1}{k^{3/2}}$. Problem No. 1 of Appendix 3 applies this result to an example which would seem to illustrate precisely the effect mentioned by Dr. Liebmann.

With regard to Fig. 19A, this has been confirmed on several occasions. In all cases the cathodes were run at very high temperatures so as to ensure space charge limited conditions. The variation in shape of the curves with modulation cannot therefore be ascribed to the physics of the cathode material, but must be due to changes in the field distribution.

Mr. Shelton's analysis based on the theory of the equivalent diode certainly helps to explain the form of my Fig. 17*A*, since, in the latter, variations of p are quite small.

I must agree with Mr. Shelton's implication that Equ. (29) is a bold stroke ; the only justification is that it is within moderate agreement with the observed phenomena.

Mr. Beatty touches on a delicate point. I'm afraid it must be confessed that no very marked *over-all* improvements have resulted from the theoretical analysis given. The reasons will be apparent from a study of what I have termed the " Constant brightness theorem " (Appendix 1). Over-all improvements could be effected only by reduction of lens aberrations, with which this paper is not concerned. On the other hand, our far greater knowledge of the basic principles of gun

design now enables us to adjust suitably the various *compromises* of performance, so that the cathode ray tube becomes closely matched to the operational requirements. In practice, the effect of this process can be almost as valuable to a user as can more fundamental progress in over-all performance. In this sense, therefore, the logical analysis has been of the utmost value.

Dr. Moss (Communicated reply to Mr. Hamburger): In general, the beam current in a cathode ray tube will be less than the total cathode current, since part of the latter is frequently wasted by the masking effect of beam limiting diaphragms. To compute the beam current, then, we may first use (35) to obtain the total cathode current and then estimate what fraction of this cathode current passes through the smallest stop used on the system.

The second part of this calculation involves a knowledge of the variation of beam density with beam radius, i.e., a knowledge of the shape of curves of the form of Fig. 19A. By determining the volume of revolution of assigned portions of such curves about the central vertical axis, it is possible to compute what fraction of the current will pass through a symmetrically placed beam stop.

A very approximate solution to this problem is as follows. The working is based on the assumption that from any infinitesimal element of the cathode surface, the emission density varies as the 3/2 power of the *grid drive for that element*. Thus, if the element in question, at distance y from the centre of the cathode surface, starts to emit when the grid drive is $\bar{E}_d$, we assert that its emission density when the grid drive is increased to E_d is proportional to $(E_d - \bar{E}_d^{3/2})$.

Consider the current contributed by a portion of the cathode surface bounded by two circles of radii y, and y + dy. This current will be given by $i = 2.\pi.K(E_d - \bar{E}_d)^{3/2}.y.dy$. The total current emitted from a circular portion of radius Y would be given by integrating this expression from O to Y. Making the bold approximation that the "mu" of the system is constant over the whole cathode face, the K may be taken outside the integral sign and the total current I is given by

$$I = 2.\pi.K\int_o^Y (E_d - \bar{E}_d^{3/2}).y.dy \quad \ldots\ldots\ldots\ldots (h)$$

Applying equ. (36) to the definition of E_d gives

$$\frac{2.y}{D} = \frac{\bar{E}_d}{E_c} \quad \ldots\ldots\ldots\ldots\ldots\ldots (i)$$

whence (h) can be written

$$I = 2.\pi.K.E_d^{3/2}\int_o^Y (1 - k.y)^{3/2}.y.dy \quad \ldots\ldots (j)$$

where

$$k = \frac{2.E_c}{D.E_d}.$$

This integral is evaluated by the successive substitutions $k.y = \sin^2\theta$ and $u = \cos\theta$. The result, after insertion of the limits, is

$$I = \frac{\pi.K.D^2}{35.E_c^2}E_d^{7/2}\left[(1 - k.Y)^{5/2}(5.k.Y + 2) - 2\right] \quad \ldots\ldots\ldots\ldots (k)$$

(k) gives us the value of the current contributed by *any* circular portion of the cathode of radius Y. For the purposes of computation it is generally convenient to eliminate Y by expressing it in terms of the grid drive necessary to produce an emitting diameter of Y. Let $\dot{E}_d$ be the grid drive necessary to produce an emitting diameter of Y. Then by (36) $2.Y/D = \dot{E}_d/E_c$, so that (k) becomes

$$I = -\frac{\pi.K.D^2}{35.E_c^2}E_d^{7/2}\left[\left(1 - \frac{\dot{E}_d}{E_d}\right)^{5/2}\left(5.\frac{\dot{E}_d}{E_d} + 2\right) - 2\right] \quad \ldots\ldots\ldots\ldots (l)$$

It will be noted that (l) has a physical significance only when $E_d > \dot{E}_d$. If in any problem $E_d < \dot{E}_d$, then we make the substitution $\dot{E}_d/E_d = 1$. This is the case when the emission from the cathode does not fill the aperture stop corresponding to emitting diameter of 2.Y.

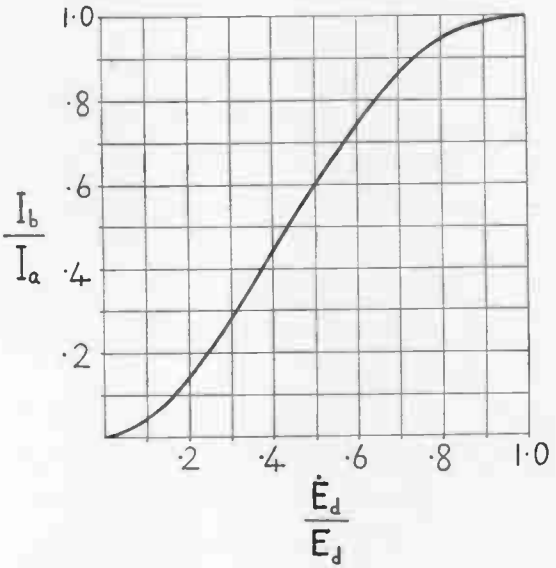


Fig. 4

For this condition equ. (l) reduces to

$$I = \frac{2.\pi K.D^2}{35.E_c^2}E_d^{7/2} \quad \ldots\ldots\ldots\ldots\ldots\ldots (m)$$

The form of equ. (m.) agrees well with the experimentally derived equ. (35). This provides some justification for the bold step of taking K outside the integral sign. Dividing (1) by (m) gives us the required ratio of beam current to total cathode current at any grid drive E_d when the smallest beam stop present in the system is filled for grid drive $\dot{E}d$.

Denoting this ratio by I_b/I_a, we finally obtain

$$\frac{I_b}{I_a} = -\frac{1}{2}\left[\left(1-\frac{\dot{E}_d}{E_d}\right)^{5/2}\left(5\cdot\frac{\dot{E}_d}{E_d}+2\right)-2\right] \quad ..(n)$$

Fig. 4 shows a plot of the ratio I_b/I_a against E_d/E_d. This curve, together with equations (35) and (36), permits an estimate to be made of the shape of the *beam* current/grid voltage curve provided we know the value of $\dot{E}_d$, i.e., the value of the grid drive at which the limiting beam stop is just filled.

NORTH WESTERN SECTION DISCUSSION

Dr. Jacob : The paper brings out in a remarkable way how solutions to many difficult and sometimes baffling problems are obtained by building upon the basis of simple fundamental structures, with suitably designed experiments to help in interpretation and further elucidation. As an example of what I mean, let me cite the reasoning in section 2.2.8, which relates to the effect of modulation on crossover size. If the simplifying idealised postulates of (*a*) freedom from space charge (*b*) no lens aberrations, are taken to hold, then the crossover size should be independent of modulation. This is derived on purely theoretical grounds which seem reasonable. If this be so, then measured values as determined by experiment should support the theory. Fig. 25A goes some way to confirm this, and it would appear on first sight that this conclusion is true in practice. However, the discussion in the next column draws attention to the fact that the deductions from the curves only apply over a certain restricted range. I am glad that this was inserted, since my own experiments have led me to believe that the crossover size does vary in an actual gun as the modulation voltage is varied—that is, if I keep to my own method of spot size measurement. In some ways, on qualitative grounds, this would be expected to happen, certainly as between the extinction point of the beam and the full modulation.

There is one point which causes me some worry. It relates to the method of focussing the beam. As far as I can gather, most of the electron beams with which the paper deals were focussed by E.S. lenses in which field stops were inserted. This is bound to have some effect on the distribution, depending on the diameter of the stop, since it is usually placed close to the region of the field-free space and traps the " surplus " part of the beam.

This point has been commented on by the author (p. 14) and it might have been expected that some of the distributions, particularly where crossover was involved, would be different, using a magnetic focussing lens with little or no aberration. All my work was done with such a lens, and this difference might account for our differences in interpretation.

The question of the effect of modulator hole was one that interested me. As far as I can remember from my own work on this matter, the cubic relation which Dr. Moss puts forward seems to have dwindled to a square law. I am not quite clear whether mine was the result of investigating the condition for a certain gradient at the cathode rather than that for a certain cut-off. In the first case, if you vary the hole diameter, then the gradient at the cathode surface appears to vary as the square of the diameter ; in the second, for the cut-off point, the field penetration varies inversely as the (diameter)^2. In any case, it seems that a square law is involved, at any rate as a first approximation.

As regards the triode portion of the tube, the relation which Dr. Moss shows to have at least an empirical foundation—$3E_d^{7/2}.E_c^{-2}$ is a very helpful one for design purposes. It tells you in effect what the characteristic curve of the tube will be like under different kinds of conditions ; and how, for any cut-off voltage, one is able to predict to within very close limits what the beam current is likely to be. An interesting problem on the theoretical side is to find the justification for the factor of 3. It seems so like π that I wouldn't be surprised if somebody dropped a pie or forgot it on the slide rule. Still, it would be nice to know why it should be there.

On certain theoretical grounds it might be possible to consider all the variations imposed on the triode portion as arising from an " equivalent diode." If this view is adopted, then the remarkable fact emerges that the " equivalent anode " of the diode is situated at the crossover for the system. The cathode, modulator and accelerator which comprise the beam forming system, if at suitable potentials, give rise to a certain field gradient at the cathode. This can be considered as produced by the equivalent anode at the crossover and the current into the beam is related to this gradient in a simple way. It is easy to see that if this gradient happens to be constant, as the crossover moves about with changes in geometry, the current will remain constant and this is, perhaps, the reason why as Dr. Moss points out, the beam current is uniquely defined by the cut-off voltage. This point of view appears to fit many observations I have made on crossover properties. I will not elaborate it further here, as I am preparing a paper on immersion systems where I shall have something more to say about it, however, I felt that another way of looking at the triode, would not be out of place.

In the slide shown (P.1) of the emitting grid in the cathode, it is quite true from the evidence you have submitted that the image area of the cathode is proportional to the modulation voltage. I notice, however, that the pattern of the illuminated grid, or rather, " self-luminous " grid, remained undistorted throughout the whole modulation range, i.e., the distance between squares remained constant.