The Marconi Review

o 123 4th QUARTER 1956 Vol. XIX

CONTENTS:

ARCONI'S WIRELESS TELEGRAPH COMPANY LIMITED
et Office, Marconi House, Chelmsford Telephone, Chelmsford 3221 Telegraphic Address, Expanse, Chelmsford

THE MARCONI GROUP OF COMPANIES IN GREAT BRITAIN

Registered Office : Marconi House,
Strand,
London, W.C.2.

Telephone : Covent Garden 1234.

MARCONI'S WIRELESS TELEGRAPH COMPANY, LIMITED

Marconi House,
Chelmsford,
Essex.

Telephone : Chelmsford 3221.
Telegrams : Expanse, Chelmsford.

THE MARCONI INTERNATIONAL MARINE COMMUNICATION COMPANY, LIMITED

Marconi House,
Chelmsford,
Essex.

Telephone : Chelmsford 3221.
Telegrams : Thulium, Chelmsford.

THE MARCONI SOUNDING DEVICE COMPANY, LIMITED

Marconi House,
Chelmsford,
Essex.

Telephone : Chelmsford 3221.
Telegrams : Thulium, Chelmsford.

THE RADIO COMMUNICATION COMPANY, LIMITED

Marconi House,
Chelmsford,
Essex.

Telephone : Chelmsford 3221.
Telegrams : Thulium, Chelmsford.

THE MARCONI INTERNATIONAL CODE COMPANY, LIMITED

Marconi House,
Strand,
London, W.C.2.

Telephone : Covent Garden 1234.
Telegrams : Docinocram.

MARCONI INSTRUMENTS, LIMITED

St. Albans,
Hertfordshire.

Telephone : St. Albans 6161/5.
Telegrams : Measurtest, St. Albans.

SCANNERS LIMITED

Woodskinners Yard,
Bill Quay,
Gateshead, 10,
Co. Durham.

Telephone : Felling 82178.
Telegrams : Scanners, Newcastle-upon-Tyne.

THE MARCONI REVIEW

No. 123 Vol. XIX *4th Quarter, 1956*

Editor : L. E. Q. WALKER, A.R.C.S.

MIS-FOCUSING AND THE NEAR-FIELD OF MICROWAVE AERIALS

BY D. H. SHINN, M.A., Ph.D.

Curves are presented showing (a) the theoretical radiation patterns of a mis-focused lens or mirror with a circular boundary and (b) the near-field of the same aerial correctly focused. Some practical applications of these are briefly discussed.

Introduction

A MICROWAVE aerial system consisting of a feed-horn and a reflector or lens is generally designed so that it is correctly focused. However, it may sometimes be convenient to use a mis-focused aerial. For this reason, and also in order to present results applicable to the near-field of a correctly focused aerial, some computations have been done by previous workers on the effect of a quadratic phase error on the radiation pattern of an aerial.

This work, by Milne[1] and Meredith[2], is applicable to line sources or horns. It is the purpose of this paper to present results which are applicable to a circular aperture and to make some deductions from these results. The results can also be applied to physical optics and some mention will be made of these applications.

Mis-focusing causes a quadratic error in phase, i.e. if the feed-horn is on the axis of a circular aperture, then the phase error across the aperture due to mis-focusing of the source is proportional to r^2, where r is the distance from the centre of the aperture. In fact, mis-focusing also introduces terms in r^4, r^6, etc., but we shall assume that the effect of these terms is negligible. It can be shown that this is so for the largest phase error considered in this paper, 5·5 radians, provided that the focal length is greater than about half of the diameter of the aperture.

The Circular Aperture (Radiation Pattern)

The radiation pattern of a circular aperture of radius b, with an aperture illumination $f(r)$ is given approximately by

$$g(v) = \int_0^b f(r)\, r\, J_0(vr)\, dr$$

or

$$g(\gamma) = \int_0^1 f(x)\, x\, J_0(\gamma x)\, dx \tag{1}$$

where $x = \frac{r}{b}$, $\gamma = bv = \frac{2\pi b \sin\theta}{\lambda}$, and J_0 is the usual Bessel Function[3].

For an aerial subject to a quadratic phase error the aperture illumination $f(x)$ is $T(x) \exp(j\alpha x^2)$, where $T(x)$ is real (the amplitude distribution) and α is the maximum error in phase. The radiation pattern is then given by

$$g(\gamma) = \int_0^1 T(x) \exp(j\alpha x^2)\, x\, J_0(\gamma x)\, dx \tag{2}$$

For constant illumination $T(x) = 1$, and

$$g(\gamma) = \int_0^1 \exp(j\alpha x^2)\, x\, J_0(\gamma x)\, dx \tag{3}$$

This integral has been discussed and tabulated by Robinson[4]. The value of $|g(\gamma)|$ is of considerable interest in optics; one of its applications is discussed and some numerical values presented by Rayleigh[5].

However, apertures of aerials are generally not uniformly illuminated. In fact, circular apertures are generally lenses or mirrors illuminated by a horn at or near the focus.

The design of such aerials is discussed by Crompton[6]. He shows that the maximum overall gain of the system assuming that the illumination of the aperture is of the form $\frac{\sin kx}{kx}$, is given when the field strength at the edge of the aperture is 0·27 of that at the centre. However, the overall gain remains within 0·5 dB of its maximum value provided that the illumination at the edge lies between 0·05 and 0·53 of that at the centre. We consider it convenient here to carry out the computation for the field strength at the edge equal to one third of that at the centre, since this criterion has often been used in designing aerials; the overall gain for this case is, according to Crompton, about 0·03 dB less than its maximum.

Following Crompton's work we should therefore assume that $T(x) = \sin 2{\cdot}28(x)/2{\cdot}28(x)$. However, the computation is much simpler if we take $T(x) = 1 - 2x^2/3$; this is quite good enough for practical purposes.

According to Robinson's notation[4]

$$A + jB = \int_0^1 \exp(j\alpha x^2)\, J_0(\gamma x)\, x\, dx \tag{4}$$

The radiation pattern corresponding to an aperture illumination

$$f(x) = \left(1 - \frac{2}{3}x^2\right) \exp(j\alpha x^2)$$

is given by

$$E + jF = (A + jB) - \frac{2}{3}(C + jD) \tag{5}$$

where

$$C + jD = \int_0^1 \exp(j\alpha x^2)\, J_0(\gamma x)\, x^3\, dx \tag{6}$$

C and D were computed from Robinson's tables, using the equations

$$\left.\begin{aligned} C &= \frac{\partial B}{\partial \alpha} \\ D &= -\frac{\partial A}{\partial \alpha} \end{aligned}\right\} \tag{7}$$

The results of computations of $\sqrt{E^2 + F^2}$ are presented in Figs. 1, 2 and 3. In Fig. 1(a) the patterns are normalized so that $\sqrt{E^2 + F^2}$ is unity for $\gamma = 0$. In Fig. 1(b) this normalization has not been carried out; the curves here therefore indicate directly how the field strength at a point varies as the aerial is misphased.

In Fig. 2 the gain in the forward direction is plotted against α.

In Fig. 3(a) one of the patterns in Fig. 1(a), for $\alpha = 4$, is compared with the pattern for an aperture with the same shape of amplitude distribution (i.e., tapering to one third at the edge) but constant phase, whose size is adjusted so that it has the same gain in the forward direction. Fig. 3(b) shows a similar comparison for $\alpha = 5$. It will be seen from these figures that the radiation pattern of the misphased aperture differs only slightly from that of the smaller in-phase aperture, at least as far as the main beam is concerned. Therefore, if it is desired to substitute for a large reflector or lens an aerial of less gain and greater beamwidth, this can be done either by moving the feed away from the focus or by using a smaller reflector or lens with the feed at the focus. Each of these devices has roughly the same effect, up to a decrease in gain of 10 dB (beamwidth increased by $\sqrt{10}$), except that, if a split-beam technique is being used, the misphased aerial may behave strangely, both because of ripples in its amplitude pattern, and also, possibly, because of peculiarities in its phase pattern. The latter is not presented here, but can be deduced from information in Fig. 4(b) and equation (8).

The Circular Aperture (Near-Field)

The near-field of a circular aperture of radius b is given approximately by

$$n(\gamma) = 2\beta \exp\left(j\frac{\gamma^2}{4\beta}\right) \exp\left(j\frac{2\pi d}{\lambda}\right) \int_0^1 f(x) \exp(j\beta x^2)\, x\, J_0(\gamma x)\, dx \tag{8}$$

where d = distance from aperture.

β = maximum phase error due to proximity $= \dfrac{\pi b^2}{d\lambda}$

$\gamma = \dfrac{2\pi b\eta}{d\lambda}$, where η = distance off axis

$f(x)$ = aperture illumination.

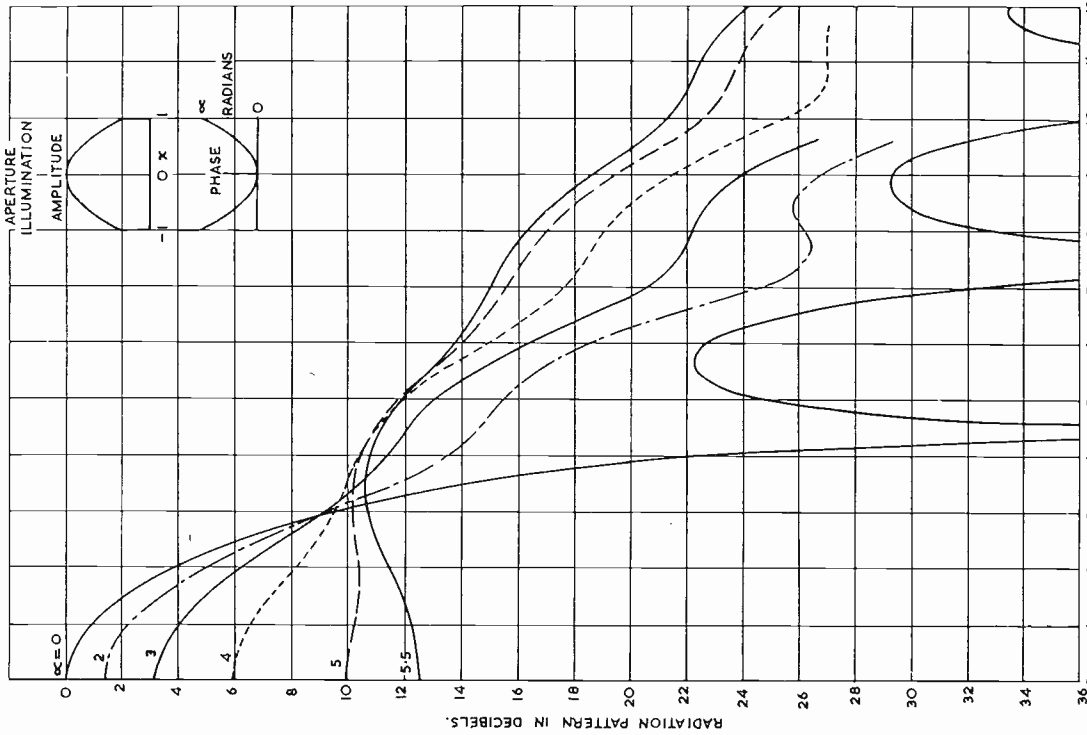


Fig. 1(a)

Radiation Patterns of a Circular Aperture with illumination $(1-\frac{2x^2}{3})^{\alpha}$ *exp* (jax^2).

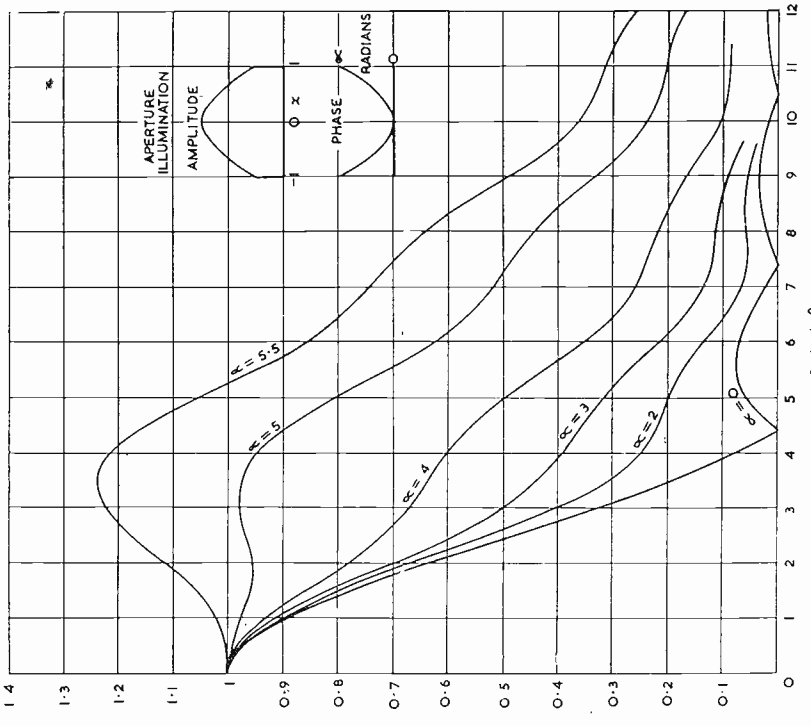


Fig. 1(b) on right

Patterns as in Fig. 1(a) *with field strength in decibels.*

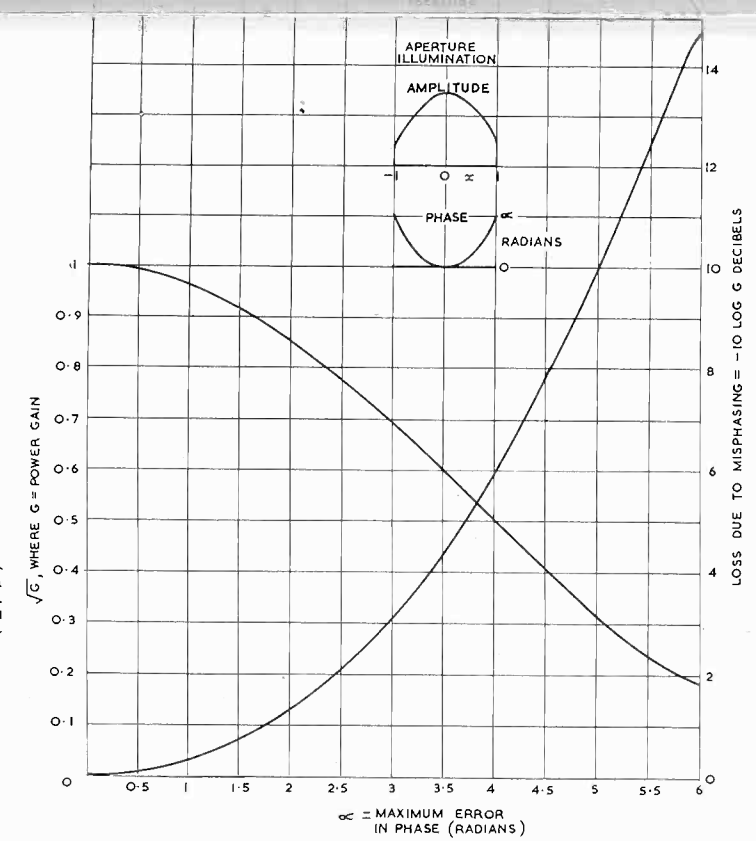
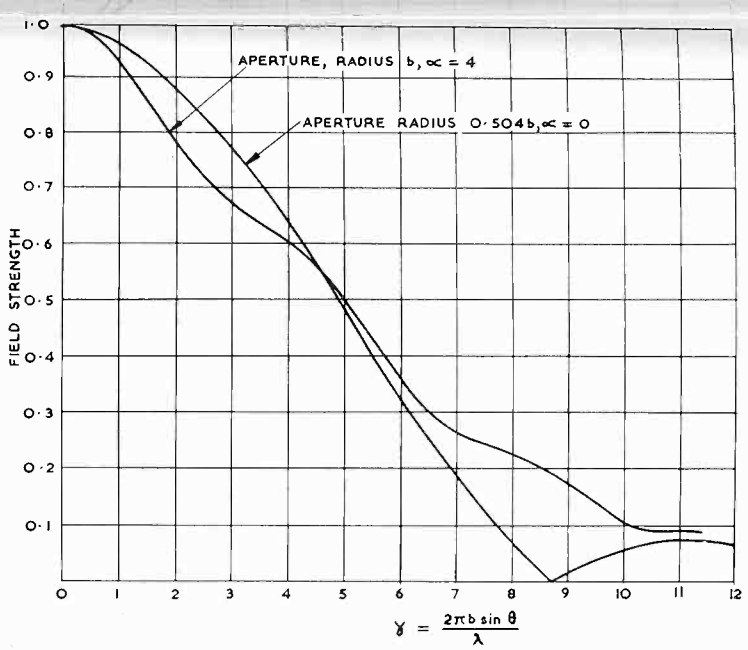


Fig. 2

Gain in forward direction of circular Aperture with illumination $(1-\frac{2x^2}{3})$ *exp* (jax^2).



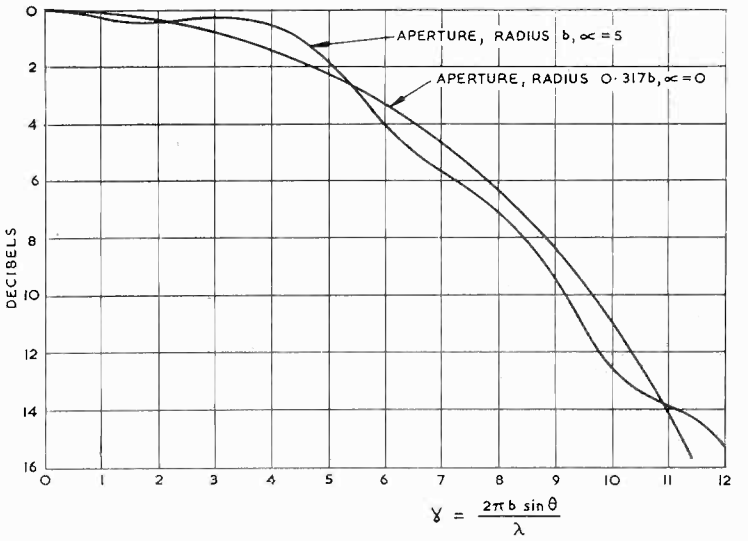


Fig. 3(a) and (b) on right

Patterns of Fig. 1 *compared with patterns of constant phase apertures of the same forward gain.*

It follows that the results of the previous section can be applied directly to the near-field. The field distribution across a plane parallel to the aperture and at a

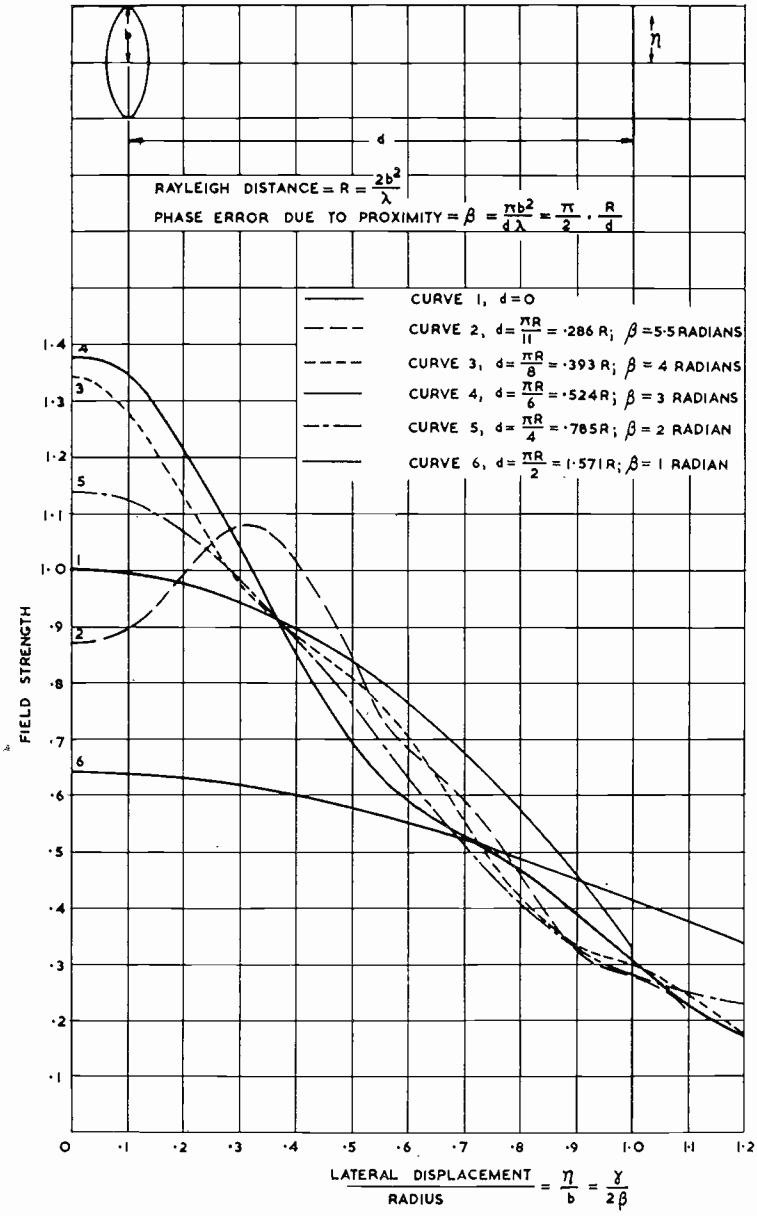


Fig. 4(a)
Near-field of a circular aperture with illumination $(1 - 2\eta^2/3b^2)$.

distance d from it is plotted in Figs. 4 and 5 for $f(x) = 1 - \frac{2}{3}x^2$ and for $f(x) = 1$, respectively, for values of d from $\pi R/11$ to $\pi R/2$, where $R = 2b^2/\lambda$.

The quantity R can conveniently be called the " Rayleigh distance ". It is that distance from the aperture at which β, the maximum phase error due to proximity,

as equal to one. It corresponds to a certain length of paper. Each length states that the best definition for a particular aperture of lens, focal length, is obtained when the aperture is adjusted so that the focal length is also equal

The Rayleigh

1. The length

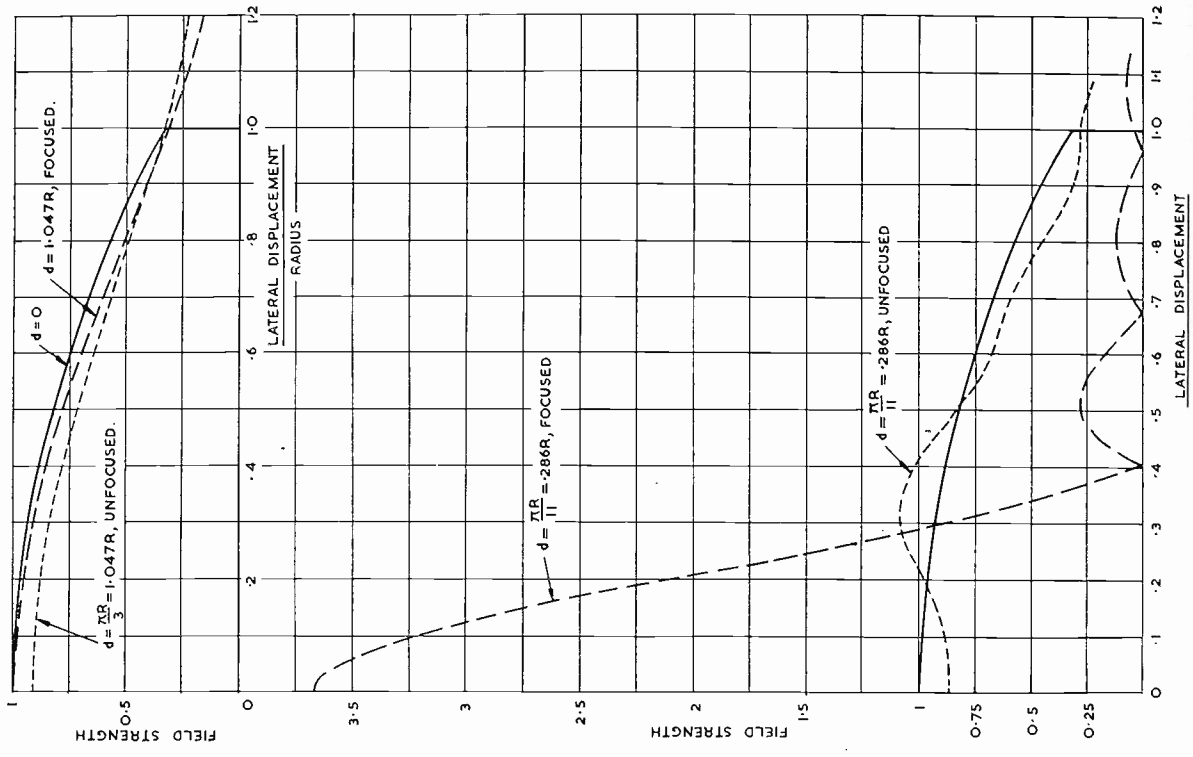


Fig. 6
Near-field of a circular aperture, as in Fig. 4(a), showing the effect of focusing

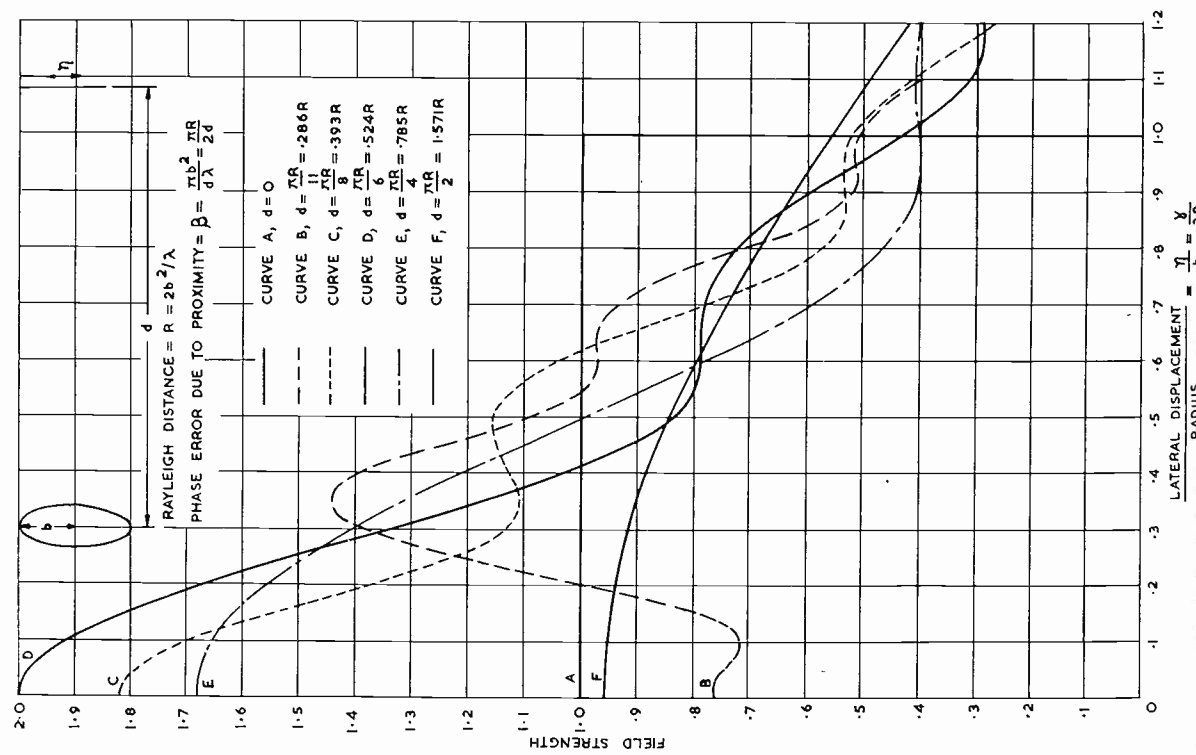


Fig. 5
Near-field of a uniformly illuminated circular aperture.

(ii). If two similar aerials, separated by less than their Rayleigh distance, are pointed directly at each other, most of the power radiated by one aerial is received by the other, i.e., the coupling loss between the aerials is only a few decibels above the sum of the spill-over, and similar losses in each aerial.

(iii). It is impossible to focus the radiation from an aerial to a point whose distance from the aerial is greater than its Rayleigh distance. This is illustrated in Fig. 6. Alternatively, geometrical optics is reliable only well within the Rayleigh distance.

Radiation from focused horns is discussed in detail by Meredith[2]. He also mentions the approximations used in deriving equation (8). Broadly speaking this equation is valid provided that the aperture subtends an angle of not more than about 10° at the point at which the field is measured. Meredith defines the Rayleigh distance as $4b^2/\lambda$, i.e., twice our Rayleigh distance.

Some diagrams corresponding to Fig. 5 are given by Bachynski and Bekefi[7], Fig. 2. They also discuss the approximations which lead to equation (8).

Conclusion

The results of this investigation, together with the results of Milne[1] and Meredith[2], are likely to be useful whenever a lens or reflector is illuminated so that the phase of the reflected wave has a quadratic type of error, or whenever the near-field of a lens or reflector is in question. Both of these circumstances may arise not only when an ordinary aerial is purposely or accidentally mis-focused or in the near-field of an ordinary aerial, but also when a large plane reflector is illuminated by an aerial, either purposely (for use as a " passive reflector "), or accidentally, e.g. when the performance of a radar or direction finder suffers due to reflections from an aircraft hangar or other large building. The curves in this paper may be used to predict the radiation pattern due to such a reflection, but the actual amplitude of the reflected signal requires further computation.

Acknowledgment

Thanks are due to Miss M. A. Millidge, of the Mathematical Section of the Marconi Research division, who did most of the computations.

References

(1) K. Milne, " The Effect of Phase Errors on Simple Aperture Illuminations ", Proc. of a Conference on Centimetric Aerials for Marine Navigation (Ministry of Transport, H.M.S.O. 1952). Milne's a, u, ϕ are equal to our b, v, γ respectively.

(2) R. Meredith, " Radiation Contours of Focused Aerials ". Memorandum No. 1206, Radar Research Establishment, Malvern, 1955. Note that $a = 2b$, $u = v/2\pi$, $2\pi\phi/\lambda = \alpha$ or β, and that his " Rayleigh distance " is $2x$ times our " Rayleigh distance."

(3) S. Silver, " Microwave Antenna Theory and Design ", (McGraw Hill Book Co.). The approximations made here are discussed in Chapter 6. Note especially Chapter 5, equations (122), (123), (124) and Chapter 6, equations (29), (71).

(4) C. Robinson, English Electric Co., Nelson Research Laboratories, Stafford. Report NSv45 (1952). This contains tables of g (γ) in equation (5) for $\alpha = 0$ (0·25) 6, $\gamma = 0$ (0·2) 3α or 10, whichever is the larger.

(5) Lord Rayleigh, " On Pin-Hole Photography ". Phil. Mag. Vol. 31, pp. 87-99, 1891; Collected Works, Vol. 3, p.429.

(6) J. W. Crompton, " On the Optimum Illumination Taper for the Objective of a Microwave Aerial ". Proc. Inst. Elec. Eng., Vol. 101, Pt. III, p.371, 1954.

(7) M. P. Bachynski and G. Bekefi, " Aberrations in Circularly Symmetric Microwave Lenses ". Trans. I.R.E., Vol. AP-4, p.412, 1956.

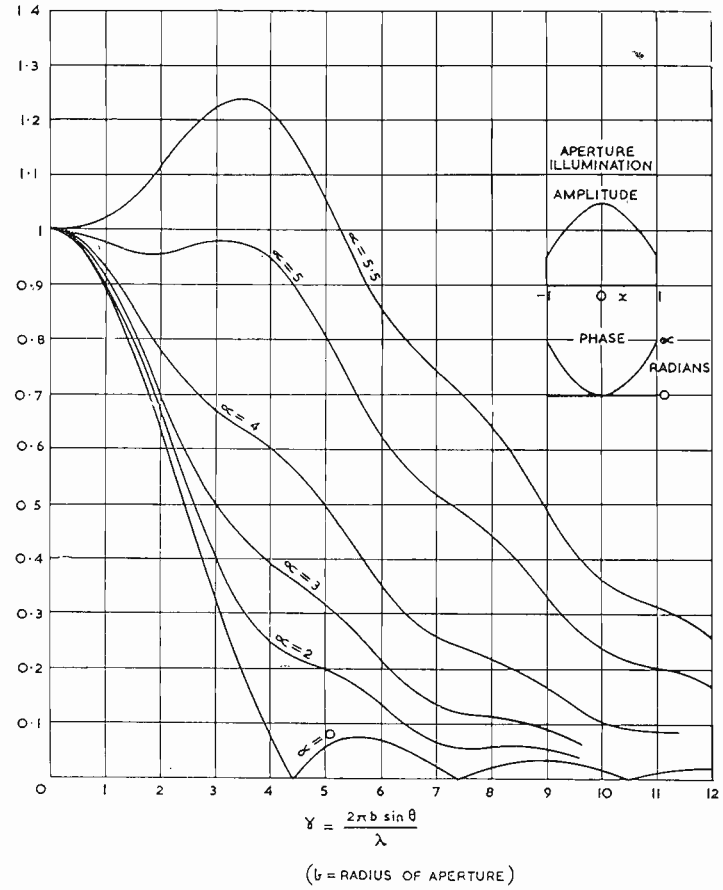


Fig. 1(a)

Radiation Patterns of a Circular Aperture with illumination $(1-\frac{2x^2}{3})$ *exp* (jax^2).

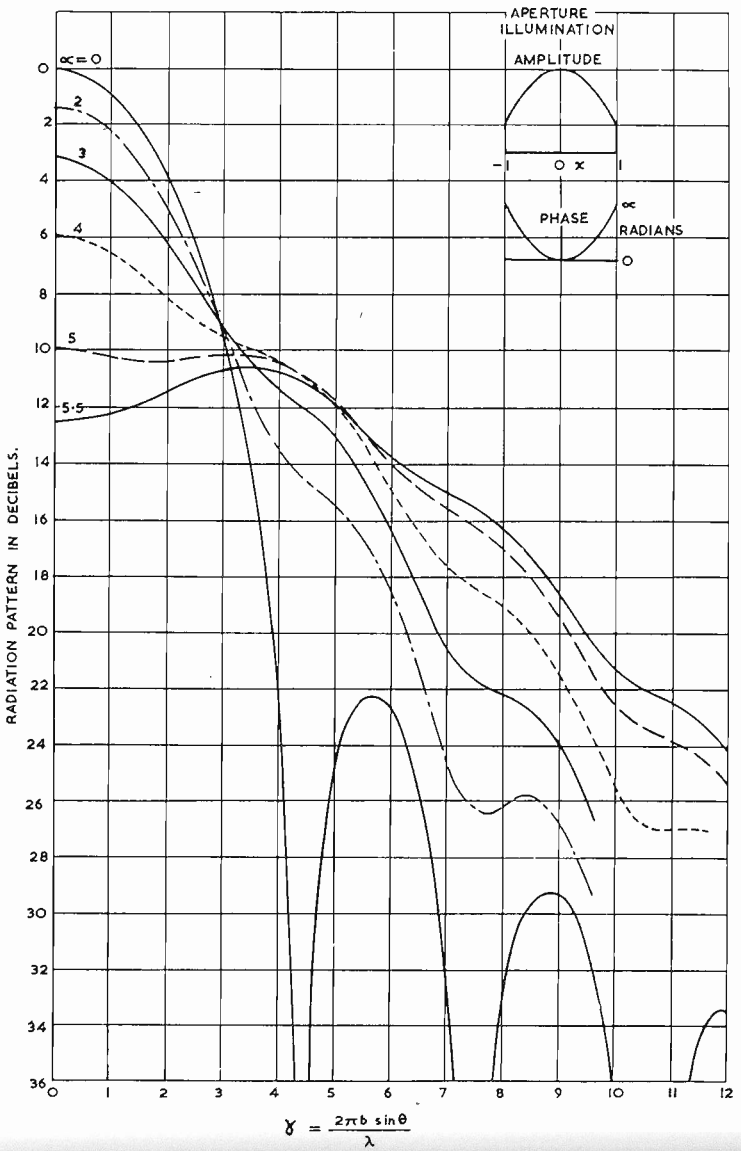


Fig. 1(b) on right

Patterns as in Fig. 1(a) *with field strength in decibels.*

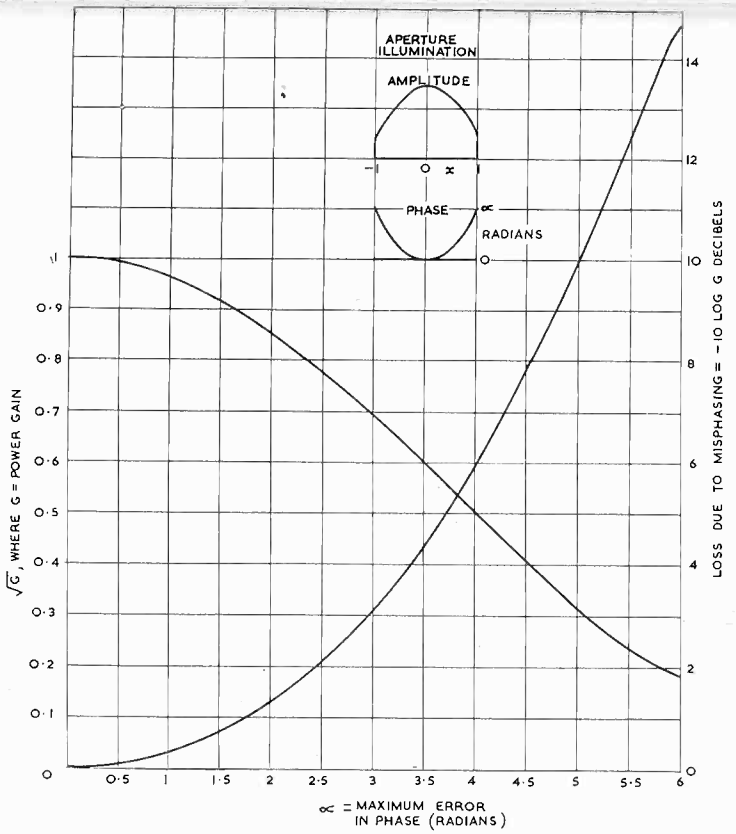
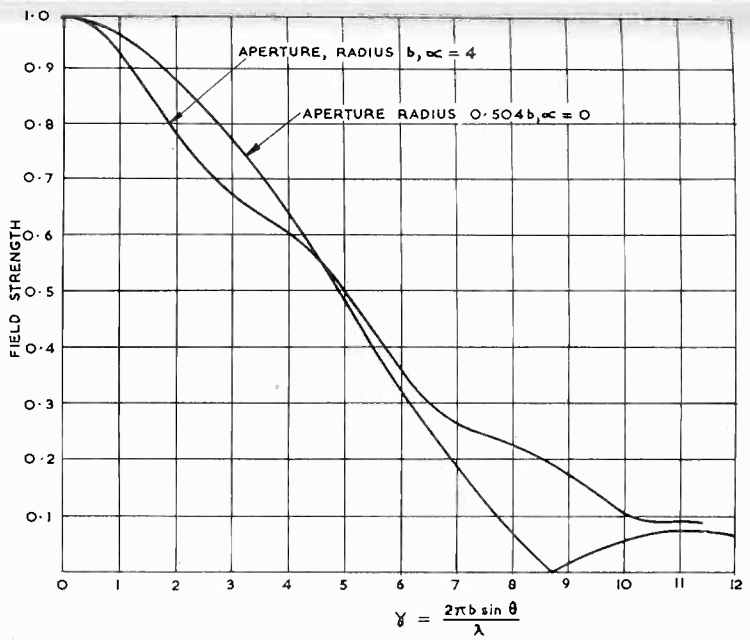


Fig. 2

Gain in forward direction of circular Aperture with illumination $(1-\frac{2x^2}{3})$ *exp* (jax^2).



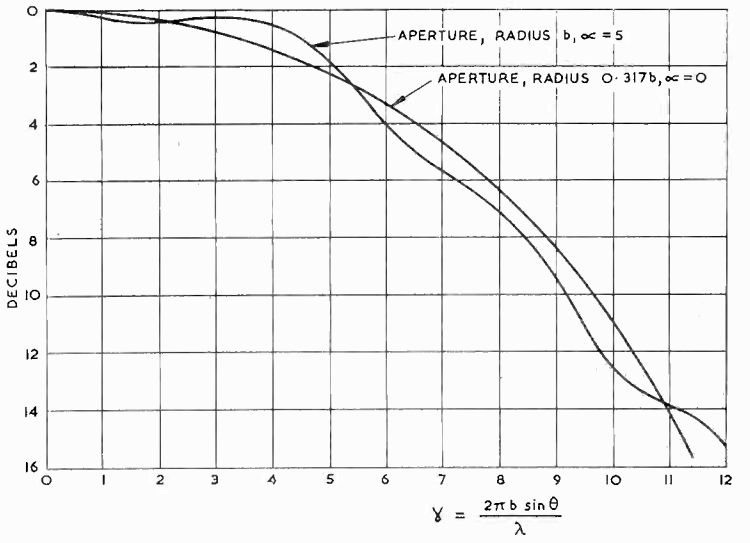


Fig. 3(a) and (b) on right

Patterns of Fig. 1 *compared with patterns of constant phase apertures of the same forward gain.*

It follows that the results of the previous section can be applied directly to the near-field. The field distribution across a plane parallel to the aperture and at a

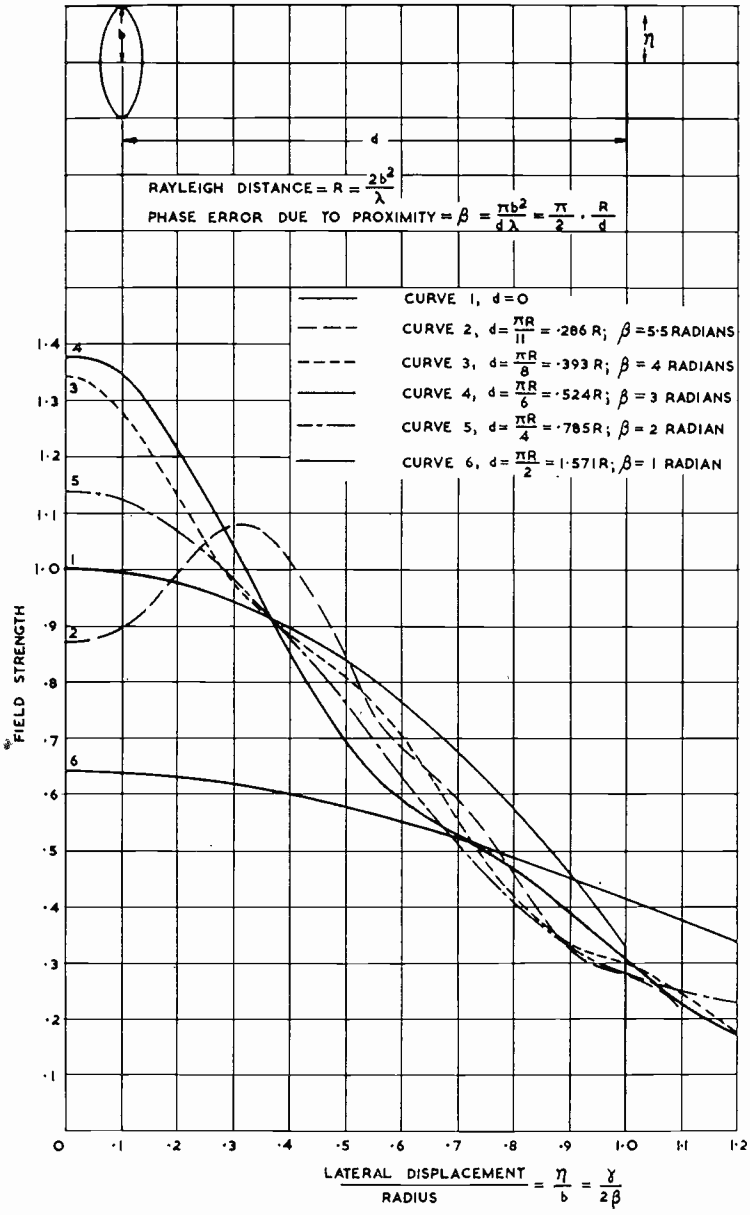


Fig. 4(a)
Near-field of a circular aperture with illumination $(1 - 2\eta^2/3b^2)$.

distance d from it is plotted in Figs. 4 and 5 for $f(x) = 1 - \frac{2}{3}x^2$ and for $f(x) = 1$, respectively, for values of d from $\pi R/11$ to $\pi R/2$, where $R = 2b^2/\lambda$.

The quantity R can conveniently be called the " Rayleigh distance ". It is that distance from the aperture at which β, the maximum phase error due to proximity,

is equal to 90°. It corresponds to $\nu = \pi$ in Rayleigh's paper[9]. Rayleigh states that the best definition for a pin-hole camera of fixed "focal length" is obtained when the aperture is adjusted so that the "focal length" is about $0{\cdot}6R$.

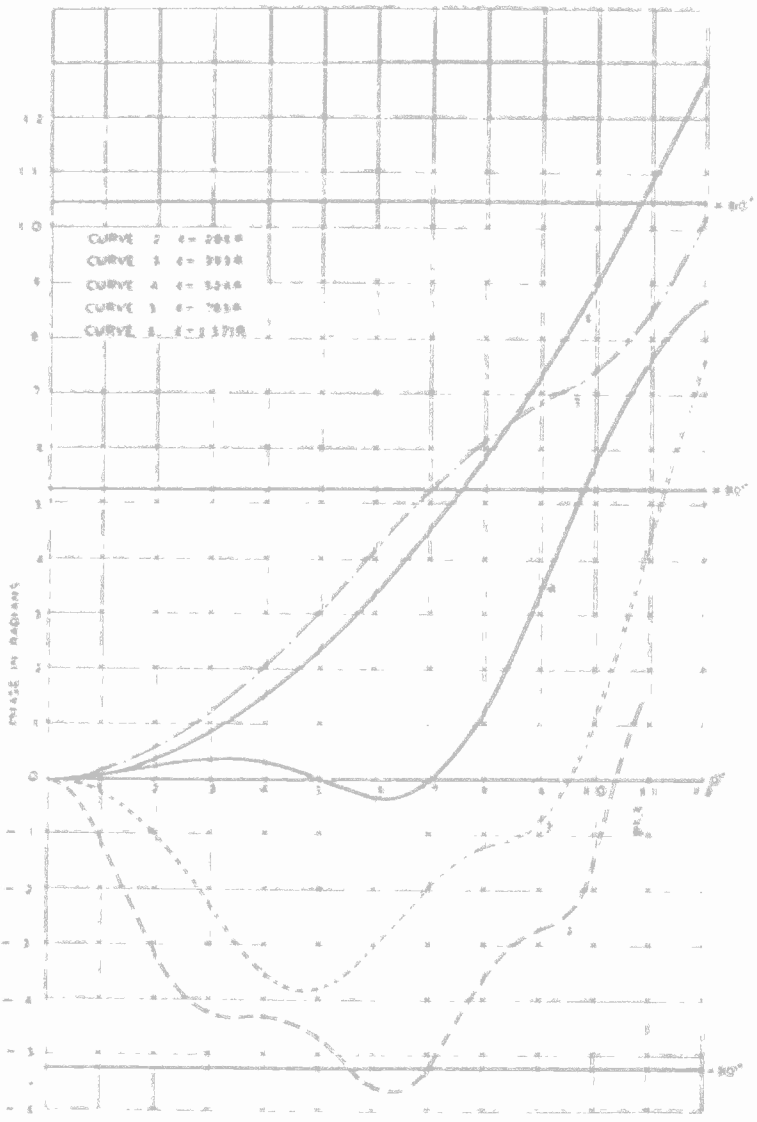


Fig. 4(b)

Phase over the near field of aperture as in Fig. 4(a)

The Rayleigh distance has other important properties.

(i) The length of a site for measuring the radiation pattern of an aerial must be greater than four times the Rayleigh distance corresponding to the maximum effective dimension of the aperture. There is something to be said for making it longer than this (see the discussion on Milne's paper[10]).

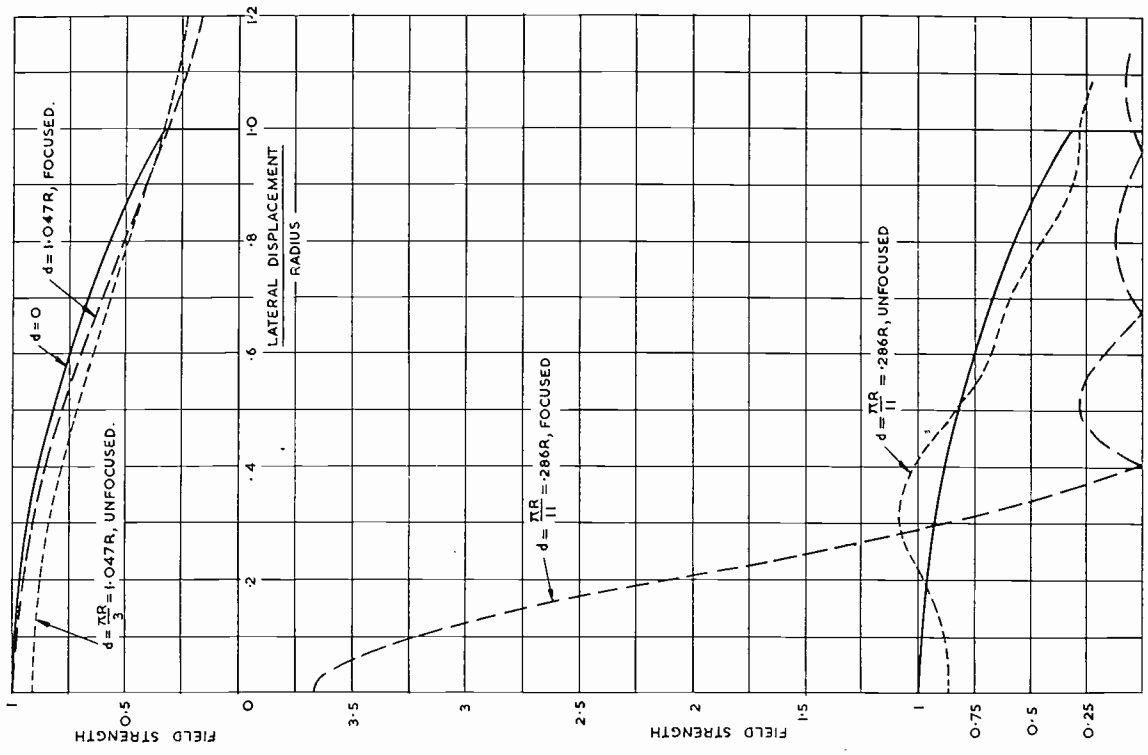


Fig. 6

Near-field of a circular aperture, as in Fig. 4(a), showing the effect of focusing.

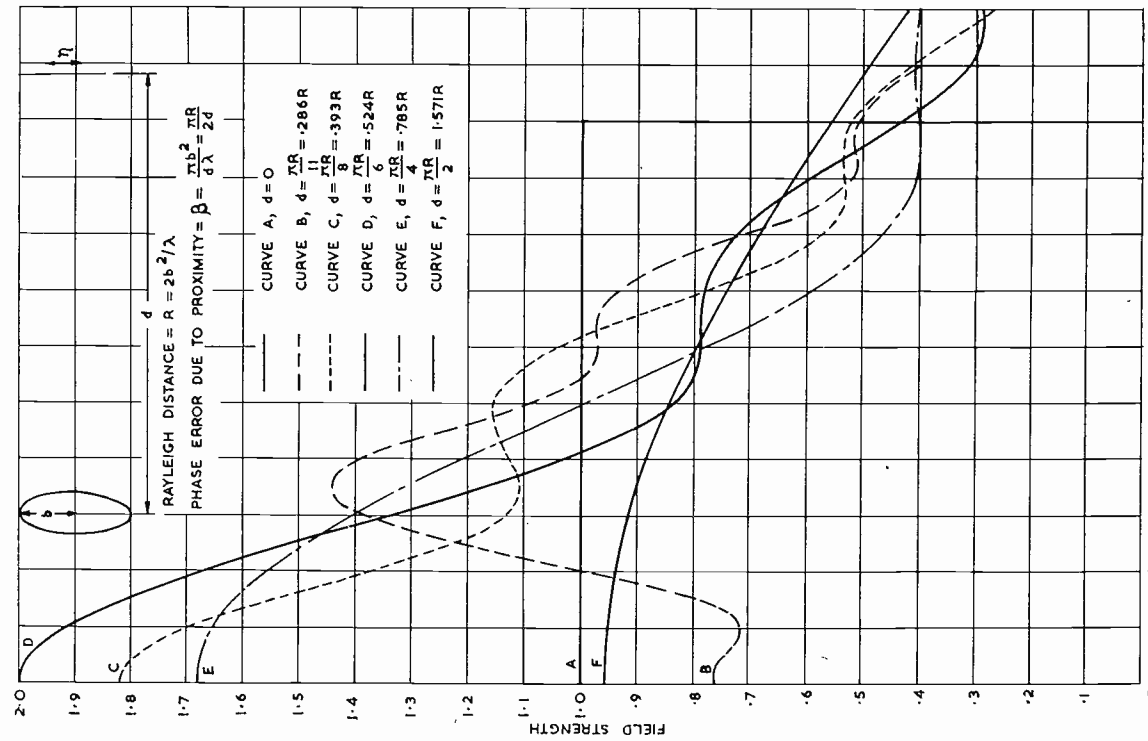


Fig. 5

Near-field of a uniformly illuminated circular aperture.

(ii). If two similar aerials, separated by less than their Rayleigh distance, are pointed directly at each other, most of the power radiated by one aerial is received by the other, i.e., the coupling loss between the aerials is only a few decibels above the sum of the spill-over, and similar losses in each aerial.

(iii). It is impossible to focus the radiation from an aerial to a point whose distance from the aerial is greater than its Rayleigh distance. This is illustrated in Fig. 6. Alternatively, geometrical optics is reliable only well within the Rayleigh distance.

Radiation from focused horns is discussed in detail by Meredith[2]. He also mentions the approximations used in deriving equation (8). Broadly speaking this equation is valid provided that the aperture subtends an angle of not more than about 10° at the point at which the field is measured. Meredith defines the Rayleigh distance as $4b^2/\lambda$, i.e., twice our Rayleigh distance.

Some diagrams corresponding to Fig. 5 are given by Bachynski and Bekefi[7], Fig. 2. They also discuss the approximations which lead to equation (8).

Conclusion

The results of this investigation, together with the results of Milne[1] and Meredith[2], are likely to be useful whenever a lens or reflector is illuminated so that the phase of the reflected wave has a quadratic type of error, or whenever the near-field of a lens or reflector is in question. Both of these circumstances may arise not only when an ordinary aerial is purposely or accidentally mis-focused or in the near-field of an ordinary aerial, but also when a large plane reflector is illuminated by an aerial, either purposely (for use as a " passive reflector "), or accidentally, e.g. when the performance of a radar or direction finder suffers due to reflections from an aircraft hangar or other large building. The curves in this paper may be used to predict the radiation pattern due to such a reflection, but the actual amplitude of the reflected signal requires further computation.

Acknowledgment

Thanks are due to Miss M. A. Millidge, of the Mathematical Section of the Marconi Research division, who did most of the computations.

References

(1) K. Milne, " The Effect of Phase Errors on Simple Aperture Illuminations ", Proc. of a Conference on Centimetric Aerials for Marine Navigation (Ministry of Transport, H.M.S.O. 1952). Milne's a, u, ϕ are equal to our b, v, γ respectively.

(2) R. Meredith, " Radiation Contours of Focused Aerials ". Memorandum No. 1206, Radar Research Establishment, Malvern, 1955. Note that $a = 2b$, $u = v/2\pi$, $2\pi\phi/\lambda = \alpha$ or β, and that his " Rayleigh distance " is 2π times our " Rayleigh distance."

(3) S. Silver, " Microwave Antenna Theory and Design ", (McGraw Hill Book Co.). The approximations made here are discussed in Chapter 6. Note especially Chapter 5, equations (122), (123), (124) and Chapter 6, equations (29), (71).

(4) C. Robinson, English Electric Co., Nelson Research Laboratories, Stafford. Report NSv45 (1952). This contains tables of $g(\gamma)$ in equation (5) for $\alpha = 0\ (0{\cdot}25)\ 6$, $\gamma = 0\ (0{\cdot}2)\ 3\alpha$ or 10, whichever is the larger.

(5) Lord Rayleigh, " On Pin-Hole Photography ". Phil. Mag. Vol. 31, pp. 87-99, 1891; Collected Works, Vol. 3, p.429.

(6) J. W. Crompton, " On the Optimum Illumination Taper for the Objective of a Microwave Aerial ". Proc. Inst. Elec. Eng., Vol. 101, Pt. III, p.371, 1954.

(7) M. P. Bachynski and G. Bekefi, " Aberrations in Circularly Symmetric Microwave Lenses ". Trans. I.R.E., Vol. AP-4, p.412, 1956.

A UNIVERSAL SCANNING CURVE FOR WIDE ANGLE MIRRORS AND LENSES

By J. F. Ramsay, M.A., M.I.E.E.

A general curve is derived showing loss of gain on scanning of a coma-corrected focusing element-lens or mirror. The curve is normalised to an aperture of 100λ *(c.* 1° *beam) with unity F—number. It can be used, however, for any aperture diameter (or beamwidth) and focal length, at any wavelength, subject to restrictions determined by the assumptions.*

Agreement with the measured scanning performance of two coma-corrected zoned mirrors, two spherical mirrors and two coma-corrected dielectric lenses is shown.

How the curve may be used to determine the scanning performance of offset fed mirrors of circular symmetry of profile is also described.

Introduction

THE question is frequently asked " What is the available scanning of such and such an aerial (mirror or lens) at some particular wavelength?" The allowable gain reduction and/or pattern deterioration is usually prescribed. Often these are left open and the question hinges on geometry. " How good is the scanning of this aerial of this size and with this focal length at this particular frequency." A further less obvious question is also involved, viz., " On what focal arc or surface has the moving feed or multifeed scanning to be placed?"

A previous paper in the *Marconi Review** gave an analysis of the scanning limitations of a corrected lens due to astigmatism. Since the publication of that paper considerable measurement has been made of the scanning performance of a variety of lenses and mirrors, largely coma-corrected, with a view to determining experimentally their useful fields. After a review of the experimental results it became evident that the 1952 theory was becoming capable of further checks. In particular, loss of gain on scanning could be obtained in a generalised form yielding engineering information for the design of wide angle elements, so frequently required by designers.

This paper therefore describes a universal scanning curve applicable to coma-corrected lenses and mirrors of circular aperture suited to any diameter, focal length or wavelength. The agreement of the curve with practice is shown.

" Loss in Gain on Scanning " Characteristic Curves

The theoretical idealisations upon which this approximate theory is based are:—

(1) That the circular lens or mirror is coma-corrected by shaping the profile, i.e., bending the lens or mirror to fulfil the Abbé sine condition.

*Marconi Review No. 107, 1952 "Scanning Aberrations of Radio Lenses," Cheston and Shinn.

(2) That in the case of a lens, the lens is infinitely thin.

(3) That the focal surface to which the performance relates is that containing points which maximise the gain.

(4) That only the lowest order of spherical aberrations is significant and that all odd phase errors are zero.

(5) That obliquity is not usually significant.

(6) That it is legitimate to consider phase errors associated with constant illumination such that the scanning will be a function of spherical aberration only, while loss in gain associated with that spherical aberration may be associated with a 10 dB (or thereabouts) illumination taper.

Key

Aerials Used for the Experimental Checks. Normalised to 100λ aperture (1° beam) and F=1.

Key	Aerial	F-No.	D/λ	Beam-width (−6db)
Δ	Zoned Mirror	1·5	70	1·35°
○	Spherical Mirror	1·0	35	2·7°
□	Zoned Mirror	1·0	35	2·8°
×	Spherical Mirror	1·5	35	2·7°
⊡	Polythene Lens	1·5	70	1·29°
⊙	Perspex Lens	1·4	35	2·75°

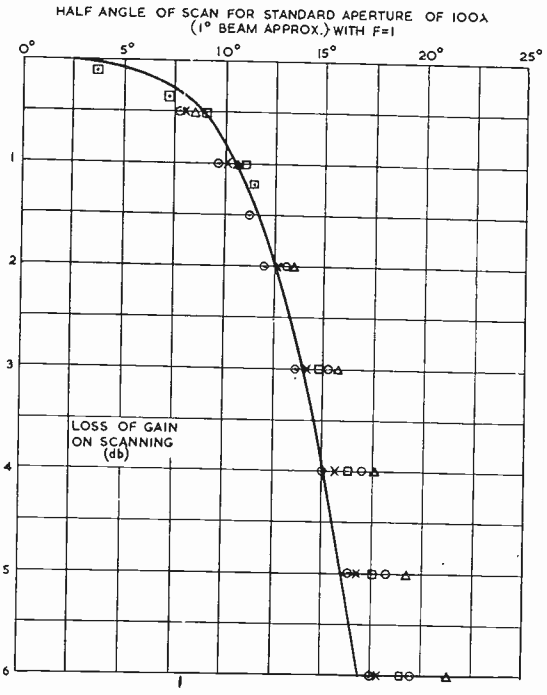


Fig. 1

Unversal Scanning Curve (Theoretical) showing experimental checks.

These were the assumptions made in 1952* and clearly limited the analysis. As the latter is, however, relatively simple and the agreement with practice satisfactory for first order estimates of the scanning, the above approximations will be retained.

Assuming that only the second order terms of the expansion of the phase error in aperture co-ordinates are significant the available scan angle* is given by

$$H.S.A. = 4\sqrt{\frac{F\delta_2}{D/\lambda}} \quad \text{radians} \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

* loc. cit. p. 176

where

$H.S.A.$ = half scan angle.
$F = \frac{f}{D}$ = F number.
D = dia. of aperture.
λ = wavelength.
δ_2 = quadratic misphasing on wavelengths.

The loss in gain due to the aberration δ_2 is derived on p. 177 of the quoted article* and is shown in Fig. 1. By eliminating δ_2 graphically the loss $L(\delta_2)$ can be expressed as a function of the scan angle.

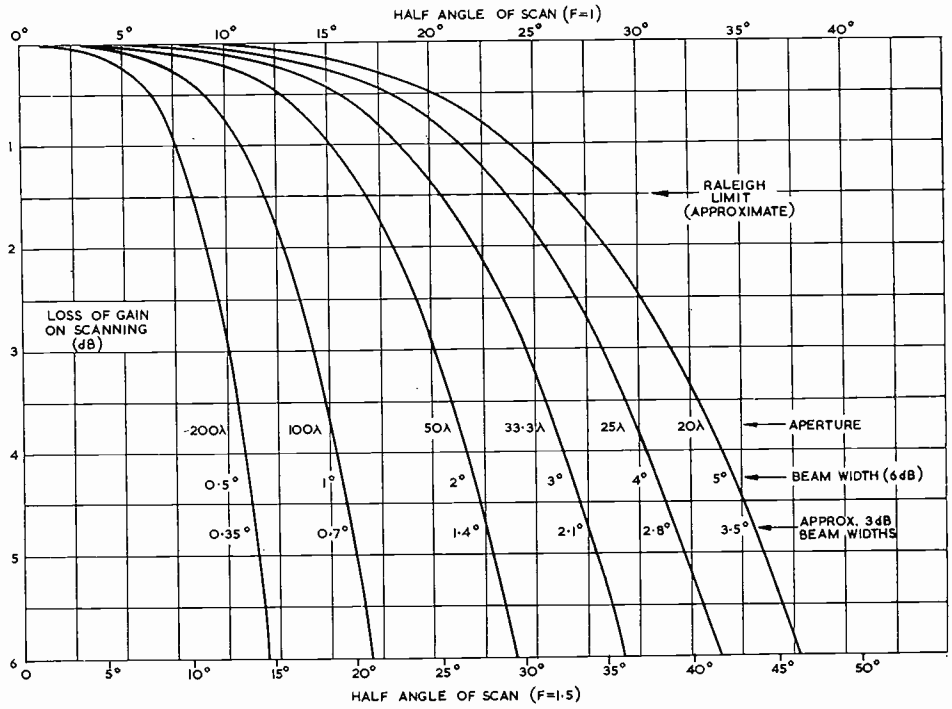


Fig. 2

Design curves available Scanning for various apertures or Beamwidths, with $F=1$ *or* $F=1{\cdot}5$.

It is, however, of practical convenience to express the relation between the loss and the scan angle in a form independent of aperture diameter, F-number and wavelength. If we therefore write

$$H.S.A.N. = H.S.A.\sqrt{\frac{D}{F\lambda}} = 4\sqrt{\delta_2}$$

we may plot $H.S.A.N.$ versus the loss $L(\delta_2)$. The scan angle is then that obtainable with unity F-number and a one wavelength aperture. Since this is only of academic interest a 100λ aperture is a preferred standard corresponding to an aerial having approximately a 1° beamwidth at the 6 *dB* level (0·7° at 3 *dB* level). The function

* loc. cit.

which is then plotted is $H.S.A.N./10$, being the normalised scan angle for an aperture of 100λ with an F-number of unity. This basic single curve is shown in Fig. 1. Also shown on the figure are experimentally determined points obtained from scanning measurements on a variety of mirrors and lenses particulars of which are given. In each case the scanning performance has been normalised to 100λ and $F = 1$.

It is evident from the figure that the scanning to the Rayleigh limit of 1·5 *dB* agrees with the simplified theory. Beyond that region there is a general tendency for the scanning to improve, especially in the case of mirrors. It is however somewhat surprising that the departure from theory is in general less than 10%, having regard to the complexity of the performance in the high loss region. It should be remarked of course that pattern deterioration is a feature of the high loss region. It is unwise to assume satisfactory patterns beyond the Rayleigh limit.

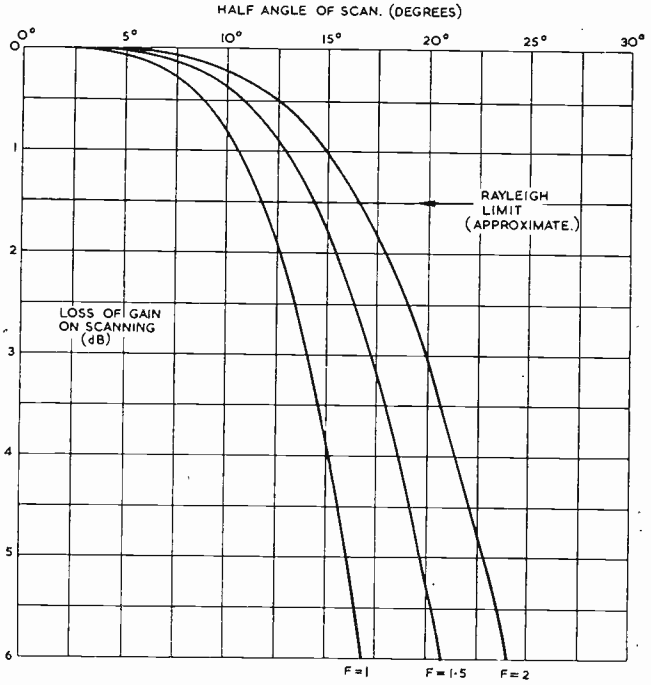


Fig. 3

Effect of changing the F number, f/D. (Standard Aerial 100λ, 1°).

The following rules apply to the use of the single universal scanning curve for determining the scanning of any given aerial assuming the latter is coma-corrected or nearly so:—

1. *Given* $F = 1$, *and an aperture* D/λ. (Fig. 2.)
   Multiply the scan angle for a given loss by
   $$\sqrt{\frac{100}{D/\lambda}}$$
   to obtain the scan angle obtained with an $F1$ aerial with aperture D/λ.
2. *Given* 100λ *aperture and* F-*number* $= F > 1$. (Fig. 3.)
   Multiply the scan angle for a given loss by
   $$\sqrt{F}$$
   to obtain the scan angle obtained with an aerial of aperture 100λ and any F-number preferably not much less than 1. (The performance cannot be guaranteed for small F-numbers.)
3. *Given the beamwidth B.W. with* $F = 1$. (Fig. 2.)
   Multiply the scan angle for a given loss by
   $$\sqrt{BW}$$
   where $BW = 6\ dB$ beamwidth
   $= 1{\cdot}4 \times 3\ dB$ beamwidth.

4. *Given the aperture* (D/λ) *or beamwidth,* (BW) *and F-number.*

Multiply the scan angle for a given loss by

$$\sqrt{F.\frac{100}{D/\lambda}} \quad \text{or} \quad \sqrt{F.\ BW}$$

again assuming a 6 *dB* beamwidth.

A convenient reference figure can be obtained from the last factor. For a coma-corrected lens or mirror (not offset), the total scan angle for 1·5 *dB* loss of gain on scanning of a 1° (100λ) aerial having $F = 1$ is 23°.

This figure thus gives the formulae

$$T.S.A. = 23^\circ \quad \sqrt{F.\ BW}$$

$$= 23^\circ \quad \sqrt{F.\frac{100}{D/\lambda}}$$

for the total scan angle to 1·5 *dB* points for any aerial of beamwidth BW, aperture D/λ and F-number $= F$.

In practice the beamwidth formulae

$$B.W. = \frac{95}{D/\lambda}$$

is probably more accurate but has been sacrificed in favour of the convenient figure of 100. With regard to beamwidth it should be remembered that narrow beams scan small angles and broad beams scan large angles.

The number of beamwidths scanned (frequently erroneously taken as a measure of scanning) is given by

$$N = \frac{T.S.A.}{BW} = 23^\circ \sqrt{\frac{F}{BW}}$$

for 1·5 *dB* loss in gain. Thus a large aerial having a narrow beam will scan more beamwidths than a small aerial with a broad beam although the larger aerial has a smaller angle of scan.

The important scanning constant is the normalised number of beamwidths scanned, i.e., the number of beamwidths scanned by a 1° aerial with $F = 1$, i.e.,

23 beamwidths

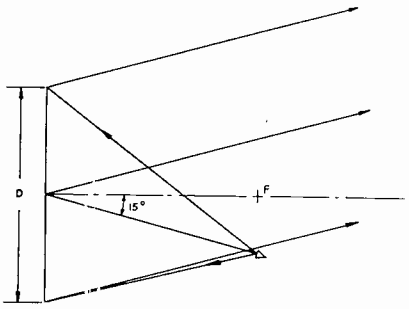
If the reader has measured any lens or mirror aerial of the kind described in this paper he should normalise the performance in the way indicated here and compare the normalised performance with the curve of Fig. 1; the scanning performance can then be assessed. At 1·5 *dB* loss the figure obtained will then agree or disagree with 23° per 1°, F 1.

How to Use the Curves for Offset Mirrors Not Specially Designed for Offsetting, i.e., Circularly Symmetric Mirrors

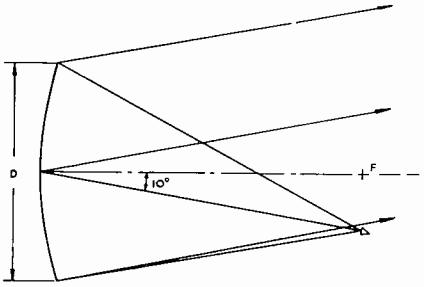
In the scanning considered so far the aerials have been taken to be coma-corrected lenses or mirrors having circular symmetry. Spherical mirrors have been included in the experimental data. Lenses in general tend to have more loss of gain on axis than mirrors, the difference usually amounting to about 2 *dB*. Other things

being equal, mirrors will be preferred. Unfortunately mirrors require offset feeding to avoid obscuration and reflection into the feed. The loss resulting from the offsetting then makes the absolute performance similar on scanning. If there is no scanning but only offsetting the mirror is still superior. If corrected when offset the performance is of course still about 2 *dB* better.

However, it is of interest to evaluate the scanning performance of a circularly symmetric mirror when used with an offset feed that just clears the emergent beam when simply offset without any other scanning.



(a) *Plane Collimating Surface* $F=1$ (*Plane Focal Service*).



(b) *Spherical Collimating Surface* $F=1{\cdot}5$ (*Plane Focal Surface*).

Fig. 4
Geometry of Offsetting.

In Fig. 4 is shown the geometry applying to this offsetting. The engineer's rule is to drop the feed by about one-quarter of the aperture. The exact amount of offsetting is a matter of allowable obscuration and reflection into the feed and of the particular geometry of the curvature of the aerial used and of the curvature of its field. Assuming as a practical measure that the objective or focusing element is flat and that the feed is dropped by $\frac{1}{4}$ the aperture to clear the beam the offset angle is

$$a = \tan^{-1}\frac{D}{4f} = \tan^{-1}\frac{1}{4F}$$

In practice we have taken the angles 10° for $F = 1{\cdot}5$ and 15° for $F = 1$ as being conveniently near the theoretical values.

With these offset angles reference to Fig. 1 shows that the loss in gain on offsetting is

$3{\cdot}9$ *dB* for $F = 1$

$0{\cdot}8$ *dB* for $F = 1{\cdot}5$

The heavy loss in gain for $F = 1$ is of course associated with a 100λ (1°) aperture and would be significantly less in the case of smaller apertures (broader beams).

Since the mirrors are circularly symmetrical the general scanning field is circularly symmetrical provided

(1) The primary illumination is circularly symmetrical.

(2) The performance is independent of the polarisation of the feed or mirror.

The first requirement is normally obtained by appropriate feed design. The second was unknown until a study was made of the effects of changing polarisation on scanning. After measurements it became clear that to the accuracy of the estimates of this paper it could be assumed that the scanning was substantially

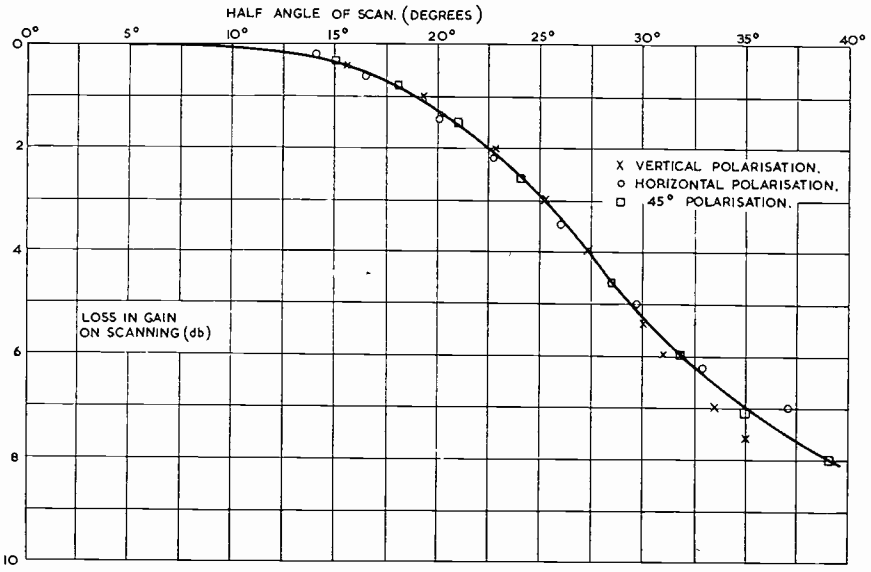


FIG. 5

Effect of Change of Polarisation on Scanning Performance.

independent of polarisation. This is illustrated in Fig. 5 which is typical of the scanning obtained with two orthogonal polarisations and one midway between them. Assuming circular symmetry of the scanning field it becomes possible to predict the available scanning of an offset fed mirror by reference to the standard curve of Fig. 1 which relates, of course, to scanning through the *focus*.

One of the assumptions made in this paper is that the feed locus of a scanning aerial is that focal surface which maximises the gain: it is the " field " of optics and in general is curved. We also assume that the direction containing the feed and the centre of the objective is the scanning direction in the case of lenses, and the image of the scanning direction in the case of mirrors.

The optimum locus for the lenses analysed in the quoted article* is given by

$$r = \frac{1}{2} f \, (1 + \cos^2\alpha)$$

where α is the scan angle. Experimentally we find that spherical mirrors have a spherical field of radius equal to focal length over the usable region while zoned coma-corrected mirrors have a flat field. As the spherical field is the most typical radio field (it suits modding reflectors) we shall restrict this illustration to this type without loss of generality. The three typical field curvatures are shown in Fig. 6.

* loc. cit.

For the sake of clarity a particular example of an offset feed aerial will be selected, viz., the standard aperture of 100λ with beamwidth *c*.1°, but with F = 1·5. If F = 1 were used, with an offset angle of 15°, the offsetting loss would be 3·9 *dB* (see Fig. 1) but with F = 1·5, i.e., 10° offset angle the loss drops to 0·8 *dB*. For smaller aerials *F* = 1 may be feasible but not for a 100λ aperture.

We assume then that the feed is offset by 10° with loss 0·8 *dB*. Next assume that a further 3 *dB* loss on scanning may be tolerated, so that referred to the gain

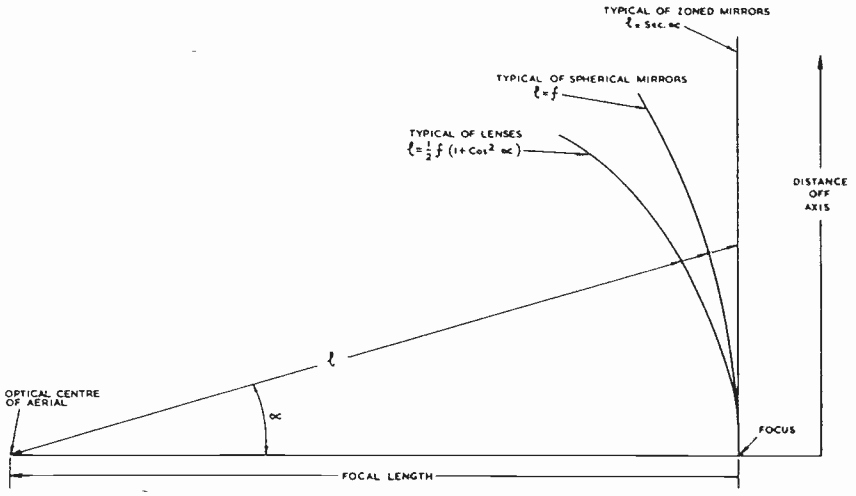


FIG. 6

Three Typical Scanning Loci.

with the feed on axis, the gain at the maximum scan angle is down by 3·8 *dB*. Three significant scanning possibilities arise:—

(1) *Scanning in the Plane of Offset.* (Fig. 7 (*a*))

In this case the feed is moved away from the standard offset position in the same plane of offset until the loss is the maximum allowable. In the example the feed would move another 5° to show 3·8 *dB* at 15° off axis (Fig. 1). Thus the scan available in this plane is 5°.

(2) *Scanning at right angles to the Plane of Offset.* (Fig. 7 (*b*))

In this case the circular symmetry of the field is used. A limit circle of maximum allowable loss may be drawn with centre the focus. Inside this circle the feed is offset in one direction to give the offset position and the offset loss. The feed is then moved at right angles to the plane of offset until it meets the limit circle. The latter feed movement gives the available scanning perpendicular to the plane of offset.

Since the feed moves on a sphere, as assumed above, the calculation of the available scan angle involves spherical trigonometry, in particular the right-angled spherical triangle formed by the offset arc on the sphere of radius equal to focal length, the arc of the scan perpendicular to the offset arc, and closing arc joining the focus to the limit circle. Given the offset

arc proportional to the offset angle A, and the closing arc proportional to the radial scan for maximum loss, C, we have to find the scan arc proportional to the available scan angle, B. This is given by

$$\cos C = \cos A \cos B$$

Note that if the angles are small, $\cos x \fallingdotseq 1 - \frac{1}{2}x^2$, and

$$C^2 = A^2 + B^2$$

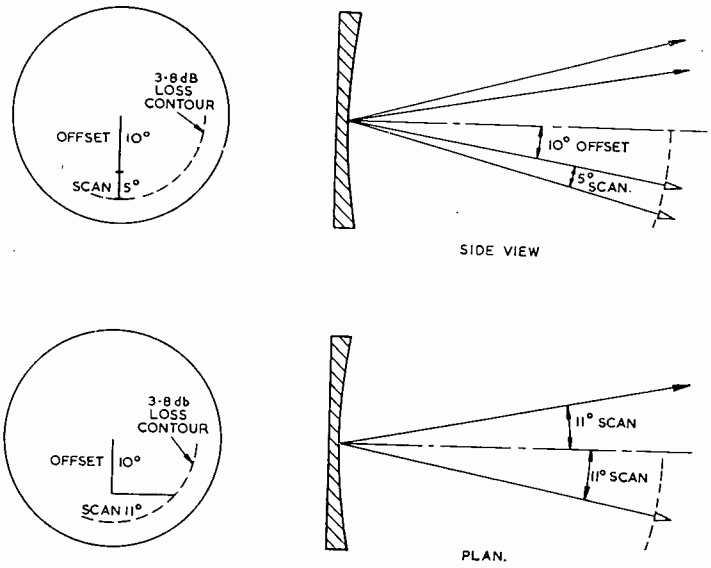


FIG. 7

Available line and volume scanning given 3dB maximum loss on Scanning.

and the scanning locus is a plane triangle with angles proportional to sides.

Applying the formula to the example gives an available total scan angle perpendicular to the plane of offset of 22°. This is markedly different from scanning in the plane of offset only, viz., 5°. It should be appreciated that *either* 5° in the plane of offset *or* 21° in the perpendicular plane are available. For a two dimensional scan both will be limited, especially the second.

This is shown below.

(3) *Volume Scan.* (Fig. 7 (c))

Assuming equal scan angles in the plane of offset and perpendicular to the plane of offset gives a " square " volume scan. For small angles the plane geometry applies and the single scan angle obtained is given by

$$B = \frac{1}{5}\left\{ 2\sqrt{\left(5C^2 - A^2\right)} - 4A \right\}$$

For the present example with $A = 10°$, $C = 15°$ we find

Total square scan $= 4{\cdot}8° \times 4{\cdot}8°$

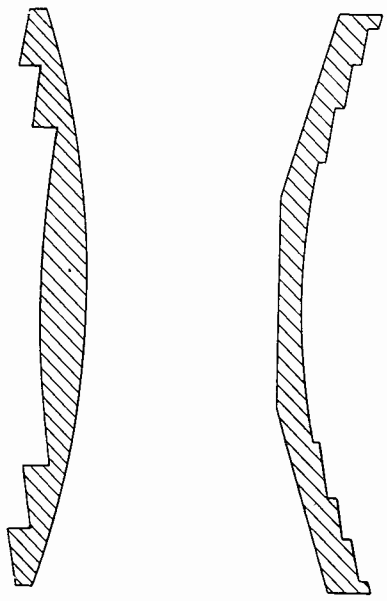


FIG. 8

Lens and mirror profiles (coma-corrected).

(a) Zoned dielectric lens. (Focus on left $F = 1{\cdot}5$*).*

(b) Zoned Paraboloidal Reflector. (Focus on right $F = 1$*).*

indicating how the scan in the plane of offset limits the scan seriously in the perpendicular plane. It can be concluded as would have been expected that scanning in the plane of offset is inefficient due to the offset requirement and the one-sided nature of this scan. The latter feature applies even if the mirror had been of the asymmetric type, fully corrected in the offset position with no loss in offsetting.

Fortunately not all mirror scanning aerials are circular and frequently have an aperture ratio. If the feed is offset in the plane of the shorter aperture dimension the scanning in the offset plane is eased and the full scanning in the perpendicular plane is realised. Such aerials can be regarded as having elliptical fields. If the aperture were elliptical, the scanning field would probably be a crossed ellipse.

Acknowledgments

Acknowledgment is made to the Board of Admiralty for permission to publish this work. The help of J. A. C. Jackson and J. Chaplin in conducting the measurements is greatly appreciated.

THE PHENOMENON OF ELECTRO-LUMINESCENCE AND ITS APPLICATION IN THE ELECTRONICS INDUSTRY

By D. W. G. Ballentyne, B.Sc.

Although the phenomenon of electroluminescence was discovered nearly twenty years ago it has not become of practical importance until recent years. Modern developments have moved away from a consideration of electroluminescence as a light source and it is the intention of the present article to review the possible uses of the effect in electronic devices. As a necessary preliminary to this discussion the article indicates the method of manufacture and the physical properties of this class of phosphors and outlines the theories which have been advanced to explain their behaviour.

It is concluded that although the brightness of panels is not, at present, great enough to allow them to be used as general lighting sources there is a possibility of using the effect in specialized ways for the display of information and in opto-electronic transducers. The effective use of these methods in many cases will, however, depend not so much on the development of electroluminescent techniques as on the production of photoconductors with very short decay times.

Introduction

PHOSPHORS have been used extensively in the electronic industry for television and radar displays for many years. A phosphor is a substance which absorbs quanta of energy from an exciting source and converts part of this energy into visible light. The primary excitant may either be a photon (e.g., ultra violet light) or a charged material particle (e.g., an electron, α-particle). The lighting industry is concerned with the manufacture of phosphors of the first type, photoluminescent phosphors, which have a strong absorption for one of the lines of the mercury arc spectrum, usually the 235·7 mμ line, whilst the electronic industry is interested in phosphors which are excited by an electron beam, i.e., cathodoluminescent phosphors. In general, phosphors of both these classes are inorganic crystalline solids consisting of a host crystal which, of course, must be transparent and which has incorporated in it by heat a small quantity of an activator.

As a generalization it may be said that the host crystal is responsible for the absorption of the exciting radiation whilst the activator is responsible for the emission of visible light. Fig. 1 shows schematically the transitions which are supposed to occur in a phosphor crystal of the class which exhibits photoconduction during excitation and emission. The filled band and conduction band refer to the electronic state of the host crystal whilst the luminescent centre L is produced by the activator ion. One quantum of exciting radiation $h\nu$ is absorbed and an electron is raised from the top of the filled band into the conduction band. The hole left in the filled band traps an electron from the luminescence centre and hence ionizes it. The free electron in the conduction band may either return directly to the luminescent centre when a quantum, of visible radiation $h\nu_1$, is emitted or it may be held in a trap such as T. These traps may be due to an ion or, more probably, to a dislocation or imperfection in the lattice. Until the electron requires sufficient thermal energy

to be returned from the trap into the conduction band it cannot contribute to the emission of the phosphor. It is these electrons which are responsible for the afterglow of a phosphor, that is, for the decay of emission with time when the exciting radiation has been removed. If the traps are all of one depth, the rate of decay, depending as it does on the probability of an electron in the trap acquiring an energy in excess of the trap depth, will be exponential. A number of trapping depths will lead to a decay law which has been loosely called bimolecular. This simplified model has a number of shortcomings and makes the explanation of stimulation and quenching difficult. When a phosphor which has been excited with say, ultra violet radiation and has then been allowed to decay, is irradiated with infra red, the light emission may either be stimulated or quenched but the radiation which produces these effects is of different wavelengths. Stimulation is easily explained on the band model by the assumption that electrons are ejected from the traps so that the rate of re-combination of electrons with ionized centres is increased and thus the light emission enhanced.

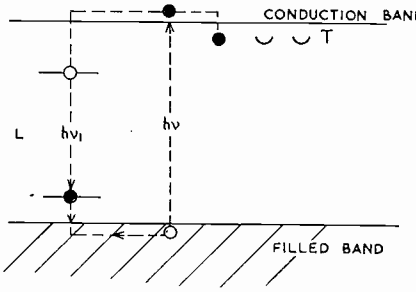


Fig. 1
The mechanism of photoluminescence.

The quenching of the emission is more difficult to understand[1]. It has been suggested that, in quenching, the radiation causes emptying of the traps close to the conduction band but at the same time raises electrons from the filled band into the ground state of the activator thus filling them by non-radiative processes.

It is intended to discuss in this article a relatively newly investigated type of luminescence produced by the application of an alternating field. As long ago as 1920 Gudden and Pohl[2] noticed that the decay of an ultra violet excited zinc sulphide phosphor activated with copper, ZnS [Cu], was affected by the application of a steady field and during the intervening years much work has been done on this phenomenon which has become known as electrophotoluminescence. In 1938 G. Destriau, however, noticed that sustained emission of light could be obtained from a previously unexcited phosphor by the application of an alternating field to the phosphor suspended in the dielectric of a condenser. This phenomenon is known as electroluminescence. This work was reported fully in a series of articles in 1947[3]. At the time, however, the efficiency of the phosphors was so low and the electroluminescent device so cumbersome as to be of no practical importance. In 1950, however, E. C. Payne *et al*[4] indicated how electroluminescence could be put to practical use and, since this time, an increasing volume of work has gone into the production of more efficient phosphors and their incorporation in panels for use as lighting sources. Unfortunately, however, at the voltages and frequencies available domestically the light output is poor and electroluminescent panels are only of use as large area low level light sources. The same limitations in voltage and frequency do not exist in electronic equipment where the brightness of the emission is only limited by the breakdown voltage of the dielectric in the electroluminescent capacitor and thus it may be that this phenomenon will be found to be more useful in flat displays or light amplifiers for television and radar.

Before discussing the possible utilization of this phenomenon it will be of interest to discuss the manufacture of phosphors and panels to indicate the physical properties

of the phosphors and to review qualitatively the theories which have been put forward to explain this phenomenon.

Electroluminescent Phosphors

A large number of different inorganic compounds have been used as the host crystals for luminescent materials but only one of these classes has been found to be efficient when excited by an alternating voltage. All electroluminescent phosphors used industrially are zinc or cadmium sulphides or selenides activated with copper. The original electroluminescent phosphors prepared by Destriau were mixed crystals of zinc oxide-zinc sulphide containing copper but as has already been pointed out, the light output of these phosphors was feeble.

Zinc sulphide can exist in two crystalline modifications, i.e., in a cubic and a hexagonal form, the transition temperature of the two forms being at 1024°C. The manufacture of a good green, orange and blue cubic zinc sulphide phosphor is described by H. H. Homer, R.M. Rulon and K. H. Butler[5], although in this case the system is a little more complex because of the inclusion of a trace of lead in the phosphor.

Thus a phosphor containing between 1×10^{-4} and 4×10^{-4} gms Cu/1 gm ZnS in the presence of a large quantity of chloride ion is green. The addition of 1×10^{-2} gm Mn/gm ZnS to the green phosphor mixture produces an orange phosphor. In each case 4×10^{-3} gms Pb/gm ZnS is added to the mixture before firing but the phosphor only contains between 10^{-5} and 10^{-6} gm Pb/gm ZnS due to the volatility of lead chloride. Firing of these phosphors is carried out either in a counter current of nitrogen or in a covered crucible. An improvement can be made in the efficiency of the phosphor by treating the fired product with warm acetic acid in order to remove zinc oxide which is found to be detrimental to its performance.

The preparation of hexagonal zinc sulphides of the same colours has been described by H. C. Froehlich[6] and by P. Zalm, G. Diemer and H. A. Klasens[7]. In both these cases an excess of copper above that which will go into solid solution in zinc sulphide was employed. According to Froehlich the optimum content of copper is of the order 1×10^{-3} gms/1 gm ZnS for a green phosphor. He showed that if the phosphor containing this quantity of copper was fired at 1200°C in a stream of dry hydrogen sulphide, the phosphor contained mainly green and red luminescent centres and was not electroluminescent. If, however, water vapour is added to the hydrogen sulphide in the proportion of one part in thirty-three the phosphor was a brilliant green and electroluminescent.

Zalm, Diemer and Klasens also reported the production of good green and blue electroluminescent hexagonal zinc sulphide phosphors when the phosphor composition was such that all the added copper could not be accommodated in the lattice. Thus they find that zinc sulphide activated by 10^{-3} gms Cu/gm ZnS and 8×10^{-4} gms Al/gm ZnS, fired at 1200°C in an atmosphere of H_2S, shows green electroluminescence. They state that the efficiency is effected by the rate of cooling and the atmosphere used. Thus the phosphor mentioned above is slowly cooled to room temperature in H_2S and the efficiency as an electroluminescent phosphor is improved. Under ultra violet excitation, however, the luminescence is poor indicating the presence of what they term quencher centres, usually known as " red centres ", in the lattice. If the phosphor is refired in air for a few minutes at 550°C most of these centres may be destroyed without seriously affecting the copper rich layers. This process again enhances the electroluminescence. Although it is agreed that such treatment does

enhance the emission we have found that in general reheating in air tends to produce phosphors which are much more susceptible to the effects of water vapour. These workers also report that when a non-electroluminescent zinc sulphide phosphor is treated with copper sulphate so that a layer of copper sulphide is formed on the surface it shows electroluminescence after drying. The efficiency is increased by firing in an atmosphere of H_2S. A similar result may be obtained by evaporating a layer of copper onto the surface. This will again result, of course, in the production of a film of copper sulphide in close contact with the ZnS [Cu] phosphor.

In our work we have been mainly concerned with hex ZnS. In general, these phosphors have a composition in the range

1 gm ZnS, (1×10^{-3} to 1×10^{-4} gm) Cu, (1×10^{-3} to 1×10^{-4} gm) *Al*. 1100°C

and they are fired in wet hydrogen sulphide, although this atmosphere may be replaced by other gentle oxidizing atmospheres, for instance carbon dioxide.

The essential condition that a sulphide of this type is electroluminescent appears to be that a large amount of copper is present in relation to the copper content of conventional phosphors. As is well known, the function of the aluminium in this type of phosphor is one of co-activation, that is, according to accepted theory, it aids the inclusion of the copper in the zinc sulphide lattice. It is significant that electroluminescence only appears when the ratio of copper to aluminium becomes greater than 1. (Fig. 2.) It is a corollary of this observation that electroluminescence seems to become possible only when there exists in the sulphide lattice a large number of the so-called red centres (Fig. 3). Thus, the absolute value of the copper content in these phosphors is not of the greatest importance but it appears to be necessary that sufficient copper be added to produce a second phase of a sulphide of copper in the zinc sulphide which, of course, itself contains some copper that gives rise to the luminescent centres, the source of the emitted light.

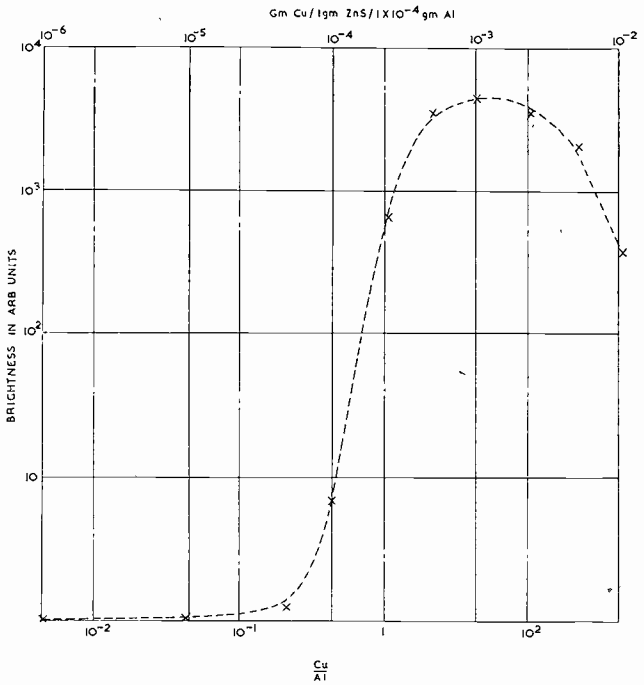


Fig. 2

Variation of the brightness of electroluminescence of a ZnS[Cu] (Al) phosphor with the molar ratio of copper to aluminium.

If the wet hydrogen sulphide is allowed to flow over the phosphor mix during firing the resultant phosphor is green whereas firing a sample of identical composition in an atmosphere of stagnant hydrogen sulphide converts the green phosphor to a blue one. The efficiency of the blue phosphor may be increased if a halogen acid is added

to the atmosphere. This addition is most easily achieved by the incorporation of an ammonium halide in the phosphor powder or by placing it in another container in the firing tube. The effect of this halide inclusion does not depend upon the ammonia or hydrogen formed by dissociation but appears to be a property of the halogen atom. The most likely effect of such an addition, therefore, would appear to be to enrich the atmosphere with sulphur vapours. This view is supported by the fact that an identical effect is observed if the ammonium halide is replaced by iodine but in this case the luminescent efficiency of the resultant phosphor is low.

Froehlich pointed out that an orange electroluminescent phosphor could be prepared by replacing the aluminium in part by manganese and firing under the same conditions as above. It has been shown in our work that phosphors ranging from green to pale yellow in colour may be prepared by making a phosphor of the general composition:—

$$1 \text{ gm } [x\% \text{ ZnS } y\% \text{ CdS}],\ 5\times10^{-3} \text{ gm Cu},\ 1\times10^{-3} \text{ Al}$$

where x varies from 100—40% and y from 0—60%. The essential condition for the preparation of phosphors of this type being the co-precipitation of the zinc sulphide-cadmium sulphide mixture.

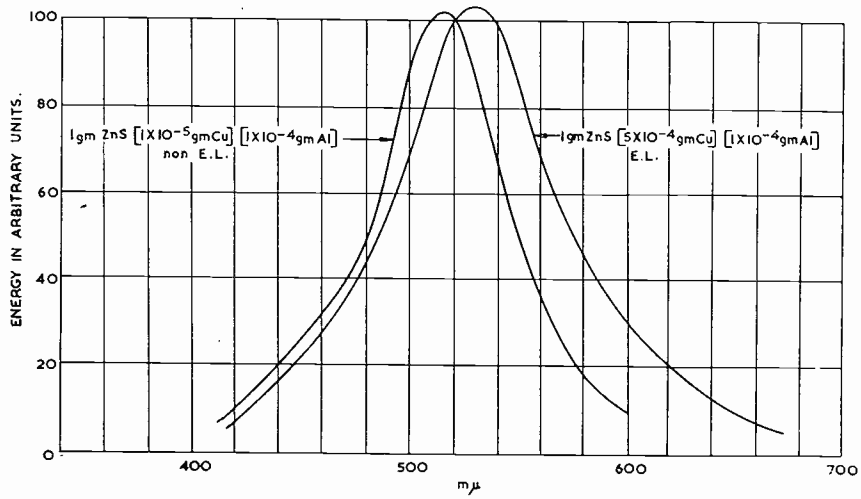


Fig. 3

Spectral distribution of a photoluminescent and an electroluminescent ZnS [Cu] (Al) phosphor.

The following table summarizes the composition and conditions of preparation of the different electroluminescent phosphors.

Phosphor No.	Composition	Firing	Colour
M.166	1 gm ZnS 5×10^{-3} gms Cu $1{\cdot}5\times10^{-3}$ gms Al	1100°C wet H_2S	Green
M.182	1 gm ZnS $1{\cdot}5\times10^{-3}$ gms Cu 7×10^{-3} gms Mn	,, ,, ,,	Orange
M.222	0·8 gms ZnS 0·2 gms CdS 5×10^{-4} gms Cu 1×10^{-3} gms Al	,, ,, ,,	Yellow
M.291	1 gm ZnS 5×10^{-3} gms Cu $1{\cdot}5\times10^{-3}$ gms Al	1100°C H_2S+NH_4I Stagnant.	Blue
M.346	1 gm M.166 1 gm M.182 2 gms M.291	—	White

It is obvious that an additive white may be obtained by mixing together a blue, green and orange phosphor. A better rendering of white could be obtained if an efficient red phosphor could be prepared but at the moment the red phosphors, usually containing zinc selenide, are not efficient.

Preparation of Panels

The electroluminescence effect was first demonstrated by Destriau[3] in a cell shown in Fig. 4 (*a*). A suspension of the phosphor in castor oil placed upon a copper block was covered with a mica sheet. The upper surface of the mica sheet was wetted

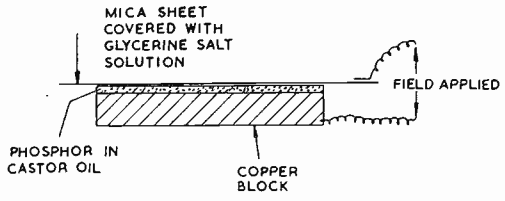
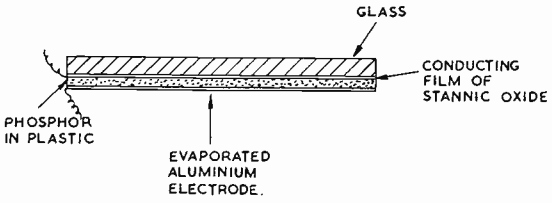


Fig. 4 (a)
Destriau's electroluminescent cell.



(b)
A conventional electroluminescent lamp.

with glycerined salt solution and the field was applied between this electrode and the copper. This device could not be of practical importance.

The practical light source generally used is shown in Fig. 4 (*b*). The phosphor suspended in a dielectric is applied to a sheet of conducting glass, another opaque conducting layer being applied to the back of this film. Conducting glass is easily prepared by the application of stannic chloride vapour to the glass sheet, heated to a temperature just below the softening point of the glass when atoms of tin diffuse into the surface and form a semi-conducting layer. The conducting back electrode

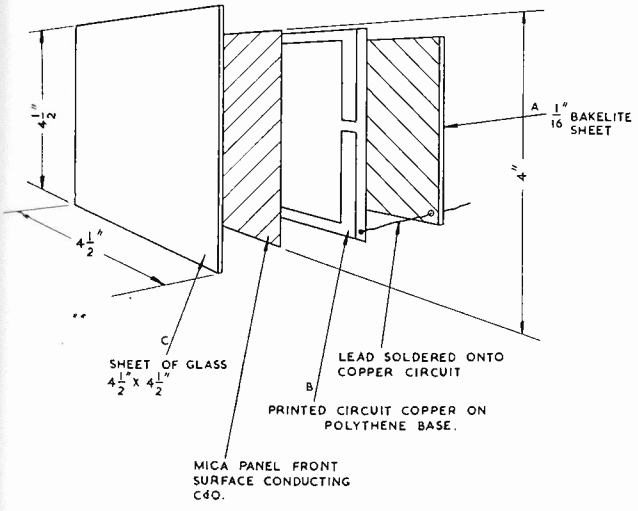
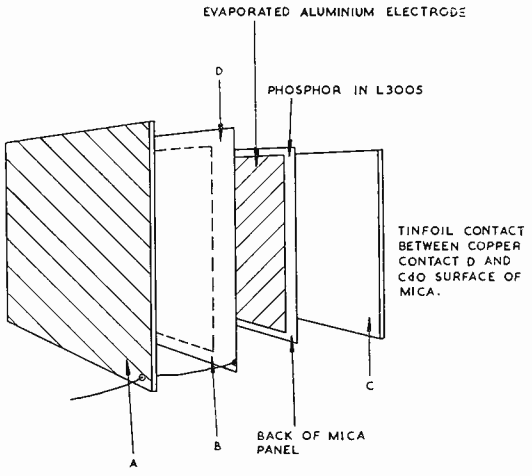


Fig. 5 (a)



(b)

A Mica electroluminescent panel.

can either be an evaporated film of aluminium, or a sprayed layer of silver or graphite. The last two layers have the disadvantage that they absorb part of the light whilst, of course, an aluminium layer reflects the light and produces, therefore, an apparent increase in brightness.

The continued life of such panels depends upon the dielectric strength of the plastic layer and although such panels are easily prepared for use at low voltages[9] we feel that a more robust panel may be necessary for continued operation at higher voltages. The design of the panel used by us is shown in Fig. 5 (*a*) and (*b*) in an exploded view. The phosphor suspended in a silicone resin (L.3005) is applied to one side of a mica sheet, and an aluminium electrode is evaporated onto this film after it has been cured. A conducting layer is then applied to the other side of the sheet. A number of substances have been used for this layer. The best film for the purpose appears to be a sputtered layer of cadmium oxide. This layer is susceptible to finger-marking and should be protected by a film of silicone lacquer. Contact is made to both the front and back electrodes by means of a printed circuit on a flexible base. The whole cell is clamped together and beeswax is poured between the bakelite backing sheet A and the glass C. Panels constructed in this fashion can be excited by voltages of up to 1,500 but at 230 volts and 50 c/s are of the same order of efficiency as commercially available panels.

Properties of Electroluminescent Phosphors

In this section we shall consider the effect of voltage frequency, temperature and time on the brightness and efficiency of electroluminescent phosphors. The observations will apply to sinusoidal alternating fields.

It was shown by Destriau that the application of an alternating voltage to an electroluminescent panel produced a light output whose amplitude varied with time. The exact nature of the variation depends to some extent upon the phosphor. Thus Fig. 6 shows the variation of the amplitude of the brightness with time for various voltages for a zinc sulphide phosphor activated with copper. In this case it can be seen that there are two primary brightness maxima and two so-called secondary maxima for each cycle of voltage. It can be seen that the two main peaks are of unequal amplitude and two explanations have been advanced for this discrepancy. According to A. N. Ince[10] the two waves viewed from each side of a symmetrical cell with two transparent electrodes were in phase, that is the primary and secondary peaks are emitted simultaneously from each side of the cell.

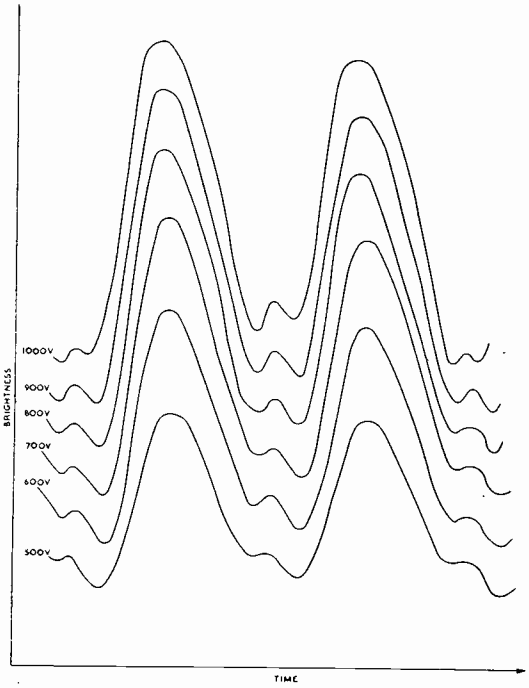


Fig. 6

Brightness waves for a green ZnS [Cu] (Al) phosphor excited at various voltages.

He showed that for a 6% second harmonic content the difference in peak height was 40% whereas for a 3% harmonic content the difference was negligible. According to Zalm, Deimer and Klasens, however, the peak with the larger amplitude occurs when the cathode is nearer the photomultiplier. They consider that the inequality in peak height may be explained by the hypothesis that each crystal only emits light on the side facing the cathode; self absorption and scattering of light by the inter-adjacent layer or by individual particles are then responsible for the asymmetry.

These two explanations are, of course, not mutually exclusive and there is no doubt that an asymmetric voltage waveform accentuates the peak difference but that absorption also plays an important part. Thus we have noticed that the difference in the peaks is greatly accentuated when the aluminium backing electrode, which is reflecting, is replaced by sprayed metal of much greater absorption. It is concluded, therefore, that the observed peak difference is not a real property of the phosphor panel but is dependent upon its construction and the purity of the applied voltage waveform.

The position of the secondary peak depends to a large extent on the temperature and on the frequency. Thus J. Mattler[11] has shown that for a green ZnS[Cu] phosphor the secondary peak moves from the descending side of the brightness wave to the ascending side of the wave as the temperature increases, though it is not clear from his results whether the minimum at the end of the secondary peak still corresponds to a voltage node. It has also been shown that, as the frequency increases, the secondary peak usually observed with a ZnS[Cu], ZnS[Ag] phosphor tends to disappear. The brightness wave for a ZnS[Cu] [Mn] phosphor is of a much simpler shape and in this case there appears to be no secondary peak. An explanation of this observation suggests itself. As has already been explained above the zinc sulphide phosphors may be divided into two classes. (*a*) Those which exhibit photo-conduction during emission and in which luminescence requires the movement of electrons in the conduction band and (*b*) those phosphors in which the transitions responsible for luminescence occur in the electron shells of the activator atom. Electroluminescent phosphors of the first class exhibit a secondary peak whilst phosophors of the second class do not. It becomes at once apparent that the secondary peak must be due to the release of electrons trapped at a great distance from the ionized luminescence centres. As we have pointed out elsewhere, it is interesting to note that phosphors which give a secondary peak appear, at low frequencies, to give it between the nodes of the voltage applied to the cell and the current passing through it and thus there appears to be some correlation between the electric power waveform and the brightness wave.

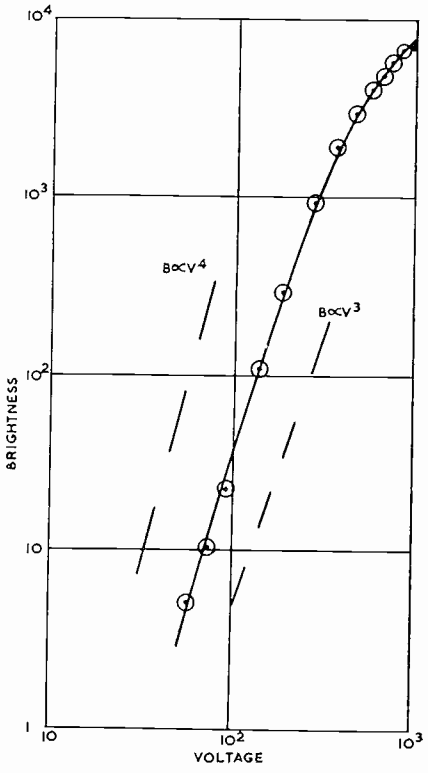


Fig. 7

The variation of the mean brightness of an electroluminescent panel with voltage.

For all practical purposes the variation of brightness with mean applied voltage can be expressed by a relation

$$B = a\,V^{n} \tag{1}$$

where n lies between 3 and 4 (Fig. 7). It is obvious that this relation is only true at low voltages and, in fact, the variation of voltage with brightness more exactly follows a law of the form

$$B = a\,V^{n} \exp\,(-b/V) \tag{2}$$

where a, b and n are constants and n has a value between 1 and 3 depending upon the phosphor but which is generally equal to approximately 2.

The marked dependence of brightness on voltage is of great importance for, as the voltage applied to a panel need not be limited in an electronic device, a great improvement in the brightness of the panels can easily be obtained. The ultimate limit to the applied voltage will be set, as pointed out above, by the strength of the dielectric layer.

According to Destriau the exact nature of the dependence of brightness on frequency varies from phosphor to phosphor. In general, at low frequencies and for singly activated phosphors most workers agree that the variation is linear. At higher frequencies some saturation occurs and even in some extreme cases a maximum is observed. Multiple activated phosphors behave in a different fashion, for the intensity of light due to the emission band of each activator may vary differently with frequency and thus the colour of the phosphor may change as the frequency increases. Hence a ZnS [Cu][Mn] phosphor is orange at low frequencies due to the manganese emission and green at high frequencies when the colour is determined by the copper emission. This dual dependence on frequency leads, in general, to a variation of brightness with frequency which is not linear[12].

The temperature dependence of electroluminescence is complicated. The first results to be published were by Destriau but they were inconclusive as they referred only to one phosphor and to a small temperature range[3]. A great number of experiments have been made by J. Mattler[13] on powder cells. He realized that, in order to obtain significant results the properties of the dielectric should remain effectively constant over the temperature range investigated and he showed a satisfactory approximation to this condition could be made with a mica panel. With ZnS[Cu], ZnS[Ag] and ZnS[Cu][Pb] phosphors the brightness increases, passes through a maximum, and then decreases as the temperation increases from -100°C. to $+120$°C. With other phosphors, however, the brightness appeared to decrease over this range. The thermoluminescence curves for the phosphors are not reported so that it is difficult to correlate this maximum to any specific physical condition such as the release of electrons from a trapping level. G. F. Alfrey and J. B. Taylor[14] have shown that, for a single crystal of zinc sulphide, the emission is substantially constant at low temperatures and at high temperatures but increases rapidly with intermediate values of temperature.

P. D. Johnson, W. W. Piper and F. E. Williams[15] have made a series of measurements on powder phosphor panels, i.e., ZnS [Cu] (Al) phosphors suspended in a dielectric of perspex and also on single crystals of ZnS [Cu] grown in vacuo. The thermoluminescence curves were obtained at a heating rate of 5°C./min. whilst the brightness temperature waves for electroluminescence were determined statically, the sample being held at a fixed temperature for a few minutes before the measurement of the brightness was made. The glow curves obtained for the phosphor powder before incorporation in the dielectric matrix are the same as those obtained after it has been treated in this way. The phosphor possesses two main trapping states in

the powdered form but only the traps corresponding to the major peak at 123°K are present in the single crystal. The electroluminescent response in each case has a well defined minimum at approximately this temperature. Increasing the voltage of excitation only shifts the thermoelectroluminescence curve upwards whereas an increase of the frequency of excitation moves the minimum towards higher temperatures. According to these workers electroluminescence is a field dominated phenomenon. Large variations in brightness are to be expected if trapped electrons can, in certain temperature ranges, alter significantly the formation of the high field zone.

The efficiency of an electroluminescent panel is not affected appreciably by increasing frequency. At low frequencies there is a slight increase in efficiency whilst at higher frequencies the efficiency decreases slightly probably due to the impedance of the conducting glass film. The variation of electroluminescence efficiency with voltage is more difficult to understand. At low voltages the efficiency increases as a third power of the voltage but it reaches a maximum at voltages of the order of 500 volts/mil and then decreases. This effect is most noticeable at low frequencies.

Many attempts have been made to devise an equivalent circuit which will explain the various properties of these phosphors. Some of these circuits[7] are most complicated. Probably the simplest circuit which will represent the behaviour of an electroluminescent panel is a resistor in series with a capacitor which has another resistor in parallel with it. In a previous paper[16] it has been demonstrated that if the parallel resistor is assumed to be voltage dependent, a relationship between brightness and voltage of approximately the correct form can be deduced providing the assumption is made that the brightness is proportional to the power available for dissipation. A calculation of the efficiency to be expected under these conditions leads to the result that at low frequencies the efficiency should approach unity and it should decrease as the voltage increases. The reason for the experimentally observed initial increase in efficiency is not clear.

Theories of Electroluminescence

A number of theories have been advanced by various workers. The methods used to prepare efficient electroluminescent zinc sulphides suggest that the phosphor particles are not homogeneous. Chemically the best electroluminescent phosphors appear to consist of two phases, one of which may either be a sulphide of copper, manganese or lead or an oxide, e.g. ZnO in the case of ZnS phosphors. It has been postulated that light is emitted from crystals only on the side on which the field is becoming negative. K. H. Butler and J. F. Waymouth[17] consider that the inhomogeneity in the crystal may lead to the whole field on the crystal being concentrated across a barrier and they believe, therefore, that two mechanisms can be advanced to explain electroluminescence. (1) Electrons from donor centres are freed either thermally or by the field and enter the conduction band. There they are accelerated by the field until they acquire sufficient energy to ionize the activator centres. (2) Electrons in the bound states of the activator centres may be raised into the conduction band by high fields by the process of wave mechanical tunnelling.

In general, modern theories tend to the view that collision rather than direct excitation is responsible for the ionization of the centres for W. W. Piper and F. E. Williams[18] estimate that for direct field excitation of the centres in ZnS [Cu] a field of 2×10^7 V/cm is required and this field is greatly in excess of the breakdown voltage of ZnS.

D. Curie[19] proposed a mechanism which depended upon a three stage process. The primary action of the field is to transfer electrons from the donor levels into the conduction band. This process may, of course, be aided by thermal excitation of these electrons. The applied field will then accelerate the electrons in the conduction band until they acquire sufficient kinetic energy either to ionize the luminescent centres or to cause the formation of electron-hole pairs in the host crystal lattice. Curie did not consider the transfer of electrons from donors to the conduction band as he did not regard it as the rate determining step. From a consideration of the last two stages he arrives at the experimental variation of brightness with voltage. On his model it is required that the electron gains an energy in excess of $h\nu$ by acceleration over a certain path length. Such a mechanism requires that the mean free path of the electron be long so that it can acquire sufficient kinetic energy during acceleration. If the electroluminescence depends upon a long path length, it would be expected that the light output would be affected by the application of an intense magnetic field. A. N. Ince[20] showed, however, that a field of 130,000 oersteds does not affect electroluminescence. It would appear, therefore, that this mechanism cannot be correct.

Piper and Williams[18] made a much more ambitious attempt to suggest a theoretical mechanism for this effect. They consider, as has already been discussed above, that direct field ionization is not possible and advance two other mechanisms.

I. An appropriate injected minority carrier could be captured by the activator system and hence ionize it. Either a *p–n* junction or a suitable surface state is necessary for this mechanism. Piper and Williams consider, however, that this mechanism would lead to a threshold voltage of only 10^2V/cm. Also, as the appropriate minority carrier would seem to be a hole, light would be expected to appear at that electrode which was becoming positive instead of, as observed by other workers, at the cathode. Also Piper and Williams maintain that there is no evidence of either a *p–n* junction or a suitable surface state in ZnS[Cu].

II. As an alternative they consider that the impact excitation or ionization mechanism is in accord with electroluminescence and related properties of ZnS[Cu]. They state that traps consisting of electron levels 0·5 e.v. below the conduction band could be field excited to the conduction band at reasonable field strengths and that they would then have some chance of acceleration to kinetic energies sufficient for impact excitation or ionization of activators. These traps are deep enough not to be thermally emptied during the voltage cycle. It is agreed that the breakdown field for pure zinc sulphide is greater than 10^5V/cm and less than 10^6V/cm, the breakdown field for the alkali halides. It is also agreed that the probability of an electron attaining sufficient kinetic energy at less than 1/10th of the breakdown field to ionize the centre is poor. The average field in the bulk of the crystal has been shown by Roberts[21] to be less than 10^4V/cm. Thus, for the above mechanism to be correct, a local concentration of field is necessary. Some sort of barrier is immediately suggested. The idea of a chemical barrier is discarded and a physical barrier of the Mott-Schottky exhaustion type is adopted. For simplicity the theory is applied to a single crystal with plane parallel sides on which electrodes are evaporated. The area of the electrodes is large in comparison to the thickness so that edge effects may be neglected. Donors next to the surface will ionize until a sufficiently wide space charge has been created to equalize the Fermi levels in the phosphor and in the metal. Now, if the crystal lies between two electrodes of different potential, electrons will move across the crystal and out at the anode easily but the cathode barrier will prevent the ingress of electrons. The exhaustion layer will broaden until the whole

potential difference is across it. As the potential is increased the thickness of the field next to the cathode is so small that it can be penetrated by electrons from the cathode. A direct current will flow through the crystal above this voltage but below it electrons must originate in the crystal. The alternating potential applied to the crystal will, in turn, deplete and replenish the exhaustion barrier at the electrode.

After the first few cycles the electron traps will behave like donors and if these donor levels are situated at different depths the deeper ones will be field ionized and the conduction electrons from sufficiently deep donors will be in strong enough fields to be accelerated to sufficiently high kinetic energies. The relatively low efficiency of electroluminescence is explicable when it is remembered that although all the conduction electrons absorb energy from the field only those from deep donors which experience relatively few collisions with lattice vibrations contribute to the impact excitation of the donors.

A similar mechanism has been considered by a number of workers. Thus Howard, whose derivation differs in detail from Piper and Williams, arrives at an expression

$$B = a\, V^{\frac{3}{2}} \left(1 - \frac{3}{C}\, V^{\frac{1}{2}} \right) e^{-\frac{C}{V^{\frac{1}{2}}}}$$

where a and C are constants.

This equation is in good agreement with experiment over five decades of brightness for a phosphor suspended in a dielectric. In this calculation, no account is taken of the sinusoidal form of the voltage. Because of the rapid variation of electroluminescence with voltage most of the emission occurs at the peak of the wave where the variation in V is small so that calculations based on the peak voltage or the R.M.S. voltage are sufficiently accurate for practical purposes.

Alfrey and Taylor[14] have also considered the mechanism of electroluminescence. They again take the view that to account for the appearance of electroluminescence at low fields it is necessary to postulate the existence of high local fields in surface barriers. They admit that so far as is known the barrier in zinc sulphide can either be of a physical or chemical type, but they point out that the electrons which are responsible for the impact excitation of centres can arise either by field excitation or by thermal ejection. They have shown, however, that thermal ejection of electrons from donor levels leads to incorrect results.

More recently theories of electroluminescence have tended to move away from a consideration of the physical barrier as its source to a consideration of the chemical barrier. The objections originally advanced to an explanation of this nature are no longer valid as it does not appear that the threshold voltage of electroluminescence is as high as was at one time believed as it has been shown that the apparent threshold is dependent to a large extent on the sensitivity of the light detectors and that, in fact, light emission occurs at very low voltages. From what has been said above it becomes apparent that electroluminescence of, for instance, ZnS[Cu] phosphors depends upon the separation of a discrete secondary phase of a sulphide of copper so that the phosphor as a whole might be represented as ZnS[Cu] ¦ Cu_xS (leaving in abeyance the question of the exact nature of the copper sulphide). Again the original assertion that light emission occurs at the electrode which is becoming cathodic is questionable. According to Zalm, microscopic examination of an emitting single crystal shows spots of light emitted in the bulk of the crystals. There appears to be no reason to suppose that electroluminescence cannot be explained by the transport of a charge at a p–n junction.

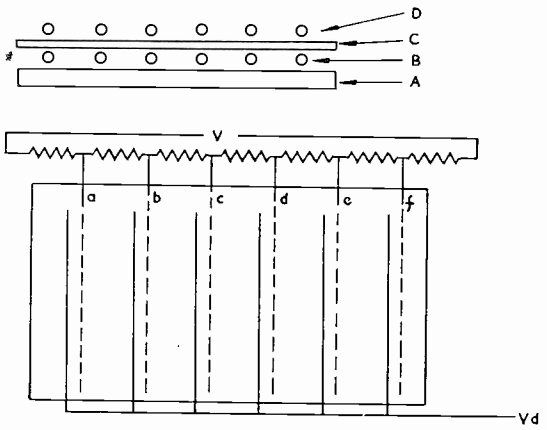
Uses of Electroluminescence

Destriau's original conception of a capacitative electroluminescent light source has not been advanced significantly since the announcement of a practical panel in 1950[4]. In general, until the brightness of the panels is increased by an order of magnitude when excited at 230V and 50 c/s they will not constitute a serious threat to the more conventional types of light source. There are, however, a number of uses to which panels have been put which depend upon two unique properties of these lights—their thinness and the fact that no heat is dissipated during operation. Hence electroluminescence is suitable for low level lighting where space is restricted, for lighting in silhouette and for luminous signs and dials.

A specialized and important application of this phenomenon is in photographic processing. As is well known, only non-actinic lights can be tolerated in the dark room and, conventionally, such lighting is provided by the light transmitted by a suitable filter placed in front of a light box. Because white light must not be allowed to escape from the box, cooling presents a difficult problem and inefficient cooling results in a rapid deterioration of the filter. An electroluminescent panel is a considerable improvement on the conventional light box.

Illuminated signs in cinemas and theatres are another application of this type of illumination for exit signs, indicating obstructions, for instance the risers of a staircase, and in general for lighting the gangways.

Electroluminescence offers an improved method of lighting instrument dials as the lettering on the dial itself is illuminated and thus distortion and errors in reading caused by parallex are reduced to a minimum. Uses of this kind are increasing in the aircraft industry where a 500 c/s supply is available in the aircraft with its intendent increase in brightness of the panel. A comprehensive review of these uses is given by Bowtell and Bates[9].



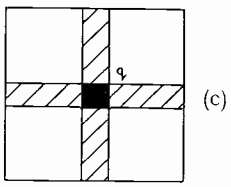


Fig. 8

A black spot electroluminescent display. The set of wires B in 8(a) *have been displaced in* 8(b).

The most obvious use for a light source which can be made to emit by the application of a field is for the display of information. Display of information can be accomplished in two ways, either by applying a voltage sufficient to make the phosphor luminesce at a point (x, y) or to apply this voltage to the whole panel and to remove it at the point (x, y) by the application of a biassing voltage.

The first method, which requires that the electrodes of the panel shall consist of two sets of orthogonal strips, is in essence the easier way but the problems of commutation of the strips is formidable. Mechanical commutation, although

possible, is obviously so cumbersome as to be an unattractive solution of the problem. A line of investigation worthy of study would appear to be the use of a scanned cathodoluminescent phosphor in conjunction with a set of photoconductor elements, one element actuating each strip in the grid. Such a method waits, however, on the preparation of a photoconductor with a rapid decay as the persistence of the image on the display screen will be determined by the rate of decay of the photoconductor or if this is very fast, by the decay time of the cathodoluminescent phosphor. Such a decay is essential when using electroluminescence to display information as electroluminescent phosphors themselves have an indetectable persistence.

A device of the other type has been described[22]. Figs. 8(*a*) and 8(*b*) show one half of such a device. A is a printed circuit on which the conductors B are laid down; C is a piece of mica on which the phosphor has been sprayed whilst D is another set of transparent conductors evaporated onto the other surface of the mica sheet. The voltage at *a*, *b*, *c*, etc. is determined by the resistors and varies from V_a to V_f, whilst the voltage applied to the continuous electrode can be set between V_a and V_f but is always 180° out of phase with the voltage on the other strips. In the diagram it has been set at V_d. The voltage at *d* is zero whilst a residual voltage appears on each of the other strips and light will be emitted from all strips except *d*. If it is supposed that another such cell is placed in contact with the first cell but with the conductors orthogonally disposed to those on the first cell, a black spot will appear at *q* Fig. 8(*c*). Experiment has shown that such a device will only become feasible with phosphors of greatly increased efficiency which will give an appreciable amount of light at low fields otherwise the resolution of the device becomes poor. An increase in resolution is limited by the dielectric strength of the material, for a voltage approaching $V \times f$ is applied to strip *a* when strip *f* is extinguished and for a device with a reasonable number of strips using phosphors available to-day such a voltage is prohibited by the dielectric strength of the panel.

The consideration of an electroluminescent panel in series with a photoconductor leads to many new devices of great importance.

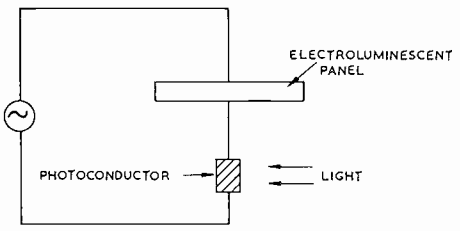


Fig. 9
An elementary light amplifier.

Suppose that a source of light is focused on to the photoconductor in Fig. 9; the impedance of the circuit will decrease, current will flow and light will be emitted. The intensity of the emitted light will be independent of the intensity of the exciting light. The construction of a light amplifier now becomes possible. A number of such devices have been described and amplifications of between 5 and 24 have been claimed. In essence such devices are simple to construct and a light amplifying panel such as that shown in Fig. 10 should produce an image of a projected picture. The amplifier consists of a sheet of conducting glass A onto which a phosphor is applied, layer B, suspended in a resin. An opaque layer, C, is applied to the baked resin and then a thick layer

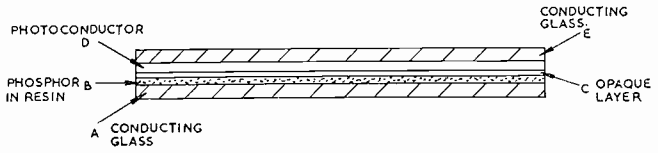


Fig. 10
A light amplifier.

(20 mils) of a photoconducting cadmium sulphide, D, is laid down. Finally a transparent backing electrode, E, usually another sheet of conducting glass, is applied to the back of the device. Light incident on the photoconductor will reduce its resistance and therefore the voltage applied to the phosphor layer will be determined by the intensity of the light. Under these conditions a picture projected onto the photoconductor will be reproduced on the electroluminescent panel and, as the brightness of the emitted light is independent of the exciting light and is only dependent on the exciting voltage, light amplification may be achieved. In fact, of course, the resolution of such a simple device is poor but many more complicated systems have been devised to overcome this defect and some of them are reviewed by B. Kazan and F. H. Nicholl[23]. A promising field of application of the light amplifier is as an image converter to convert ultra violet or infra red radiation into visible radiation.

E. E. Loebner[24] has discussed the use of electroluminescent panels in series with crystalline cadmium sulphide photoconductors in various devices to which the generic name opto-electronic transducers has been given. The physical dimensions of these devices are small. They consist of an electroluminescent panel ($\frac{1}{4}$ in. $\times \frac{1}{4}$ in.) in contact with a cadmium sulphide crystal to which electrical contact is made by a cats whisker. There are many uses for devices of this type—for instance, in the construction of storage cells and shift registers for computors and so on. Fig. 11 shows diagrammatically an elementary storage cell. The information is supplied in the form of a pulse by connecting O to A. This causes panel P_1 to light for a short period of time. Panel P_1 illuminates detector D_1 and causes its resistance to decrease causing panel P_2 to light. The illumination from P_2 reduces the resistance of D_2 and thus, although the resistance of D_1 returns to its static value, panel P_2 remains on. Thus the information is stored. If B is connected to O panel P_4 becomes illuminated causing the resistance of D_3 to fall. The system is so arranged that a fall in the resistance of D_3 or D_4 on its own is not sufficient to light the panel P_3. Thus if no information is stored in P_2 the connection of O to B will not produce a light pulse at P_3 but if P_2 is on, P_3 will light up and a voltage will appear across CD. Simultaneously P_5 will light up, reducing the resistance of D_5 to a small value and shorting the voltage source V_2. Panel P_2 is now extinguished and the whole system returns to an unlit condition ready to receive further information.

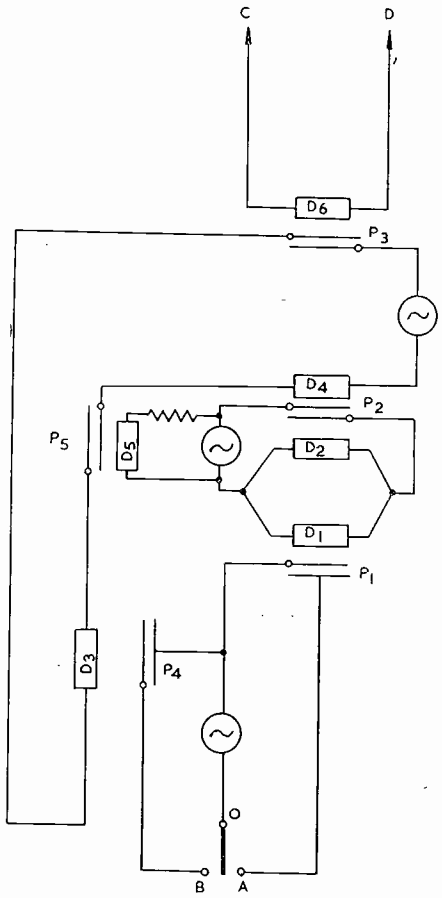


Fig. 11

A storage device using opto-electronic transducers.

From a consideration of this elementary circuit the endless possibilities of devices of this type are apparent. For most uses, however, it is necessary that the decay time of the detector elements should be very rapid and at present this cannot be achieved practically.

References

(1) G. F. J. Garlick. Luminescent Materials, Oxford 1949, pp. 21 and 166.
(2) B. Gudden and R. W. Pohl, Zeits f. Physik. 1920, 192.
(3) G. Destriau, Phil. Mag. 38, 1947, 700, 774, 800.
(4) E. C. Payne, E. L. Mayer and C. W. Jerome, Illum. Engng. 45, 1950, 668 Sylvania Technologist, Jan. 1951, 2.
(5) H. H. Homer, R. M. Rulon and K. H. Butler, J. Electrochem. Soc. 100, p.556, 1953.
(6) H. C. Froehlich, J. Electrochem. Soc. 100, 1953, 496.
(7) P. Zalm, G. Diemer and H. A. Klasens, Philips Res. Rep. 9. 1954, 81.
(8) H. C. Froehlich, J. Opt. Soc. Amer. 43, 1953, 320.
(9) J. N. Bowtell and M. C. Bate, Illum. Engng. 7. 1955, 3.
(10) A. N. Ince, Brit. J. App, Phys. Supp. 4, 1955, 54.
(11) J. Mattler, Luminescence Colloquium, Paris, May 1956, paper 23.
(12) G. Destriau and H. F. Ivey, Proc. I.R.E., 12, 1955, 1911.
(13) J. Mattler, Comptes Rendus 239, 1954, 1116.
(14) G. F. Alfrey and J. B. Taylor, Proc. Phys. Soc. 68, 1955, 5.
(15) P. D. Johnson, W. W. Piper and F. E. Williams, J. Electrochem. Soc. 103, 1956, 221.
(16) D. W. G. Ballentyne, Luminescence Colloquium, Paris, May 1956, paper 28.
(17) K. H. Butler and J. F. Waymouth, Brit. J. App. Phys. Supp. 4, 1955, 33.
(18) W. W. Piper and F. E. Williams, Brit. J. App. Phys. Supp. 4, 1955, 39.
(19) D. Curie, Journal de Physique et Le Radium, 14, 1953, 510.
(20) A. N. Ince, Proc. Phys. Soc. Vol. 67, 1954, 870.
(21) S. Roberts, J. Opt. Soc. Amer. 42, 1952, 850.
(22) L. E. Q. Walker and R. J. Kemp, Brit. Patent No. 704, 166.
(23) B. Kazan and F. H. Nichol. Proc. I.R.E. 43, 12, 1955, 1888.
(24) F. E. Loebner, Proc. I.R.E. 43, 12, 1955, 1897.

PULSE SHAPING TO A GIVEN MONOTONIC DISCHARGE FUNCTION

MRS. K. G. FANCOURT, B.A., and J. K. SKWIRZYNSKI, B.Sc., A.R.C.S.

The problem discussed below arose during the study of the characteristics of the non-magnetic radio-frequency mass spectrometer. The method of realization of a required monotonic discharge function proved very satisfactory in this particular case and it is hoped it will be of use in similar problems.

Introduction

A RESISTANCE-CAPACITANCE network is used to produce a given discharge function having a steep initial slope, the input being a square wave function. A typical input (continuous line) and output (broken line) are shown in Fig. 1. In the particular problem discussed the voltage transfer ratio was given as an algebraic function, but a graphical function is equally suitable for the purpose of realization. A sum of exponentials is used to obtain an approximation in the time domain, and not in the frequency domain as in most standard methods. As equal weight need not be given to all points on the curve the degree of approximation over different parts can be controlled as desired and the steep initial slope can be realized as accurately as the flat part of the curve. The degree of accuracy depends on the number of exponential terms used, so this also is easily controlled. The approximating function is easily transformed to the network transfer function and realized as an RC network.

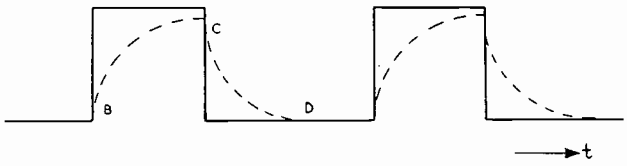


Fig. 1

The input and output waveforms.

The Problem

During the study of the characteristics of the non-magnetic radio-frequency mass spectrometer it was decided to sweep the frequency over a wide range according to a predetermined function, in order to provide a linear mass scale.

An oscillator was built such that its frequency was proportional to the applied sweep voltage:

$$f \propto V_s$$

The RF potential was applied to the spectrometer analysers giving an output such that

$$M \propto \frac{1}{f^2}$$

where M=mass number of ion. The output was displayed on an oscilloscope. To obtain a linear mass scale, i.e., $M \propto t$, it is necessary that

$$\frac{1}{f^2} \propto t$$

Therefore,

$$\frac{1}{V_s^2} \propto t$$

therefore

$$V_s = \sqrt{\frac{k}{t}}$$

Now the range of V_s was such as to give a frequency range of 7 Mc/s down to 511 kc/s and the time of sweep was 0·02011 secs. Normalize the voltages by dividing by 511 kc/s. Then at $t = 0{\cdot}02011$ secs., $V_s = 1$, and hence $k = 0{\cdot}02011$.

$$V_s = \sqrt{\frac{0{\cdot}02011}{t}} \qquad (1)$$

At $t = 0$ this function goes to infinity. Now the actual voltage falls from $\frac{7000}{511} = 13{\cdot}7$ as its highest value. Substituting this value in (1) gives $t_0 = 0{\cdot}000107$ secs.; i.e., V_s should start from $t_0 = 0{\cdot}000107$ secs. and not from $t=0$; but t_0 is small enough ($\frac{1}{2}\%$ of a scan) to be neglected, and the function will be assumed to start at $t = 0$ with the value $V_s = 13{\cdot}7$.

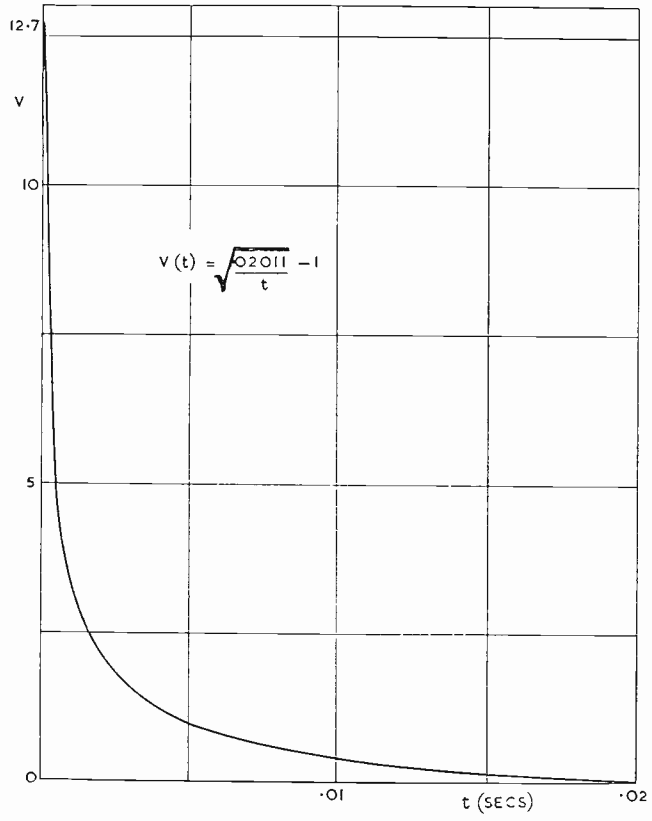


Fig. 2

The required discharge function.

The method given below applies to a function which approaches zero asymptotically as $t \to \infty$, and is effectively already zero at the end of the scan; but the function (1) is unity at this point. Therefore the function

$$V(t) = \sqrt{\frac{0{\cdot}02011}{t}} - 1 \qquad (2)$$

adjusted so that $V(t) = 12{\cdot}7$ at $t = 0$, was realized (see Fig. 2). The additional unit voltage was obtained by means of a constant DC bias.

Approximation to a Given Monotonic Discharge Function

The method consists of finding a sum of terms of the form $Ae^{-\alpha t}$ to give a suitable approximation to the required function. The curve is built up from the larger values of time t by choosing terms with successively greater α's, until the last term completes the realization of the steep slope near $t = 0$.

The first step is the selection of a point $t = t_1$ on the curve $V(t)$, at which a function $A_1e^{-\alpha_1 t}$ is fitted by equating values and slopes. $A_1e^{-\alpha_1 t}$ must give a suitable approximation to that part of the curve where $t \geqslant t_1$, since the later terms, having greater α's, will have negligible effect here. If the approximation is not good enough a greater t_1 must be chosen (but this cannot proceed indefinitely; there will be an optimum point beyond which the approximation will get worse). If the approximation is better than necessary a smaller t_1 may be chosen, which will mean that either a better approximation may be made to the steep part of the curve, or fewer terms may be used in the approximation. The point finally selected depends, of course, on the requirements of the particular case, i.e. whether actual or fractional deviation from the required function is considered the more important, and which parts of the curve require most accurate realization. In the problem considered here the actual deviation was used as a criterion, and it was endeavoured to keep this deviation constant over the whole range.

The second exponential, $A_2e^{-\alpha_2 t}$, is fitted to the function $V(t) - A_1e^{-\alpha_1 t}$, by value and slope, at a point $t = t_2 < t_1$, and this process is continued for points successively nearer $t = 0$.

In the particular problem considered here the final, nth term, was equated to the value of $V(t) - A_1e^{-\alpha_1 t} - \ldots\ldots - A_{n-1}e^{-\alpha_{n-1} t}$ at $t = 0$ and at a point near $t = 0$, since the slope of $V(t)$ was not determined in this region. Provided it was small the particular value of t chosen had a negligible effect on the result. This method may be useful in other problems, or it may be more convenient to fit to value and slope at one point $t = t_n$ as in finding the first $(n-1)$ terms, where t_n is either zero or very small. In fact the method used will probably be immaterial.

The number of terms employed depends on the shape of the function and the degree of accuracy required. It can only be found for a particular case by trial and error. Two solutions of our problem are given, one employing two and one three terms; the second, of course, gives the closer approximation, but both are satisfactory.

Realization of the Function

The approximation function is in the form

$$V(t) = \sum_{q=1}^{n} A_q e^{-\alpha_q t} \tag{3}$$

This defines a decay curve (CD in Fig. 1). The corresponding pulse build-up (BC in Fig. 1) is

$$G(t) = \sum_{q=1}^{n} A_q - \sum_{q=1}^{n} A_q e^{-\alpha_q t} \tag{4}$$

The network transfer function is obtained from this by means of a Laplace transformation:

$$H(p) = \sum_{q=1}^{n} A_q - \sum_{q=1}^{n} \frac{A_q p}{p + \alpha_q} \tag{5}$$

This may be expressed in the form

$$H(p) = \text{constant} \times \frac{p^{n-1} + a_{n-2}\, p^{n-2} + \ldots\ldots + a_1 p + a_0}{p^n + b_{n-1}\, p^{n-1} + b_{n-2}\, p^{n-2} + \ldots\ldots + b_1 p + b_0} \tag{6}$$

where the a's and b's are functions of the A's and α's. It will be noticed that the coefficient of p^n in the numerator vanishes identically, since it equals

$$\sum_{q=1}^{n} A_q - A_1 - A_2 \ldots\ldots - A_n$$

Now consider the " chain-matrix " equation of a two-terminal pair network:

$$\begin{Vmatrix} E_1 \\ I_1 \end{Vmatrix} = \begin{Vmatrix} \mathcal{A} & \mathcal{B} \\ \mathcal{C} & \mathcal{D} \end{Vmatrix} \begin{Vmatrix} E_2 \\ -I_2 \end{Vmatrix} \tag{7}$$

When the output termination is an open circuit, $I_2 = 0$, and

$$H(p) = \frac{E_2}{E_1} = \frac{1}{\mathcal{A}} \tag{8}$$

The matrix of the network shown in Fig. 3 is

$$\begin{Vmatrix} 1 & r \\ 0 & 1 \end{Vmatrix} \begin{Vmatrix} 1 & 0 \\ pC_1 & 1 \end{Vmatrix} \begin{Vmatrix} 1 & 0 \\ \dfrac{pC_2}{pR_2C_2 + 1} & 1 \end{Vmatrix} \cdots\cdots \begin{Vmatrix} 1 & 0 \\ \dfrac{pC_n}{pR_nC_n + 1} & 1 \end{Vmatrix} \begin{Vmatrix} 1 & 0 \\ \dfrac{1}{R} & 1 \end{Vmatrix}$$

$$= \begin{Vmatrix} 1 + r\left(pC_1 + \displaystyle\sum_{q=2}^{n} \frac{pC_q}{pR_qC_q + 1} + \frac{1}{R}\right) & r \\ pC_1 + \displaystyle\sum_{q=2}^{n} \frac{pC_q}{pR_qC_q + 1} + \frac{1}{R} & 1 \end{Vmatrix} \tag{9}$$

i.e., comparing with (7)

$$\mathcal{A} = 1 + \frac{r}{R} + prC_1 + r \sum_{q=2}^{n} \frac{pC_q}{pR_qC_q + 1} \tag{10}$$

This leads to

$$\frac{1}{\mathcal{A}} = \text{constant} \times \frac{p^{n-1} + c_{n-2}\,p^{n-2} + \ldots\ldots + c_1\,p + c_0}{p^n + d_{n-1}\,p^{n-1} + d_{n-2}\,p^{n-2} + \ldots\ldots + d_1\,p + d_0} \qquad (11)$$

where the c's and d's are functions of r and the R's and C's. This is the same equation as (6). Hence $H(p)$ may be realized as the RC network of Fig. 3 by equating the coefficients $c_q = a_q$, $d_q = b_q$.

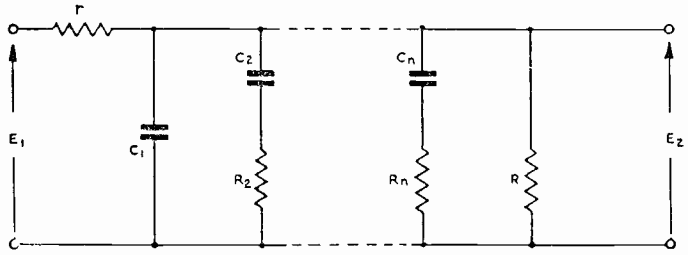


Fig. 3

The RC network used for realization.

Realization of the given function, using two terms

The required function is given by (2) and shown in Fig. 2.

In the realization using two terms, $t_1 = 0{\cdot}003$ is chosen as the first fitting point. Then $A_1 = 3{\cdot}6$, $\alpha_1 = 270$, to give the required value and slope at this point. The deviation from the required function, for $t > t_1$, is given effectively in Fig. 4, which shows the difference between the complete approximation function and the required function. A higher value of t_1 would give a closer approximation to the part of the curve where $t > t_1$, but would necessitate the use of more than one further term to obtain an equally good approximation to the steep initial slope. As can be seen in Fig. 4 the two term function obtained by choosing $t_1 = 0{\cdot}003$ gives an approximation equally good for the steep initial slope and for the flat part of the curve. The final approximation function depends largely on the value chosen for t_1. Once this point and hence the first term, are determined, the second term is fixed within fairly narrow limits. The same is true for the three term approximation.

The second term is fitted by value at $t = 0$ and $0{\cdot}0005$, giving $A_2 = 9{\cdot}1$, $\alpha_2 = 2900$. The approximating function is thus

$$F_2(t) = 3{\cdot}6\,e^{-270t} + 9{\cdot}1\,e^{-2900t} \qquad (12)$$

Comparing with (5) and (6) the network transfer function is

$$H(p) = 12{\cdot}7 - \frac{3{\cdot}6\,p}{p + 270} - \frac{9{\cdot}1\,p}{p + 2900}$$

$$= \text{constant} \times \frac{p + 363{\cdot}43}{p^2 + 3170\,p + 783000} \qquad (13)$$

From (10) the transfer function of the network of Fig. 3, for $n = 2$, is given by

$$\frac{1}{H(p)} = 1 + \frac{r}{R} + p\,r\,C_1 + \frac{p\,r\,C_2}{p\,R_2\,C_2 + 1} \qquad (14)$$

This is readily converted to the form of (6) and (11) and by equating the coefficients with those of (13) the following results are obtained:

$$\left.\begin{aligned} &\text{Voltage ratio } K = \frac{R}{R+r} \\ &C_1 = 0{\cdot}464.10^{-3}\,\frac{1}{Kr} \text{ farads} \\ &C_2 = 0{\cdot}833.10^{-3}\,\frac{1}{Kr} \text{ farads} \\ &R_2 = 3{\cdot}30\,Kr \text{ ohms.} \end{aligned}\right\} \qquad (15)$$

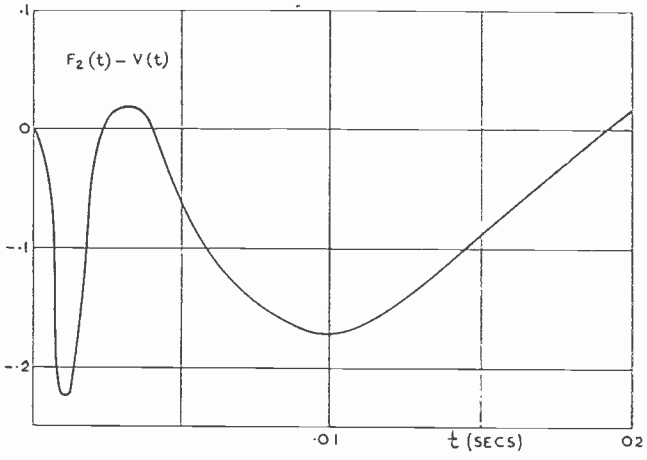


Fig. 4
Difference between two term approximation and required function

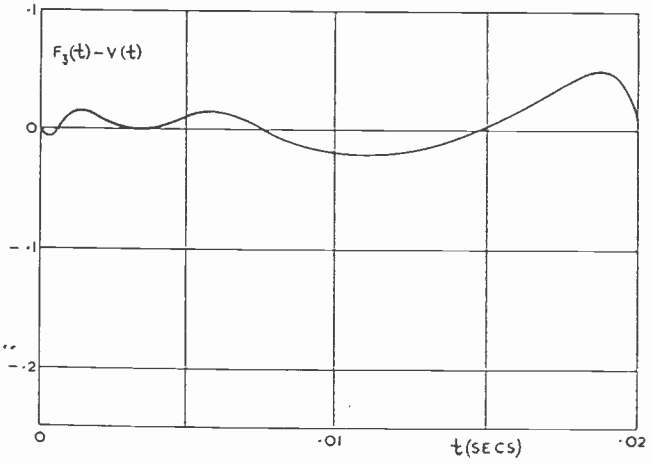


Fig. 5
Difference between three term approximation and required function.

Realization of the given function, using three terms

By using three terms a much closer approximation is obtained over the whole range, as is seen by comparing Figs. 4 and 5.

The first fitting point is $t_1 = 0{\cdot}006$. This gives the best possible approximation to the flat part of the curve. This is shown by that part of the curve of Fig. 5 for which $t > 0{\cdot}006$, where it can be seen that the error is distributed equally above and below the time axis. A smaller t_1, e.g., $t_1 = 0{\cdot}003$, used in the last section, gives a greater error below the axis, as can be seen in Fig. 4. As t_1 is progressively increased the error curve moves upwards over the axis, to the optimum position at $t_1 = 0{\cdot}006$. If t_1 is increased beyond 0·006 the error above the axis at the end of the range increases rapidly. Thus, since we seek to fit the approximating function to the given curve equally well along its whole length our value of $t_1 = 0{\cdot}006$ is the best, or near the best, that can be chosen. The addition of two further terms gives an equally good approximation to the steep initial slope (see Fig. 5) and so there would be no point in employing more than three terms.

Having chosen t_1, the other terms follow almost automatically. The second fitting point is $t_2 = 0{\cdot}002$, and the final term is fitted by value at $t = 0$ and 0·0005. Then the approximating function is

$$F_3(t) = 2{\cdot}50\ e^{-184t} + 3{\cdot}79\ e^{-1080t} + 6{\cdot}41\ e^{-4030t} \qquad (16)$$

Equation (16) is realized as a network by the methods indicated above. The network constants are

$$\left.\begin{aligned} &\text{Voltage ratio } K = \frac{R}{R+r} \\ &C_1 = 0{\cdot}418.10^{-3}\,\frac{1}{Kr} \text{ farads} \\ &C_2 = 0{\cdot}203.10^{-3}\,\frac{1}{Kr} \text{ farads} \\ &C_3 = 0{\cdot}850.10^{-3}\,\frac{1}{Kr} \text{ farads} \\ &R_2 = 3{\cdot}29\ Kr \text{ ohms} \\ &R_3 = 5{\cdot}25\ Kr \text{ ohms} \end{aligned}\right\} \qquad (17)$$

Conclusion

In the particular problem considered a good approximation is obtained by the use of three terms, and one satisfactory for most practical applications by the use of two terms. In either case the approximation is realized by an RC network, which is quickly obtained by the solution of simple algebraic equations. An increase in the number of terms would greatly increase the complexity of the algebra, but two or three terms should give a satisfactory approximation to most monotonic decreasing functions. Then a certain amount of labour is needed to find an approximating function, as this can only be done by trial and error, but the final realization of this function is quick and simple.

Acknowledgment

The authors wish to thank Mr. C. J. Bennett for proposing the problem and for the information about its practical significance.

BOOK REVIEWS

Commercial Broadcasting in the British West Indies. Produced by Central Rediffusion Services Ltd. Butterworths Scientific Publications. Price 5/-.

This book describes in detail the development of Commercial Broadcasting in Barbados, British Guiana, Jamaica, Trinidad and Tobago since agreements were signed with Central Rediffusion Services, Ltd., mainly in the years following the Second World War. The low standards of living and the limited Colonial Government funds available presented serious economic problems in the equipping and running of a broadcasting service. These problems were largely overcome by the finances which became available as the result of the commercial nature of the programmes. Based on the conservative figure of five listeners for each receiver or loudspeaker the total listenership has been built up from 156,500 in 1947 to 795,000 in 1954 out of a total population of 2,875,000.

Considerable sums of money have been spent by Central Rediffusion Services for the capital equipment of studios, transmitters and wired cable routes. Coverage has been obtained by wired, rediffusion, V.H.F., F.M., medium and short-wave transmitters, the latter being used in conjunction with vertical incidence shortwave arrays to give overall blanket coverage.

Complete programme schedules are given in the appendices and these show that commercial interests have not been allowed to interfere with the overall quality and balance of the programmes. Local talent and events are used whenever possible but B.B.C., N. American and other transcripted material forms a large part of the programmes.

In communities where illiteracy is rife broadcasting assumes great importance since in many cases it is the only means of large scale communication. It has, therefore, proved invaluable, for instance in giving warning of the hurricanes which are prevalent in the Caribbean.

Wireless Servicing Manual by W. T. Cocking, M.I.E.E. Ninth Edition—Iliffe and Sons Ltd. Price 17/6.

The appearance of a ninth edition of this deservedly popular " Manual " reminds one of the debt that the listening public owes to the Service Engineer amateur and professional. In the twenty years since the book was launched we have seen a steady trend towards uniformity of design of domestic receivers, with a simultaneous decline in the number and frequency of dramatic technical developments such as we were accustomed to in the thirties.

The introduction of a V.H.F. broadcast service, however, and the current interest in " high fidelity " reproduction have brought renewed interest to a subject that appeared to be on the way to technical stagnation, and in this context the revised " Manual " should maintain its popular appeal.

The author treats servicing problems in a completely general manner, the emphasis being on principles; unlike some other servicing handbooks, no specific reference to current receivers is to be found in the book, and it is largely to this fact that its timeless nature and its reappearance in edition after edition (like the Ford car) can be attributed.

The broad layout of the book is unaltered in the new Edition, but format and type are changed for the better, and nomenclature has been brought up to date. Chapters on V.H.F. receivers, Push-pull Amplifiers and Negative Feedback have been introduced, while the chapter on Television Receivers (a subject deserving of a book to itself) has rightly been dropped.

The book is commendably practical in outlook and, except in one or two instances, free from serious error.

Radio and Electronic Components. Sir Isaac Pitman and Sons, Ltd., London.
Vol. I. Fixed Resistors; G. W. A. Dummer, 28/–
Vol. II. Variable Resistors; G. W. A. Dummer, 30/–

The British Components Industry is well served by the publication, for the first time in this country, of a series of volumes under the general title of Radio and Electronic Components. The two volumes now available admirably fulfil the objectives of the series in assisting the user to choose the appropriate type of component for his purpose and to understand the fundamental characteristics of the various forms.

The author, who has been associated for some years with the development of components in the Ministry of Supply, is to be congratulated on the thoroughness with which he has undertaken the preparation of these volumes. Volume I, Fixed Resistors, not only contains all the information usually available in catalogues but also a wide range of technical data including physical properties of materials, high frequency performance and specifications; manufacture is described, though in no great detail, with the aid of block schematic diagrams; non-linear resistors receive adequate mention; a chapter devoted to experimental types provides some indication of the efforts being made to exploit new techniques in manufacture, the end products frequently having attractive properties over a relatively narrow range but making them suited to specific applications. Not the least valuable features of this volume are an 18-page bibliography sub-divided under 14 headings which must be one of the most comprehensive yet published, and a 24-page " chart " in which mechanical, electrical and climatic characteristics of representative types are presented, which facilitates comparison or selection for a particular application. It is thought that Fig. 3.5 representing an instrument for measuring high-value resistors is over simplified and could with advantage be more fully described in the text, whilst the horizontal scale in Fig. 4.1 requires explanation or correction. Throughout Chapter 4 the term " wattage rating " is used where " power rating " would be more appropriate.

A few pages and tables appearing in Vol. I are included for convenience and with advantage in Vol. II, Variable Resistors; here the ground is less familiar and the degree of precision attainable with the high grade wire wound types described may come as a surprise to those not accustomed to using them. The general pattern is similar to that of Vol. I—properties of wire, particularly contact wear and noise; measurement, with stress on linearity; curve matching, or reproducing a desired law by combination of fixed and variable resistors; types, and methods of manufacture including prevision toroidal winding machines; sine—cosine potentiometers; experimental forms exploiting unusual manufacturing techniques; again an excellent though shorter bibliography and to conclude, a 24-page comparison chart.

Following the chart and liable to be overlooked is a suggested " standard " terminology which, in the absence of any generally recognized terms and definitions, could well appear in a more prominent position, and be mentioned in the index. In Fig. 7.4(a) the section of the former, described as rectangular, is not so depicted, whilst in Fig. 8.10 the upper input step (not stop as in text opposite) should be labelled 25° not 20°. These criticisms as well as those above are of relatively minor importance for the volumes, each of about 180 pages, are well produced and arranged, and the figures, diagrams and curves are excellent. Further volumes in this series are in preparation and are awaited with interest.

THE MARCONI REVIEW

VOLUME XVII

JANUARY 1954 — DECEMBER 1954

NOS. 112—115

INDEX

VOLUME XVII

NOS. 112 — 115