WIRELESS ENGINEER

Vol. XXVI SEPTEMBER 1949 No. 312

The Intrinsic Impedance of Space

IN the *Revue Générale de l'Électricité* of March (p. 120), there is a strange account of a note by E. Brylinski presented to the Academie des Sciences by Louis de Broglie. According to this note one obtains two different values for the intrinsic impedance of space or of any medium, depending on the system of units employed. If expressed in the unrationalized m.k.s. system the value is stated to be $Z_0 = 30\Omega$, but if expressed in the rationalized system the value becomes $Z_0 = 477\Omega$. The author of the note says that this duality is a source of confusion and cannot be accepted; he then gives a very peculiar calculation in support of the 30 Ω value. Although the value 477Ω occurs several times throughout the note we feel sure that if MM. de Broglie and Brylinski were to get together and go into the matter more thoroughly they would agree that $4\pi \times 30$ is 377 and not 477. As they rightly say, the value of the impedance cannot really depend on the system of units employed, and whether one bases the intrinsic impedance on the ratio of voltage to current or on the ratio $\mathcal{E}/H$ one must obtain the same result if the units are properly employed.

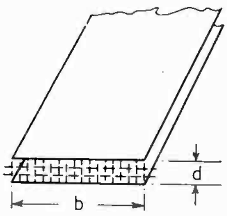


Fig. 1.

Probably the simplest way of approaching the problem is to consider a long line consisting of two parallel flat strips of considerable width b and separated by a small distance d* (Fig. 1).

Neglecting fringing, the inductance per metre length is $(4\pi/10^7)d/b$ henry, and the capacitance per metre length $(b/4\pi d)/(9 \times 10^9)$ farad. Then for a long line

$$Z_0 = \frac{V}{I} = \sqrt{\frac{L}{C}} = 120\pi \times \frac{d}{b} \text{ ohms.}$$

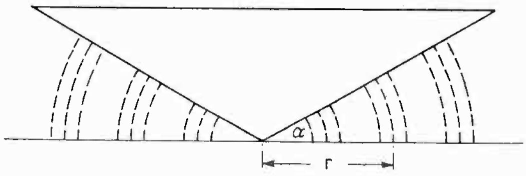


Fig. 2.

Putting $\mathcal{E} = V/d$ volts per metre and $H = I/b$ amperes per metre we have $\mathcal{E}/H = 120\pi$ ohms for the intrinsic impedance of the medium. In this example both $\mathcal{E}$ and H are uniform fields; it is immaterial whether they are expressed per metre or per cm. The intrinsic impedance is the impedance of a column of cross-section 1×1, whatever the units. It is important to note, however, that if $\mathcal{E}$ is in volts per cm and H in oersteds the latter must be converted into amperes per cm if the quotient is to be in ohms, and we then have $4\pi/10 \times \mathcal{E}/H = 120\pi$ ohms.

An interesting application of this intrinsic impedance of space is the calculation of the radiation resistance of the inverted cone transmission line which simulates the radiation from an aerial.†. We showed that the angle α in Fig. 2 is about 35°, depending somewhat on the type of aerial which is to be simulated. In radians α is about 0.6. At a distance r from the transmitter $V = \mathcal{E} \times \alpha r$ and $I = H \times 2\pi r$ because the length d of the path of the electric field is αr and the

* *Electrical Review*, September 26, 1913.

† Loc cit and *Wireless Engineer*, July 1944, p. 305.

length b of the path of the magnetic field is $2\pi r$. $\mathcal{E}$ is in volts per unit length and H in amperes per unit length. We can neglect the small differences in the value of $\mathcal{E}$ and in the length of the magnetic path at different points between the earth and the cone ; for this small value of α these only amount to a few per cent. Hence

$$\frac{V}{I} = \frac{\mathcal{E}}{H} \times \frac{\alpha r}{2\pi r} = 120\pi \times \frac{\alpha}{2\pi} = 60\alpha = 36\Omega$$

In *Wireless Engineer* of April 1945 we showed that the radiation resistance of a quarter-wave earthed vertical aerial is 36.6 ohms.

Having thus illustrated an application of the intrinsic impedance of the transmitting medium, we return to the French note in which the authors consider a bifilar line of which the conductors have a radius a and unit permeability. They then say ' In order not to be confused by the fact that the two currents flow in opposite directions, we consider a circuit composed of only one of them and a neutral fictitious wire half way between them, which can be very fine as it carries no current and produces no magnetic field. If b is the shortest distance between this neutral wire and the surface of the actual conductor, the expression for the inductance will be $L = \mu_0 (0.5 + 2 \log_e(a + b)/a)$. Now we can reduce b to a very small value (0.001 cm for example) and let a increase indefinitely, so that $\log_e(a + b)/a = \log_e(1 + b/a) = b/a \rightarrow 0$ and one will have $Z_c = \mu_0 c/2$. As the conductor only occupies half the space it will be necessary to double this value and this gives for Z_0, the intrinsic impedance of space, the value $\mu_0 c$.' They say that this gives a result in favour of the value $Z_0 = 30\Omega$. This is obviously obtained by putting $c = 30 \times 10^7$ m/sec and $\mu_0 = 10^{-7}$ in accordance with the unrationalized m.k.s. system, but what all the above has to do with the intrinsic impedance of space is beyond our comprehension.

Even if two parallel conductors are made so large that they almost touch, the combined inductance will be $(1 + 4 \log_e 2a/a)$ and that of each considered separately $(0.5 + 2 \log_e 2)$. A fictitious fine wire placed between them cannot be made an excuse for stopping the integration half way, and then merely doubling the result. It would also appear that they have filled the space, the intrinsic impedance of which is to be determined, with a conductor. But apart from this the authors appear to have a wrong idea of what is usually understood by the term ' the intrinsic impedance ' of space or of a medium, a term that we believe was first employed and clearly defined by Schelkunoff in 1938.‡ This trouble goes back a long way, for in Heaviside's "Electrical Papers," Vol. II, p. 377, we read ' Since the line integral of H is electric current and the line integral of $\mathcal{E}$ is electromotive force, the ratio of $\mathcal{E}$ to H is the resistance-operator of an infinitely long tube of unit area ; a constant measurable in ohms, being 60 ohms in vacuum, or 30 ohms on each side. Why it is a constant is simply because the waves cannot return, as there is no reflecting barrier in the infinite dielectric.'

‡ *Bell Syst. tech. J.*, January 1938.

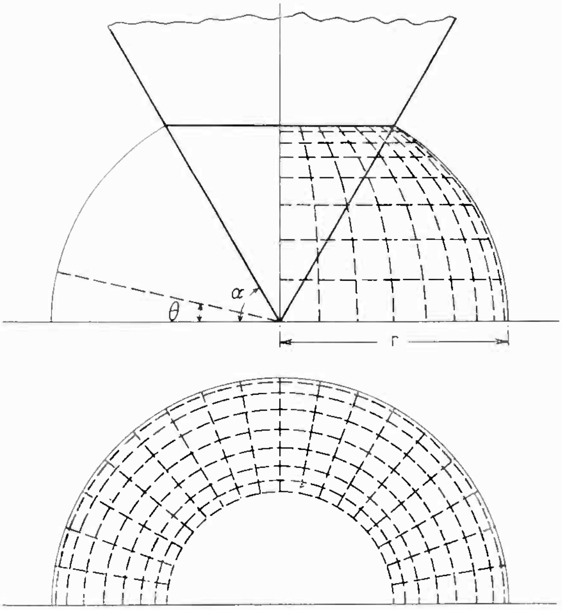


Fig. 3

He was considering ' a plane of sheet impressed force in a non-conducting dielectric ;—the disturbance is then propagated both ways undistorted at the speed of $v = (\mu\kappa)^{-\frac{1}{2}}$.' If the transmission from the source is in both directions, they are in parallel, and if the resistance on each side is 30 ohms, the combined resistance will be 15 and not 60 ohms. A few pages earlier (p. 369) Heaviside shows that for a distortionless telegraph circuit

$$Z = Lv \text{ where } v = 1/\sqrt{LC}$$

He then says $Lv = L \times 30$ ohms = the impedance of the circuit at A. With reference to this Dr. G. A. Campbell in a memorandum written in 1932 says ' Apparently Heaviside's 30 ohms was in ordinary ohms and not in Heaviside's own units, as Nichols quite naturally assumed. The correct explanation of the 30 ohms seems to be that Heaviside's ' resistance-operator of an infinitely long tube of unit area ' was not intended to be the characteristic impedance as I define it.' It certainly looks as if Heaviside's resistance-operator is to blame for the 30-ohms muddle in which Brylinski and de Broglie have become entangled.

It is easily seen that 30 ohms is the characteristic impedance of a line having unit inductance per unit length (that is, 1 c.g.s. unit of inductance per cm of length) since if $L = 1$, $1/CL = v^2 = 9 \times 10^{20}$ and $C = 1/9 \times 10^{20}$ e.m. units. Hence for such a line $Z_0 = \sqrt{L/C} = 3 \times 10^{10}$ c.g.s. units = 30 ohms, and for any line $Z_0 = 30L$ ohms where L is the inductance per cm in c.g.s. units. The dielectric is, of course, assumed to be space.

Having disposed of the 30 ohms, we shall now apply the 377-ohms intrinsic impedance to an inverted cone in which the angle α is so large that the approximation made in considering Fig. 2 is no longer permissible. In Fig. 3 the angle α is 60 degrees. The electric-field lines are drawn at the radius r for each 10° of the circumference; the magnetic-field lines are circles drawn at such distances apart that the spherical surface is divided up into squares; that is, as the electric lines converge on approaching the cone, the magnetic lines are drawn closer together, H and $\mathcal{E}$ being equal at every point. Since the latitudinal distance between adjacent electric lines varies as cos θ, the longitudinal distance between the magnetic lines must vary in the same way and the number of squares per unit angle must vary as $1/\cos\theta$. The total number of squares in angle α will be

$$6 \times \frac{3}{\pi}\int_0^{\alpha} d\theta/\cos\theta = \frac{18}{\pi}\log_e \tan\left(\frac{\alpha}{2}+\frac{\pi}{4}\right)$$
$$= (18/\pi) \times 1.315$$

in which $\alpha = 60°$. This gives 7.53 squares between the ground and the cone. As each square subtends 10 degrees horizontally, each horizontal layer contains 36 squares. Each square is a cross-section of a pyramid extending from the apex of the cone out into space with an intrinsic impedance of 377 ohms. As in Fig. 1, we determine the characteristic impedance of the conical line by multiplying the intrinsic impedance by the number of squares in the direction d of the electric field, and dividing by the number of squares in the direction b of the magnetic field. Hence

$$Z_0 = 377 \times \frac{7.53}{36} = 79 \text{ ohms}$$

which agrees with the value calculated by the accurate formula (see *Wireless Engineer*, July 1944) viz.

$$Z_0 = 60 \log_e \tan\left(\frac{\alpha}{2}+\frac{\pi}{4}\right) = 79 \text{ ohms.}$$

We trust that this has removed any uncertainty as to the exact meaning of intrinsic impedance and as to its value. G. W. O. H.

CONSTANT PHASE-SHIFT NETWORKS

By R. O. Rowlands, B.Sc.

(B.B.C. Engineering Training Department)

SUMMARY.—It is shown that for every phase-shift network there can be found a corresponding attenuation network whose attenuation is directly related to the phase shift of the former network. The technique of designing attenuation networks is more straightforward than that of designing phase-shift networks, and so the requirements of the phase-shift network are first translated into those of an attenuation network. This network is then designed and from its parameters, the parameters of the phase-shift network are calculated.

Introduction

In a passive linear network made up of positive inductors and capacitors the phase shift is an increasing function of frequency and so a network of this type cannot be used to produce a substantially constant shift over a given range of frequencies. A network made up wholly of negative inductors and capacitors will have a phase characteristic which is a decreasing function of frequency. By using both these types of sections in building up a composite network it is possible to obtain a phase shift which is oscillating about the required constant value over a given frequency range. The amount by which the phase deviates from the constant value depends upon the number of sections used.

MS accepted by the Editor, December 1948

Theory

The problem of designing $\pi/2$ phase-shift networks will be considered first; and afterwards, the modifications necessary to produce a phase shift other than $\pi/2$.

Consider the network shown in Fig. 1. The impedance of the network is given by

$$Z = \sqrt{(j\omega L)/(j\omega C)} = \sqrt{(L/C)}$$

and is therefore constant at all frequencies. The propagation function of this network is imaginary at all frequencies and is given by

$$\gamma = j\beta = \log_e \frac{1 + \sqrt{(j\omega L \cdot j\omega C)}}{1 - \sqrt{(j\omega L \,.\, j\omega C)}}$$

where β is the phase shift in radians.

If $\sqrt{LC} = K$

Then $j\beta = \log_e \dfrac{1 + j\omega K}{1 - j\omega K}$

It is possible to plot the graph of β against ω for varying values of K and to select those curves whose sum meets the requirements of a substantially-constant phase shift. This method is tedious and difficult compared with the problem of designing attenuation networks where usually it is only necessary to secure a certain minimum attenuation over a given frequency range; for whereas in the former problem we have to be within a maximum and a minimum limit, in the latter problem we are working to a minimum limit only. This suggests that if the phase-shift network can be linked to an allied attenuation network the problem will be greatly simplified.

If a number of basic sections are connected in tandem the phase shift of the composite network will be the sum of the phase shifts of the individual sections;

i.e., $j\beta = j\beta_1 + j\beta_2 + \dots + j\beta_n$

$$= \log_e \frac{1 + j\omega K_1}{1 - j\omega K_1} + \log_e \frac{1 + j\omega K_2}{1 - j\omega K_2} + \dots + \log_e \frac{1 + j\omega K_n}{1 - j\omega K_n}$$

$$= \log_e \frac{(1 + j\omega K_1)\,(1 + j\omega K_2) \dots (1 + j\omega K_n)}{(1 - j\omega K_1)\,(1 - j\omega K_2) \dots (1 - j\omega K_n)}$$

$$= \log_e \frac{1 + j\omega S_1 + j^2\omega^2 S_2 + \dots + j^n\omega^n S_n}{1 - j\omega S_1 + j^2\omega^2 S_2 - \dots (-j)^n\omega^n S_n} \quad \dots \quad (1)$$

where $S_1 = K_1 + K_2 + \dots + K_n$

$S_2 = K_1K_2 + K_1K_3 + \dots + K_{n-1}K_n$

$S_n = K_1K_2K_3 \dots K_n$

Now $e^{j\beta} = \cos\beta + j\sin\beta$

i.e., $j\beta = \log_e(\cos\beta + j\sin\beta)$

$\therefore$ if $\beta = \dfrac{\pi}{2}$,

$$j\frac{\pi}{2} = \log_e j = \log_e j\frac{(1-j)}{(1-j)}$$

$$= \log_e \frac{1+j}{1-j} = \log_e \frac{a + ja}{a - ja}$$

The condition that $\beta = \pi/2$ then is that equation (1) takes the above form. It may be seen by inspection that the real part of the numerator is equal to the real part of the denominator and the imaginary part of the numerator is the negative of the imaginary part of the denominator. The remaining condition is that the real and imaginary parts of the numerator are numerically equal;

i.e., that $1 - \omega^2 S_2 + \omega^4 S_4 - \dots = \omega S_1 - \omega^3 S_3 + \omega^5 S_5 \dots \quad \dots \quad (2)$

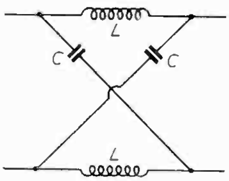


Fig. 1. *Basic section.*

This is a polynomial of degree n in ω, and may be solved for n frequencies at which the phase shift is $\pi/2$. At intermediate frequencies the phase shift deviates from $\pi/2$. In order to determine the values to be assigned to the coefficients S_r—where r takes the values 1 to n—so that the deviation of the phase shift from $\pi/2$ at these intermediate frequencies is a minimum, we shall endeavour to relate them to other coefficients whose values may be determined by using the theory of attenuation networks.

Consider then an allied network the physical configuration of which is immaterial and whose propagation function is real and is given by

$$\gamma = \log_e \frac{1 + h\omega}{1 - h\omega}$$

In this expression h is a factor determining the frequency at which the attenuation becomes infinite;

i.e., if ω_∞ corresponds to this frequency

$1 - h\omega_\infty = 0$

or $h = 1/\omega_\infty$

The propagation function of n sections of this network is given by

$$\gamma = \alpha = \log_e \frac{1 + h_1\omega}{1 - h_1\omega} + \log_e \frac{1 + h_2\omega}{1 - h_2\omega} + \dots + \log_e \frac{1 + h_n\omega}{1 - h_n\omega}$$

$$= \log_e \frac{1 + R_1\omega + R_2\omega^2 + \dots + R_n\omega^n}{1 - R_1\omega + R_2\omega^2 - \dots (-1)^n R_n\omega^n}$$

where $R_1 = h_1 + h_2 + h_3 + \dots + h_n$

$R_2 = h_1h_2 + h_1h_3 + \dots + h_{(n-1)}\, h_n$

$R_n = h_1h_2h_3 \dots h_n$

α is infinite when the denominator is zero;

i.e., when $1 - R_1\omega + R_2\omega^2 + \dots (-1)^n R_n\omega^n = 0$

or $1 + \omega^2 R_2 + \omega^4 R_4 + \dots = \omega R_1 + \omega^3 R_3 + \omega^5 R_5 \dots \quad \dots \quad (3)$

If we make $R_1 = S_1$; $R_2 = -S_2$; $R_3 = -S_3$; $R_4 = S_4$; $R_5 = S_5$; $R_6 = -S_6$; etc. .. (4)

then equations (2) and (3) are identical and may be solved for n frequencies at which the phase shift of the first network becomes $\pi/2$ and the attenuation of the second network becomes

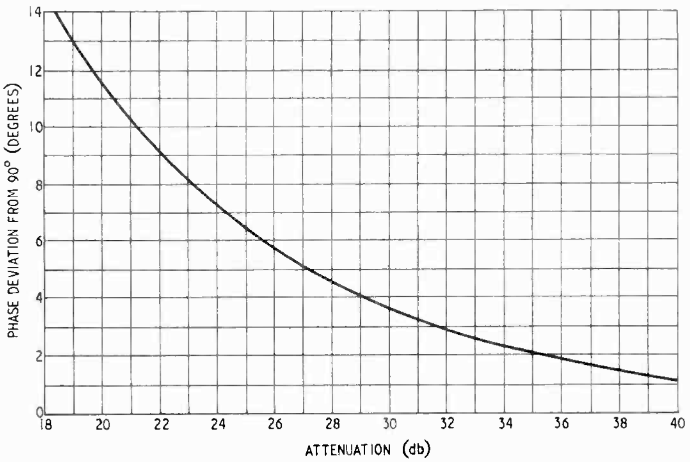


Fig. 2. *Relation between phase-shift of one network and attenuation of the derived network.*

infinite. Again if the left-hand sides of equations (2) and (3) are designated x and the right-hand sides y then the phase shift of the first network is given by

$$j\beta = \log_e \frac{x + jy}{x - jy} = 2j \tan^{-1} \frac{y}{x} \quad .. \quad .. \quad (5)$$

and the attenuation of the second network by

$$\alpha = \log_e \frac{x + y}{x - y} = 2 \tanh^{-1} \frac{y}{x} \quad .. \quad .. \quad (6)$$

and so

$$\frac{y}{x} = \tan \frac{\beta}{2} = \tanh \frac{\alpha}{2} \quad .. \quad .. \quad .. \quad (7)$$

This means that there is an exact relationship between the phase shift of the first network and the attenuation of the second. This relation is shown in graphical form in Fig. 2.

Design Procedure

So far the S coefficients of the phase-shift network have been related to R coefficients of an attenuation network. Again the phase-shift requirements of the former network may now be translated into attenuation requirements of the latter network. To obtain the values of the R coefficients which satisfy these requirements consider the attenuation function of a single section of this network given by

$$\alpha = \log_e \frac{1 + h\omega}{1 - h\omega} = 2 \tanh^{-1} h\omega$$

$$\text{i.e., } \tanh \frac{\alpha}{2} = h\omega$$

$$\text{and } \log_e \tanh \frac{\alpha}{2} = \log_e h\omega$$

By choosing a log scale for frequency and a log-tanh scale for attenuation the graph for this type of section will be a pair of straight lines as shown in Fig. 3.* The effect of varying h is to shift the graph laterally. It is a comparatively simple matter, therefore, by plotting graphs and adding the attenuations of individual sections, to determine the number of sections and their h values necessary to give the required performance. The values of R_1 to R_n are then determined. The K values of the sections of the corresponding phase-shift network are the roots of the polynomial

$$K^n - S_1K^{n-1} + S_2K^{n-2} - S_3K^{n-3} + S_4K^{n-4} - \ldots . = 0$$

$$\text{or } K^n - R_1K^{n-1} - R_2K^{n-2} + R_3K^{n-3} + R_4K^{n-4} - \ldots . = 0$$

Some of these roots are positive and some are negative. The positive roots correspond to sections with positive components and the negative roots to sections with negative components.

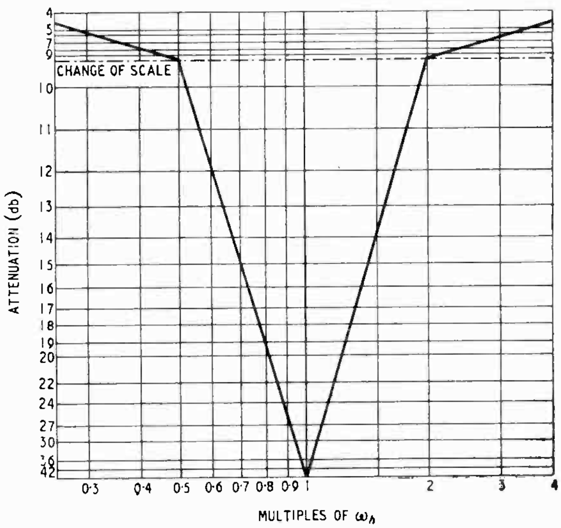


Fig. 3. *Network attenuation relation in terms of ω_h.*

* The discontinuity in the lines is entirely due to the change in the attenuation scale.

Phase Shifts other than $\pi/2$

In certain applications it may be necessary to produce a phase shift other than $\pi/2$. In such a case the requirements may be related to those of a network producing a phase shift of $\pi/2$ and the design equations modified in the following manner. From equation (5)

$$\tan\frac{\beta}{2} = \frac{y}{x}$$

As β oscillates about the value $\pi/2$ so y/x oscillates about the value 1, and the angle β' defined by $\tan\frac{\beta'}{2} = \frac{my}{x}$ oscillates about the value θ

$$\text{where } \tan\frac{\theta}{2} = m \quad \ldots \quad (8)$$

For any particular value of y/x let β become $\pi/2 - \epsilon$ and β' become $\theta - \delta$

$$\text{then } \frac{\tan\frac{\theta-\delta}{2}}{\tan\frac{\theta}{2}} = \frac{\frac{my}{x}}{m} = \frac{y}{x} = \tan\left(\frac{\pi}{4} - \frac{\epsilon}{2}\right)$$

$$\therefore \frac{\tan\frac{\theta}{2} - \tan\frac{\delta}{2}}{\tan\frac{\theta}{2}\left(1 + \tan\frac{\theta}{2}\tan\frac{\delta}{2}\right)} = \frac{1 - \tan\frac{\epsilon}{2}}{1 + \tan\frac{\epsilon}{2}}$$

Let $t = \tan\frac{\theta}{2}$; $d = \tan\frac{\delta}{2}$; and $e = \tan\frac{\epsilon}{2}$; then

$$\frac{1 - d/t}{1 + dt} = \frac{1 - e}{1 + e}$$

whence $e\{2 + d(t - 1/t)\} = d(t + 1/t)$

If d and e are small we may replace them by $\frac{\delta}{2}$ and $\frac{\epsilon}{2}$ and also ignore the terms involving their product.

$$\text{We have then } \epsilon = \left(\tan\frac{\theta}{2} + \cot\frac{\theta}{2}\right)\frac{\delta}{2} \quad \ldots \quad (9)$$

The procedure is then to evaluate the R terms in the manner described above for a network with an attenuation corresponding to a deviation ϵ and then evaluate the K terms corresponding to the equation $my/x = 1$;

$$\text{viz., } K^n - mS_1K^{n-1} + S_2K^{n-2} - mS_3K^{n-3} + \ldots = 0 \quad \ldots \quad (10)$$

Example

Suppose that it is desired to design a network giving a phase shift of $60 \pm 10°$ over the frequency range 300 c/s to 1200 c/s. If we call the geometrical mid-band frequency ω_0 then the range will extend from $\frac{1}{2}\omega_0$ to $2\omega_0$. Putting $\theta = 60°$ and $\delta = 10°$, in equation (9) we have that

$$\begin{aligned}\epsilon &= (\tan 30° + \cot 30°) \times 5 \\ &= (0.5774 + 1.7321) \times 5 \\ &= 2.3 \times 5 \quad = 11.5°\end{aligned}$$

Referring to Fig. 2 we see that a deviation of 11.5° corresponds to a minimum attenuation of 20 db. From Fig. 3 we see that a single section of the attenuation network will give an attenuation of just over 9 db at the extremities of the frequency range and so it looks as if two sections will be adequate. The requirements can easily be met but the optimum performance is obtained from the arrangement shown in Fig. 4, where the minimum attenuation at the midband is made equal to that at the extremities of the band and is of the order of 25 db. With this arrangement the values of $\omega_0 h$ come to 0.625 and 1.6.

From equations (4) we have that

$$\omega_0 R_1 = 0.625 + 1.6 = 2.225 = S_1\omega_0$$
$$\text{and } \omega^2{}_0 R_2 = 0.625 \times 1.6 = 1 = -S_2\omega^2{}_0$$

Again from (8) $m = \tan 30° = 0.5774$

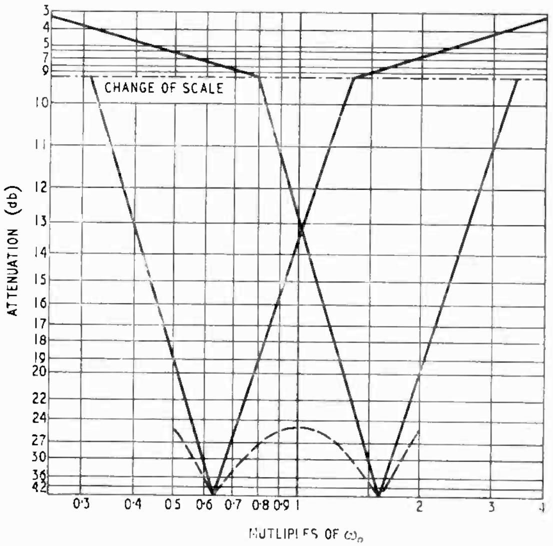


Fig. 4. *Two sections in tandem with peaks at* $0.625\,\omega_0$ *and* $1.6\,\omega_0$ *(solid lines) give the response shown dotted.*

So equation (10) becomes $K^2 - m\,S_1K + S_2 = 0$

$$\text{i.e., } K^2 - \frac{1.3}{\omega_0}K - \frac{1}{\omega_0{}^2} = 0$$

$$\text{or } K_1 = \frac{1.845}{\omega_0} = \sqrt{L_1C_1}$$

$$\text{and } K_2 = -\frac{0.545}{\omega_0} = \sqrt{L_2C_2}$$

Knowing the impedance of the circuit the values of L_1, C_1, L_2 and C_2 may be worked out. In consequence of K_2 being negative, minus signs have to be assigned to L_2 and C_2. The realization of these negative components is discussed in the next section.

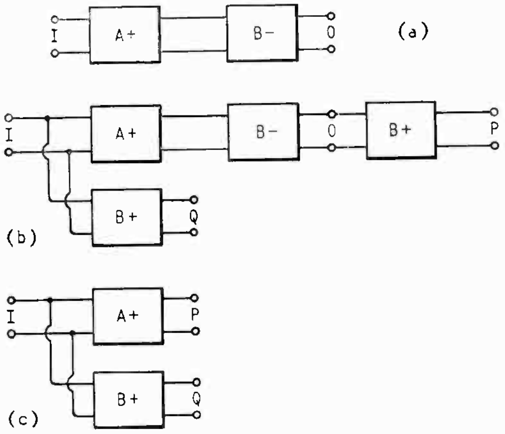


Fig. 5. *Sections having positive and negative elements are grouped as at (a). The addition of two B+ networks (b) leaves the phase difference between P and Q unaltered. The B+ and B− networks cancel to give (c).*

Alternative Circuit

We have derived a method of designing a network which produces a constant phase shift between its input and output terminals. Such a network however will have negative components in those sections corresponding to negative values of K. In most practical applications, however, such as in a single-sideband modulator the requirements are two pairs of terminals with a constant phase difference between them. An arrangement satisfying this condition can be built up entirely of positive components, as will be shown.

The network designed as above consists of the tandem connection of a number of sections corresponding to the various values of K. If the sections corresponding to positive values of K are segregated to one end of the network and the ones corresponding to negative values are segregated to the other end, then the network will be as shown in Fig. 5(a), where the former sections are represented by the block A + and the latter by the block B −. If now two similar networks B +, whose components are equal and opposite in sign to those in the B − sections are inserted, one in tandem with the output terminals and the other in parallel with the input terminals as shown in Fig. 5(b), the phase difference between the output terminals of the two B + networks still meets the required performance.

The phase shift between terminals I and O will be substantially $\pi/2$.

Let the phase shift in B + be ϕ

Then the phase shift from I to P $= \pi/2 + \phi$

And the phase shift from I to Q $= \phi$

So the phase difference between P and Q $= \pi/2 + \phi - \phi = \pi/2$.

The B − and B + sections in tandem, however, cancel each other out as is shown in the Appendix. The network may therefore be realized as the sections A + and B + in parallel as shown in Fig. 5(c).

Acknowledgment

The author is indebted to Dr. Sturley for his valuable suggestions when reading the manuscript of the paper.

APPENDIX

Let two networks be connected in tandem, the networks being identical except for the components in one being the negative of those in the other. Let them be expanded in the form of a ladder. This is always possible so long as we admit of negative components. The networks will then be as shown in Fig. 6. First, the impedances Z_1 and $-Z_1$ in series at the junction of

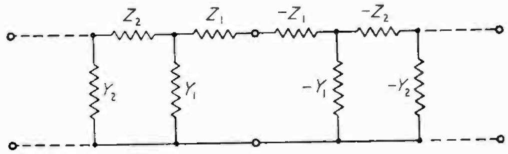


Fig. 6. *This diagram illustrates how positive and negative networks in tandem cancel out.*

the two networks cancel each other out. This leaves the admittances Y_1 and $-Y_1$ in parallel, but these now cancel each other. Proceeding along the networks in a similar manner we find eventually that there is complete cancellation of the two networks.

ELECTRICAL PROPERTIES OF WATER

Reflection Characteristics of Water Surfaces at V.H.F.

By J. A. Saxton, Ph.D., B.Sc., A.M.I.E.E.

(Communication from the National Physical Laboratory)

SUMMARY.—The variation of the dielectric properties of water as a consequence of anomalous dispersion is discussed. It is shown that this dispersion occurs mainly between the frequencies of 10^3 and 10^6 Mc/s, over which interval the permittivity of water falls from 80 to 5.5. The ionic conductivity of fresh water is of importance only for frequencies less than 10^3 Mc/s, and in sea water the ionic conductivity loses its significance when the frequency exceeds 2×10^4 Mc/s. The results are given of some calculations which illustrate the effect of the anomalous dispersion on the reflection coefficient of fresh water surfaces for radio waves.

1. Introduction

In the consideration of the propagation of very short radio waves a knowledge of the reflection coefficient of the earth's surface is necessary. Further, at wavelengths less than 10 cm the scattering and absorption of radio waves by water drops in the atmosphere is of importance. Since water is an important constituent of the earth's surface, and in view of the increasing use of very short-wave radiation in radio communication and navigation, it has been suggested that a review is desirable of the existing data, both experimental and theoretical, relating to the dielectric properties of water at these very-short wavelengths. This paper presents such a review.

It has usually been the practice to regard water as a reasonably good conductor for electromagnetic waves longer than a certain value of the order of 10 or 20 metres, say, for sea water, and about 3,000 metres for fresh water. It has further been common to consider fresh water as behaving mainly as a dielectric at wavelengths less than about 1 metre, but this is by no means true except in a very restricted range of wavelength. The latter qualification arises since the water molecule is electrically polar, as a result of which water exhibits anomalous dispersion, and the region of the radio-frequency spectrum in which this dispersion is manifest is principally between the wavelengths of 10 cm and 1 mm (frequencies 3,000 to 300,000 Mc/s). Sea water, in which the ionic conductivity is much greater than in fresh water, cannot be regarded as even a moderately good dielectric at any wavelength in the radio-frequency spectrum.

Now although the major part of the anomalous dispersion occurs between the wavelengths of 10 cm and 1 mm, some small variation of the permittivity exists up to a wavelength of 30 cm, while a variation in the absorption coefficient exists up to even slightly longer wavelengths. Above the wavelength of 30 cm the refractive index is given adequately by the square root of the low-frequency relative permittivity, which is 80 at 20° C, but even at a wavelength of 1 metre the effective dipolar conductivity, with which the dispersion is associated, is rather greater than the ionic conductivity produced in ordinary fresh water by the normally occurring amounts of dissolved salts. A reasonable average value for the ionic conductivity of fresh water is about 10^8 e.s.u. or 1.1×10^{-2} mho/m. It will be seen later that there is no appreciable difference between the electrical characteristics of pure and fresh water for wavelengths less than about 30 cm (frequencies greater than 1,000 Mc/s). In sea water the ionic conductivity may be taken as about 4×10^{10} e.s.u., and it will be seen that the dielectric properties of sea water differ noticeably from those of pure water at wavelengths greater than 3 cm (frequencies less than 10,000 Mc/s).

MS accepted by the Editor, September 1948

2. Dielectric Properties of Pure Water

Extensive measurements have recently been made by Saxton and Lane[1(a)] of the dielectric properties of water at wavelengths of 1.24 and 1.58 cm, and as a function of temperature in the range 0° C to 40° C. The results of this work have since been substantially confirmed by Collie, Hasted and Ritson[2].

An examination[1(b)] of this experimental data in the light of the theory of anomalous dispersion in polar liquids has shown that from it the dielectric properties of water may be predicted satisfactorily over the wavelength range 1 to 10 cm. The present work gives an extension of the theoretical calculations to longer wavelengths, and shows that the predictions are in good agreement with the known data.

2.1. *Theoretical basis of calculations*

The original theory of dispersion in polar liquids is due to Debye[3], and in its original form it accounts satisfactorily for the observed behaviour of many simple solutions of polar substances in non-polar solvents. It has been necessary, however, to modify the theory somewhat in the consideration of pure polar liquids such as, for example, water, because of the strong interaction fields between the molecular dipoles in such liquids which are neglected in the Debye theory. Nevertheless, for any theory in which the internal field in the dielectric is assumed to be a linear function of the polarization, the form of the dependence of the complex permittivity on frequency remains the same. The actual expression used for the internal field determines the value of the dipole relaxation time, an important parameter in the theory.

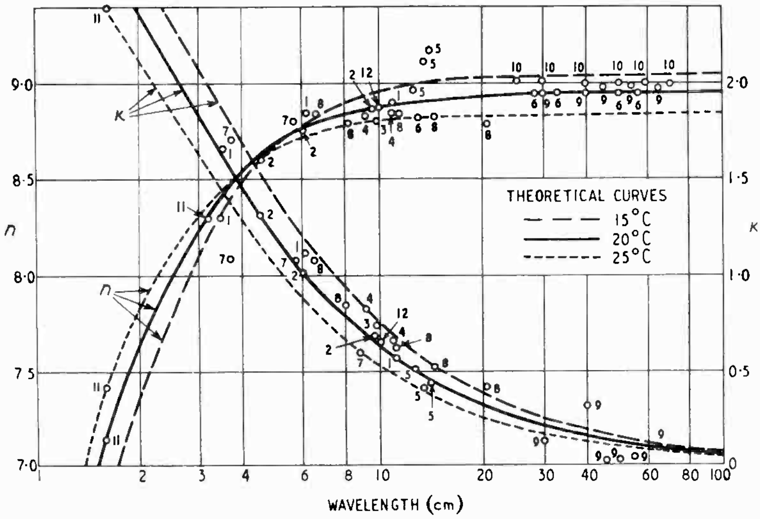


Fig. 1. *Dielectric properties of water in the wavelength range 1 cm to 1 m; key to experimental observations:— 1. Abadie 15° C; 2. Esau & Bäz 19° C; 3. Conner & Smyth 25° C; 4. Cooper 22° C; 5. Seeberger 16° C; 6. V. Ardenne 18° C; 7. Eckert 15°—20° C; 8. Knerr 22° C; 9. Weichmann 17° C; 10. Frankenburger 17° C; 11. Saxton 20°, 25° C; 12. V. Hippel 20° C.*

If Onsager's[4] treatment of the internal field is accepted, a procedure supported by the work of Fröhlich and Sack[5] in the case of low viscosity liquids like water, then we find that the complex permittivity, ϵ may be expressed thus:—

$$\epsilon = \epsilon' - j\epsilon'' = \epsilon_0 + (\epsilon_s - \epsilon_0)/(1 + j\omega\tau) \quad (1)$$

where ϵ_s = static permittivity

ϵ_0 = permittivity at the high frequency end of the region of dipolar dispersion

τ = relaxation time

$\omega = 2\pi f$, f being the frequency in c/s.

It has been usual to regard ϵ_0 as being equal to the square of the optical refractive index, which amounts to saying that it arises from that part of the total polarization due only to the electronic polarization. This approximation is not satisfactory if the atomic polarization, which arises from atomic vibrations in the infra-red region of the spectrum, is relatively large, as has been shown to be the case in water[1(b),2] since ϵ_0 in fact should represent the effects of both electronic and atomic polarizations. Thus ϵ_0 for water has been found to be 5.5—as compared with 1.8 on the basis of the electronic polarization only.

The assumptions underlying equation (1) are justified in the case of water[1(b)], and the behaviour of water may be described in terms of a single relaxation time at any given temperature, the relaxation time being a function of temperature and decreasing as the temperature increases. ϵ_0 is assumed not to vary appreciably with temperature, and the variation of ϵ_s with temperature is well known.

There are several quantities in terms of which the electrical properties of a lossy dielectric may be described, and the expressions relating these quantities may be conveniently summarized thus:—

$$\epsilon = \epsilon' - j\epsilon'' = \epsilon' - 2j\sigma/f = (n - jk)^2 \quad (2)$$

where σ is the conductivity in e.s.u., n is the refractive index and κ the absorption coefficient.

From equations (1) and (2) we may then derive the following relations:—

$$\epsilon' = \frac{\epsilon_s + \epsilon_0 x^2}{1 + x^2} \quad (3)$$

$$\epsilon'' = \frac{x(\epsilon_s - \epsilon_0)}{1 + x^2} \quad (4)$$

where $x = \omega\tau$

and also $2n^2 = [\epsilon'^2 + \epsilon''^2]^{\frac{1}{2}} + \epsilon' \quad (5)$

$$2\kappa^2 = [\epsilon'^2 + \epsilon''^2]^{\frac{1}{2}} - \epsilon' \quad (6)$$

$$n\kappa = \sigma/f = \epsilon''/2 \quad (7)$$

It may be noted that equations (3) and (4) are identical in form with those derived originally by Debye except that according to his theory

$$x = \left(\frac{\epsilon_s + 2}{\epsilon_0 + 2}\right)\omega\tau.$$

2.2. *Comparison between Theory and Experiment*

As a consequence of the mechanism involved in dipolar dispersion the dielectric properties of water vary with temperature as well as with wavelength. Early observers were more concerned with variations of the properties with wavelength, and the importance of the temperature factor was perhaps not sufficiently appreciated. The majority of experimental results obtained before the last few years refer to water at about room temperatures, and the various observations of refractive index (n) and absorption coefficient (κ)[1, 6–15] given in Fig. 1 are

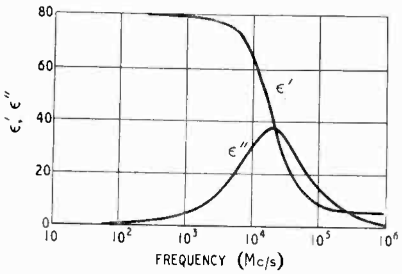


Fig. 2. *Components ϵ' and ϵ'' of the complex dielectric constant of pure water, $\epsilon' - j\epsilon''$, at a temperature of 20°C.*

confined to the temperature range 15° C to 25° C. Three theoretical curves for temperatures of 15, 20 and 25° C, calculated as indicated above, and based on relaxation times determined by the author[1(b)] are shown for comparison with the experimental data. It is apparent that the theoretical curves agree well with the experimental observations from centimetre wavelengths up to 1 metre. When the effect of varying temperature is taken into account, the variations amongst the experimental results would certainly not justify any important shift in the theoretical curves.

3. Complete Region of Dipolar Dispersion in Water

The agreement between theory and experiment shown in Fig. 1 makes it reasonable to calculate

TABLE I

Temperature (°C)	ϵ_s	$\tau \times 10^{12}$ (sec)
0	88	19.05
5	86	14.60
10	84	11.85
15	82	9.60
20	80	8.10
25	78.2	6.80
30	76.4	5.95

the electrical characteristics of water over an even wider range of wavelengths. The values of ϵ_s and τ given in Table I, together with equations (3) to (7) enable one to calculate the dielectric properties of pure water at any frequency, and also at any temperature in the range 0° C to 30° C, though for the remainder of this work we shall confine our attention to water at 20° C.

The values of ϵ' and ϵ'' from 20° C calculated in this manner as a function of frequency from 10 to 10^6 Mc/s are shown in Fig. 2. This frequency range contains the entire region of dipolar dispersion in water, in fact almost the whole of the dispersion takes place in the range 10^3 to 10^6 Mc/s. ϵ' still has substantially its static value at 10^3 Mc/s, but ϵ'' continues to vary a little as the frequency is reduced from 10^3 to 10^2 Mc/s.

Although the mechanism of energy absorption in the liquid arising from dipole relaxation is somewhat different from that concerned in the case of ionic conductivity, we may consider the dipolar loss as being due to an effective conductivity which is calculable from equations (4) and (7). This effective conductivity is shown as σ_p in Fig. 3: it is of the order of 10^7 e.s.u. at a frequency of 100 Mc/s, it rises to nearly 10^{12} e.s.u. at 10^5 Mc/s and then changes little up to 10^6 Mc/s.

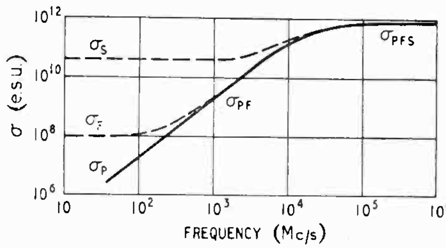


Fig. 3. *Conductivities σ_p, σ_F and σ_s of pure, fresh and sea water; temperature 20° C.*

4. Effect of Ionic Conductivity

Dissolved salts in water lead to ionic conductivity, and it might at first be thought that this ionic conductivity should simply be added to the dipolar conductivity σ_p in order to obtain the total effective conductivity. To a first approximation this is in fact true[9], and especially so if the salt concentration is very low as, for example, in fresh water. It is known[1(c), 16], however, that the presence of appreciable quantities of an electrolyte in water causes a change in the dipole relaxation time, and in strong solutions this is a factor which cannot be neglected. Although this factor is just beginning to be of importance for the concentration of ordinary salt (about 4%) occurring in sea water, its significance is not such that any great error results if the total effective conductivity of sea water is also estimated simply by adding arithmetically the ionic and dipolar contributions—at any rate for wavelengths greater than a few centimetres. We shall therefore neglect the influence of salt on the dipole relaxation time in the few calculations relating to sea water made in the present section.

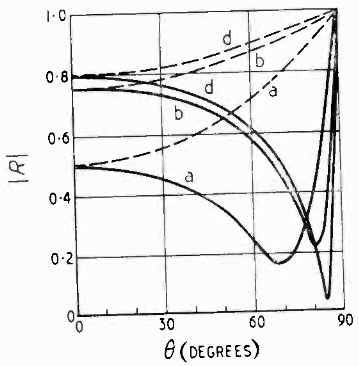


Fig. 4. *Modulus of reflection coefficient $|R|$ of fresh water for various wavelengths as a function of angle of incidence (θ); temperature = 20° C. Curves a, b and d are for wavelengths of 1 mm, 1 cm and 6 m respectively; – – – – – = polarization perpendicular to plane of incidence (R_H), ——— = polarization parallel to plane of incidence (R_V). Note. The values of $|R|$ for a wavelength of 10 cm differ only slightly from those of a wavelength of 6 m and could not easily be distinguished from the latter on the scale of Fig. 4.*

If we take 10^8 e.s.u. as an average value for the ionic conductivity of fresh water and combine it with the dipolar conductivity we obtain a total conductivity for fresh water σ_F, as indicated in Fig. 3. It will be seen that for frequencies less than 100 Mc/s the ionic conductivity is the only term which matters, while for frequencies exceeding 1,000 Mc/s only the dipolar conductivity term is significant.

TABLE II

Frequency (Mc/s)	$\epsilon = \epsilon' - j\epsilon''$	
	Fresh Water	Sea Water
1	$80 - 200j$	$80 - 80{,}000j$
10	$80 - 20j$	$80 - 8{,}000j$
10^2	$80 - 2j$	$80 - 800j$
10^3	$79 - 4j$	$79 - 80j$
10^4	$65 - 30j$	$65 - 40j$
10^5	$8 - 15j$	$8 - 15j$

In sea water, on the other hand, where we may take the ionic conductivity to be about 4×10^{10} e.s.u., it is apparent from curve σ_s, Fig. 3, that this is the significant term up to frequencies in the region of 2,000 Mc/s, and that it is not until the frequency is as high as 20,000 Mc/s that the dipolar term is completely predominant. Table II gives a comparison of the complex dielectric constants of fresh water and sea water as a function of frequency.

These values show quite clearly that only in the range of radio frequencies 50 to 1,000 Mc/s (wavelengths 6 metres to 30 cm) is it reasonable to regard fresh water as behaving mainly like a dielectric, while sea water never behaves so at any point in the radio spectrum. For frequencies less than 0.1 Mc/s (wavelengths greater than 3,000 metres) fresh water has chiefly the characteristics of a conductor, though not a very good one, whereas sea water behaves mainly as a conductor at frequencies less than 30 Mc/s (wavelengths greater than 10 metres). The same value of ϵ' has been assigned to sea water as to fresh (and pure) water in Table II but, although this is not strictly correct[1(c), 16], it is a satisfactory approximation for the present comparison.

5. Reflection Coefficients of Water Surfaces at V.H.F.

The reflection coefficients calculated in this section are for plane waves incident at plane surfaces of fresh water. It is apparent from the foregoing discussion that the reflection coefficient of sea water will not be greatly different from that of fresh water when the frequency exceeds 10^4 Mc/s, so that the manner in which dipolar dispersion influences the reflection coefficient may be largely seen from a consideration of fresh-water surfaces only.

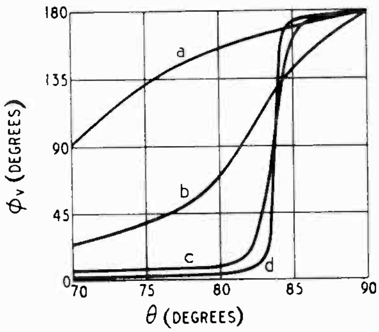


Fig. 5. *Phase retardation, ϕ_V, on reflection at a fresh-water surface of waves polarized in the plane of incidence as a function of angle of incidence θ; temperature 20° C. Curves a, b, c and d are for wavelengths of 1 mm, 1 cm, 10 cm and 6 m respectively.*

The reflection coefficient is not only a function of the complex dielectric constant of the water, and of the angle of incidence, but is also dependent upon the polarization of the incident waves. We shall denote by R_V and R_H the reflection coefficients for waves polarized with the electric vector in, and perpendicular to, the plane of incidence respectively. When the dielectric constant of the reflecting medium is complex the Fresnel reflection coefficients are also complex, and in general waves suffer a change in phase or reflection different from the normal values of either 0 or π which obtain when ϵ'' is zero. Thus in general R_V and R_H are of the form $|R_V|e^{j\phi_V}$ and $|R_H|e^{j\phi_H}$, where ϕ_V and ϕ_H are the phase changes on reflection for the two types of polarization.

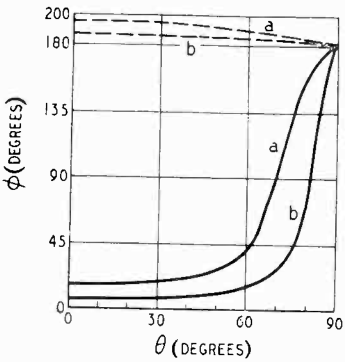


Fig. 6. *Phase retardation on reflection at a fresh-water surface; temperature 20° C. Curves a and b are for wavelengths of 1 mm and 1 cm respectively.* ------ *polarization perpendicular to plane of incidence* (ϕ_H), ——— *polarization parallel to plane of incidence* (ϕ_V).

The Fresnel formulae in the form given by McPetrie[17] have been used to calculate $|R_V|$, $|R_H|$, ϕ_V and ϕ_H for wavelengths of 6 metres, 10 cm, 1 cm and 1 mm, (frequencies 50, 3,000, 3×10^4 and 3×10^5 Mc/s) as a function of θ, the angle of incidence, and the results are shown graphically in Figs. 4, 5 and 6. The modulus of the reflection coefficient is given in Fig. 4, while Figs. 5 and 6 illustrate the phase change on reflection.

It is well known that for a pure dielectric the reflection coefficient R_V falls to zero at the Brewster angle of incidence given by $\theta = \tan^{-1}\sqrt{\epsilon}$, and further that the phase change on reflection, ϕ_V, is zero for $\theta < \tan^{-1}\sqrt{\epsilon}$ and π for $\theta > \tan^{-1}\sqrt{\epsilon}$. It will be seen from Fig. 4 that as the wavelength is reduced (in the range under consideration) and the term ϵ'' increases in importance, so the Brewster angle becomes less well defined, and that in fact $|R_V|$ never falls to zero but only reaches a certain minimum value depending upon the relative values of ϵ' and ϵ''. Also, as shown in Figs. 5 and 6, although the phase retardation on reflection does change rapidly through the region of this pseudo Brewster angle, it does so continuously, and in fact passes through the value $\pi/2$ close to the point where $|R_V|$ has its minimum value. It may be noted that for $\theta < 70°$ the values of ϕ_V are less than 1° and 3° for the wavelengths of 6 metres and 10 cm respectively. Further, for the wavelength of 6 metres ϕ_H varies from 180.5° at $\theta = 0$ to 180° at $\theta = 90°$; and for the wavelength of 10 cm the corresponding range of ϕ_H is 181° to 180°.

6. Conclusion

The anomalous dispersion of water, arising from the permanent electric polarity of the water molecule, is of major importance in the frequency range 10^3 to 10^6 Mc/s. The value of the complex dielectric constant has been calculated over the frequency range 10 to 10^6 Mc/s, and the effects of the small ionic conductivity of fresh water and the much larger ionic conductivity of sea water are discussed. The behaviour of fresh water is similar to that of pure water for frequencies greater than 10^3 Mc/s, whereas this condition does not obtain for sea water until the frequency exceeds 2×10^4 Mc/s. The reflection coefficient of fresh water surfaces has been calculated as a function of angle of incidence for frequencies of 50, 3×10^3, 3×10^4 and 3×10^5 Mc/s, (wavelengths of 6 metres, 10 cm, 1 cm and 1 mm) and it is shown how the Brewster Angle becomes ill-defined in the region of dipolar dispersion.

7. Acknowledgments

The work described above was carried out in the Radio Division of the National Physical Laboratory as part of the programme of the Radio Research Board, and this paper is published by permission of the Department of Scientific and Industrial Research.

REFERENCES

[1] (a) J. A. Saxton and J. A. Lane, "Meteorological Factors in Radio Wave Propagation," Conference Report, *Phys. and Roy. Met. Socs.*, 1946, p. 278. (b) J. A. Saxton, ibid., p. 292. (c) J. A. Saxton, ibid., p. 316.

[2] C. H. Collie, J. B. Hasted and D. M. Ritson, "The Dielectric Properties of Water and Heavy Water," *Proc. phys. Soc.*, 1948, Vol. 60, p. 145.

[3] P. Debye, "Polar Molecules," The Chemical Catalog Co. Inc., 1929, New York.

[4] L. Onsager, "Electric Moments of Molecules in Liquids," *J. Amer. chem. Soc.*, 1936, Vol. 58, p. 1486.

[5] H. Fröhlich and R. Sack, "Dielectric Properties of Dipolar Substances," *Proc. roy. Soc.*, 1944, Vol. 182A, p. 388.

[6] P. Abadie, "Extension of the Measurements of Dispersion and Absorption by Liquids, to the Region of Centimetric Radio Electric Waves," *Trans. Faraday Soc.*, 1946, Vol. 42A, p. 143.

[7] A. Esau and G. Bäz, "Reflection and Absorption Measurements on Water and Some Organic Liquids," *Phys. Zeits.*, 1937, Vol. 38, p. 774.

[8] W. P. Connor and C. P. Smyth, "The Dielectric Dispersion and Absorption of Water and some Organic Liquids," *J. Amer. chem. Soc.*, 1943, Vol. 65, p. 382.

[9] R. Cooper, "The Electrical Properties of Salt Water Solutions over the Frequency Range 1–4,000 Mc/s," *J. Instn elect. Engrs*, 1946, Part III, Vol. 93, p. 69.

[10] M. Seeberger, "Investigation of Dispersion with Undamped Ultra-Short Waves," *Ann. d. Physik*, 1933, Vol. 16, p. 77.

[11] H. von Ardenne, O. Groos and G. Otterbein, "Dispersion Measurements in the Region of Decimetre Waves," *Phys. Zeits*, 1936, Vol. 37, p. 533.

[12] F. Eckert, "On Anomalous Dispersion in the Region of Short Electric Waves," *Verh. d. Physik. Gesell.*, 1913, Vol. 15, p. 307.

[13] H. W. Knerr, "The Electric Spectrum of Liquid Water from Five to Twenty-Five Centimetres," *Phys. Rev.*, 1937, Vol. 52, p. 1054.

[14] R. Weichmann, "The Band Absorption Spectrum of Water at Wavelengths of Several Decimetres," *Ann. d. Physik.*, 1921, Vol. 66, p. 501.

[15] E. Frankenburger, "Measurement of the Refractive Index of Water between Twenty-Three and Seventy-Three Centimetres Wavelength," *Ann. d. Physik*, 1929, Vol. 1, p. 948.

[16] J. B. Hasted, D. M. Ritson and C. H. Collie, "Dielectric Properties of Aqueous Ionic Solutions," *J. chem. Phys.*, 1948, Vol. 16, p. 1.

[17] J. S. McPetrie, "The Reflection Coefficient of the Earth's Surface for Radio Waves," *J. Instn elect. Engrs*, 1938, Vol. 82, p. 214.

HIGH-POWER CATHODE-RAY TUBES

For Fixed Station P.P.I. Display

By Hilary Moss, Ph.D., B.Sc., A.M.I.E.E., M.Brit.I.R.E.

(Electronic Tubes, Ltd.)

SUMMARY.—This paper gives a preliminary survey of the design of large-screen cathode-ray tubes suitable for direct-viewing purposes. The tubes in question have screen diameters of the order of 30 inches. It is shown that the design is critically dependent on the form of the relation which is assumed to exist between the response of cascade screens and the beam voltage. Two distinct calculations are made—one assuming a linear response to voltage and the second a square-law response. Solutions in these two cases differ widely and this fact indicates the necessity for careful investigation of phosphor response as a function of beam voltage before further progress can be made. In particular, information is required on afterglow behaviour for which no data appears to be available over the voltage ranges concerned. Nevertheless, the paper is a guide to the probable form and operating conditions of these large-screen tubes. The difficulties of mechanical design are largely ignored, but a few notes on this matter are included in an appendix.

Introduction

THE problem of producing a large-screen p.p.i. display has been of interest for many years, but so far no very satisfactory solution has been found. The work has been almost entirely conducted on projection systems, which have the familiar drawbacks of indifferent brightness and contrast so well known to television engineers. This paper investigates the alternative approach, namely that of constructing large direct-viewing tubes. It is appreciated that these present considerable mechanical difficulties, but the investigation is felt to be worthwhile, since it is almost certain that their performance can be made much higher than that of any projection system.

The methods employed for the deduction of the operating conditions and tube dimensions belong to the province of scaling theory which the author has treated in a previous paper[1]. Familiarity on the part of the reader with that paper will be assumed, and the underlying theory will not be recapitulated here.

Statement of the Problem

The problem is to design a cathode-ray tube with a screen diameter of approximately 30 inches and intended for p.p.i. working. The natural approach is to consider some suitable prototype tube with which a considerable amount of experience has been gained and perform on it a simple scaling operation ; Fig. 1 indicates the essential beam dimensions of such a prototype. For the moment we are concerned with the general method and we do not therefore introduce at this stage any numerical values. The derived tube is to have a screen diameter k times that of the prototype. We shall also postulate that both the central and deflected spot sizes of the derived tube are the same as those of the prototype. In addition, since the spot writing speed in both radial and circumferential directions on the derived tube will be each k times as large as on the prototype, it can be shown that the screen excitation of the derived tube needs to be k^2 times that on the prototype. More precisely, this statement has the following meaning. Consider a portion of the screen of the prototype tube of area δS. In time T the beam energy delivered over it will be EIT and the mean energy per unit area will be $EIT/\delta S$. Now on the derived tube the corresponding area scanned in time T will be $k^2\delta S$ and the mean energy per unit area will be $E'I'T/k^2\delta S$.

MS accepted by the Editor, November 1948

LIST OF SYMBOLS

I = beam current
E = anode voltage
k = constant of scale
r_2 = size of spot
u = crossover to focusing lens distance
v = screen to focusing lens distance
B = beam width in plane of focusing lens
α = scaling constant for triode only
S = screen diameter
Primed symbols refer to the derived tube, and unprimed symbols to the prototype.

Hence for equality of excitation per unit area (that is equality of surface brightness) we have

$$E'I' = k^2EI \quad \quad (1)$$

The considerations necessary to preserve equality of the spot sizes at the screen centre and at the screen edge are essentially distinct and must be treated separately. (Ref. 1, Section 1.5).

Conditions for Equal Deflection Defocusing

Since the exact relationship between the degree of deflection defocusing and the scanning angle is complicated, it is generally advisable to keep this angle constant when using scaling theory.

The most direct method, therefore, of deriving from Fig. 1 a tube having a screen diameter of kS is merely to multiply every dimension on Fig. 1 by k. This obviously will result in a tube in which the degree of deflection defocusing (that is, difference in spot sizes at screen edge and centre) is also multiplied by k. Now it can also be shown that the deflection defocusing is proportional to the beam width in the deflectors.* anode voltage.† Since the magnification in the derived tube of Fig. 2 is k times as great as the prototype of Fig. 1, it follows that

$$y_2' = ky_2\sqrt{E/E'} \qquad (3)$$

Substituting from (2) into (3) immediately gives

$$y_2' = y_2$$

so that the condition for constant central-spot size is automatically satisfied in this case.

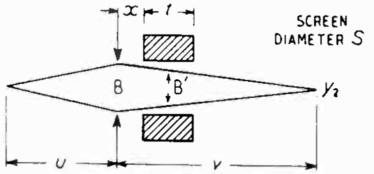
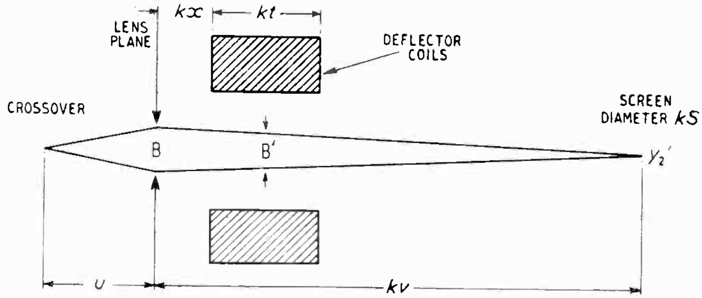


Fig. 1 (above). *Essential beam dimensions of prototype tube.*

Fig. 2 (right). *Beam dimensions of the derived tube.*



Hence if, after the scaling operation on Fig. 1 already referred to, we subsequently reduce the beam width to $1/k$ (i.e., beam width kept constant as in Fig. 1) the deflection defocusing will be held constant. Furthermore, the deflection defocusing is quite independent of conditions on the left-hand side of the focusing lens, with the result that we may keep the distance u constant. Fig. 2 now shows the essential beam dimensions of the derived tube, which will have an unchanged performance so far as deflection defocusing is concerned. It will be noted that the scale of the deflectors has been multiplied by k and that at all corresponding points in the deflector fields of the derived and prototype tubes, the beam widths are equal.

Conditions for Constant Size of Central Spot

For the moment, the possible perturbations due to space charge will be ignored. These are subject to later investigation. The central-spot size is a function of the operating voltage of the tube and since the latter quantity will be defined by considerations of screen brightness, we have to commence by making reasonable postulates as to the way in which screen brightness is related to the beam voltage and current.

Assumption 1. *Screen brightness and afterglow proportional to beam current and proportional to beam voltage.* In this case, postulating that the beam currents in the two tubes shall be the same, it immediately follows from equation (1) that

$$E' = k^2E \qquad (2)$$

Again, spot size in a cathode-ray tube of constant geometry varies inversely as the root of the final

In order to preserve constancy of modulation characteristic on the derived tube, it can be shown that it is necessary and sufficient for its cathode-to-grid spacing to be adjusted to preserve constancy of cut-off voltage, although the anode voltage is multiplied by k^2. (Ref. 2, Section 2.2.3). This may be achieved merely by multiplying the cathode-to-grid spacing on the derived tube by k^2.†

Table I summarizes the relative proportions and operating conditions of the derived tube for any scale factor change. The numerical values in brackets are the special case, based on the prototype VCR 516 and assuming a scale factor of $k = 3$. A peak beam current of 150 μA has been assumed, which is the order of current employed in the VCR 516.

Assumption 2. *Screen brightness and afterglow proportional to beam current and proportional to beam voltage squared.* Again we shall postulate constancy of beam current in the two tubes, whence it immediately follows that

$$E'^2 = k^2E^2 \qquad (4)$$

so that $E' = kE$. In contradistinction to Assumption 1, this last equation obviously means that a further change has to be effected in order to preserve constancy of central-spot size in the derived system of Fig. 2. This is most easily brought about by multiplying all dimensions of the triode by α. Then

$$y_2' = k\alpha y_2\sqrt{E/E'}$$

and substituting from (4) for $\sqrt{E/E'}$, then gives

$$y_2' = k\alpha y_2/\sqrt{k}$$

so that the condition $y_2' = y_2$ requires that

* This is probably only an approximation valid over restricted ranges.

† In a multi-anode tube, we are assuming that the anode voltage *ratios* are held constant; i.e., all anode voltages are multiplied by k.

$\alpha = 1/\sqrt{k}$. This defines the factor by which the triode of the derived tube must be scaled in order to produce constancy of central-spot size.

Again, constancy of modulation characteristic requires constancy of cut-off voltage, so that the cathode-to-grid spacing of the derived tube after the triode has been scaled must be further multiplied by k to allow for the increased anode voltage. As compared with the prototype tube, therefore, the cathode-to-grid spacing of the derived tube is multiplied by $\sqrt{k}$. Table II summarizes the parameters and operating conditions of the derived tube. The numerical values in brackets again give parameters in terms of the VCR 516 with a scaling factor $k = 3$.

These solutions postulate constant beam current, the cathode loading being allowed to look after itself. In the solution of Table I the cathode loading remains unchanged. In the second solution it is multiplied by the reciprocal of the change in cathode area; i.e., by k.

For the treatment where the cathode loading is held constant and the beam current allowed to vary, the reader is referred to another paper.[5]

TABLE I

Characteristics of Derived Tube. Assuming phosphor response linearly proportional to beam voltage.

Screen diameter ..	$\times k$	600 mm
Lens-to-screen distance	$\times k$	930 mm
Scanning angle ..	$\times 1$	26°
Crossover-to-lens distance	$\times 1$	150 mm approx.
Beam width at corresponding points	$\times 1$	Beam diameter in lens plane 4.5 mm
Scale of deflector coils and neck diameter	$\times k$	Neck diameter 105 mm
Spot size at all points	$\times 1$	1 mm diameter approx.
Beam current ..	$\times 1$	150 μA
Anode voltage ..	$\times k^2$	45 kV
Scanning power ..	$\times k^2$	
Cathode-to-grid spacing	$\times k^2$	Depends on triode design
Cut-off voltage and grid drive.. ..	$\times 1$	Cut-off — 50 target. Grid drive approx. 35 V max.
Triode dimensions ..	$\times 1$	Grid hole diameter 1 mm

Numerical values refer to derived tube based on VCR 516 with scale factor $k = 3$.

Space-Charge Perturbation

The foregoing reasoning has ignored possible deviations due to space-charge repulsions at the screen. Accordingly, Fig. 3 gives space-charge-limited trajectories, calculated by the method of Thompson and Headrick[3,4]. These curves are plotted for the worse case (i.e., that based on Assumption 2) where the beam voltage is lower and where space-charge effects are therefore

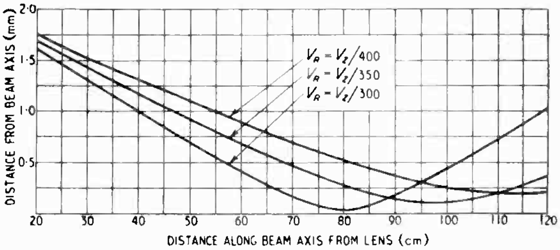


Fig. 3. *Calculated space-charge-limited trajectories for $E = 15$ kV, $I = 150$ μA and a beam diameter in the focusing lens of 4 mm.*

more severe. It will be seen from Fig. 3 that the limiting spot size due to space charge is very small in comparison with that to be expected from the usual considerations of thermal-emission spread and lens aberrations. Hence, we are entitled to assume that space charge at the screen will not vitiate the previous working.

It is unnecessary to consider space-charge at the crossover, since the anode voltage has been raised three times and we can, therefore, be certain that the space-charge influence in that region in the derived tube will be smaller than that in the prototype. (It is proved in Ref. 1, Appendix 2, that scaling the triode by the factor α causes the crossover size also to scale by factor α even in the presence of space-charge.)

Hence, we conclude that the results of Tables I and II are not upset by space-charge.

TABLE II

Characteristics of Derived Tube. Assuming phosphor response proportional to beam voltage squared.

Anode voltage ..	$\times k$	15 kV
Scanning power ..	$\times k$	
Cathode to grid spacing	$\times k$	Depends on triode design*
Triode dimensions ..	$\times \frac{1}{\sqrt{k}}$	Grid hole diameter approx. 0.6 mm

Numerical values refer to derived tube based on VCR 516 with scale factor $k = 3$. Only characteristics which differ from those of Table 1 are included.

* Adjustment made additionally to change in whole-scale of triode.

APPENDIX

Notes on Mechanical Design

The major problem is to achieve a design which will withstand the very high atmospheric pressure on the screen face and the bulb generally. It is probably not impossible to attain this end by the use of a conventional glass envelope, but apart from the obvious difficulties of weight, this would inevitably involve a very thick screen which in turn would cause severe parallax errors.

For these reasons, it is felt that a more promising approach would be to employ a glass-metal construction with continuous evacuation. Fig. 4 gives a diagrammatic sketch of a possible construction. The conical portion of the cathode-ray tube is shown at A and is made of brass, tin-plated to ensure vacuum tightness. (In order to avoid having to fabricate the large cone in brass, an alternative would be to make this portion of the bulb of a series of cylinders of decreasing diameter, these cylinders being linked together by soldering them into suitably machined diameter-reducing flanges.) The end of the bulb terminates with an accurately ground flat surface T, on to which the screen plate P of glass is waxed. Reduction of the stress in the glass plate P due to atmospheric pressure is brought about by the use of two internal metal spiders shown at B and C. These spiders are disposed at right-angles to each other and are both diameters of the screen end. They are ground on the outer surface so as to be exactly level with the

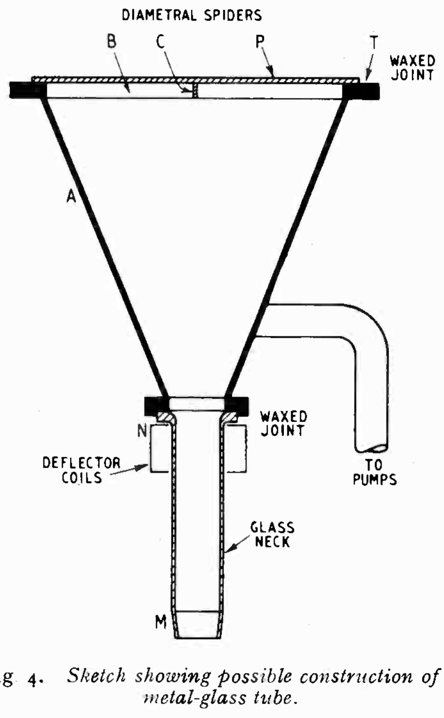


Fig 4. *Sketch showing possible construction of a metal-glass tube.*

ground flange T. Their inevitable defect of obstructing a portion of the screen area is reduced by making them thin, say 1 to 2 mm wide, in the direction of the screen surface, while the mechanical strength is ensured by making them deep in the direction of the beam. It is apparent that some experiments would be necessary to determine the degree of strengthening brought about by these spiders. For a 27-in diameter screen it is probable that a glass thickness of about $\frac{5}{8}$ in would be suitable.

On account of obstruction of the screen area, the use of such support spiders might not be admissible. In that event the screen thickness for an overall screen diameter of about 30 in, would need to be approximately 2 in. This result follows from a simple formula illustrated by Fig. 5. Here the circular disc P, of thickness t and and diameter d, is freely supported round its periphery.

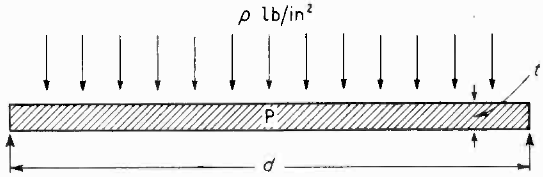


Fig. 5. *Calculation of screen thickness.*

It is supposed to be subjected to a uniform pressure of p lb/in² over its surface. Then the maximum stress induced in the disc can be shown to vary as d^2/t^2. This equation is homogeneous in d and t whence we deduce that the maximum stresses in all discs of the same *form* (i.e., same d/t ratio) are the same. We infer that the fracture point is uniquely defined by the maximum stress. Experience has shown that a disc $\frac{3}{8}$-in thick and of 6-in diameter will withstand atmospheric pressure. Thus for a 30-in diameter the required thickness appears to be just under 2 in. This argument postulates that the elastic properties of the larger disc can be made the same as for the smaller one—in particular in respect of homogeneity. Doubtless this is a postulate easier to state than to achieve, but the answer gives us a guide.

Reverting to Fig. 4, the neck of the bulb is made of glass and is terminated in a flange joint at N and a cone joint at M. Any of the standard waxed or rubber gasket joints may be made at N. The joint at M would preferably be made non-permanent, using vacuum grease, to permit ready interchange of electron guns.

It will be noted from Tables I and II that the neck diameter of the derived tubes is quite large, so that the use of an electrostatically-focused gun would seem attractive, employing that system of construction where the neck forms the final part of the two-cylinder lens. With such large neck diameters, voltage insulation is quite easy. and lenses of low aberration can be readily made. Ion burn, thus concentrated at the screen centre, is of no account on a centred p.p.i. system.

Aluminization of the screen would be necessary to avoid 'piling.' Apart from the difficulties associated with the size of the area to be covered, this technique should be relatively easy, since the screen is merely a flat disc, immediately detachable.

The general conclusion is that the constructional problems are almost wholly economic. Given proper facilities few scientific difficulties would seem likely.

REFERENCES

[1] "Simplification of Cathode Ray Tube Design by Application of Similitude Theory", Hilary Moss, *J. Televis. Soc.*, March 1946.

[2] "The Electron Gun of the Cathode Ray Tube—Part 2", Hilary Moss, *J. Brit. Instn Radio Engrs*, June 1946.

[3] "Space Charge Limitations in the Focus of Electron Beams", B. J. Thompson and L. B. Headrick. *Proc. Inst. Radio Engrs*, July 1940.

[4] "A Space Charge Problem", Hilary Moss, *Wireless Engineer*, July 1945.

[5] "Television Cathode Ray Tubes", Hilary Moss, *Wireless World*, June and July 1949.

FEEDBACK AMPLIFIER DESIGN

Conditions for Flat Response

By Hans Mayr

SUMMARY.—It is usual to treat the performance of feedback amplifiers inside the original pass-band in terms of linearization, outside of it in terms of stability; in this paper, the performance of such amplifiers is treated in the whole frequency band from zero to infinity in terms of frequency response.

After deriving a simple general equation, which gives the frequency response of the complete amplifier as a function of the response of the amplifier without feedback and of the frequency characteristics of the feedback network, a complete treatment is given for the special case of amplifiers with up to four stages of resistance-capacitance or tuned-circuit coupling, with constant feedback and equal centre frequencies for all stages.

Design formulae are given which make it possible to determine the characteristics and the values of the component parts of such amplifiers with definite pre-selected frequency response.

Introduction

THE advantages of negative feedback are well known: decreased frequency distortion, decreased harmonic distortion, decreased background noise and increased stability, inasmuch as the amplification is much less dependent on valve characteristics and supply voltages. Its drawbacks are the loss of amplification and a certain instability at very high or very low frequencies, which may lead even to self-oscillation.

Now the loss of amplification, inherent in the system of negative feedback, is the price we must pay for its other advantages; but since the available output power is not diminished, it is generally quite easy to provide for the necessary additional amplification.

On the other hand, the instability at extreme frequencies is not at all an inevitable characteristic of negative feedback, but merely a consequence of the unsuitable design method ordinarily employed. In fact the usual procedure is, first, to design an amplifier with somewhat poor performance, but with a gain in excess of the desired one by the amount of the expected feedback; secondly, to apply to this amplifier negative feedback until the desired amplification is reached; thirdly, if instability occurs, to apply some stabilizing means and to adjust it by cut and try methods until the whole device becomes sufficiently stable. What really happens is this: the application of negative feedback to the amplifier enlarges its original pass-band; the central part of the response curve is straightened, but in the neighbourhood of its limits there may appear peaks; if these peaks reach infinite gain, self-oscillation occurs. Stabilization requires the response curve of the original amplifier to be modified so as to lower the resulting peaks.

MS accepted by the Editor, August 1948

Using a correct design method it is not only possible to obtain perfectly stable amplifiers, but even to realize any pre-selected response curve and bandwidth compatible with the type of amplifier in question. Such a design method should start with the choice of the desired response curve, then determine the characteristics of the amplifier and feedback network necessary to obtain this response and at last give the values of the component parts needed for these characteristics. Obviously it is first necessary to know all the response curves that can be realized with the particular type of amplifier in question, as functions of some suitable parameters; then, to know the relation existing between these parameters and the characteristics of the amplifier; finally, to determine the values that the various parts of the amplifier must have in order to obtain the desired characteristics.

In the following we shall at first derive a general equation which gives the response curve of a feedback amplifier if the response of the amplifier without feedback and the frequency attenuation of the feedback network are known. Then we shall treat in detail the special case of an amplifier with up to four stages of resistance-capacitance or tuned-circuit coupling, with constant feedback; all stages are supposed to be tuned to the same frequency. The amplifier will be designed to give the most uniform amplification possible. Though apparently a very special case, it is probably applicable to the majority of practical problems, especially in the field of measurements.

Response of Feedback Amplifiers

Denoting by V_i, V_o, I_i, I_o respectively the input voltage, the output voltage, the input current and the output current of an amplifier, its amplification may be expressed in terms of any one of the following four ratios:

V_o/V_i . . . voltage gain
I_o/I_i . . . current gain
V_o/I_i . . . mutual impedance
I_o/V_i . . . mutual admittance.

Of these four expressions the first two are generally complex numbers, while the others are, as indicated, a complex impedance and a complex admittance.

In the following we shall use the term 'amplification' indifferently for any one of the four ratios; by 'input' and 'output' we denote the corresponding quantities. For instance, whenever 'amplification' is intended to mean the mutual impedance, 'input' means input current and 'output' means output voltage.

Negative feedback consists in feeding a fraction of the output through a suitable network back to the input terminals, so as to oppose the action of the input. In order to maintain the original output level, the input must now be increased by an amount equal to the fraction of the output which has been fed back. Denoting by α the amplification without feedback, by A the amplification with feedback and by β the fraction of the output fed back to the input terminals, we obtain the well known equation

$$A = \frac{\alpha}{1 + \alpha\beta} \quad \ldots \quad (1)$$

Since α as well as β are functions of frequency, equation (1) gives not only the change in the value of the amplification, but in its frequency characteristic, too. By choosing a suitable set of reference values α_0, β_0, A_0 corresponding to the same reference frequency $\omega_0/2\pi$, the two effects may be separated; we find

$$A_0 = \frac{\alpha_0}{1 + \alpha_0\beta_0} \quad \ldots \quad (2)$$

$$\frac{A}{A_0} = \frac{\alpha}{\alpha_0} \cdot \frac{1 + \alpha_0\beta_0}{1 + \alpha\beta} \quad \ldots \quad (3)$$

Now, the reference value of amplification with feedback A_0 is reduced with respect to the value of amplification without feedback α_0 by the factor $1 + \alpha_0\beta_0$. We call this factor the degree of negative feedback and denote it by n; it is generally a complex constant.

Introducing this constant in the equations (2) and (3) we obtain after a simple transformation

$$A_0 = \frac{\alpha_0}{n} \quad \ldots \quad (4)$$

$$\frac{A_0}{A} = 1 + \frac{1}{n}\left(\frac{\alpha_0}{\alpha} - 1\right) + \frac{n-1}{n}\left(\frac{\beta}{\beta_0} - 1\right) \quad (5)$$

Defining as frequency response the frequency characteristic of amplification referred to some arbitrary reference value, and defining further as frequency attenuation the reciprocal of the frequency response, we see that equation (5) gives directly the frequency attenuation of the amplifier with feedback, if the frequency attenuation of the amplifier without feedback and the frequency attenuation of the feedback network are known.

For the mathematical treatment, it is preferable to use the frequency attenuation instead of the frequency response, because it avoids the cumbersome fractions; but, the response being a more immediate expression for the performance of an amplifier, as is readily recognized by observing the graphical representation, it is advisable to express the results in terms of response.

An inspection of equation (5) reveals immediately two well known relations: the influence of the term $\frac{\alpha_0}{\alpha} - 1$, due to the amplifier proper, is reduced n times, indicating the degree of stabilization obtained by the feedback; furthermore, if n approaches infinity, equation (5) reduces to

$$\frac{A_0}{A} \approx \frac{\beta}{\beta_0} \quad \ldots \quad (5a)$$

indicating that in the limit the performance of the amplifier is determined by the feedback network only and completely independent of the characteristics of the amplifier itself.

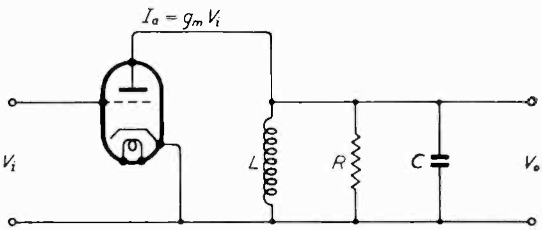


Fig. 1. *Tuned amplifier stage.*

Amplifiers with Constant Feedback

An amplifier is usually intended to give uniform response within its pass-band. As equation (5a) shows, this condition is approached, at least for the higher degrees of feedback, by making the feedback constant and independent of frequency. In this case equation (5) simplifies to

$$\frac{A_0}{A} = 1 + \frac{1}{n}\left(\frac{\alpha_0}{\alpha} - 1\right) \quad \ldots \quad (6$$

In order to determine the frequency attenuation of the amplifier with feedback it is now sufficient to know the frequency attenuation of the amplifier without feedback and the degree of feedback applied.

Frequency Attenuation of Amplifier Stage

Fig. 1 gives the schematic diagram of a one-stage amplifier with tuned-circuit coupling. The resistance R is supposed to lump the load, the anode resistance of the valve and the losses of the tuned circuit.

The frequency attenuation of this stage may be expressed by

$$\frac{g_m V_i}{V_0} = \frac{1}{R} + j\omega C + \frac{1}{j\omega L} \quad \ldots \quad \ldots \quad (7)$$

Introducing the abbreviations

$$\alpha = \frac{V_0}{V_i} \quad Q = R\sqrt{\frac{C}{L}} \quad \omega_0 = \frac{1}{\sqrt{LC}} \quad \ldots \quad (8)$$

we obtain the following simple expressions:

$$\alpha_0 = g_m R \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (9a)$$

$$\frac{\alpha_0}{\alpha} = 1 + jQ\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right) \quad \ldots \quad \ldots \quad (9b)$$

The two parameters of equation(9), ω_0 and Q, are theoretically unlimited.

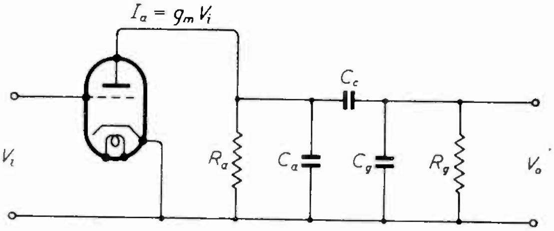


Fig. 2. *RC-coupled amplifier stage.*

Fig. 2 gives the schematic diagram of a one-stage amplifier with resistance-capacitance coupling. Here the anode resistance of the valve is lumped with R_a.

The frequency attenuation is given by

$$\frac{g_m V_i}{V_0} = \frac{1}{R_a}\left(1 + \frac{C_g}{C_c}\right) + \frac{1}{R_g}\left(1 + \frac{C_a}{C_c}\right) + j\omega\left(C_g + C_a + \frac{C_g C_a}{C_c}\right) + \frac{1}{j\omega C_c R_g R_a} \quad (10)$$

Introducing the following set of abbreviations

$$\left.\begin{aligned} &R = \sqrt{R_a R_g}; \\ &C = \sqrt{C_g C_c + C_g C_a + C_c C_a}; \quad \omega_0 = \frac{1}{CR} \\ &b = \sqrt{\frac{R_g}{R_a}}; \quad d = \frac{C_g}{C}; \quad k = \frac{C_c}{C}; \\ &p = \frac{C_a}{C}; \quad Q = \frac{1}{b(d+k) + \frac{1}{b}(p+k)} \end{aligned}\right\} (11)$$

we have

$$\alpha_0 = k\,Q \,.\, g_m R \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (12a)$$

$$\frac{\alpha_0}{\alpha} = 1 + jQ\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right) \quad \ldots \quad \ldots \quad (12b)$$

Between the quantities d, k, p exists the relation

$$dk + dp + kp = 1 \quad \ldots \quad \ldots \quad \ldots \quad (13)$$

Of the two parameters of equation(12), ω_0 is still theoretically unlimited; but Q is confined to values lower than 1/2, as may easily be verified by differentiating the expression for Q with respect to its variables and equating to zero.

The interesting result is that the frequency attenuation of the tuned-circuit amplifier and the resistance-capacitance coupled amplifier are identical and completely determined by only two parameters, ω_0 and Q.

Frequency Attenuation of Multi-stage Amplifiers with Feedback

The frequency attenuation of a multi-stage amplifier is found by simply multiplying the attenuation characteristics of its single stages. The result is particularly simple if the various stages have the same resonant frequency $\frac{\omega_0}{2\pi}$.

Introducing the result in equation (6) we find the frequency attenuation of a multi-stage amplifier with constant feedback. We obtain for amplifiers with up to four stages

$$\frac{A_0}{A} = 1 + j\frac{1}{n}Q\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right) \quad \ldots \quad \ldots \quad (14a)$$

$$\frac{A_0}{A} = 1 + j\frac{1}{n}(Q_1 + Q_2)\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right) - \frac{1}{n}Q_1Q_2\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right)^2 \quad \ldots \quad \ldots \quad (14b)$$

$$\frac{A_0}{A} = 1 + j\frac{1}{n}(Q_1 + Q_2 + Q_3)\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right) - \frac{1}{n}(Q_1Q_2 + Q_1Q_3 + Q_2Q_3)\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right)^2 - j\frac{1}{n}Q_1Q_2Q_3\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right)^3 \quad \ldots \quad (14c)$$

$$\begin{aligned}\frac{A_0}{A} = 1 &+ j\frac{1}{n}(Q_1 + Q_2 + Q_3 + Q_4)\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right) \\ &- \frac{1}{n}(Q_1Q_2 + Q_1Q_3 + Q_1Q_4 + Q_2Q_3 + Q_2Q_4 + Q_3Q_4)\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right)^2 \\ &- j\frac{1}{n}(Q_1Q_2Q_3 + Q_1Q_2Q_4 + Q_1Q_3Q_4 + Q_2Q_3Q_4)\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right)^3 \\ &+ \frac{1}{n}Q_1Q_2Q_3Q_4\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right)^4 \quad (14d)\end{aligned}$$

For practical applications it is better to pass to polar co-ordinates and determine the modulus and phase angle of attenuation separately. Using an appropriate set of abbreviations it is possible to express these quantities in quite a simple manner. We find, in fact, the following equations:

for one stage:

$$\left|\frac{A_0}{A}\right| = \sqrt{1 + x^2} \quad \ldots \quad (15a)$$

$$\phi = -\tan^{-1} x \quad \ldots \quad (15b)$$

$$x = \frac{Q}{n}\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right) \quad \ldots \quad (15c)$$

for two stages:

$$\left|\frac{A_0}{A}\right| = \sqrt{1 + a_1 x^2 + x^4} \quad \ldots \quad (16a)$$

$$\phi = -\tan^{-1} \frac{b_1 x}{1 - x^2} \quad \ldots \quad (16b)$$

$$x = \sqrt{\frac{Q_1 Q_2}{n}}\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right) \quad \ldots \quad (16c)$$

$$a_1 = \frac{Q_1^2 + Q_2^2 - 2(n-1)Q_1 Q_2}{n Q_1 Q_2} \quad \ldots \quad (16d)$$

$$b_1 = \frac{Q_1 + Q_2}{\sqrt{n Q_1 Q_2}} \quad \ldots \quad (16e)$$

for three stages:

$$\left|\frac{A_0}{A}\right| = \sqrt{1 + a_1 x^2 + a_2 x^4 + x^6} \quad \ldots \quad (17a)$$

$$\phi = -\tan^{-1} \frac{b_1 x - x^3}{1 - b_2 x^2} \quad \ldots \quad (17b)$$

$$x = \sqrt[3]{\frac{Q_1 Q_2 Q_3}{n}}\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right) \quad \ldots \quad (17c)$$

$$a_1 = \frac{Q_1^2 + Q_2^2 + Q_3^2 - 2(n-1)(Q_1 Q_2 + Q_1 Q_3 + Q_2 Q_3)}{\sqrt[3]{n^4 Q_1^2 Q_2^2 Q_2^3}} \quad \ldots \quad (17d)$$

$$a_2 = \frac{Q_1^2 Q_2^2 + Q_1^2 Q_3^2 + Q_2^2 Q_3^2}{\sqrt[3]{n^2 Q_1^4 Q_2^4 Q_3^4}} \quad \ldots \quad (17e)$$

$$b_1 = \frac{Q_1 + Q_2 + Q_3}{\sqrt[3]{n^2 Q_1 Q_2 Q_3}} \quad \ldots \quad (17f)$$

$$b_2 = \frac{Q_1 Q_2 + Q_1 Q_3 + Q_2 Q_3}{\sqrt[3]{n Q_1^2 Q_2^2 Q_3^2}} \quad \ldots \quad (17g)$$

for four stages:

$$\left|\frac{A_0}{A}\right| = \sqrt{1 + a_1 x^2 + a_2 x^4 + a_3 x^6 + x^8} \quad (18a)$$

$$\phi = -\tan^{-1} \frac{b_1 x - b_3 x^3}{1 - b_2 x^2 + x^4} \quad \ldots \quad (18b)$$

$$x = \sqrt[4]{\frac{Q_1 Q_2 Q_3 Q_4}{n}}\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right) \quad \ldots \quad (18c)$$

$$a_1 = \frac{Q_1^2 + Q_2^2 + Q_3^2 + Q_4^2 - 2(n-1)(Q_1 Q_2 + Q_1 Q_3 + Q_1 Q_4 + Q_2 Q_3 + Q_2 Q_4 + Q_3 Q_4)}{\sqrt{n^3 Q_1 Q_2 Q_3 Q_4}} \quad (18d)$$

$$a_2 = \frac{Q_1^2 Q_2^2 + Q_1^2 Q_3^2 + Q_1^2 Q_4^2 + Q_2^2 Q_3^2 + Q_2^2 Q_4^2 + Q_3^2 Q_4^2 + 2(n-1)Q_1 Q_2 Q_3 Q_4}{n\, Q_1 Q_2 Q_3 Q_4} \quad \ldots \quad (18e)$$

$$a_3 = \frac{Q_1^2 Q_2^2 Q_3^2 + Q_1^2 Q_2^2 Q_4^2 + Q_1^2 Q_3^2 Q_4^2 + Q_2^2 Q_3^2 Q_4^2}{\sqrt{n Q_1^3 Q_2^3 Q_3^3 Q_4^3}} \quad \ldots \quad (18f)$$

$$b_1 = \frac{Q_1 + Q_2 + Q_3 + Q_4}{\sqrt[4]{n^3 Q_1 Q_2 Q_3 Q_4}} \quad \ldots \quad (18g)$$

$$b_2 = \frac{Q_1 Q_2 + Q_1 Q_3 + Q_1 Q_4 + Q_2 Q_3 + Q_2 Q_4 + Q_3 Q_4}{\sqrt{n Q_1 Q_2 Q_3 Q_4}} \quad \ldots \quad (18h)$$

$$b_3 = \frac{Q_1 Q_2 Q_3 + Q_1 Q_2 Q_4 + Q_1 Q_3 Q_4 + Q_2 Q_3 Q_4}{\sqrt[4]{n Q_1^3 Q_2^3 Q_3^3 Q_4^3}} \quad \ldots \quad (18i)$$

Since for acoustical reproduction the phase characteristic is of less importance than the amplitude characteristic, we base our considerations on the latter.

The amplitude characteristics of the multi-stage amplifiers may be expressed by one common general formula; indeed, equations (15a), (16a), (17a) and (18a) may all be expressed by

$$\left|\frac{A_0}{A}\right| = \sqrt{1 + a_1 x^2 + a_2 x^4 + \ldots + a_{r-1} x^{2(r-1)} + x^{2r}} \quad \ldots \quad (19)$$

where r is the number of stages.

Since equation (19) is an even function of x, all the characteristics are symmetrical with respect to $x = 0$ or $\omega = \omega_0$. In order to find further details of the shape of the attenuation characteristics, we determine the maximum and minimum values of (19). This is done by differentiating equation (19) with respect to x and equating to zero. We get

$$x = 0 \quad \ldots \quad (20)$$

and

$$a_1 + 2a_2 x^2 + \ldots + (r-1)\, a_{r-1} x^{2(r-2)} + r\, x^{2(r-1)} = 0 \quad \ldots \quad (21)$$

The roots of (21) correspond to the maximum or minimum values of (19). Expressing (21) as the product of its root factors, we may write the identity

$$a_1 + 2a_2 x^2 + \ldots + (r-1) a_{r-1} x^{2(r-2)} + r\, x^{2(r-1)} \equiv r\,(x^2 - x_1^2)(x^2 - x_2^2) \ldots (x^2 - x_{r-1}^2) \quad \ldots \quad (22)$$

from which we get immediately

$$a_{r-1} = -\frac{r}{r-1}(x_1^2 + x_2^2 + \ldots + x_{r-1}^2) \quad \ldots \ldots \ldots \ldots (23)$$

Equation (23) shows that there exists at least one pair of complex roots, if a_{r-1} is positive. But from equations (16) to (18) we learn that with the exception of a_1, which may be either positive, zero or negative, all other coefficients are always positive. Since in the case of four stages equation (21) is of the sixth degree, it has a total of six roots, of which at least two are complex and consequently not more than four real; in addition we have the real root at $x = 0$. Now, between two consecutive maximum values there must be a minimum value and therefore of the five possible real roots three at most may correspond to maximum values of the response curve; in this case we have a peak at the centre frequency and two other peaks near the limits of the pass band, one at a very low and the other at a very high frequency.

In the same way it may be shown that for three and two stages there cannot exist more than two peaks near the limits of the pass band, one at a very low and the other at a very high frequency; for one stage, there exists only one peak at the centre frequency. We may therefore conclude that for amplifiers with up to four stages, the response curve cannot have more than two peaks near the limits of the pass band, one near the upper and the other near the lower limit.

We have seen that, with the exception of a_1, all coefficients of equation (21) are positive. If a_1 is also positive, or zero, equation (21) has no real roots at all and the response curve has no peaks except that at the centre frequency.

The behaviour of the response curve near the centre of the pass-band may be shown by expanding equation (19) in a power series. Expanding first in binomial series and then in Maclaurin's series we find:

$$\left|\frac{A_0}{A}\right| = \sqrt{1 + a_1x^2 + a_2x^4 + \ldots}$$

$$= 1 + \tfrac{1}{2}a_1x^2 - \tfrac{1}{8}(a_1^2 - 4a_2)x^4 + \ldots$$

$$= 1 + \frac{1}{2}\frac{d^2}{dx^2}\left|\frac{A_0}{A}\right| \cdot x^2 + \frac{1}{24}\frac{d^4}{dx^4}\left|\frac{A_0}{A}\right| \cdot x^4 + \ldots \quad \ldots \ldots \ldots (24)$$

Consequently we obtain for

$$a_1 = 0 \qquad \frac{d^2}{dx^2}\left|\frac{A_0}{A}\right| = 0 \quad \ldots \ldots (25a)$$

$$a_1 = a_2 = 0 \quad \frac{d^4}{dx^4}\left|\frac{A_0}{A}\right| = 0 \quad \ldots \ldots (25b)$$

and so on.

The result is that the more coefficients of equation (19) vanish, the flatter becomes the response curve near the centre of the pass band. But as we have already seen, only a_1 can be made to vanish, while all other coefficients are always positive. In order to obtain the flattest response curve possible, it is therefore necessary to make these other coefficients as small as possible.

In conclusion we may say that by making a_1 zero and the other coefficients as small as possible we obtain perfectly smooth response curves without any spurious peaks; since these curves are very flat in the neighbourhood of the centre frequency and rather steep near the limits of the band, they represent a very good approximation to the rectangular response curve usually considered ideal. Furthermore, it is evident that amplifiers with such a response are perfectly stable in the whole frequency band from zero to infinity.

Feedback Amplifiers with Flat Response

Applying these principles to amplifiers with one stage, we see at once from equation (15a) that there are no arbitrary coefficients at all and that therefore the response curve when expressed as a function of the variable x is the same for any one-stage amplifier.

In the case of two-stage amplifiers, we learn from equation (16a) that there is but one arbitrary constant, a_1. According to (16d) this may be made to vanish, if the following relation is satisfied:

$$\frac{Q_1}{Q_2} = n - 1 \pm \sqrt{n(n-2)} \quad \ldots \ldots (26)$$

Equation (26) is valid only if n is larger than 2, since it is of course physically impossible to realize complex values of Q.

For three-stage amplifiers, we have according to equation (17a) two coefficients a_1 and a_2. With the aid of equation (17d), a_1 may be made to vanish, yielding

$$1 + \frac{Q_2^2}{Q_1^2} + \frac{Q_3^2}{Q_1^2} - 2(n-1)\left(\frac{Q_2}{Q_1} + \frac{Q_3}{Q_1} + \frac{Q_2Q_3}{Q_1^2}\right) = 0 \quad \ldots (27)$$

For a_2 we find from equation (17e)

$$a_2 = \sqrt[3]{\frac{Q_2^2Q_3^2}{n^2Q_1^4}}\left(1 + \frac{Q_1^2}{Q_2^2} + \frac{Q_1^2}{Q_3^2}\right) \quad \ldots (28)$$

which is always positive. We see that a_2 is a function of two variable parameters, Q_2/Q_1 and Q_3/Q_1. By means of equation (27) Q_3/Q_1 may be expressed as a function of Q_2/Q_1 and so there remains but one independent variable, Q_2/Q_1. In order to find the smallest value that a_2 can assume, the ordinary procedure would be to differentiate (28) with respect to the variable Q_2/Q_1 and then to equate to zero. This would give the value of Q_2/Q_1; inserting this in (27) would then give Q_3/Q_1.

However, in this special case there is a much simpler way of finding the answer.

Putting $Q_2/Q_1 = \xi$ and $Q_3/Q_1 = \eta$ we may write for equation (27)

$$\Psi(\xi,\eta) = 0 \quad \ldots \quad (27a)$$

and for equation (28)

$$a_2 = F(\xi,\eta) \quad \ldots \quad (28a)$$

where Ψ and F are symmetrical in ξ and η. We introduce now a new co-ordinate system with the straight line $\xi = \eta$ as u-axis and with the v-axis normal to it. We get

$$\Psi(\xi,\eta) = \psi(u,v) \quad \ldots \quad (29)$$

$$F(\xi,\eta) = f(u,v) \quad \ldots \quad (30)$$

The necessary, though not sufficient, conditions for a minimum value of a_2 are now

$$\frac{\partial f}{\partial v} + \frac{\partial f}{\partial u}\cdot\frac{du}{dv} = 0 \quad \ldots \quad (31)$$

The relation $\psi(u, v) = 0$ represents, because of its symmetry in ξ and η, a curve which is symmetrical about the u-axis. At its intersection with this axis therefore the first-order derivative, taken in a direction normal to the axis vanishes. This yields

$$\left(\frac{du}{dv}\right)_{v=0} = 0 \quad \ldots \quad (32)$$

The expression $f = F(\xi, \eta)$ represents in the coordinate system F, ξ, η a surface which is symmetrical to the plane $\xi = \eta$; therefore the expression $f(u, v)$ is symmetrical about the plane $v = 0$. In this plane therefore the first order derivative taken in a direction normal to the plane vanishes and we get

$$\left(\frac{\partial f}{\partial v}\right)_{v=0} = 0 \quad \ldots \quad (33)$$

From (32) and (33) it follows that equation (31) is fulfilled for $v = 0$, so that $f(u, v)$ is a minimum. Therefore $F(\xi, \eta)$ is a minimum for $\xi = \eta$ and, finally, a_2 is a minimum for $Q_2 = Q_3$.

The strict proof that the value found by this method is truly a minimum and not a maximum, or flexure point, would be rather laborious; but since all sections of a_2 parallel either to ξ or to η possess a minimum value, but no maximum or flexure point, it is quite evident that our value is a minimum.

Putting in equation (27) $Q_3 = Q_2$ we find the following relation:

$$\frac{Q_1}{Q_2} = 2(n-1) + \sqrt{2n(2n-3)} \quad \ldots \quad (34)$$

The procedure for four-stage amplifiers is analogous to that used for three-stage amplifiers. Of the three coefficients present in equation (18a) a_1 can be made to vanish, while a_2 and a_3 are always positive, as is evident from equations (18e) and (18f); the latter two coefficients assume simultaneous minimum values if $Q_2 = Q_3 = Q_4$. This may be proved following the same lines as in the case of three stages, with the difference that recourse to four dimensional geometry is necessary.

Inserting $Q_2 = Q_3 = Q_4$ in equation (18d) and equating to zero yields

$$\frac{Q_1}{Q_2} = 3(n-1) + \sqrt{3n(3n-4)} \quad \ldots \quad (35)$$

We have thus found the conditions necessary to obtain the flattest possible response curves for amplifiers with one, two, three and four stages; applying these results to equations (15) to (18) we obtain the following formulae;

for one stage:

$$\left|\frac{A_0}{A}\right| = \sqrt{1 + x^2} \quad \ldots \quad (36a)$$

$$\phi = -\tan^{-1} x \quad \ldots \quad (36b)$$

$$x = \frac{Q}{n}\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right) \quad \ldots \quad (36c)$$

for two stages:

$$\left|\frac{A_0}{A}\right| = \sqrt{1 + x^4} \quad \ldots \quad (37a)$$

$$\phi = -\tan^{-1}\frac{\sqrt{2}x}{1 - x^2} \quad \ldots \quad (37b)$$

$$x = Q_2\sqrt{\frac{Q_1}{nQ_2}}\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right) \quad \ldots \quad (37c)$$

$$\frac{Q_1}{Q_2} = (n-1) + \sqrt{n(n-2)} \quad \ldots \quad (37d)$$

for three stages:

$$\left|\frac{A_0}{A}\right| = \sqrt{1 + a_2x^4 + x^6} \quad \ldots \quad (38a)$$

$$\phi = -\tan^{-1}\frac{b_1x - x^3}{1 - b_2x^2} \quad \ldots \quad (38b)$$

$$x = Q_2\sqrt[3]{\frac{Q_1}{nQ_2}}\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right) \quad \ldots \quad (38c)$$

$$a_2 = \frac{1}{\sqrt[3]{n^2}}\left(\sqrt[3]{\frac{Q_2^4}{Q_1^4}} + 2\sqrt[3]{\frac{Q_1^2}{Q_2^2}}\right) \quad \ldots \quad (38d)$$

$$b_1 = \frac{1}{\sqrt[3]{n^2}}\left(\sqrt[3]{\frac{Q_1^2}{Q_2^2}} + 2\sqrt[3]{\frac{Q_2}{Q_1}}\right) \quad \ldots \quad (38e)$$

$$b_2 = \frac{1}{\sqrt[3]{n}}\left(\sqrt[3]{\frac{Q_2^2}{Q_1^2}} + 2\sqrt[3]{\frac{Q_1}{Q_2}}\right) \quad \ldots \quad (38f)$$

$$\frac{Q_1}{Q_2} = 2(n-1) + \sqrt{2n(2n-3)} \quad \ldots \quad (38g)$$

$$Q_2 = Q_3 \quad \ldots \quad (38h)$$

for four stages:

$$\left|\frac{A_0}{A}\right| = \sqrt{1 + a_2x^4 + a_3x^6 + x^8} \quad .. \quad (39a)$$

$$\phi = -\tan^{-1}\frac{b_1x - b_3x^3}{1 - b_2x^2 + x^4} \quad .. \quad (39b)$$

$$x = Q_2\sqrt[4]{\frac{Q_1}{nQ_2}}\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right) \quad .. \quad .. \quad (39c)$$

$$a_2 = \frac{1}{n}\left[2(n-1) + 3\frac{Q_1}{Q_2} + 3\frac{Q_2}{Q_1}\right] \quad .. \quad (39d)$$

$$a_3 = \frac{1}{\sqrt{n}}\left[\sqrt{\frac{Q_2^3}{Q_1^3}} + 3\sqrt{\frac{Q_1}{Q_2}}\right] \quad .. \quad (39e)$$

$$b_1 = \frac{1}{\sqrt[4]{n^3}}\left[\sqrt[4]{\frac{Q_1^3}{Q_2^3}} + 3\sqrt[4]{\frac{Q_2}{Q_1}}\right] \quad .. \quad (39f)$$

$$b_2 = \frac{1}{\sqrt{n}}\left[3\sqrt{\frac{Q_1}{Q_2}} + 3\sqrt{\frac{Q_2}{Q_1}}\right] \quad .. \quad (39g)$$

$$b_3 = \frac{1}{\sqrt[4]{n}}\left[\sqrt[4]{\frac{Q_2^3}{Q_1^3}} + 3\sqrt[4]{\frac{Q_1}{Q_2}}\right] \quad .. \quad (39h)$$

$$\frac{Q_1}{Q_2} = 3(n-1) + \sqrt{3(3n-4)} \quad .. \quad (39i)$$

$$Q_2 = Q_3 = Q_4 \quad .. \quad .. \quad .. \quad (39j)$$

From equations (36) and (37) we see that the frequency response of one-stage and two-stage amplifiers as a function of x is independent of the degree of negative feedback, n. It is therefore possible to represent the modulus as well as the phase angle of response, for these cases, each by a single curve. This has been done in Figs. 3 and 4. For three and four stages, however, the response curves are different for different values of feedback. Since the influence of the degree of feedback on the shape of the response curve is not very great, it is sufficient to give these curves for the limit values of n; these limits are, for three stages, $n = 3/2$ and $n = \infty$; for four stages, $n = 4/3$ and $n = \infty$. These curves are given in Figs. 5 to 8.

The expressions for the modulus of the frequency response are even functions of x and therefore symmetrical about the axis $x = 0$. The figures reproduce only the right half of the curves, for positive values of x. The response values for the symmetrical left half of the curve, corresponding to negative values of x, are obtained by simply changing the sign of the abscissae.

The expressions for the phase angle of the frequency response are odd functions of x and therefore symmetrical about the origin $x = 0$, $\phi = 0$. The figures reproduce likewise only the right half of the curves, for positive values of x. The phase angle values for the left half of the curve, corresponding to negative values of x, are obtained by changing the sign of the abscissae as well as of the ordinate. The phase angles of the frequency response for negative values of x are therefore always positive.

Design of Feedback Amplifiers with given Response

With the equations (36) to (39) and the curves of Figs. 3 to 8 we have attained our first object, namely to determine the flattest frequency response curves that can be realized with amplifiers with up to four stages of resistance-capacitance or tuned-circuit coupling, using constant feedback and having all the stages tuned to the same frequency. The parameters used are: the degree of negative feedback, n; the Qs of the various stages; and the common centre frequency, $\omega_0/2\pi$.

The next step is to find the numerical values of these parameters corresponding to a definite amplifier with a frequency response selected among the possible curves given above. This may be done with the aid of equations (36) to (39), as follows:

We choose the number of stages, r, the degree of negative feedback, n, and the response (modulus or phase angle) corresponding to two arbitrary frequencies, $\omega_1/2\pi$ and $\omega_2/2\pi$, one near the lower and the other near the upper frequency limit. Then we read from the curves the values x_1 and x_2 corresponding to the two chosen response values; the value of x_1 corresponding to the lower frequency is, of course, negative.

Now we compute the centre frequency from

$$\omega_0^2 = \omega_1\omega_2\frac{x_2\omega_1 - x_1\omega_2}{x_2\omega_2 - x_1\omega_1} \quad .. \quad .. \quad (40)$$

and the ratio Q_1/Q_2 from equations (37d), (38g) or (39i) according to the number of stages.

The Qs of the various stages are then given by:

one stage:

$$Q = n\frac{x_1}{\dfrac{\omega_1}{\omega_0} - \dfrac{\omega_0}{\omega_1}} \quad .. \quad .. \quad .. \quad (41)$$

two stages:

$$Q_2 = \sqrt{\frac{nQ_2}{Q_1}}\frac{x_1}{\dfrac{\omega_1}{\omega_0} - \dfrac{\omega_0}{\omega_1}};\ Q_1 = Q_2\cdot\frac{Q_1}{Q_2} \quad .. \quad (42)$$

three stages:

$$Q_2 = Q_3 = \sqrt[3]{\frac{nQ_2}{Q_1}}\frac{x_1}{\dfrac{\omega_1}{\omega_0} - \dfrac{\omega_0}{\omega_1}};\ Q_1 = Q_2\cdot\frac{Q_1}{Q_2} \quad .. \quad .. \quad (43)$$

four stages:

$$Q_2 = Q_3 = Q_4 = \sqrt[4]{\frac{nQ_2}{Q_1}}\frac{x_1}{\dfrac{\omega_1}{\omega_0} - \dfrac{\omega_0}{\omega_1}};$$

$$Q_1 = Q_2\cdot\frac{Q_1}{Q_2} \quad .. \quad .. \quad .. \quad (44)$$

The above design formulae show that these amplifiers are made up of one rather selective stage, corresponding to Q_1, and a number of equal broadly tuned stages, corresponding to Q_2, Q_3. . . . . With increasing feedback, the selectivity of the former must be increased, while the selectivity of the latter stages approaches a limiting value. It may be shown that this limiting value is of the order of the Q of a single-stage amplifier without feedback, having the same bandwidth as the complete amplifier with feedback.

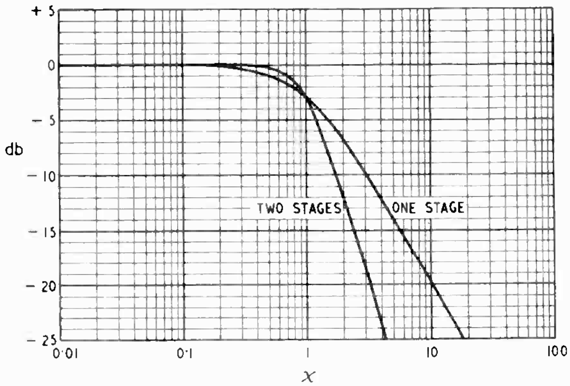


Fig. 3. *Modulus of frequency response for one and two stages.*

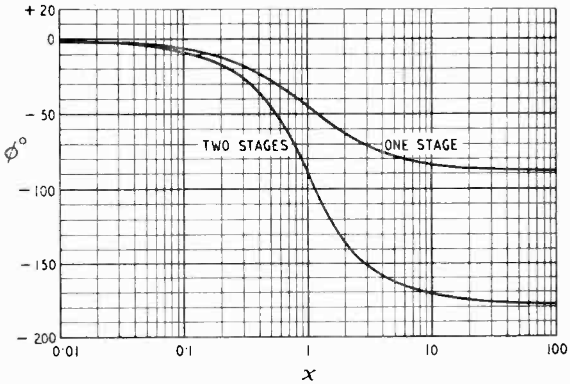


Fig. 4. *Phase angle of frequency response for one and two stages.*

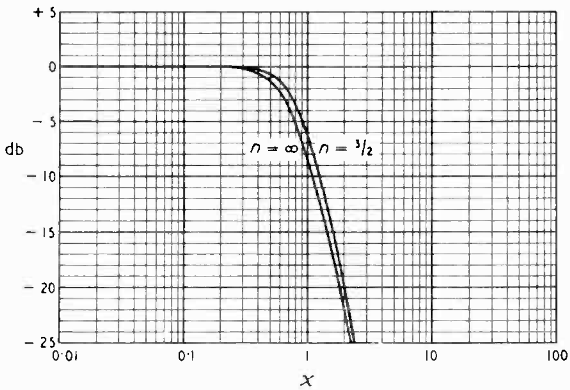


Fig. 5. *Modulus of frequency response for three stages.*

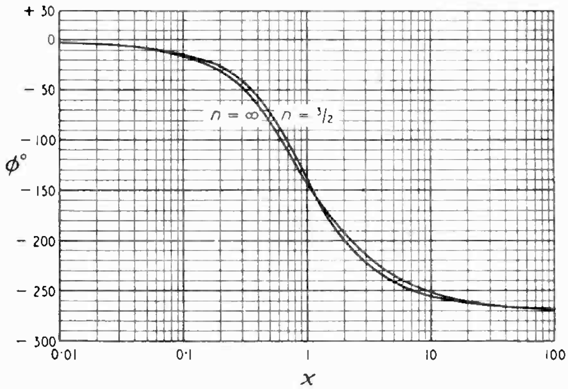


Fig. 6. *Phase angle of frequency response for three stages.*

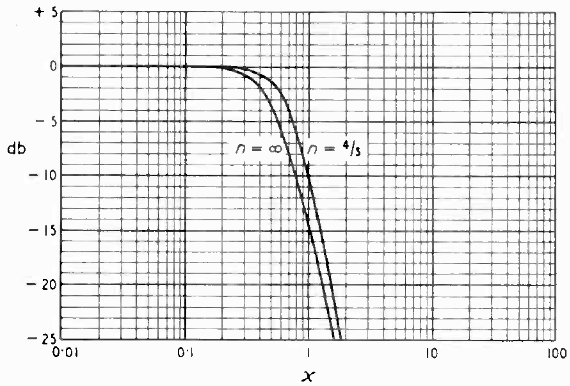


Fig. 7. *Modulus of frequency response for four stages.*

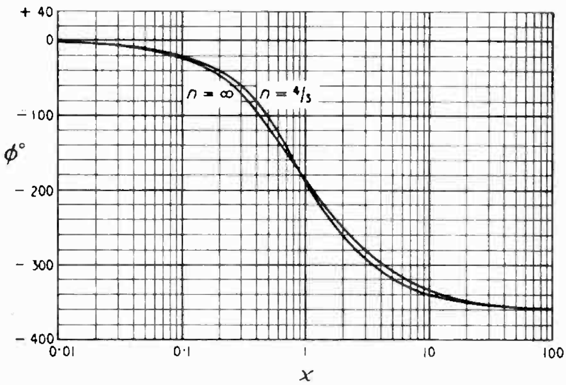


Fig. 8. *Phase angle of frequency response for four stages.*

Determination of Component Values

Given the values of Q and ω_0 for every stage of the amplifier, its frequency characteristics are completely determined. The third and last step is now to find the values of the component parts of the amplifier which yield the desired values of Q and ω_0.

Theoretically, any value of Q may be realized with a tuned-circuit-coupled amplifier stage. However, a practical upper limit is set by the highest Q that can be realized with available inductors. For low-frequency work, this limit is of the order of 50. If the desired Q is lower than 1/2, it is usually convenient to use resistance-capacitance coupling.

The amplifier is at first designed following standard methods, but neglecting all reactance elements; i.e., for tuned-circuit stages we consider only the total resistance R of Fig. 1 and for resistance-capacitance stages the anode and grid resistances R_a and R_g of Fig. 2. The amplifier is designed to have a gain approximately n times the desired final value, considering that according to equation (12a) the real gain of a resistance-capacitance coupled stage is somewhat less than the theoretical gain with a pure resistance load; for wide-band stages, however, the difference is so slight that it may be neglected.

The feedback circuit must be so designed as to decrease the gain n times. It must be independent of frequency, and the best plan is to use a resistive network. For high-frequency or wide-band amplifiers it will be necessary to shunt the low resistances by suitable capacitances in order to compensate for the inevitable stray capacitances of the higher resistors.

For tuned-circuit stages, the reactance elements are then determined from the following formulae, which may be derived from equation (8):

$$L = \frac{R}{\omega_0 Q}; \qquad C = \frac{Q}{\omega_0 R} \quad \ldots \quad \ldots \quad (45)$$

The design of resistance-capacitance coupled stages is somewhat more complicated. Since only two conditions are to be met by three capacitances, an arbitrary value may be assigned to one of these capacitances. It is convenient to fix the value of the grid capacitance C_g at its lowest possible value; i.e., the input capacitance of the following stage with some allowance for stray capacitances. The other two capacitances, C_a and C_c are then found as follows:

We first compute the auxiliary values of equation (11),

$$R = \sqrt{R_a R_g}; \; C = \frac{1}{\omega_0 R}; \; b = \sqrt{\frac{R_g}{R_a}}; \; d = \frac{C_g}{C};$$

$$p = \frac{b}{2Q}\left[1 - \sqrt{1 + 4Q\frac{d}{b} - 4Q^2\left(1 + \frac{1}{b^2} + d^2\right)}\right] \quad \ldots \quad \ldots \quad \ldots \quad (46)$$

$$k = \frac{1 - pd}{p + d}$$

and find then for the anode and coupling capacitances

$$C_a = p \, . \, C \qquad C_c = k \, . \, C \quad \ldots \quad \ldots \quad (47)$$

The amplification at the centre frequency is

$$\alpha_0 = kQ \, . \, g_m R \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (48)$$

A thorough investigation of the expressions for p and k shows that they yield physically realizable values for C_a and C_c only if for a given value of the resistance ratio b the grid capacitance remains within certain limits, which may be given as follows:

(1) $0 < Q < \dfrac{b}{2\sqrt{1 + b^2}}$;

$$\left.\begin{array}{l} 0 < d < \dfrac{1}{2bQ}\left[1 - \sqrt{1 - 4Q^2}\right] \\ \dfrac{1}{2bQ}\left[1 + \sqrt{1 - 4Q^2}\right] < d < \dfrac{1}{2bQ}\left[1 + \sqrt{(1 + b^2)(1 - 4Q^2)}\right] \end{array}\right\} \quad (49a)$$

(2) $\dfrac{b}{2\sqrt{1 + b^2}} < Q < \frac{1}{2}$;

$$\left.\begin{array}{l} \dfrac{1}{2bQ}\left[1 - \sqrt{(1 + b^2)(1 - 4Q^2)}\right] < d < \dfrac{1}{2bQ}\left[1 - \sqrt{1 - 4Q^2}\right] \\ \dfrac{1}{2bQ}\left[1 + \sqrt{1 - 4Q^2}\right] < d < \dfrac{1}{2bQ}\left[1 + \sqrt{(1 + b^2)(1 - 4Q^2)}\right] \end{array}\right\} \quad (49b)$$

Conclusion

With the determination of the values of the component parts of the amplifier we have reached the object stated at the outset: to design a feedback amplifier with a certain pre-selected response curve. Since the response curves chosen have no spurious peaks in the whole frequency band from zero to infinity, the amplifiers designed along the lines given in this paper are perfectly stable and permit the full use of all the well-known advantages of negative feedback.

GAIN OF AERIAL SYSTEMS

By D. A. Bell, M.A., B.Sc., M.I.E.E.

SUMMARY.—The maximum gain of an aerial of given aperture (measured in wavelengths) depends on the phase distribution of the illumination of the aperture. Three cases are considered, in order of increasing gain :—(i) Uniform-phase radiators (broadside arrays and 'optical' radiators). (ii) Radiators with effective phase-shift of π (end-fire aerials of all kinds). (iii) Aerials with closely-spaced phase reversals (the high-gain short aerials of Bouwkamp and de Bruijn and of La Paz & Miller).

An example of the relation between the reduction of the radiation resistance of a constituent element of an array and the increase of gain of the array is given in Appendix I and a comparison with optical laws in Appendix II.

1. Introduction

IN a recent publication[1] it has been suggested that 'It may, for instance, be found possible to develop a general and over-riding theorem connecting the directivity and selectivity of an aerial system with its physical size or other geometrical feature.' The purpose of this note is to compare a proposed theorem with the available information consisting of

(a) The calculated and observed gain of broadside aerials.

(b) The calculated and observed gain of end-fire aerials.

(c) The theoretical prediction[2, 3] that there is no upper limit to the gain obtainable from a small aerial by choice of current distribution.

A three-fold theorem is proposed for two-dimensional arrays, as follows :—

(1) The maximum gain of a two-dimensional aerial system in which all radiating elements are excited in the same phase corresponds to the optical diffraction pattern for a uniformly illuminated aperture of the same area :—

$G = 4\pi A/\lambda^2$ referred to an omni-directional radiator,

whence $G = 8.4A/\lambda^2$ referred to a $\lambda/2$ dipole.

(2) The maximum gain of an aerial in which the effective phase varies gradually, by a total amount of approximately π, from one edge of the aperture to the opposite one, is twice the gain of a uniformly-illuminated aperture.

(3) The gain of an aerial in which the phase is reversed at short intervals (radiation sources in opposite phase separated by considerably less than $\lambda/2$) is theoretically unlimited, but this method of operation involves very large amplitudes for the individual radiation sources and is hardly practicable for aerials larger than one wavelength in the directional plane.*

An extension for three-dimensional arrays is :—

(4) The directivity of a three-dimensional array can be deduced as the combination of the directivity of a two-dimensional constituent of it and the directivity of the pattern in which the two-dimensional elements are assembled in the three-dimensional array.

Part (1) covers broadside arrays of dipoles and microwave aerials of the 'optical aperture' type (e.g., horns and mirrors) and its interpretation by Fourier transforms is well-known.

Part (2) is the 'end-fire theorem' which has been found to apply to a number of specific types of aerial (end-fire array of dipoles, Yagi, polyrod, rhombic). Although this does not seem to be generally realized, the 'end-fire theorem' can also be interpreted in terms of Fourier transforms.

Part (3) is added to cover the exceptional aerial current distributions of a 'non-optical' type [see Appendix II, paragraph (i)] for which high gains have been predicted mathematically but which have had limited practical application owing to the extreme difficulty of realizing the current distribution needed for the abnormally high values of gain.

2. Use of Fourier Transforms

The principle of conservation of energy is regarded as sufficient proof that the gain of an aerial in a preferred direction can be related to its directivity, provided the latter is known throughout the solid angle into which the aerial radiates. Following Ramsay[4], the radiation pattern in a given plane of an aerial of aperture width a in that plane is related to the function $F(x')$ which defines the current† amplitude at a distance x' from the centre of the aperture by the formula

$$E_\theta = (1 + \cos\theta)\int_{-a/2}^{a/2}\left[F(x')\exp\left(j\frac{2\pi}{\lambda}x'\sin\theta\right)\right]dx' \quad .. \quad (1)$$

MS accepted by the Editor, September 1948

* Since this paper was written, the problem of aerials of unlimited gain has been discussed by P.M. Woodward and J. D. Lawson.[16]

† With an array of wire radiators, the field depends literally on currents in the aerial; but with a horn or mirror, an equivalent function of either magnetic or electric field must be taken.

The mean field-strength in the plane is then

$$E_m = \int_0^{2\pi} \frac{1}{2\pi} E_\theta d\theta \quad \ldots \quad \ldots \quad \ldots \quad (2)$$

and the directional gain is E_0/E_m where E_0 is the maximum value of E_θ corresponding to the preferred direction of radiation. If the beam is no wider than $\pm 36^\circ$ for a broadside aerial or $\pm 6^\circ$ for an end-fire aerial, the factor $(1 + \cos\theta)$ will vary by no more than 10% over the beam-width and may be regarded as a constant entering equally into the maximum and the mean value, so that for sufficiently narrow beams the directivity is represented simply by the integral. Taking a new variable $u' = \sin\theta/\lambda$, the directivity of a narrow beam is presented by

$$E_\theta = G(u') = \int_{-a/2}^{a/2} F(x') \exp(j2\pi u'x')dx' \quad (3)$$

(Ramsay[4] takes $u = -\sin\theta/\lambda$, but by omitting the negative sign equation (3) is brought into line with Campbell & Foster's notation[5]). By the reciprocal property of Fourier transforms, it can immediately be said that if the desired radiation pattern is given as $G(u')$, the necessary aerial amplitude distribution is

$$F(x') = \int_{-\infty}^{\infty} G(u') \exp(-j2\pi u'x')du \quad \ldots \quad (4)$$

where, as in (3), $u' = \sin\theta/\lambda$. The equations can be more closely related to the physics of the problem if the co-ordinate x' is transformed to a normalized co-ordinate $x = x'/\lambda$ so that

$$E_\theta = \int_{-a/2\lambda}^{a/2\lambda} f(x) \exp(j2\pi x \sin\theta)dx$$

where $f(x)dx = F(x')dx'$. Now take $\sin\theta = u$, giving

$$E_\theta = G(u) = \int_{a/2\lambda}^{a/2\lambda} f(x) \exp(j2\pi ux)dx \quad \ldots \quad (3a)$$

The reciprocal equation (4) now becomes

$$f(x) = \int_{-\infty}^{\infty} G(u) \exp(-j2\pi ux)du \quad \ldots \quad (4a)$$

$$= \int_{-\pi}^{\pi} G(\sin\theta) \exp(-j2\pi x \sin\theta) \cos\theta \, d\theta \quad (4b)$$

This makes it explicit that the transformation operates between the angular directivity pattern on the one hand and the current distribution in the aperture *measured in wavelengths* on the other. (The same relation is, of course, implicit in Ramsay's notation where the angular unit for the directivity pattern is $\pi au = -\pi(a/\lambda)\sin\theta$, instead of the angle itself.) Since the directivity pattern $G(\sin\theta)$ must necessarily repeat when θ is increased by 2π, the integral with respect to θ in (4b) can be confined to the limits $-\pi$ to $+\pi$.

For particular cases, it is known that when the desired $G(u)$ is a geometrically simple pattern (e.g., rectangle, triangle, half cosine, etc.), a restriction on the range of x results in the superposition of a ripple on the desired $G(u)$. Since Fourier transforms admit of direct addition and substraction, this can also be shown by regarding the pattern from the restricted source as the difference between the desired pattern (from an infinite source) and the pattern produced by an infinite source with an amplitude distribution which is zero over the region corresponding to the actual finite source but elsewhere has the ideal pattern.

3. End-Fire Theorem

There is a very general theorem concerning the increase of gain which can be obtained in a long end-fire array by making the relative phases of the currents in the elements differ from the values which would give simple addition of fields in the end-on direction. This was pointed out by Hansen and Woodyard[6] and a pictorial interpretation has been given by F. K. Goward.[7] The latter points out that with the modified phasing the direction in which the fields from all currents are additive is in an imaginary part of space, corresponding to $\sin\theta > 1$; and a narrow main lobe is, therefore, obtained because the part of the main lobe radiated into real space corresponds to the steeply-falling side of the polar distribution and not to the comparatively flat top. (This has been slightly differently expressed by Ratcliffe[12] who pointed out that the vectors representing the fields from the several elements are initially at cumulative angles to each other, not in a straight line, so that increase of all these angles as one moves off the beam therefore causes a rapid decrease in the magnitude of the resultant.) An end-fire theorem has been stated by G. E. Mueller and W. A. Tyrrell[8] in the form that the gain of a uniformly excited end-fire aerial is

$$g = 4A\rho$$

where g is power gain (times), ρ is the length in wavelengths of the end-fire system, and A is a factor which for optimum phasing varies between 2 for $\rho = 2$ and 1.8 for $\rho = \infty$. The optimum phasing condition is that there should be a phase difference between the first and last elements which is π plus 2π times the number of wavelengths between them. (This corresponds to the velocity of propagation along the feed to the elements being effectively less than the free-space value.) The difference of phase of feed proportional to length in wavelengths would be the necessary condition for all parts to radiate effectively in phase as seen from a distant point in the end-fire direction, so that the effective phase variation over the length of the radiator is simply the additional amount of π radians.

The above theorem is for uniform current density along the array, but it has been pointed out both by Goward for arrays[7] and by Mueller and Tyrrell for polyrods[8] that reduction of side-lobes results from a tapered distribution of current along the aerial system.

The rhombic aerial is not obviously an end-fire array, because its breadth is comparable with its length. It is, however, a varying-phase system because the increased path length along the inclined sides of the rhombic gives a greater phase delay between the two ends of the rhombic than corresponds to propagation directly along the diagonal; and taking the formulae given, for example, by Bruce, Beck and Lowry[9] the relation between side length and side angle for maximum radiation in the forward direction is

$$\sin \phi = 1 - \lambda/(2\,l)$$

where l is the length of each side of the rhombic and ϕ is the 'angle of tilt' or half the obtuse angle included between adjacent sides. Now the length along the line of shoot which corresponds to one side is $l \sin \phi = l\,\{1 - \lambda/(2\,l)\}$. The effective phase retardation for the one side is the difference between the actual side length l and the length along the line of shoot,

$$l - l\,(1 - \lambda/2\,l) = l\,(\lambda/2\,l) = \lambda/2.$$

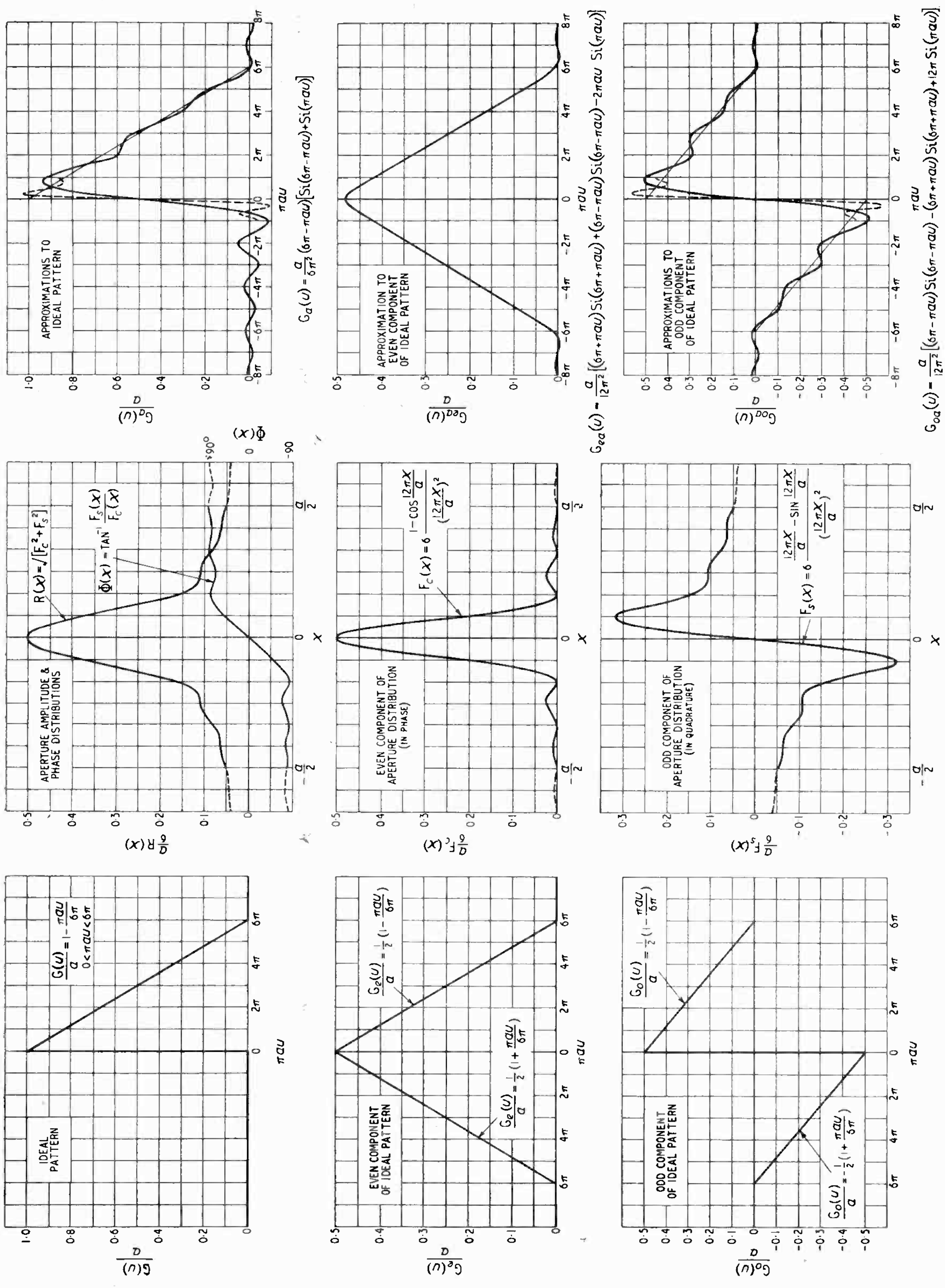
Corresponding to this path difference of $\lambda/2$ there will be a phase retardation of π so that each side of the rhombic conforms to the end-fire phasing condition.

The additional gain of twice power, or alternatively a halving of length for the same gain, by non-uniform phasing of the aperture can also be predicted from the Fourier transforms. In the figure (which is reproduced from Part V of Reference 4) the abscissae πau in the 1st and 3rd columns correspond to angle for a given aerial aperture in wavelengths, and the ordinates $G(u)/a$ represent intensity of radiated energy. In the centre column the abscissae are distances along the aerial aperture, measured from the centre, and the ordinates indicate the magnitude of aerial current at a given point in the aperture (and the phase Fig. 1.4.). Thus columns 1 and 3 are ideal and real-approximation radiation patterns, and column 2 shows the aerial current distributions which give rise to the patterns of column 3. In rows 2 and 3, we see that either the even pattern or the odd pattern from an aerial aperture of dimension $\pm\ a/2$ spreads over an angle from $\pi au = -6\pi$ to $\pi au = +6\pi$. But on combining the even and odd components, as indicated in the first row, an aerial aperture of the same dimension $\pm\ a/2$ produces an asymmetrical pattern which occupies only *half* the angular spread, from $\pi au = 0$ to $\pi au = +6\pi$. Examination of Fig. 1.4. shows that this is achieved by a current distribution of which the phase (referred to the current at the centre point of the aerial as datum) varies from $-90°$ at one end of the aperture to $+90°$ at the other end. But this is an end-to-end variation of π, which is the condition found in the end-fire theorem for doubling the gain (or halving the beam width) for a given aperture.

The use of the Fourier transform should facilitate the adjustment of phase and amplitude distribution in end-fire aerials so as to minimize side-lobes. A further point is that the condition for the doubling of gain is that the directivity pattern must be asymmetric (i.e., the aerial must not be a broadside radiator), but there seems no necessity for it to be end-fire. With dipole arrays there is, however, an advantage in the end-fire type, since a broadside or quasi-broadside wire aerial requires the addition of a reflector to suppress the backward radiation which would otherwise be equal to the forward radiation.

4. Maximum Gain for Uniform-Phase Aerials

If the mean-square aerial current is fixed (mean taken over the aperture) the maximum field in the broadside direction will be given by a distribution which is uniform in phase and amplitude. (Uniform phase is necessary in order that the contributions from all elements of the aperture may have additive effect in the broadside direction. Uniform amplitude is postulated because departure from uniformity will decrease the mean relative to the mean-square, and the field-strength in the broadside direction is proportional to the mean current distribution while the mean-square was taken as constant). But it may be that some other current distribution, which gives less maximum field, may give greater aerial gain because the total power for a given mean-square current amplitude in this distribution is decreased relative to that for uniform distribution to a greater extent than is the maximum field. This condition is indicated in a wire-aerial system by a decrease in the radiation resistance. But in terms of directional pattern, this must mean that if the new amplitude distribution produces a main lobe which is approximately as wide as that of a uniform distribution (as it will with uniform phase), the higher gain can only be obtained by transferring energy from the side-lobes to the main beam.



Graphs of Fourier-transform pairs which correspond to linear distributions of aerial current and angular distributions of radiated power. (Reproduced from "Marconi Review").

Since the uniform aperture distribution gives a directional pattern in which only a few per cent. of the total energy goes into the side lobes, no great improvement in gain is to be expected from special distributions of current amplitude. Inspection of the patterns produced by simple current distributions (as illustrated in Reference 4, for example) shows that there is little variation in the ratio of maximum field to r.m.s. current in the aerial aperture, but the possibility of a fairly considerable transfer of energy between the skirts of the main beam and the minor lobes. By virtue of the superposition of Fourier transforms, any arbitrary current distribution is capable of approximation by the sum of a number of simple components, and it is therefore improbable that a special distribution of amplitude could give a radical improvement.

5. Additional Gain from Phase-Reversing Distribution

An apparent contradiction of the general theory has been demonstrated by Bouwkamp and de Bruijn[2] who state that for a given field-strength in the equatorial plane of a straight-wire radiator, the lower limit to the total power which need be radiated is zero. This implies that for a finite radiated power the upper limit to the directional gain of the aerial is infinity. As an example, they take the radiation pattern in the form $\sin^{(2n+2)}\theta$ and find a method of approximate calculation of the aerial current distribution which corresponds to this pattern. The calculated gains for this particular type of pattern and various values of n are shown in their Fig. 3, and quantitatively the following points may be. noted :—

(i) The paper is concerned with an aerial consisting of a single wire of variable length.

(ii) The length considered is up to two wavelengths (half-length equivalent to 2π radians).

(iii) The gain of the special distribution, referred to a Hertzian doublet, is independent of aerial length.

(iv) This corresponds to the fact that the current distribution is in the nature of a 'damped oscillation' having a large amplitude at the centre of the aerial but a damping coefficient such that the amplitude is very small by the time the end of the aerial is reached. The periodicity along the aerial is such that the phase of the current reverses in a distance much less than half a wavelength.

(v) Even with n as high order as 25, the gain over a Hertzian doublet is only about 2.3 times power.

(vi) The trend of the curves shows that for large values of aerial length a uniform current distribution will give higher gain than the oscillatory type of distribution ; at an aerial half-length of 2π, constant-current distribution is already as good as about the eighth-order pattern. (i.e., the distribution giving a radiation pattern $\sin^8 \theta$).

The principle involved is that the field-strength in the optimum direction is the relatively small difference between two large components in opposite phase, so that a small phase-shift due to a small departure from the optimum angle results in a change in the relative values of the two components which is sufficient to annul the resultant. This is comparable with Ratcliffe's[12] treatment of the Yagi aerial array. The current distribution as a whole is simultaneously made such as to avoid the production of side-lobes. It follows that the maximum currents in the aerial must be much larger than in a non-reversing current distribution, which is interpreted as implying a very low radiation resistance and consequent high copper loss. The principle requires a reversal of phase in a distance much less than half a wavelength, and therefore is inapplicable to 'optical' radiators of the usual type. It is just conceivable that a similar result could be obtained from an array of open-ended waveguides with suitable phase differences between them. In any case, however, the method does not appear worth while for apertures greater than a wavelength in extent, because the gain from a Bouwkamp–de Bruijn distribution is independent of aerial length while that from a non-reversing distribution increases with length.

For a $\lambda/8$ aerial, the distribution suggested by La Paz and Miller [10] also involves a phase reversal. Their distributions, both for $\lambda/8$ and for longer aerials up to a wavelength, suffer the disadvantage of having large current amplitudes at the ends of the aerial, with a consequent emphasis of minor lobes. But the distributions are of simpler appearance, and therefore probably easier to realise, than those of Bouwkamp and de Bruijn.

6. Radiation Resistance and Directivity in Arrays

In Section 5 it was remarked that the additional gain from an oscillating-phase current distribution in a short aerial would be reflected in a reduction of radiation resistance, and this principle can be deduced quite generally, as follows.

An array of two non-interacting driven elements gives a two-fold gain in power. If now the two elements are steadily brought closer together, the limit will be reached when they coalesce, becoming one aerial and giving unity gain. At the same time, the radiation resistance of the single resulting aerial will be twice that of the original combination of two non-interacting aerials fed in parallel; and so if the double current continues to flow in the one aerial, the field-strength in the favoured direction will be the same as when the aerials were separated; but the power input will be doubled and so the gain is reduced to unity. The gain of two equal radiators over a single element is thus restored to unity when the two coalesce.

If at some intermediate position, the resistance R_0 of an aerial element in free space is reduced to R_0/m for each of two interacting elements, the combined input power for the doubled field will be reduced from $2\,i^2 R_0$ to $2\,i^2 R_0/m$ and the power gain in the direction of additive fields will be 4 divided by $2/m$, which is $2m$. It is important to note, however, that this gain will only be obtained in a direction in which the fields are additive; and with close spacing and differently phased currents in the two aerials, there may not be any such direction. An example of the way in which this works out for a 2-element array is given in Appendix I.

In general, if m is the ratio of the free-space value of radiation resistance (of each member of an n-fold array of similar elements) to the average radiation resistance of the elements when assembled in a uniformly-driven array, the *maximum* gain is $m \times n$. In uniform-phase broadside arrays, m is usually fairly close to unity. It has been calculated[11] that an array of N parallel $\lambda/2$ dipoles spaced at $\lambda/2$ intervals has a gain of $4N/3$ times, and that an array of N collinear dipoles has a gain of $2N/3$ times. Consequently 'square' broadside arrays of N dipoles total, such as those variously described as Koomans and Tannenbaum (fir-tree), have a power gain very close to N times.

7. Limits on the Applicability of the Concept of Directional Gain

It has been pointed out[13] that the normal concept of 'aerial gain' depends on the cross-section of the beam, at the measuring station, being considerably greater than the aperture of the transmitting aerial. The suggested criterion of sufficient distance r is $0.52\, r\,\lambda/d^2 \gg 1$ where d is the aperture diameter. An alternative point of view[14] is that at a distance comparable with the aperture dimension, the contributions from the different parts of the aperture will not be in phase; and the minimum distance is suggested as $r\lambda \ll d^2$. For a phase difference of $\lambda/16$ the condition[15] is $0.5 r\lambda/d^2 = 1$. (It has also been noted in the practice of aerial measurements[11,15] that the range at which directional patterns are observed should satisfy the condition $r > d^2/2\lambda$.) This may prove of practical importance in communication with millimetre waves[14] but is not usually a serious restriction in communication with centimetre waves[13].

8. Acknowledgments

This paper is based on work which was carried out for the research programme of British Telecommunications Research Ltd.

The author's thanks are due to Marconi's Wireless Telegraph Co., Ltd., for permission to reproduce the diagrams on p. 309 from *Marconi Review*, Vol. 10, No. 4, p. 158, 1947.

APPENDIX I

Gain and Radiation Resistance for 2 Parallel $\lambda/2$ Dipoles.

Mutual impedances between two parallel aerials have been computed by G. H. Brown (" Directional Antennas," *Proc. Inst. Radio Engrs*, Jan. 1937, Vol. 25, p. 78), and reference to his Figs. 11 and 12 shows that with two aerials spaced $\lambda/8$ apart and fed in opposite phase, the input resistance of each is reduced to a small fraction of its free-space value. This does not mean that the gain is high in the same ratio, because the spacing is so small that there is no direction in which the fields from the two anti-phased currents are additive. Conversely, if the two aerials spaced $\lambda/8$ are fed in the same phase, Brown's graphs show that the input resistance of each is nearly double the free-space value; and this means that at $\lambda/8$ apart the aerials are practically 'coalesced' in the sense indicated in Section 6, as is confirmed by the directional gain of this system being practically unity (Brown's Fig. 15, pattern 4, $d/\lambda = 0.125$ and $\alpha = 0°$). For the existence of a direction in which the fields are directly additive with a 2-element array, the sum of the spacing (in wavelengths) and the phase-difference in fractions of a cycle must be either zero (for a broadside array) or a whole number. This means that the combined effect, expressed as a path difference, is either zero or a whole number of wavelengths. But the spacing required for additive fields is not the condition for maximum gain, since closer spacing gives more than sufficient reduction in input resistance to counterbalance the cosine factor in the resultant of two fields having a phase-difference. Thus with $\alpha = 180°$ (end-fire) examination of Brown's Fig. 15 shows that there is a power gain of about 2.3 times (field gain of 1.6 times) at $d/\lambda = 0.125$, compared with about 1.7 times at $d/\lambda = 0.5$. At $\alpha = 180°$ and $d/\lambda = 0.125$, the factor m for reduction of input resistance is about 7.5 so that this would show a gain of 15 times; but the phase difference will reduce the power gain by a factor of $\cos^2(\pi/8) \approx 0.146$, giving a resultant gain of 2.2. The difference between this and the value of 2.3 read off from Brown's Fig. 15 is within the limits of error in scaling Brown's graphs and figure. In his Fig. 16 Brown shows that the maximum possible gain from 2 driven $\lambda/2$ aerials is about 1.85 times field-strength, which is a power gain of 3.4 times or just over 5 db. From the general theory set out in the present paper (and intended primarily for large arrays) one would predict a maximum possible gain of 4 times, or 6 db, for this case, as follows. The broadside gain of two equal radiators in the same phase is twice; and the end-fire theorem gives an additional gain of twice, making four times in all (6 db). As might be expected, the gain physically realizable from two elements, the smallest possible array, is slightly less (by about 1 db) than the maximum predicted by the application of the general theory of large-aperture arrays.

APPENDIX II

Some Relevant Analogies

(i) *Optical Resolution.*

The directivity of an aerial is comparable with the resolving power of an optical instrument, and the latter is known to be rigorously governed by the aperture measured in wavelengths. The uniform-phase aerials (cm-wave mirrors or horns and broadside arrays of wire aerials) correspond very closely to the optical case, but both end-fire arrays and the phase-reversing distributions discussed in Section 5 appear to transgress this well-established optical law. The explanation is that in optics we have no reasonable means of introducing a change of *phase* of the wave-front between adjacent points of an aperture. The additional gain from both end-fire and phase-reversing aerials is due to this control of local phase, which is possible in some radio aerials but impossible in optics.

(ii) *Pulse Duration and Bandwidth.*

The mathematics of the Fourier transform is equally applicable (a) to the relation between aerial current distribution and angular field distribution and (b) to the relation between the frequency spectrum representing a pulse and the duration of the pulse. A 'better than normal' directivity for an aerial might, therefore, be expected to have a mathematical function which could equally be applied to producing a pulse with duration substantially less than the reciprocal of the corresponding frequency bandwidth.

One notable point is that the directivity function repeats itself when the angle increases by 2π. Consequently, the pulse analogue is a train of regularly repeated pulses, and not a single isolated pulse. It still appears theoretically possible for a limited spectral distribution of energy, having the same form as Bouwkamp and de Bruijn's spatial distribution of aerial current, to produce a train of repeated pulses in which each pulse has any desired degree of narrowness.

REFERENCES

[1] "The Fundamental Research Problems of Telecommunications," D.S.I.R. Report, H.M. Stationery Office, London 1948.

[2] "The Problem of Optimum Antenna Current Distributions," C. J. Bouwkamp and N. G. de Bruijn, *Philips Research Reports*, Vol. 1, No. 2, Jan. 1946, p. 135.

[3] "Note on the Maximum Directivity of an Antenna," H. J. Riblet, *Proc. Inst. Radio Engrs*, Vol. 36, p. 620, May 1948.

[4] "Fourier Transforms in Aerial Theory," J. F. Ramsay, *Marconi Review*, Part I, 1946, No. 83, p. 139; II, 1947, No. 84, p. 17; III, No. 85, p. 41; IV, No. 86, p. 81; V, No. 87, p. 157; VI, 1948, No. 89, p. 45.

[5] "Fourier Integrals for Practical Application," G. A. Campbell and R. M. Foster, Van Nostrand, N. York, 1948.

[6] "A New Principle in Directional Antenna Design," W. W. Hansen and J. R. Woodyard, *Proc. Inst. Radio Engrs*, March 1938, Vol. 26, p. 333.

[7] "An Improvement in End-Fire Arrays," F. K. Goward, *J. Instn elect. Engrs*, Part III, Nov. 1947, Vol. 94, p. 415.

[8] "Polyrod Antennas," G. E. Mueller and W. A. Tyrrell, *Bell Syst. tech. J.*, Oct. 1947, Vol. 26, p. 837.

[9] "Horizontal Rhombic Antennas," E. Bruce, A. C. Beck and L. R. Lowry, *Proc. Inst. Radio Engrs*, Jan. 1935, Vol. 23, p. 24.

[10] "Optimum Current Distributions on Vertical Antennas," L. La Paz and G. A. Miller, *Proc. Inst. Radio Engrs*, May 1943, Vol. 31, p. 214.

[11] "Dimensioning of Directional Antennas" (according to Dr. Kurt Franz, and reported by A. M. Stevens) F.I.A.T. Final Report No. 610. H.M.S.O., London.

[12] "Aerials for Radar Equipment," J. A. Ratcliffe, *J. Instn elect. Engrs*, Vol. 1943, 1946, Pt. IIIA, No. 1, p. 22.

[13] "Intensity-Distance Law of Radiation," D. A. Bell, *Wireless Engineer*, June 1948, Vol. 25, p. 199.

[14] See reference (1) p. 24, item D8.

[15] "Microwave Antenna Measurements," C. C. Cutler, A. P. King and W. E. Kock, *Proc. Inst. Radio Engrs*, Dec. 1947, Vol. 35, p. 1462.

CORRESPONDENCE

Letters to the Editor on technical subjects are always welcome. In publishing such communications the Editors do not necessarily endorse any technical or general statements which they may contain.

Negative Feedback Amplifiers

SIR,—Replying in your July issue to my letter published in May on his article in your February issue, C. F. Brockelsby agrees with my principal point that the amount of feedback obtainable over a given band by the maximally-flat method of design is lower than that obtainable by Bode's method, and goes on to compare the two methods.

He then invites me to give a page reference to the analysis of the three-stage amplifier which I attribute to Bode. May I give a very brief account of this analysis, as it shows the advantages and drawbacks of any method of feedback-amplifier design, and the reasons why all amplifiers providing substantial feedback over wide bands must be designed and tested up to very-high frequencies.

Bode develops a law connecting the phase and attenuation characteristics of any minimum phase-shift network [ch. XIV], and applies it to the general case of a feedback amplifier of any number of stages [ch. XVIII]. [The three-stage amplifier is just one example of the general case. I am sorry if my previous letter was not clear on this point]. At very-high frequencies the load on each stage of an amplifier consists of the parasitic capacitances into which it feeds, and there will be a phase shift of $\frac{\pi}{2}$ radians between the input and output of each stage. Consequently the phase of the returned signal will not differ from that of the input by π radians as it does in the centre of the pass band, but by $\left(\pi - \frac{n\pi}{2}\right)$ radians for n stages. If $n > 2$ this will have changed sign and at some frequency between the pass band and the final asymptote the feedback will be positive. If the amplifier is not to oscillate, the loop gain at this frequency must be less than unity. The problem is to obtain the greatest possible difference of gain between this point and the edge of the pass band. Bode shows that this is done when the slope of the gain characteristic joining the edge of the pass band to the final asymptote is a uniform 12 db/octave, with slight modifications at each end. If the slope is made steeper, the phase shift will exceed π radians and the amplifier will oscillate. If it is made less steep, the amplifier will always be stable, but there will be less feedback available.

The gain/frequency characteristic of a three-stage amplifier with two narrow and one wide stage approximates to this theoretical optimum, and requires only slight modifications to provide adequate phase margins against oscillation, and to provide the changes of cut-off slope just outside the pass band and just before the asymptote is reached which are required if the maximum possible feedback is to be obtained. Of course there are other ways of getting the same gain/frequency characteristic, but this is usually the simplest, and a feedback amplifier with any other shape of loop characteristic must necessarily have less feedback over the same band.

When the loop characteristic of the maximally-flat design is compared with this maximum-feedback characteristic, the merits and disadvantages of the design are clear. Its simplicity, and the large phase margin due to the low rate of cut off (6 db/octave) over the greater part of its range are in its favour. Against these it provides less feedback, and when maximum available feedback is not required, for a given amount of feedback the design must be controlled up to a higher frequency. Mr. Brockelsby mentions this last point as a disadvantage of Bode's design, and assumes that the maximally-flat circuit will behave like three simple RC stages up to any frequency; but it is the failure of circuits to behave so predictably at the frequencies at which stray inductances and capacitances become important that makes it necessary to test all feedback amplifiers up to these frequencies. The maximally-flat circuit is no exception.

In conclusion may I say that in criticizing some of the detailed claims made for the design I am not questioning the value or the novelty of the combination of feedback with a flat external-gain characteristic.

T. S. McLeod.

Standard Telecommunications Laboratories, Ltd., Enfield.

Transit-Time Effects in U.H.F. Valves

Sir,—I wish to comment on Professor John Thomson's paper in the June 1949 issue of *Wireless Engineer*.

Bakker and de Vries[1] in 1935 correctly derived the r.f. conductance of a retarding-field diode for the case of a single velocity of emission. Furthermore they were able to deduce the whole admittance (i.e., including susceptance) by using the induced-current method of analysis instead of the kinetic-energy method which gives only the conductance.

Since the transit angle of a returning electron is closely proportional to its emission velocity, the single-velocity theory is not accurate. The recent analyses by N. Begovich[2] and J. J. Freeman[3] which take account of the Maxwellian velocity distribution are therefore a marked improvement. Freeman's analysis also gives the fluctuation noise in a non-conducting diode and shows that the equivalent temperature of the diode is equal to its cathode temperature, in agreement with the experimental observations of van der Ziel and Versnel.[4]

R. E. Burgess.

National Physical Laboratory, Teddington, Middx.

REFERENCES

[1] C. J. Bakker and G. de Vries, "On Vacuum Tube Electronics," *Physica*, 1935, Vol. 2, pp. 683–697.

[2] N. Begovich, "High Frequency Loading", *Phys. Rev.*, 1948, Vol. 74, p. 1563.

[3] J. J. Freeman, "Noise Spectrum of a Diode with Retarding Field", *J. Res. Nat. Bur. Standards*, 1949, Vol. 42, pp. 75–88.

[4] A. van der Ziel and A. Versnel, "Total Emission Noise in Diodes", *Nature*, 1947, Vol. 159, p. 640.

Sir,—Mr. Burgess's letter states that the conductance of a retarding-field diode has been correctly derived. I fear that no one of the writers on the subject would claim that the conductance as derived is correct for anything save a simple idealization. It was, indeed, one of the objects of my recent article to suggest that much remained to be done, particularly with reference to the effect of space charge.

To take a simple example, Mr. Burgess states that the transit angle of a returning electron is closely proportional to its emission velocity. That this is true where the field is constant is obvious; that it is true for small negative potentials on the anode (such as are found in many practical cases) is equally obviously false. Curiously enough, taking account of the distribution of velocities, as Begovich has succeeded in doing, makes no difference to the fundamental properties of the conductance, as may be seen by a study of Fig. 4 of Begovich's most recent paper[1]. Moreover, since the distribution is random in time, one would not expect it to do so.

I should like therefore, to re-emphasize that in the important region where the constant retarding or accelerating potential difference is small (but still large compared with the high-frequency potential difference) the present state of theory cannot be regarded as satisfactory.

Greenwich. J. Thomson.

[1] *J. appl. Phys*, Vol. 20, p. 457, 1949.

BOOK REVIEW

Microwaves and Radar Electronics

By Ernest C. Pollard and Julian M. Sturtevant. Pp. 426 + vii. Chapman & Hall, 37, Essex St., London, W.C.2. Price 30s.

This book is of American origin and covers most of the important matters pertaining to radar in descriptive form. No attempt is made to go at all deeply into the theory and practice of radar and, indeed, this would be impossible in any single volume. The book is, however, more than an elementary survey of radar and is not for those without a considerable backing of ordinary radio knowledge.

The first four chapters deal with electromagnetic fields, coaxial lines, waveguides, cavities, the production of microwaves and microwave technique. Pulse circuits, c.r. tube indicators, amplifiers and noise are covered and there are chapters on servomechanisms and computors, miscellaneous circuits and the accessories of radar. Under this last propagation is rather surprisingly included. The last chapter deals with microwaves in physical research and there are three appendices dealing with the Fourier integral, curl and Stokes' theorem and units.

The book is intended primarily for those who have a good knowledge of ordinary radio technique as used in communications but know little about radar and is meant to give them a background for further study.

W. T. C.

OVERSEAS STANDARDS

It is not always realized that the British Standards Institution acts as the United Kingdom agent for overseas national standards organizations. Enquiries about such standards should, therefore, be addressed to the British Standards Institution, 24–28, Victoria St., London, S.W.1. They should not be sent directly to overseas organizations, for these organizations will only refer them to B.S.I.

WIRELESS PATENTS

A Summary of Recently Accepted Specifications

The following abstracts are prepared, with the permission of the Controller of H.M. Stationery Office, from Specifications obtainable at the Patent Office, 25, Southampton Buildings, London, W.C.2, price 2/- each.

ACOUSTICS AND AUDIO-FREQUENCY CIRCUITS AND APPARATUS

613 141.—Variable-gain amplifier, of the compressor type, giving a constant signal/noise ratio for speech transmitted, say from an aeroplane, against a fluctuating background of noise.

Bendix Aviation Corporation. Convention date (U.S.A.) 12th March, 1945.

613 530.—Automatic gain-control system, particularly for audio-frequency amplifiers, in which a limiter valve co-operates with a resistance-capacitance network of predetermined time constant.

R. V. Howard. Convention date (U.S.A.) 19th July, 1944.

DIRECTIONAL AND NAVIGATIONAL SYSTEMS

612 250.—Triggering and time-delay circuit for generating pulsed trains of predetermined duration, particularly for use in radiolocation.

Hazeltine Corporation (assignees of J. J. Okrent). Convention date (U.S.A.) 6th June, 1945.

612 627.—Quick-acting automatic gain-control system, particularly suitable for coded pulse-modulated signals, as used in radiolocation.

Hazeltine Corporation (assignees of H. A. Wheeler). Convention date (U.S.A.) 1st June, 1945.

613 017.—Radiolocation system in which the significant time interval is recorded on a cathode-ray tube at one station, where it is scanned and re-transmitted to a distant station.

Marconi's W.T. Co. Ltd. (assignees of A. V. Bedford). Convention date (U.S.A.) 19th June, 1942.

613 143.—Radiolocation equipment, particularly for use as an altimeter, wherein a balanced pair of slot aerials are differentially coupled to two rectifiers to produce the significant beat-frequency.

Standard Telephones and Cables Ltd. and E. O. Willoughby. Application date 15th March, 1946.

613 454.—Automatic gain-control system, particularly adapted to off-set undesirable phase-shift effects produced in radiolocation equipment using rotating aerials.

Sperry Gyroscope Co. Inc. Convention date (U.S.A.) 6th March, 1945.

613 741.—High-speed aerial-switching system, particularly for use in radiolocation, in which the impedance of a transmission-line element is varied from zero to infinity.

Sadir-Carpentier (assignees of P. F. M. Gloess). Convention date (France) 31st October, 1941.

613 862.—Scanning system, of the radiolocation type, for reproducing by means of a stylus, on a recording-paper, scenes obscured by fog or darkness.

Marconi's W.T. Co. Ltd. (assignees of H. A. Iams). Convention date (U.S.A.) 21st February, 1945.

613 964.—Radiolocation system in which a land beacon, radiating a pulsed beam to associated repeating stations, furnishes an indication of obstacles to the pilot of an aircraft.

Standard Telephones and Cables Ltd. (assignees of H. G. Busignies). Convention date (U.S.A.) 5th March, 1945.

RECEIVING CIRCUITS AND APPARATUS

(See also under Television)

612 024.—Receiver which, in the absence of an incoming signal, is automatically muted, through a circuit having a time constant designed to cope with the effects of fading.

The General Electric Co. Ltd. and L. C. Stenning. Application date 15th May, 1946.

612 448.—Receiver for frequency-modulated signals, wherein the carrier is converted into a secondary wave possessing a degree of asymmetry that is proportional to the impressed signal.

Marconi's W.T. Co. Ltd. (assigness of W. L. Carlson). Convention date (U.S.A.), 22nd May, 1945.

612 472.—Interstage coupling for a short-wave amplifier, in which the two parts of a tapped inductance coil are correlated with the electrode capacitances of the coupled valves.

The British Thomson Houston Co. Ltd. Convention date (U.S.A.) 26th May, 1945.

612 536.—Gain-control voltage, derived from the frequency-discriminator of a receiver for frequency-modulated signals, and applied to facilitate tuning.

Radio Corporation of America (assignees of W. La V. Carlson). Convention date (U.S.A.) 26th April, 1945.

612 585.—Receiver for frequency-modulated signals, wherein a lock-in oscillator is associated with a tank circuit which is coupled to a resonant network in order to extend the normal range of lock-in frequencies.

Marconi's W.T. Co. Ltd. (assignees of M. S. Corrington). Convention date (U.S.A.), 29th May, 1945.

612 651.—Device for minimizing the self-inductance of the cathode-supply leads in a valve for amplifying ultra-short waves.

Philips Lamps Ltd. Convention date (Netherlands) 14th September, 1939.

612 842.—Transmission line of two parallel band-shaped conductors, for coupling a receiver to a remotely-situated single-turn frame aerial.

Philips Lamps Ltd. Convention date (Netherlands) 12th February, 1942.

TELEVISION CIRCUITS AND APPARATUS

FOR TRANSMISSION AND RECEPTION

612 438.—Television cabinet in which the viewing-screen and associated reflector are automatically brought into co-operative position by the opening of a hinged door.

Marconi's W.T. Co. Ltd. (assignees of H. McD. Rundle). Convention date (U.S.A.) 22nd March, 1945.

612 533.—Television system in which audio-frequency signals are transmitted by frequency-modulating a sub-carrier wave that is generated only during the fly-back periods of the video signals.

Marconi's W.T. Co. Ltd. (assignees of K. Schlesinger). Convention date (U.S.A.) 20th April, 1945.