SELECTIVE DISSEMINATION OF INFORMATION

ONE of the many problems associated with the 'explosion' of technical information is the growing difficulty that the individual research worker has in ensuring that he becomes aware of all the new developments in his field as they are announced. It is theoretically possible for any individual to keep himself completely abreast of the literature by reading through a large number of the lists of published and unpublished papers which already appear at weekly, fortnightly or monthly intervals and selecting for subsequent study those items that are of interest to him. In practice, no individual worker has the time to carry out this formidable task.

One way of overcoming the difficulty would be to set up a service in which every item published or issued is considered by some competent agency understanding the requirements of an individual user. Notification would then be sent to him of any items which seemed likely to be of interest. It is apparent that, with a large number of users, such a service would not be feasible using normal manual methods: modern data processing techniques must be used.

This is the thinking behind the announcement made by the Earl Mountbatten of Burma, K.G., in his address as Chairman of N.E.R.C. at the Dinner of the Conference of the Electronics Industry on 10th November 1964. N.E.R.C. proposes to investigate a system of Selective Dissemination of Information (S.D.I.) which will provide a personal service to individuals. A 'profile' of each user's interests will be compiled and stored in a computer in the form of the index terms used in the system. The documents received will be indexed using the same terms and the details held in a second store of the computer. Each week the two stores will be compared, and where the indexing of a document sufficiently matches an individual's 'profile', the appropriate information will be sent to him.

An element of 'feedback' is included in the system, the users being sent a card with each document reference which they will be asked to return, stating whether the document in question is relevant to their interests or not. This information will allow continuous adjustment of the user profiles to ensure that the user's requirements are met as exactly as possible.

In the proposed project N.E.R.C. hopes to obtain the collaboration of some 800 electronics research workers in industry, universities and technical colleges, and government research establishments. These 800 users will be chosen to be as representative as possible of all fields of electronics research and of all types of worker, from Engineer to Research Manager, and from Research Student to Professor.

In addition to the 'personal' current-awareness service the project provides for a weekly bulletin, consisting of a print-out in broad subject categories of all the information received during the week. Furthermore, it is proposed that the information accumulated shall be available for retrospective searching by computer and the compilation of bibliographies on request.

N.E.R.C. has recommended to the Department of Scientific and Industrial Research that the setting up, operation and testing of such a system should be supported as a three-year project. Electronics is a field of endeavour finding itself in urgent need of some way of dealing with the information problem and it is fitting that the electronic computer should be brought in to help.

In proposing the investigation the National Electronics Research Council is convinced that it will have significance far beyond the field of electronics and in fact will be of national value since such a system could be applied in all fields of science and technology.

G. D. C.

INSTITUTION NOTICES

Institution Premiums and Awards

The Council of the Institution announces that the following awards are to be made for outstanding papers published in *The Radio and Electronic Engineer* during 1963:

THE CLERK MAXWELL PREMIUM

"Loss Properties of Cylindrical Waveguides containing Gyromagnetic Media" by R. A. Waldron, M.A. (Member), and Mrs. D. J. Bowe, B.Sc.

"Properties of Inhomogeneous Cylindrical Waveguides in the Neighbourhood of Cut-off" by R. A. Waldron, M.A. (Member).

(Published in the April and June issues respectively.)

THE HEINRICH HERTZ PREMIUM

"Stabilization of a Variable Quantity with Single Valued Reference" by D. M. Makow, Dr.-Ing. (May).

THE J. LANGHAM THOMPSON PREMIUM

"The Development of ARCH—A Hybrid Analogue-Digital System for Computers for Industrial Control" by G. B. Cole, B.Sc.(Eng.), and S. L. H. Clarke, B.A. (Member) (July).

THE VLADIMIR K. ZWORYKIN PREMIUM

"Muscle Substitutes and Myo-electric Control" by A. H. Bottomley, M.B., B.Chir., A. B. Kinnier Wilson, M.B., M.R.C.P., and the late A. Nightingale, Ph.D. (December).

THE A. F. BULGIN PREMIUM

"An Investigation into the Acoustic Noise Produced by the Ferrite Core used in Television Line Output Transformers" by J. C. MacKellar (Associate Member) (May).

THE LORD RUTHERFORD AWARD

"The Performance of an Ammonia Maser with Two Resonators in Cascade" by D. C. Lainé, Ph.D., and R. C. Srivastava, Ph.D. (August).

THE LESLIE MCMICHAEL PREMIUM

"Microwave Branching Systems" by M. V. O'Donovan (Graduate) (May).

THE CHARLES BABBAGE AWARD

"A Computer Fixed Store using Light Pulses for Read-out" by G. R. Hoffman, Ph.D., and D. C. Jeffreys, M.Sc., Ph.D. (Associate Member) (February).

The following Premiums are withheld because either no eligible papers were published during the year or the eligible papers were not of high enough standard: the Rediffusion Television Premium, the Lord Brabazon Award, the Hugh Brennan Premium, the Marconi Premium, the Arthur Gay Premium, the Dr. Norman Partridge Memorial Premium and the J. C. Bose Premium.

The Premiums and Awards will be presented at the Annual General Meeting in London on 9th December 1964.

Joint I.E.R.E.–I.E.E. Symposium on "Microwave Applications of Semiconductors"

The Institution is to be joined by the Electronics Division of the Institution of Electrical Engineers in holding a Symposium on the above subject at University College, London, from 30th June to 2nd July 1965.

The Symposium will cover fundamental properties and the development of semiconductor devices, their incorporation into complete microwave components and their systems applications. Applications to be described will include oscillators, harmonic generators, amplifiers, detectors, frequency changers, switches and limiters, and systems made up of these component parts.

The Joint Committee invites offers of papers with provisional titles as soon as possible; synopses should be in the hands of the Committee for approval not later than 31st December 1964. Offers of papers should be sent to The Secretary, Joint Organizing Committee for the Symposium on "Microwave Applications of Semiconductors", The Institution of Electronic and Radio Engineers, 8–9 Bedford Square, London, W.C.1.

Indian Proceedings

The June issue of the *Proceedings* of the Indian Division of the Institution included a paper on "Studies on Transistor Phase-shift Oscillators" by Dr. V. V. Narasimha Rao and Professor B. Ramachandra Rao, D.Sc. (Member) of the Electronic Research Laboratories at Andhra University.

The paper uses general transformation methods and derives the maintenance and timing expressions of the oscillations as relationships between the transistor parameters and circuit characteristics. Prediction of the limit of maximum attainable frequency with such an oscillator is also made from this analysis by considering the effect of collector-base capacitance and is compared with experimental results.

Copies of the Indian *Proceedings* may be obtained from the Administrative Secretary, 7 Nandidrug Road, Bangalore. Price: 2 rupees, members; 5 rupees, non-members.

Conference on Submillimetre Waves

The British Radio Spectroscopy Group is holding a conference on "The generation, detection and properties of coherent radiation of wavelength less than one millimetre" in Malvern on 6th and 7th April 1965. Details and application forms, which must be returned by 26th February 1965, may be obtained from Mr. M. F. Kimmitt, Royal Radar Establishment, St. Andrew's Road, Malvern, Worcestershire.

Near-field Measurements and the Determination of Aerial Patterns

By

J. R. BLAKEY, M.Sc. †

Summary: Measurements carried out in the near-field region of aerial systems can be used to predict radiation patterns. While near-field methods are unlikely to be used where conventional aerial measurements are satisfactory they become of increasing interest under those situations where normal procedures are either inadequate or impossible to apply.

Near-field techniques can be broadly divided into two classes, namely those in which the radiation pattern is simulated directly within the near-field region and those in which the radiation pattern is computed directly from the near-field measurements.

Several methods representative of each of these two classes are discussed (e.g. those due to Cheng, Bickmore and Jull) and the range of validity and likely applicability of each is indicated.

List of Symbols

D	major dimension of an aerial	μ_0	permeability of free space
λ	wavelength	ε_0	permittivity of free space
f	focal length	Z_0	characteristic impedance of free space
δ	displacement of primary feed	E_z	component of electric field
R	distance from origin of co-ordinates	H_y	component of magnetic field
r	distance between any two points	$S = \sin \theta$	
dA_1, dA_2	elementary aerial surfaces	$B(S)$	angular spectrum of plane waves

$$k = \frac{2\pi}{\lambda}$$

G_y near-field component of magnetic field

$H_n^{(2)}$ Hankel function of second kind and order n

$$k_0 = \frac{2\pi}{\lambda} \sqrt{\mu_0 \varepsilon_0}$$

a_n coefficients of cylindrical modes

1. Introduction

With the increasing use of high-resolution radiating systems the consequent increase in the far-field limiting distance $2D^2/\lambda$ makes the selection of suitable sites for conventional aerial pattern determinations difficult and renders existing aerial ranges inadequate. For instance with a high-resolution movement control radar operating at $\lambda = 4$ mm the aerial system is a linear array of length $D = 800\lambda$: the $2D^2/\lambda$ criterion shows that the far-field distance is greater than 5 km. Under these and similar circumstances the prediction of aerial patterns from near-field measurements may be the only practicable alternative to conventional testing methods.

† Department of Physics, Battersea College of Technology, London, S.W.11.

2. Nature and Characteristics of the Near Field

The characteristic feature of the near field is the variation of the measured angular response pattern with distance from the aerial. This behaviour can be readily understood when the aerial is considered as an array of a large number of infinitesimal radiating elements. The field at any point is the vector sum of the fields due to each of these elements. At large distances the paths from each element to the point of measurement are virtually parallel. The relative phases of each component of the field due to path-length differences depend solely on the angle between the direction of the point of measurement and the normal to the aerial system. Under these conditions the far-field pattern will be observed. As the distance between the aerial and the point of measurement is reduced the various paths will no longer be parallel

and the relative phases of the component fields will depend on distance as well as angle. The distance which divides these two regions (the far-field limit) is conventionally taken to be $2D^2/\lambda$.

As the distance is reduced the following changes in the radiation pattern are generally observed (Fig. 1). The amplitudes of the minima between the side-lobes increase, the effect being greatest close to the main beam. At the same time the side-lobe levels closest to the main beam also increase but side-lobes further away may actually decrease. As the distance is further reduced the minima continue to fill up until finally the first and eventually other side-lobes are engulfed and become part of the main beam. At a distance of about $0.2D^2/\lambda$ the main beam itself may undergo modification by splitting into two components. The complicated field pattern which is observed very close to the aerial is not due solely to the radiation field. It is the superposition of several effects such as evanescent modes, surface waves and multiple scattering between the various components of the system and its physical supports. These contributions
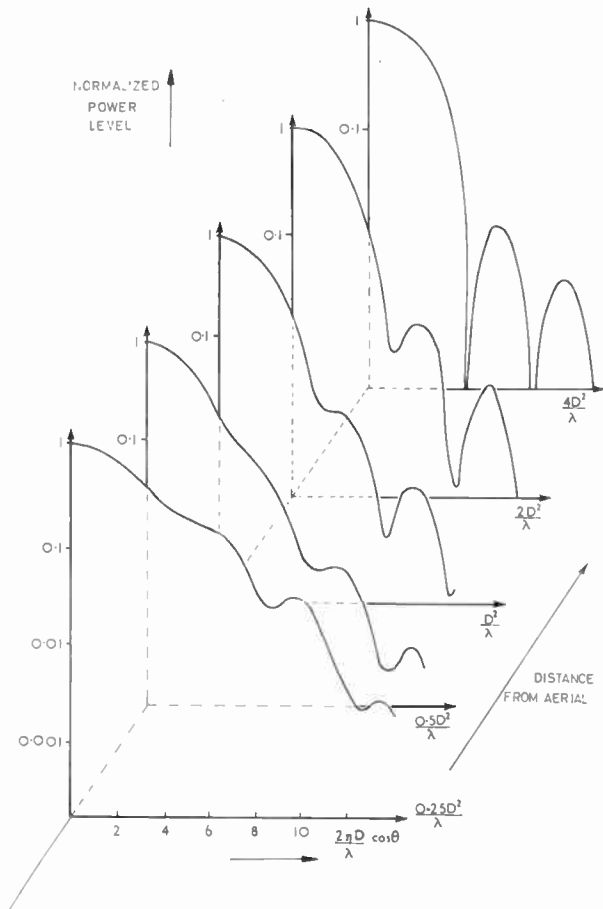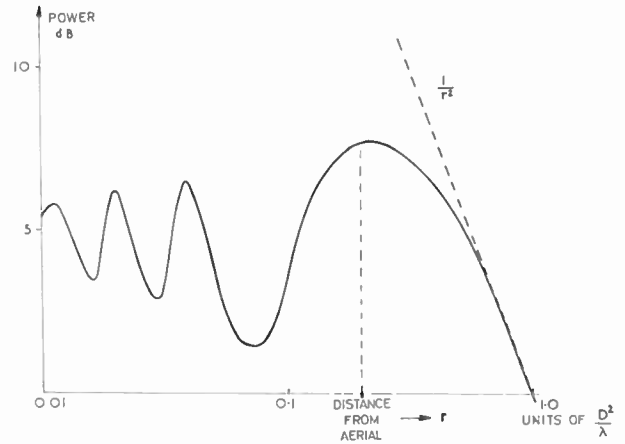
Fig. 2. Typical variation of intensity in the direction of the main-lobe maximum.

generally become insignificant compared with the radiation field at a distance of the order of the maximum dimension of the aerial. For this reason the region very close to the aerial is specifically excluded from the discussion. Also the power density in the direction of the main beam maximum varies in a characteristic way (Fig. 2). At distances greater than D^2/λ the power density varies as the inverse square of the distance. For distances less than D^2/λ the power density increases less rapidly than the inverse square and reaches a maximum value at approximately $0.2D^2/\lambda$.

As the distance to the aerial is progressively reduced the power density successively attains a series of maximum and minimum values. This behaviour is completely analogous to the Fresnel zone phenomena of physical optics and can be explained in a similar manner.[1]

3. Simulating Far-field Patterns within the Near-field Region

Since near-field patterns arise from the variation in path distances from each point of the aerial to the point of measurement it follows that the far-field pattern could be observed in the region of that point if the phases of each element of the aerial were modified so that the path distance variations were exactly compensated. In mirror-type radiating systems the primary feed is located at the focus of the reflector. This produces constant phase surfaces in planes perpendicular to the axis of the system. The aperture phase distribution is most simply modified by displacing the primary feed away from the focal point. By displacing the feed a suitable distance the modified aperture phase distribution should permit the observation of the far zone pattern. Owing to the geometry of real systems this ideal
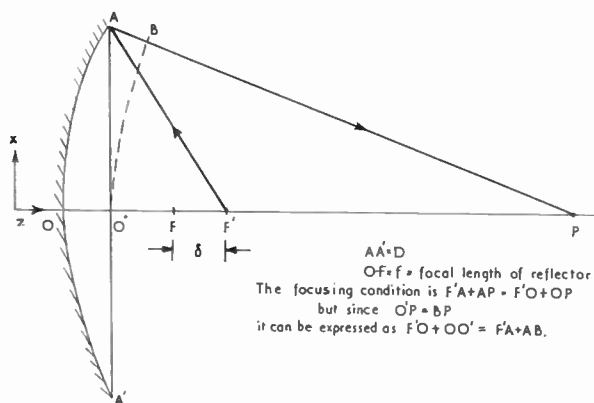
Fig. 1. Typical variation of the normalized angular power pattern of an aerial with distance.

Fig. 3. Ray geometry for a displaced feed focusing system.

$AA' = D$
$OF = f =$ focal length of reflector
The focusing condition is $F'A + AP = F'O + OP$
but since $O'P = BP$
it can be expressed as $F'O + OO' = F'A + AB$.

situation is rarely achieved. However, by assuming that the primary feed can be treated as a point source, it is not difficult to calculate the required displacement to focus the radiation at a given distance R for the case of a parabolic reflector. If the total path for the two rays emerging from the apex of the parabola and the edge of the aperture are made equal, it follows from Fig. 3 that

$$F'O + OO' = F'A + AB \qquad \ldots\ldots(1)$$

Assuming that $f^2/R \ll 1$ it can be deduced from eqn. (1) that the required displacement is given by

$$\delta \simeq \frac{f^2}{R}\left[1 + \frac{f}{R} + \left(\frac{D}{4f}\right)^2\right] \qquad \ldots\ldots(2)$$

where f is the focal length of the reflector.

This value is only correct, however, for the two extreme rays which were chosen. For all intermediate rays the actual paths are longer. Worse still, these intermediate rays do not all focus at the same point and therefore give rise to various geometrical aberrations analogous to those encountered in optical systems. The only way in which all optical paths could be made equal and be brought to a focus at a common point would be to use an ellipsoidal reflector. Such a reflector would not normally be used in a conventional radiating system but the concept enables a more useful displacement criterion to be established. The equation of an ellipse in the x-z plane is

$$z = \left(\frac{f_1 + f_2}{2}\right)\left[1 - \sqrt{1 - \frac{x^2}{f_1 f_2}}\right] \qquad \ldots\ldots(3)$$

where f_1, f_2 are the foci of the ellipse. If $x^2/f_1 f_2 \ll 1$, expansion of the root in eqn. (3) leads to the following approximation:

$$z = \frac{(f_1 + f_2)}{4 f_1 f_2} x^2 \qquad \ldots\ldots(4)$$

But this is the equation of a parabola of focal length f, where $f = (f_1 f_2)/(f_1 + f_2)$. Thus, to the accuracy

with which the parabola approximates to the ellipse, a point source placed at f_1 will have its radiation focused at f_2. Since $f_2 = R$ the required displacement is thus

$$\delta = f_1 - f = \frac{f^2}{R - f} \qquad \ldots\ldots(5)$$

In eqn. (5) the displacement is seen to be independent of the dimensions of the radiating aperture. This criterion is to be preferred in practice since geometrical aberrations will be minimized.

This technique extends the usefulness of existing pattern ranges but in order to maintain the uniformity of illumination over the aperture under test the dimensions of the transmitting aerial must be reduced. It can be shown that provided the aerial under test is in the far field of the transmitter the consequent reduction in signal level due to the lower gain of the transmitter is completely compensated by the reduced distance of measurement. An extension of this argument shows that the interaction between the aerials remains unchanged. The advantages of the method are paid for by the reduction of the angular range over which the radiation pattern is accurately simulated. This aspect is considered in connection with the deformation technique but the analysis and resulting criterion (eqn. (13)) can be applied directly to the present discussion.

The displacement technique obviously cannot be applied to non-focused systems such as linear and curtain arrays but the principle of phase compensation can still be applied. Bickmore's deformation technique[3,4]—wherein the required phase compensation is achieved by physically deforming the array to a prescribed surface or curve—is a good example of this. The requisite curve can be deduced in the following way. A system consisting of two aerials A_1 and A_2 separated by a distance R is designed to give maximum power transfer in the near-field region (Fig. 4). It is assumed that no interaction occurs between the aerials. Let A_1 be the transmitter. The radiation
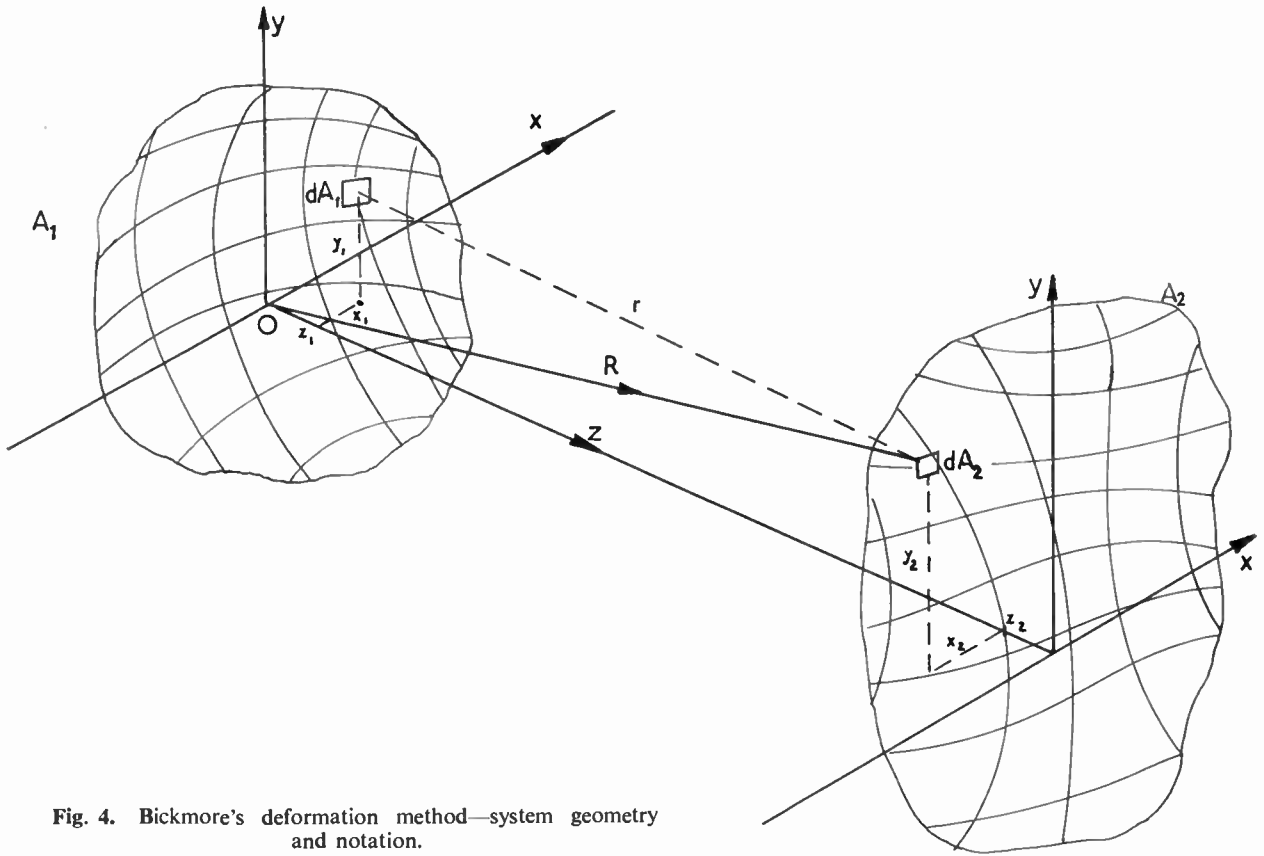
Fig. 4. Bickmore's deformation method—system geometry and notation.

from an element dA_1, located at the point (x_1, y_1, z_1) will induce a current dI in an element of the receiving aerial dA_2 located at (x_2, y_2, z_2). The elements dA_1 and dA_2 are a distance r apart and provided that $1/r \ll 2\pi/\lambda$ the induced current is given by

$$dI = \frac{EF_1 F_2}{Z_0} \frac{e^{jkr}}{r} dA_1 dA_2 \qquad ......(6)$$

where F_1, F_2 are proportionality constants for the respective elements and Z_0 is the characteristic impedance of free space. Provided that the polarization of each element dA_1 is in the same direction and if F_1 and F_2 are assumed to be constant for every element dA_1 and dA_2 then the total current induced in A_2 is

$$I = \frac{EF_1 F_2}{Z_0} \int\limits_{A_1} \int\limits_{A_2} f_1(x_1, y_1) f_2(x_2, y_2) \frac{e^{jkr}}{r} dA_1 dA_2 \qquad ......(7)$$

where $f_1(x_1, y_1); f_2(x_2, y_2)$ are the corresponding aperture field distributions. This integral is a familiar one in aerial theory and if the usual Fresnel region approximations are made[1] it can be written in the form

$$I = \frac{EF_1 F_2}{Z_0} \frac{e^{jkR}}{R} \int\limits_{A_1} \int\limits_{A_2} (f_1, f_2) e^{jkr} dA_1 dA_2 ......(8)$$

By applying the Cauchy inequality

$$\left| \int f(z) \, dz \right| \leq \int \left| f(z) \right| \, dz$$

to eqn. (8) it follows that

$$I \leq \left| \frac{EF_1 F_2}{Z_0} \cdot \frac{e^{jkR}}{R} \right| \int\limits_{A_1, A_2} \int \left| f_1, f_2 \right| e^{jkr} dA_1 dA_2 \quad ...(9)$$

The integral is then maximized by choosing a suitable geometry for the aerials. The relevant aerial surfaces are determined by the application of the variational calculus.[9] It is then found that the optimum geometry for maximum power transfer consists of two-plane parallel arrays. Since a simulated far-field pattern is required each aerial must be regarded as an effective point source with respect to the other and for maximum gain in this latter case the optimum geometry becomes two spherical surfaces of radius R, the centre of curvature for A_1 being at $(R, 0, 0)$ and that of A_2 being at the origin of the co-ordinate system. This is obviously a reasonable result since all points

on A_1 are equidistant from the point $(R, 0, 0)$. It is necessary to know how closely the observed pattern simulates the far-field pattern in the vicinity of the point $(R, 0, 0)$. This is readily determined from eqn. (8). The current induced in an elementary aerial dA_2, at any point (x_2, y_2, z_2) by the spherical array A_1 is given by

$$I = \frac{EF_1 F_2}{Z_0} \frac{e^{jkR}}{R} dA_2 \times$$
$$\times \int_{A_1} f_1 \exp\left[\frac{jk(x_1 x_2 + y_1 y_2)}{R}\right] \exp\left[\frac{jkz_1(z_2 - R)}{R}\right] dA_1$$
$$\quad\quad\quad ......(10)$$

An identical expression is obtained for the equivalent source field due to a spherical wavefront of radius R, so that the following argument can be applied equally to both curved arrays and focused systems. The integral (eqn. (10)) can be interpreted as the far-field pattern modified by a phase error function $\exp[jkz_1(z_2 - R)/R]$. From this it can be seen that, within the limits of validity of the Fresnel approximation, the far-zone pattern is accurately simulated in the plane $z_2 = R$. In practice measurements are not usually carried out in a given plane or along a given line but rather at a constant distance from the aerial. It is therefore necessary to evaluate the phase error function to determine the angular range over which the phase error in the simulated pattern does not exceed a specified value. For any point (x_2, y_2, z_2) at a distance R from the origin and subtending an angle θ with the z-axis

$$(z_2 - R) = -R \sin\theta \tan\frac{\theta}{2} \quad ......(11)$$

If a circular array with diameter D in the plane $z = 0$ deformed into a segment of a sphere of radius R whose centre is located at $(R, 0, 0)$, then the maximum displacement of the edge of the array is given by

$$z_1 \simeq \frac{D^2}{8R} \quad\quad ......(12)$$

Combining eqns. (11) and (12) gives the required angular range in which the phase error will not exceed a specified value, i.e.

$$\left|\frac{kz_1(z_2 - R)}{R}\right| \le \frac{kD^2 \sin\theta \tan\theta/2}{8R} \quad ......(13)$$

A convenient value for the maximum permissible phase error is $\pi/8$ radians. If this value is used in eqn. (13) then measurements taken over the corresponding angular range θ will give a simulated far-field pattern that is at least as accurate as that obtained from conventional measurements at a distance $2D^2/\lambda$. It can be seen from eqn. (13) that as R is decreased the angular range in which the pattern is simulated to the required accuracy decreases also. The deformation technique is perhaps most readily applied to

linear arrays. Bickmore verified his analysis by comparing the far-field patterns of undeformed slotted waveguide arrays with those obtained under focused conditions. For example, the pattern of a 250λ array ($\lambda = 8$ mm) was measured at a distance of 65 ft, the corresponding far-field limit being 4386 ft. The results were in excellent agreement with those measured at a distance of $1.37 D^2/\lambda$ for the undeformed array. Equation (13) shows that the pattern was accurately simulated for an angular range of $\pm$ 10 deg. This was more than adequate in this case since the 3-dB width of the main beam was only 14 minutes of arc. Despite the short distance involved in the simulated measurement no trouble was experienced from multiple scattering between the transmitter aerial and the array.

4. Computation of Aerial Patterns from Near-field Measurements

It is well known that the aperture distribution of a reflector aerial and its far-field pattern together constitute a Fourier transform pair.[1] This result is readily interpreted by applying the concept of an angular spectrum of infinite plane waves.[7] The field distribution over the aperture plane can be thought of as the result of a superposition of a very large number of infinite plane waves, each travelling in a particular direction θ. In other planes parallel to the aperture each component wave will arrive with a phase which depends upon its direction. The superposition of all the waves in a particular plane will build up a complex amplitude distribution which will differ from that in the aperture plane and will also be different in planes at different distances from the aperture. At sufficiently large distances from the aperture the observed patterns will tend more and more closely to the far-zone pattern. It can be shown[5] that in the far zone the amplitude in the direction θ is due solely to the strength of the plane-wave travelling in that direction. Since each distribution of the near field arises from the same angular spectrum it should be possible to deduce a transform relationship between the near-field distributions and the far-field pattern which is analogous to the aperture field–far field transform. This can be done by considering the phase relationships between the near-field distributions and the aperture distribution. For simplicity the two-dimensional case is considered in which the aerial is located in the plane $z = 0$. An infinite plane-wave travelling in a direction θ with respect to the z-axis in free space has a transverse magnetic component H_y given by

$$H_y = B \exp[-jk_0(x \sin\theta + z \cos\theta)] \quad ......(14)$$

where
$$k_0 = \frac{2\pi}{\lambda}\sqrt{\mu_0 \varepsilon_0}$$

and B is a constant. Putting $\sin \theta = S$ for convenience, eqn. (14) becomes

$$H_y = B \exp\left[-jk_0\{Sx + z(1-S^2)^{\frac{1}{2}}\}\right] \quad \ldots\ldots(15)$$

Thus the field in any z-plane will be obtained by summing the effect of the waves given by eqn. (15) travelling in all directions S.

$$H_y(x,z) = \int_{-\infty}^{\infty} B(S) \exp\left[-jk_0\{Sx + z(1-S^2)^{\frac{1}{2}}\}\right] dS$$
$$\ldots\ldots(16)$$

The integration is performed over all directions S. This includes the 'complex' directions for values of $|S| > 1$. Only waves travelling in directions $|S| < 1$ contribute to the radiated field while those for which $|S| > 1$ are confined to the radiating surface and are evanescent in the z-direction. When z is put equal to zero in eqn. (16) the aperture field distribution is obtained.

$$H_y(x,0) = \int_{-\infty}^{\infty} B(S) \exp\left[-jk_0 Sx\right] dS \ldots\ldots(17)$$

Now eqn. (17) is a Fourier transform and hence $B(S)$ is interpreted as the far-field pattern. Using the transform-pair notation eqn. (17) can be written as

$$H_y(x) \leftrightarrow B(S) \quad \ldots\ldots(18)$$

A plane-wave travelling in the direction θ suffers a phase change of $k_0 z_0 (1 - S^2)^{\frac{1}{2}}$ travelling from the origin to the plane $z = z_0$. For angles of 30 deg or less the bracket can be expanded to give, with little error, $k_0 z_0 - \frac{1}{2} k_0 z_0 S^2$. The first term is the phase change incurred in travelling a distance z_0 and the second term is the contribution to the phase change due to the direction of travel. The required transform pair is obtained by substituting this expression for the phase into eqn. (17). Denoting the resulting near-field distribution by $G_y(x)$ this becomes

$$G_y(x) \leftrightarrow B(S) \exp\left(\frac{jk_0 S^2 z_0}{2}\right) = B(S) Y(S) \ldots(19)$$

also

$$Y(S) \leftrightarrow y(x) = \exp\left(\frac{jk_0 x^2}{2z_0}\right) \quad \ldots\ldots(20)$$

By using these results it is possible to calculate the aerial pattern from near-zone measurements of amplitude and phase carried out in an infinite plane parallel to the aperture. Since any practical measurement would have to be carried out over a finite plane this would give rise to errors of uncertain magnitude in the computed far-zone pattern. In any case, the phase approximation used in the derivation limits its usefulness to angles of less than 30 deg from the beam centre, with 15 deg probably being a more realistic estimate.

The method is of some use in the many cases of practical interest where the detailed pattern is not required but where a knowledge of side-lobe levels is important. These levels can be calculated in the case of linear arrays from near field amplitude measurements alone, provided that certain reasonable assumptions are made.[6] The application of the transforms (eqns. (19) and (20)) enables the relative side-lobe levels to be calculated for the near field for a given far-field pattern. Bates and Elliot carried out such calculations for aerial patterns of the Dolph-Chebychev and $(\sin \theta)/\theta$-type because these represent the extremes of a range of far-field patterns which includes those most commonly employed. The calculations assume that the far-field side-lobe maxima occur at the same (normalized) directions with respect to the main beam for both patterns and that these directions are the same for the near-field pattern. The evaluation of true side-lobe levels is quite simple, only requiring a measurement of the side-lobe levels relative to the main beam maximum and the angular width between the first minima beside the main beam, as well as the knowledge of the aerial dimensions and wavelength used. With these quantities the curves prepared by Bates and Elliot can be entered to give the true side-lobe level. It follows from the approximations used that this method can only be applied at distances within the first Fresnel zone where there is a recognizable 'quasi-far-field' pattern.

A completely different approach is needed when the complete radiation pattern is required. This has been provided by the recent work of Jull.[7, 8] In the two-dimensional case the aerial is taken to lie entirely within a cylinder of radius r_0. Provided that the only component of the electric field is E_z, a general solution of Maxwell's equations in circular cylindrical coordinates which also satisfies the radiation condition at infinity is

$$E_z(r, \theta) = \sum_{n=-\infty}^{\infty} a_n H_n^{(2)}(kr) \exp(jn\theta) \quad \ldots\ldots(21)$$

where $H_n^{(2)}(kr)$ is a Hankel function of the second kind and argument (kr). Now for very large values of r the use of the leading term in the asymptotic expansion of $H_n^{(2)}(kr)$ leads to the following expression for the far-field.

$$E_z(r, \theta) \simeq \left(\frac{2}{\pi kr}\right)^{\frac{1}{2}} \exp\left[-j\left(kr - \frac{\pi}{4}\right)\right] \times$$
$$\times \sum_{n=-\infty}^{\infty} a_n \exp\left[jn\left(\theta + \frac{\pi}{2}\right)\right] \quad \ldots(22)$$

Since only the angular variation of the relative amplitude is needed the aerial pattern can be written as

$$B(\theta) = \left| \sum_{-\infty}^{\infty} a_n \exp\left[jn\left(\theta + \frac{\pi}{2}\right)\right] \right| \quad \ldots\ldots(23)$$

This expression is essentially a Fourier series. The purpose of the near-field measurement is to determine the values of the coefficients a_n. For measurements carried out at a distance b from the aerial it follows from eqn. (21) that

$$E_z(b,\theta) = \sum_{n=-\infty}^{\infty} a_n H_n^{(2)}(kb) \exp(jn\theta) \quad \text{......(24)}$$

The coefficients of this Fourier series are given by

$$a_n H_n^{(2)}(kb) = \frac{1}{2\pi} \int_0^{2\pi} E_z(b,\theta) \exp(-jn\theta)\, d\theta \quad \text{...(25)}$$

Since Hankel functions are quite well tabulated and can also be readily computed, the coefficients a_n can be found from eqn. (25). These values of a_n are then substituted into eqn. (23) to give the far-zone pattern.

For normal aerials (i.e. non-supergain arrays) the coefficients a_n will normally be insignificant for $n > kr_0$; where $r_0 = D/2$. Thus a measurement of amplitude and phase at each of $(2kr_0 + 1)$ points equally spaced around the circle of radius b, should be sufficient to determine all a_n and hence the far-field pattern.

From this analysis it follows that the complete pattern can be determined from measurements taken at any distance and involves no approximations. In practice the shortest distance employed would depend upon the availability of tabulated Hankel functions or the amount of computational effort required to derive them but should be useful for distances in the range $(2D^2/\lambda)/16$ upwards. For highly directive arrays it would be necessary in any case for measurements to be taken at a large number of points. Thus the computation of radiation patterns by this method may well require the use of a large-capacity electronic computer.

The method just discussed is perfectly general and is applicable in principle to three-dimensional problems for which analogous solutions to Maxwell's equations in the appropriate spherical harmonics can be obtained.

5. Conclusions

Several methods for determining the radiation patterns from near-field measurements have been discussed. In most cases only a portion of the total pattern can be determined with reasonable accuracy and in some methods the useful reductions of the measuring range is limited, either by physical considerations or by the nature of the approximations used. Only Jull's method permits, in principle, the total pattern to be determined from measurements taken at any distance from the aerial without approximations, but this method finds its limitation in the amount of computational effort which may be involved.

6. References

1. S. Silver, "Microwave Antenna Theory and Design", (McGraw-Hill, New York, 1949).
2. D. K. Cheng, "On the simulation of Fraunhoffer radiation patterns in the Fresnel region", *I.R.E. Trans. on Antennas and Propagation*, No. PGAP 5, p. 399, October 1957.
3. R. W. Bickmore, "Fraunhoffer pattern measurements in the Fresnel region", *Canadian J. Phys.*, **35**, p. 1290, 1957.
4. R. W. Bickmore, "Measurements of a focused radiating system", *Canadian J. Phys.*, **35**, p. 1299, 1957.
5. H. G. Booker and P. C. Clemow, "The concept of an angular spectrum of plane waves and its relation to that of polar diagram and aperture distribution", *J. Instn Elect. Engrs*, **97**, Part III, p. 11, January 1950.
6. R. H. J. Bates and J. Elliot, "The determination of the true side lobe level of long broadside arrays from radiation pattern measurements made in the Fresnel region", *Proc. Instn Elect. Engrs*, **103C**, p. 307, March 1956.
7. E. V. Jull, "The Predictions of Radiation Patterns from Near-field Measurement". Ph.D. Thesis, London University, July 1960.
8. J. Brown and E. V. Jull, "The prediction of aerial radiation patterns from near-field measurements", *Proc. Instn Elect. Engrs*, **108B**, p. 635, November 1961.
9. G. Joos, "Theoretical Physics", p. 75, 3rd Edn, (Blackie, London, 1958).

DISCUSSION

Under the Chairmanship of Mr. P. F. Mariner

Mr. J. Welsh: Mr. Blakey has described three methods of determining the far-field radiation pattern of a large aerial, all involving making measurements within the Rayleigh distance. These were briefly: (a) by moving the feeder, in the case of a horn and dish type aerial, by a calculated amount; (b) by bending the collimator into an arc of a circle of known radius, mainly applicable to the linear array; and (c) by making measurements within the

Rayleigh distance on the correctly adjusted aerial and then applying some rather lengthy mathematics.

One of the principal reasons for wanting to measure radiation patterns is to assist the setting up of an aerial, i.e. finding the optimum feeder position or adjusting the phase and amplitude of individual radiators to produce the required pattern.

How would the above procedures be used to assist setting up an aerial when the correct feeder position, for example, is not known and, in the case of large aerials, where the above techniques would seem to be most applicable, is bending the collimator or linear array a practical proposition?

Mr. J. R. Blakey (*in reply*): In the case of a 'horn and dish' system the displacement technique could be applied in reverse by displacing the feed the required distance δ from its designed position, then adjusting the position of the displaced feed to give the optimum simulated pattern. The final feed position for normal operation will then be at a distance δ towards the dish from the position for optimum simulated pattern.

The bending of arrays with large physical dimensions is almost certainly impractical except in the case of wave-guide arrays operating at high microwave frequencies. The method may be of value in finding the optimum adjustment of individual elements of a plane array if these are mounted on a special curved collimator.

Mr. R. F. Kyle: The author dealt with the method developed by J. Brown and Jull for calculating far-field patterns knowing the phase and amplitude round a closed surface, which as I understand it requires considerable computational effort. I would like the author to expand on his method of using an analogue computer in the form of a filter network and its improvement on the above method. It would seem that to get reasonably accurate far-field diagrams, except for very short arrays, the design of the filter network will be complex, and the setting up procedure will be lengthy; would the author please comment.

Mr. Blakey (*in reply*): The proposed technique is essentially an extension of Brown and Jull's work. The periodic signal obtained in the near field is operated on directly by a filter with a suitable response such that its output is proportional to the far-field pattern. Preliminary calculations show that the filter characteristic is quite a simple function and the filter synthesis procedure should be quite straightforward and measurements at a distance of $\frac{1}{20}$ of the far-field limit appear to be feasible. The setting up procedure might well be tedious but would normally be done in the laboratory. The potential advantage of the method is the virtually direct determination of the pattern.

Mr. P. F. Mariner: One of the problems in determining performance of aerials is in knowing, with a good degree of accuracy, what are the multiple side-lobe levels and what is the level of all-round radiation, and sometimes in particular of back-lobe radiation. It seems to me that, in this method which has been discussed, there will be some difficulty in predicting side-lobe levels, even if you are able to make the measurements through 360 deg of rotation of the aerial to be tested. The mathematical computations involved and the experimental equipment required would appear to be about as complicated in the method you have suggested as those involved in measuring the phase amplitude next to the aperture of the aerial, where a very large proportion of the radiated energy can be sampled,

and where it is consequently possible to deduce side-lobes to a very wide angle indeed. I agree, however, that in the method discussed there is unlikely to be any interaction between the aerial and the measuring probe, which may occur in the case I have cited.

Mr. Blakey (*in reply*): The question of complexity either in technique or computation seems inevitable in the near-field approach. The justification for its use is surely that one is prepared to accept the complexities in order to obtain an answer at all. While the method discussed may not do anything to reduce complexity, it may offer greater convenience in its application.

Mr. R. Graham: In the Brown and Jull method of determining far-field radiation pattern by measuring the field at a moderate distance from the aerial, is it necessary to measure the field pattern in both amplitude and phase, or is the phase assumed to be constant?

If the phase measurement is required, how would Mr. Blakey suggest that this could be made?

Mr. Blakey (*in reply*): In general both amplitude and phase must be measured in the near-field region. The standard technique for determining the phase involves the comparison of the signal from the test aerial with one derived from the transmitter by means of a r.f. bridge. A typical example of the method is described in Brown and Jull's paper.

Mr. C. G. F. Cooper: Mr. Blakey, when he discussed how one might obtain the far-field pattern of a straight array by making relatively close pattern measurements on a curved array, considered only broadside arrays. Would he care to extend the discussion to arrays firing at other angles?

Mr. Blakey (*in reply*): The method can be extended, in principle, to arrays firing at other angles and appropriate curves could be determined. For angles not too far from broadside the simple circular array would probably be adequate. Equation (13) can be used to determine the permissible angular deviation from broadside.

Mr. J. Lait: Many radar equipments employ electronic scanning, in which progressively changing radiation patterns are produced by varying the phase distribution across the aerial aperture. Designers would welcome methods of near-field measurement which would enable them to assess how much pattern deterioration may be expected near the edges of the scan or how great frequency shift may be introduced before reaching the greatest tolerable pattern distortion.

Can the author's proposals be extended to cover cases of this kind, in which the direction of maximum radiation does not coincide with the normal to the plane of the aperture, so that on-axis measurements and the use of cylindrical co-ordinates are no longer appropriate?

Mr. Blakey (*in reply*): Although I am not qualified to comment about scanning arrays, near-field methods would seem to be applicable to such systems if each angle of fire is considered separately. In the most general case when

the radiated field cannot be represented by a super-position of cylindrical modes the expansion can be achieved in spherical modes, but the computational effort in this case would be very great indeed.

Dr. D. J. Neale: The first two methods of correcting for the phase error across the aperture, outlined in the paper, are only valid for an extremely narrow segment of the radiation pattern. Can the transform method cope with a wider range of angles, e.g. for a broadcast array where

angles of up to 45 deg or more are of considerable interest?

Mr. Blakey (*in reply*): Dr. Neale's first statement requires qualification. The two methods described will only be valid for an extremely narrow segment of the radiation pattern if measurements are made at extremely short distances. The normal Fourier transform will cope with angles of $\pm$ 90 deg about broadside. The modified transform described necessarily introduces greater uncertainties at wide angles.

STANDARD FREQUENCY TRANSMISSIONS

(Communication from the National Physical Laboratory)

Deviations, in parts in 10^{10}, from nominal frequency for **October 1964**

October 1964	GBR 16kc/s 24-hour mean centred on 0300 U.T.	MSF 60 kc/s 1430–1530 U.T.	Droitwich 200 kc/s 1000–1100 U.T.	October 1964	GBR 16 kc/s 24-hour mean centred on 0300 U.T.	MSF 60 kc/s 1430–1530 U.T.	Droitwich 200 kc/s 1000–1100 U.T.
1	− 150·7	− 150·9	+ 7	17	− 150·8	− 150·6	− 11
2	− 151·0	− 151·2	+ 7	18	− 150·4	− 150·2	− 11
3	− 151·1	− 150·5	+ 8	19	− 150·7	− 151·7	− 11
4	− 150·3	− 150·7	+ 9	20	− 151·7	− 150·8	− 10
5	− 150·5	− 150·8	− 14	21	− 150·7	− 149·7	− 8
6	—	− 150·3	− 14	22	− 150·1	− 149·8	− 7
7	—	− 151·0	− 14	23	− 149·6	− 149·6	− 8
8	− 150·8	− 150·8	− 12	24	− 149·7	− 150·3	− 6
9	− 150·2	—	− 14	25	− 151·4	− 152·1	− 6
10	− 149·6	− 149·5	− 12	26	− 151·7	− 151·1	− 5
11	− 149·5	− 149·5	− 11	27	− 151·1	− 149·8	− 6
12	− 149·7	− 151·3	− 11	28	− 149·8	− 150·1	− 5
13	− 150·6	− 150·4	− 12	29	− 150·2	− 150·9	− 5
14	− 150·5	—	− 10	30	− 150·8	− 151·1	− 5
15	− 151·0	− 150·7	− 12	31	− 149·9	− 149·8	− 4
16	− 150·8	− 151·8	− 10				

Nominal frequency corresponds to a value of 9 192 631 770 c/s for the caesium F,m (4,0)–F,m (3,0) transition at zero field.

CHANGES IN RADIO BROADCASTS BY NATIONAL BUREAU OF STANDARDS

In accordance with the policy of co-operation between N.B.S. and the U.S. Naval Observatory, changes in the time ticks broadcast by stations of both were made to improve their agreement. The ticks broadcast by N.B.S. stations WWV (Greenbelt, Md.), WWVH (Maui, Hawaii), and WWVB (Fort Collins, Colo.) were retarded 1·0 milli-seconds at 0000 U.T. on 1st October 1964. At the same time the time ticks broadcast by Navy stations were advanced 1·6 milliseconds. This brings into closer agree-ment epochs of the time standards broadcast by the United States Government.

Station WWVH discontinued its forecasts of propaga-tion in the North Pacific on 1st November 1964. These propagation forecasts were formerly given hourly in Morse code, during the first half of the fifth minute of each hour. The predictions were supplied by the N.B.S.

North Pacific Radio Warning Service at Anchorage, Alaska, which will be known in future as the C.R.P.L. High Latitude Space Environment Monitoring Station. It will concentrate on the monitoring and rapid reporting of geophysical events of interest not only to h.f. radio propagation users, but also to those interested in changes in the space environment. Thus, though h.f. propagation notices will no longer be given on WWVH, it is anticipated that at a later date, as the need develops, some other state-ment upon solar and geophysical activity will be dis-seminated in their place.

Propagation forecasts will continue to be given every five minutes from station WWV. Although the WWV forecasts are for the North Atlantic area, predictions of disturbed conditions throughout the northern auroral region will be generally applicable to the North Pacific area also.

The Significance of the Excess Charge Product in Drift Transistors

By

G. G. BLOODWORTH, M.A.,
D.U.S. (*Associate Member*)†

Summary: The net current due to diffusion and drift in the base of a transistor is shown to be everywhere equal to that produced by drift alone due to a hypothetical distribution of potential equal at any point to the product of the densities of majority and excess minority carriers. Using this analogy, the large-signal equivalent circuit is obtained without any restriction of the geometrical shape or impurity variation in the base.

List of Symbols

A	surface area
C	distributed capacitance in the analogue
C_E, C_C	capacitances in the equivalent circuit
D_n, D_p	diffusion constants of electrons and holes
E	electric field intensity
f_c	cut-off frequency
G, G_s	conductances distributed in the bulk and at the surface
I_E, I_C	emitter and collector currents
I_{EB0}, I_{CB0}	emitter and collector saturation currents
J	current density in the analogue
J_n, J_p	current densities due to electrons and holes
J_s	component of J_p normal to the surface
k	Boltzmann's constant
n	density of electrons
n_N	equilibrium density of electrons in N-type material
n_e, n_{Ne}	values of n and n_N adjacent to the emitter junction
n_c, n_{Nc}	values of n and n_N adjacent to the collector junction
n_s, n_{Ns}	values of n and n_N adjacent to the surface

$p, p_N, p_e, p_{Ne},$ p_c, p_{Nc}, p_s, p_{Ns}	corresponding densities of holes
q	electronic charge
Q_B	charge stored in the base
r	bulk recombination constant
R_C, R_D, R_E	resistances in the equivalent circuit
S	surface recombination velocity
t	time
T	absolute temperature
$v_{e(b)}, v_{c(b)}$	emitter and collector bias voltages
V	potential analogous to ϕ
V_s	potential at the surface
$\mathscr{V}$	volume
α_N, α_I	normal and inverse current gains with grounded-base
γ	emission efficiency
μ_n, μ_p	mobilities of electrons and holes
$\phi \equiv q(p-p_N)n_N$	excess charge product
ϕ_e, ϕ_c	values of ϕ adjacent to the emitter and collector junctions
ϕ_s	value of ϕ adjacent to the surface
$\phi_0 \equiv q n_N p_N$	
ρ	resistivity in the analogue
τ	lifetime of excess holes in the base
θ	angle between J_p and J_s

1. Introduction

The transport of excess minority carriers across the base of a bipolar junction transistor is entirely due to diffusion if the concentration of impurities—the 'doping level'—is everywhere the same and the maximum density of minority carriers is negligible compared with the density of majority carriers. As the diffusion current is proportional to the gradient

of the excess density of minority carriers, this density can be treated as analogous to an electrical potential which would produce the same current distribution by ohmic conduction (drift) in a resistive medium occupying the same space. By dividing the hypothetical distributed resistance into three lumped resistances,[1] the current division between the terminals of the transistor can be represented. The large signal equivalent circuit is completed by introducing at each

† Department of Electronics, University of Southampton.

junction an appropriate non-linear transformation between the hypothetical voltage representing excess density and the real bias voltage.

At high levels of injection, when the density of the minority carriers becomes comparable with that of the majority carriers, the distribution of majority carriers is altered significantly and an electric field appears which assists the passage of minority carriers by producing additional drift current. This system is non-linear[7] and the base can no longer be represented by linear resistances.

If the density of minority carriers is everywhere negligible compared with that of the majority carriers, but a drift field is present due to variations in doping level, then the system is again linear because the field is independent of the minority carrier distribution. The analysis of this case is well-known for a one-dimensional transistor, i.e. a transistor with plane parallel junctions infinite in extent and with no variations in doping level parallel to the junctions. Descriptions of this model[2, 8] usually assume simple forms of doping variation across the base, for example, an exponential decrease which produces a uniform field.

In the present paper an analysis is given for an arbitrary shape in three dimensions, and all restrictions on the doping variation and the excess charge distribution are removed except those needed to maintain linearity. The doping level may vary arbitrarily across the faces of the junctions as well as throughout the base, and it is not even assumed that the diffusion constant for the minority carriers does not vary with the doping level. The same familiar equivalent circuit is obtained, and it is shown that the equations given by Ebers and Moll[10] for the diffusion transistor are valid in this more general case.

It will first be shown that the vector sum of the minority carrier diffusion and drift currents at any point is proportional to the gradient of the product of the densities of majority and excess minority carriers. This product, the excess charge product, is recognized as analogous to an electrical potential which would produce the total current by drift alone. The hypothetical system of distributed resistances can be represented by lumped components, as before. If the majority carrier density is constant in some region, because the doping is constant, the analogy adopted here reduces to the familiar one for the diffusion transistor since then the only variable is the excess minority carrier density.

To preserve linearity, it is again necessary to assume that the field in the base is entirely due to the distribution of majority carriers in thermal equilibrium, produced by the static doping variation. The effects of base width modulation are also disregarded.

2. Current Flow in the Base

The analysis will be presented for a pnp transistor, and it will be assumed, until a later section, that each junction has unity emission efficiency. This means that, when forward biased, each junction injects holes into the base, but the electron flow out of the base through the junction is negligible. The holes leaving a junction are transported by drift and diffusion either to the other junction or to recombination with electrons at the surface or in the bulk of the semiconductor. Electrons flow from the base terminal to provide the recombination current.

The net current density at any point due to holes is given by the vector sum of the drift and diffusion current densities, i.e. by

$$J_p = q\mu_p pE - qD_p \,\mathrm{grad}\, p \qquad \ldots\ldots(1)$$

where p is the density of holes with mobility μ_p and diffusion constant D_p, E is the electric field, and q is the magnitude of the electronic charge. Similarly, the net current density due to electrons is given by

$$J_n = q\mu_n nE + qD_n \,\mathrm{grad}\, n \qquad \ldots\ldots(2)$$

The basic assumption to be made is that J_n is sufficiently small for the distribution of field E experienced by the holes to be given accurately by equating J_n to zero:

$$E = -\frac{D_n}{\mu_n} \cdot \frac{1}{n} \,\mathrm{grad}\, n \qquad \ldots\ldots(3)$$

This assumption can be justified in a general way for the regions of interest by comparing the magnitudes of the terms in eqns. (1) and (2). Consider first the active region of the base near to and between the junctions, where J_p is relatively large and J_n is much smaller because only a small proportion of holes recombine. Suppose that the hole current in some small volume is predominantly due to drift, so that $J_p \simeq q\mu_p pE$. The drift component of J_n will be much greater than J_p because $\mu_n n$ exceeds $\mu_p p$ due to the n-type impurity. The net current J_n must therefore be negligible compared with its own drift component; the two components in eqn. (2) must cancel almost exactly, and eqn. (3) follows immediately. On the other hand, if the hole flow is primarily by diffusion the use of an approximate expression for E is clearly allowable.

This argument cannot be applied to the more remote regions of the base because the outward flow of holes changes progressively to an inward flow of electrons and J_p eventually becomes smaller than J_n. The current density J_n should not, however, rise significantly because the increase in current is balanced by the increase in physical dimensions. The expression given for E should therefore remain valid throughout the remote regions as a whole. It may fail for the region immediately adjacent to the base terminal

where the electron current originates and its density may be high, particularly if the effective contact area is small. As only a tiny fraction of the excess holes are in this region the validity of the model to be described for the hole flow should not be seriously affected.

Using Einstein's relation $\dfrac{D_n}{\mu_n} = \dfrac{D_p}{\mu_p}$ and eliminating E from eqns. (1) and (3), the hole current density becomes

$$J_p = -qD_p\left[\frac{p}{n}\,\mathbf{grad}\,n + \mathbf{grad}\,p\right] = -\frac{qD_p}{n}\,\mathbf{grad}\,pn$$

It is further assumed that E is independent of p, because no net space charge is introduced by the injection from the junctions. The excess charges are equal:

$$p - p_N = n - n_N$$

where p_N and n_N are the thermal equilibrium densities of holes and electrons at the same point in the N-type material. The essential requirement is that the excess charge density produced at any point due to injection from the junctions must be negligible compared with n_N, so that $n \simeq n_N$. With high level injection this assumption fails: n then becomes a function of the excess hole density $(p - p_N)$ and so the drift field E (eqn. (3)) depends on p and J_p (eqn. (1)) is not a linear function of p. Ignoring this effect, J_p is given by the linear equation

$$J_p = \frac{-qD_p}{n_N}\,\mathbf{grad}\,pn_N \qquad \ldots\ldots(4)$$

If there is no excess charge $(p = p_N)$, the current density is expected intuitively to be zero:

$$0 = \frac{-qD_p}{n_N}\,\mathbf{grad}\,p_N n_N \qquad \ldots\ldots(5)$$

This result also follows directly from the fact[3] that the product of the equilibrium densities of minority and majority carriers is independent of the doping level, and therefore is constant throughout the base.

The current density can be expressed in terms of the excess density of holes, by combining eqns. (4) and (5).

$$J_p = \frac{-qD_p}{n_N}\,\mathbf{grad}\,(p - p_N)n_N$$

There must be a net flow of holes into any small volume due to bulk recombination and, under transient conditions, to any increase in charge storage. Therefore J_p must satisfy the continuity equation

$$-\,\mathrm{div}\,J_p = \frac{q(p - p_N)}{\tau} + q\frac{\partial p}{\partial t}$$

where τ is the lifetime of the excess holes. It can reasonably be assumed that $\tau n_N = r$, where r is some constant, because the probability of recombination for holes is proportional to the density of electrons n_N. However, the model to be developed does not actually require r to be independent of n_N, although it is essential that τ be assumed independent of p to maintain the linearity of the equations in p.

As p_N and n_N are independent of t, the continuity equation can be expressed in the form

$$-\,\mathrm{div}\,J_p = \frac{q(p - p_N)n_N}{r} + \frac{q}{n_N}\frac{\partial}{\partial t}(p - p_N)n_N$$

The equations to be satisfied by J_p are therefore

$$\left.\begin{aligned} -\,\mathrm{div}\,J_p &= \frac{\phi}{r} + \frac{1}{n_N}\frac{\partial\phi}{\partial t} \\ -J_p &= \frac{D_p}{n_N}\,\mathbf{grad}\,\phi \end{aligned}\right\} \qquad \ldots\ldots(6)$$

where $\phi \equiv q(p - p_N)n_N$ may be termed the 'excess-charge product'.

3. The Analogy between Excess-charge Product and Potential

The equations satisfied by the current I and voltage V at any point x on a transmission line are similar to those relating J_p and ϕ. If ρ is the series resistance, G the shunt conductance, and C the shunt capacitance, each per unit length, then

$$-\frac{\partial I}{\partial x} = GV + C\frac{\partial V}{\partial t}$$

$$-I = \frac{1}{\rho}\frac{\partial V}{\partial x}$$

Generalized to three dimensions, these equations become

$$\left.\begin{aligned} -\,\mathrm{div}\,J &= GV + C\frac{\partial V}{\partial t} \\ -J &= \frac{1}{\rho}\,\mathbf{grad}\,V \end{aligned}\right\} \qquad \ldots\ldots(7)$$

where J is the current density corresponding to any distribution of potential V. ρ is now the volume resistivity and G and C represent distributed conductance and capacitance linking each volume element with some external terminal, each per unit volume. Such a system is not, of course, realizable in the ordinary way because G and C would have to be connected via some fourth dimension. However, a comparison of eqns. (6) and (7) suggests that the current J produced entirely by drift under the action of the potential distribution V may be identified with the net current in the transistor due to both drift and diffusion, providing the following analogies are drawn:

(a) The potential V represents the excess charge product ϕ.

(b) The conductance G represents the reciprocal of the recombination constant r.

(c) The capacitance C represents the reciprocal of the electron density n_N.

(d) The resistance ρ represents the ratio of n_N to the diffusion constant D_p.

This model is viable for the transistor in only three dimensions because electrons flow from the base terminal to supply the recombination current and to neutralize any increase in the density of holes.

The current J_p must be everywhere normal to 'equipotential' surfaces in ϕ, as illustrated in Fig. 1 for the case where current from the emitter divides between the collector and the surface of the base. Elemental tubes containing the current flow can be constructed, as shown, but it should be noted that the current varies along each tube due to the effects of G and C. If the walls of each tube are everywhere parallel to the current flow, and the cross-sectional area is so small that transverse variations in G, C and ρ can be ignored, then each tube becomes a one-dimensional system equivalent to a non-uniform transmission line. The behaviour of the entire system is determined by the boundary conditions at the junctions and the free surface of the base: these will now be considered.

4. The Emitter and Collector Junctions

The excess density of holes produced in the base by an emitter bias voltage $v_{e(b)}$ (positive for forward bias), under steady-state conditions, is given by[4]

$$p_e - p_{Ne} = p_{Ne}\left(\exp\frac{qv_{e(b)}}{kT} - 1\right)$$

where k is Boltzmann's constant and T is the absolute temperature. The additional suffix e is used here to distinguish the hole densities adjacent to the emitter junction. The excess charge product in this region is

$$\phi_e = qn_{Ne}(p_e - p_{Ne}) = qn_{Ne}p_{Ne}\left(\exp\frac{qv_{e(b)}}{kT} - 1\right)$$

As the value of the product $n_N p_N$ is the same throughout the base, the values of ϕ adjacent to the emitter and collector junctions are given respectively by

$$\phi_e = \phi_0\left(\exp\frac{qv_{e(b)}}{kT} - 1\right)$$
$$\phi_c = \phi_0\left(\exp\frac{qv_{c(b)}}{kT} - 1\right) \quad \ldots\ldots(8)$$

where $\qquad \phi_0 \equiv qn_N p_N$

Disregarding variations in temperature, these equations show that the action of each junction is to produce an equipotential surface in ϕ at the edge of the depletion layer with a value uniquely defined by the bias voltage. As ϕ is positive for a forward-biased junction, and negative for a reverse-biased junction, the potential analogy shows that current flows from the former to the latter.

As the relation between ϕ_e and $v_{e(b)}$ is independent of n_{Ne}, variations in impurity concentration and hence in n_{Ne} across the face of the junction are immaterial in this context. Such variations would introduce transverse components in E and hence in the drift component of J_p, but these would be balanced by transverse diffusion so that J_p flows normal to the face. The crucial point is that the face of the depletion layer must be an equipotential surface in ϕ because the bias voltage $v_{e(b)}$ is constant right across the junction. Variations in n_{Ne} must be automatically compensated by variations in the thickness of the depletion layer and in the level of injected charge p_e.
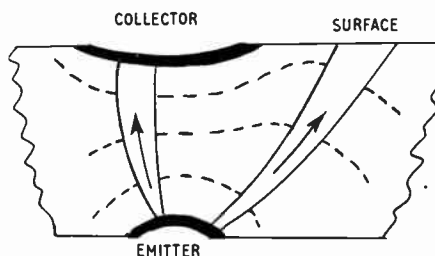
Fig. 1. Tubes of current flow. The broken lines are sections of equipotential surfaces (ϕ constant).

5. Surface Recombination

Holes which reach the free surface of the base under the combined action of drift and diffusion disappear there by recombination, and thus constitute a positive current into the surface from the bulk material. A corresponding current of electrons flows from the base terminal through the bulk material to the surface.

The hole current J_p does not in general flow into the surface at right angles because it is clearly not an equipotential surface in ϕ as it is contiguous to the junctions and $\phi_e \neq \phi_c$. However, the component of J_p normal to the free surface is given by[9]

$$J_s = q(p_s - p_{Ns})S \qquad .$$

where $(p_s - p_{Ns})$ is the excess density of holes just beneath the surface and S, which represents the average velocity of these holes into the surface, is known as the surface recombination velocity. The value of S may vary over the surface, but for the sake of linearity it is assumed to be independent of the hole density at any particular point. An elemental tube of current flow from the emitter to the surface is shown in Fig. 2.
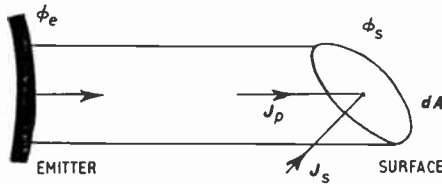
Fig. 2. Surface recombination.

The excess charge product adjacent to the surface is given by

$$\phi_s = q n_{Ns}(p_s - p_{Ns})$$

where n_{Ns} is the electron density in this region.

The current density normal to the surface is therefore given by

$$|J_p| \cos \theta = |J_s| = \frac{|S|}{n_{Ns}} \phi_s$$

where θ is the angle between J_p and J_s. If ϕ is to be replaced by a corresponding potential distribution, surface recombination can be allowed for by incorporating a system of conductances each connecting an element of surface dA with the base terminal. The current flowing in such a conductance $G_s |dA|$ due to the electrical potential V_s at the surface must equal the recombination current $J_s.dA$ due to ϕ_s.

Thus

$$V_s G_s |dA| = J_s.dA = \phi_s . \frac{S}{n_{Ns}} . dA$$

As V_s is to be identified with ϕ_s, the required conductance per unit area is given by $G_s = \frac{|S|}{n_{Ns}}$.

6. Summary of Equations

The currents flowing into the various regions of the base are as follows:

surface $$\int_A J_p.dA = \int_A \frac{\phi}{n_{Ns}} S.dA$$

bulk $$-\int_r \operatorname{div} J_p.d\mathscr{V} = \int_V \left(\frac{\phi}{r} + \frac{1}{n_N}\frac{\partial\phi}{\partial t}\right)d\mathscr{V}$$

junction $$\int_A J_p.dA = -\int_A \frac{D_p}{n_N}\operatorname{grad}\phi.dA$$

The sum of the left-hand sides of these equations is zero, according to Gauss' Theorem, if the integrals are taken over the complete surface including the junctions and throughout the volume. This represents conservation of charge, because it means that the net flow of holes into the base through the junctions at any instant equals the total sum of the recombination currents and the rate of increase of stored charge.

The above system of equations for the current distribution within the base can be solved, at least in principle, for any geometrical shape and any given values of ϕ_e and ϕ_c, as determined by the bias voltages $v_{e(b)}$ and $v_{c(b)}$. However, the essential features of the system can be elucidated by a physical argument.

7. A Two-dimensional Model

The main characteristics of the current distribution in the base can be visualized in terms of familiar electrical components if ϕ is treated as potential and one dimension is ignored for the moment. This simplification makes the physical conception relatively easy and does not invalidate any of the results obtained.

For a two-dimensional transistor the base becomes a conducting sheet with non-uniform surface resistivity ρ. The emitter and collector junctions become portions of the perimeter held at potentials ϕ_e and ϕ_c relative to the base terminal, which can be represented by a perfectly conducting ground plane as shown in Fig. 3. Recombination in the bulk and at the surface in the actual transistor is now represented by shunt conductances distributed across the sheet (G) and around the free edges (G_s). The effects of charge storage in the distributed shunt capacitance C will be considered in a later section after the equivalent circuit for the steady state has been obtained.

Suppose that the emitter is forward biased, and the collector bias is zero; then ϕ_e is a positive voltage and ϕ_c is zero. The transition between these values can be represented by a map of equipotential lines drawn across the conducting sheet. All potentials and currents are proportional to ϕ_e, and the total emitter current divides between the recombination conductances and the short circuit at the collector in a ratio which is determined by the magnitudes of the distributed shunt and series resistances. The terminal currents can therefore be divided between two imaginary lumped resistances R_E and R_D: if I_C and I_E are the currents flowing into the emitter and collector respectively, as shown in Fig. 3, then

$$\phi_e = -R_D I_C = R_E(I_E + I_C) \qquad (\phi_c = 0)$$

If ϕ_e is negative (reverse bias) and ϕ_c remains zero, all currents and potentials are reversed in sign and the same equation applies. A similar equation can be written for the currents produced when ϕ_e is zero and a potential ϕ_c is applied at the collector:

$$\phi_c = -R_D I_E = R_C(I_E + I_C) \qquad (\phi_e = 0)$$

The same equivalent resistance R_D appears, as a direct consequence of the 'Principle of Reciprocity'. The current produced in the emitter lead, when short-circuited, by a voltage generator in the collector lead must equal the current produced in the short-circuited collector lead by the same voltage generator trans-
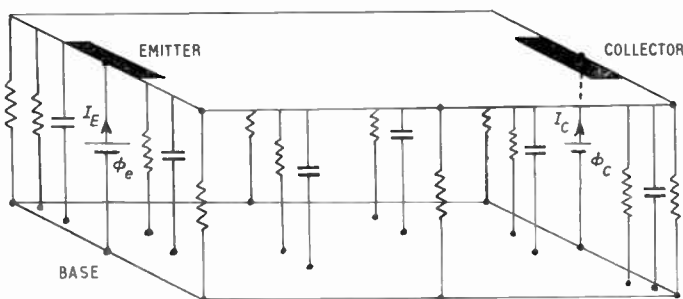
Fig. 3. Two-dimensional base with shunt admittances.

ferred to the emitter. The lack of physical symmetry, both in the sizes of the junctions and the distribution of resistances, produces different base currents in the two cases, and so $R_E \neq R_C$ in general.

The terminal currents produced by ϕ_e and ϕ_c applied separately are therefore as follows:

If $\phi_c = 0$: $\quad I_C = -\dfrac{\phi_e}{R_D} \quad I_E = \phi_e\left(\dfrac{1}{R_E} + \dfrac{1}{R_D}\right)$

If $\phi_e = 0$: $\quad I_E = -\dfrac{\phi_c}{R_D} \quad I_C = \phi_c\left(\dfrac{1}{R_C} + \dfrac{1}{R_D}\right)$

According to the Principle of Superposition, the net currents produced in a linear network by two voltage generators are each the sum of the corresponding currents produced by the generators separately. The terminal currents are therefore, in general

$$I_E = \frac{(\phi_e - \phi_c)}{R_D} + \frac{\phi_e}{R_E}$$
$$I_C = \frac{(\phi_c - \phi_e)}{R_D} + \frac{\phi_c}{R_C} \qquad \ldots\ldots(9)$$

The equivalent circuit shown in Fig. 4 can therefore be drawn for the transistor, and used in conjunction with the non-linear transformations eqn. (8) to relate the various currents to the actual bias voltage applied.

The above results can be obtained without reducing the transistor to two dimensions, taking the electrical analogue, and invoking the Principles of Reciprocity and Superposition. Ebers and Moll[10] obtained the essential result

$$I_E = a_{11}\phi_e + a_{12}\phi_c \qquad \ldots\ldots(10)$$
$$I_C = a_{21}\phi_e + a_{22}\phi_c$$

for the case of zero drift field throughout the base, and showed that $a_{12} = a_{21}$ by applying Green's theorem to the vector equations. The vector equations obtained in the present paper are equivalent to those of Ebers and Moll, with excess charge product

replacing excess hole density: this generalization does not invalidate the proof of eqn. (10). If the excess charge products ϕ_e and ϕ_c are treated finally as analogous to voltages, these equations transform to eqn. (9) and the equivalent circuit is obtained.
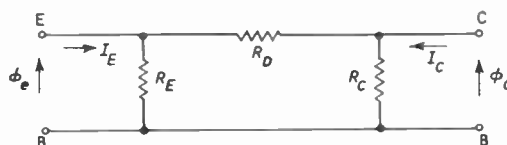
Fig. 4. Equivalent circuit for the steady state.

8. The Large-signal Equivalent Circuit

The transformations (eqn. (8)) between the imaginary voltages ϕ_e and ϕ_c in the equivalent circuit and the real voltages $v_{e(b)}$ and $v_{c(b)}$ depend only on the value of voltage ascribed to ϕ_0. This value is clearly arbitrary, and might be taken simply as one volt. The values of the three resistances R_C, R_D and R_E can then be determined for any particular transistor by measuring the terminal currents under suitable conditions.

The equivalent circuit gives very simple equations if ϕ_e or ϕ_c is zero, but in practice the significant currents are usually given for the cases where the junctions are alternately reverse biased, not short-circuited. If the collector is reverse biased so that

$$-v_{c(b)} \gg \frac{kT}{q}\left[\simeq 25 \text{ mV at room temperature}\right]$$

then $\qquad\qquad \phi_c = -\phi_0$

and therefore

$$I_E = \phi_e\left(\frac{1}{R_D} + \frac{1}{R_E}\right) + \frac{\phi_0}{R_D}$$

$$I_C = -\frac{\phi_e}{R_D} - \phi_0\left(\frac{1}{R_D} + \frac{1}{R_C}\right)$$

Eliminating ϕ_e yields

$$I_C = -\frac{R_E}{R_E+R_D} \cdot I_E - \phi_0\left(\frac{1}{R_C} + \frac{1}{R_E+R_D}\right)$$

The grounded-base current gain in the normal direction and the collector saturation current are therefore given by

$$\alpha_N = \frac{R_E}{R_E+R_D}, \qquad I_{CB0} = \phi_0\left(\frac{1}{R_C} + \frac{1}{R_E+R_D}\right)$$

Similarly, for inverse operation,

$$\alpha_I = \frac{R_C}{R_C+R_D}, \qquad I_{EB0} = \phi_0\left(\frac{1}{R_E} + \frac{1}{R_C+R_D}\right)$$

The three resistances can be found by measuring any three of the above quantities and solving the appropriate equations. Incidentally, the well-known relation

$$\alpha_N I_{EB0} = \alpha_I I_{CB0}$$

follows directly from these equations.

9. Charge Storage in the Base

The charge stored in the shunt capacitance in any region of the base in the two-dimensional transistor is given by the sum of the charges produced by ϕ_e and ϕ_c applied separately, invoking again the Principle of Superposition. As the steady-state potential at any point is proportional to the voltage of the generator applied, the charge stored in any small region is also proportional to the generator voltage. The total charge, in the steady state, is therefore of the form

$$Q_B = C_E\phi_e + C_C\phi_c$$

This equation will also be applicable during transient changes if ϕ_e and ϕ_c change so slowly that the charge distribution follows with negligible delay. The equivalent circuit can therefore be amended as shown in Fig. 5 to allow for slow changes.
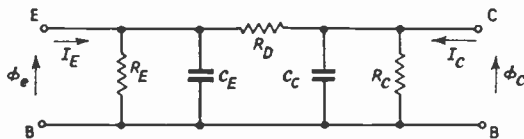
Fig. 5. Equivalent circuit including stored charge.

A single time-constant is thus introduced when the transistor is turned on without saturating, because if the collector remains sufficiently reverse-biased ϕ_c does not change and so C_C has no effect. For sinusoidal drive, the grounded-base current gain becomes

$$\alpha_N = \frac{\alpha_{N0}}{1+\mathrm{j}\dfrac{f}{f_c}}$$

where

$$\alpha_{N0} = \frac{R_E}{R_E+R_D} \quad \text{(as before)}$$

and the cut-off frequency is given by

$$f_c = \frac{R_E+R_D}{2\pi C_E R_E R_D}$$

The values of C_E and C_C can therefore be found by measuring, for example, α_N and α_I at any frequency where the stored charge has an appreciable effect.

10. Emission Efficiency

The model so far developed accounts for the division between the terminals of a pnp transistor of the currents carried across the junctions by holes. The electron currents will now be considered. The fraction of each junction current carried by holes, known as the emission efficiency, is usually near to unity because the emitter and collector regions are more strongly doped than the base. However, the electron current flows directly between each junction and the base terminal and forms an appreciable fraction of the base current.

As the emission efficiency is independent of current at low injection levels,[11] it follows that the total electron current from each junction must vary with bias voltage in the same way as the hole current, that is in direct proportion to the excess charge product. It is not necessary to assume that the emission efficiency γ has the same value all over the face of the junction. The total electron current at the emitter is given by

$$\int_A J_n \cdot \mathrm{d}A = \int_A \left(\frac{1}{\gamma} - 1\right) J_p \cdot \mathrm{d}A$$

where the integral is taken over the junction area. If ϕ_c is zero, the hole current density J_p at the emitter is everywhere proportional to ϕ_e, i.e. $J_p = C\phi_e$ where C may vary across the junction but is independent of ϕ_e. Thus the electron current

$$\int_A J_n \cdot \mathrm{d}A = \int_A \left(\frac{1}{\gamma} - 1\right)\phi_e \cdot C \cdot \mathrm{d}A$$

$$= \phi_e \int_A \left(\frac{1}{\gamma} - 1\right) \cdot C \cdot \mathrm{d}A$$

is proportional to ϕ_e as stated, because ϕ_e is constant across the junction.

The electron current at the emitter is, however, independent of ϕ_c because there is no coupling between the electron flows at the two junctions. Similarly, the electron current at the collector is proportional to ϕ_c and independent of ϕ_e. These current components flow directly to the base terminal

(using the conventional direction of current flow) and can clearly be allowed for in the equivalent circuit by introducing additional conductances in parallel with R_E and R_C respectively.

The separate components of base current, due to recombination and to emission efficiency less than unity, cannot be distinguished in the resistance values obtained by measurement of α_N, etc. Although the equivalent circuit obtained for any transistor provides a complete description of the relationship between the bias voltages and terminal currents it is not possible to determine the current carriers involved without additional information about the physical dimensions, hole lifetime, doping variations, etc.

The charge carried by the capacitances C_E and C_C is always equal to the excess charge of holes in the base, regardless of emission efficiency, because there is no storage of majority carriers.

11. Non-linear Effects

The main effects which invalidate the linear representation of the base at high current levels may be summarized as follows.

It has been assumed that $p \ll n_N$ and hence that the field E is independent of the distribution of holes. When p becomes comparable with n_N the additional field assists the transfer of holes: in effect, R_D decreases.

Secondly, the emission efficiency γ has been taken as independent of current. At high currents γ decreases[11] and so R_E and R_C are effectively reduced.

Thirdly, the voltage drop associated with the net flow of electrons from the base terminal into the active region between the junctions has been ignored. At heavy currents this voltage becomes appreciable and the forward-bias voltage in the middle of the emitter junction becomes less than that around the edges.[12] In this situation ϕ_e is no longer constant across the junction.

12. The Effects of Non-uniform Base Doping

The equivalent circuit developed in the previous sections, and others that are essentially the same, are commonly used to represent diffusion transistors, where it is assumed at the outset that n_N is substantially constant throughout the base. Instead of

ϕ_e and ϕ_c, the excess charge densities $q(p_e - p_{Ne})$ and $q(p_c - p_{Nc})$ are treated as analogous to voltages applied to the π-equivalent circuit. The argument given in the previous sections shows that this analogy is valid provided only that $n_{Ne} = n_{Nc}$, i.e. the levels of doping immediately adjacent to the junctions must be equal. The much more sweeping assumption that the doping is everywhere uniform, and that consequently there is no drift field, is not necessary.

The equivalent circuit can be amended to allow for the inequality of n_{Ne} and n_{Nc}, and yet maintain the familiar analogy between excess hole density and voltage, by adding a current generator as shown in Fig. 6. This circuit has been derived[5] directly for a one-dimensional model with the drift field E constant across the base, but it also holds in the general case under examination, without these assumptions. The components must clearly be chosen so that the currents I_E and I_C in the two circuits (Figs. 5 and 6) are equal for all values of ϕ_e, ϕ_c, n_{Ne} and n_{Nc}. A satisfactory solution can be obtained by first equating the currents in the shunt admittances:

$$R'_E = \frac{R_E}{n_{Ne}} \qquad C'_E = n_{Ne}C_E \qquad R'_C = \frac{R_C}{n_{Nc}} \qquad C'_C = n_{Nc}C_C$$

Let
$$I_D = G_e \cdot \frac{\phi_e}{n_{Ne}} - G_c \cdot \frac{\phi_c}{n_{Nc}}$$

The coefficients G_e and G_c, together with R'_D, must be chosen to satisfy the equation

$$\frac{\phi_e - \phi_c}{R_D} = \left(\frac{\phi_e}{n_{Ne}} - \frac{\phi_c}{n_{Nc}}\right) \cdot \frac{1}{R'_D} + G_e \frac{\phi_e}{n_{Ne}} - G_c \frac{\phi_c}{n_{Nc}}$$

i.e.
$$\frac{\phi_e}{R_D} = \frac{\phi_e}{n_{Ne} R'_D} + \frac{G_e \phi_e}{n_{Ne}}$$

and
$$\frac{\phi_c}{R_D} = \frac{\phi_c}{n_{Nc} R'_D} + \frac{G_c \phi_c}{n_{Nc}}$$

Any one of the three conductances G_e, G_c and $1/R'_D$ can be taken arbitrarily equal to zero. For example, if $G_c = 0$

then
$$R'_D = \frac{R_D}{n_{Nc}}$$

and
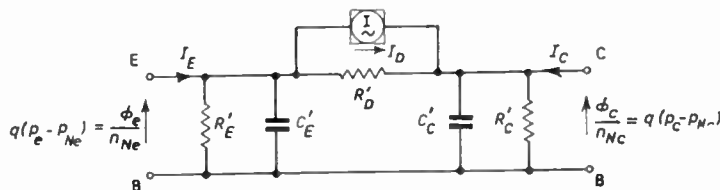$$G_e = \left(\frac{n_{Ne}}{n_{Nc}} - 1\right) \cdot \frac{1}{R'_D}$$

Fig. 6. Equivalent circuit with voltage analogous to excess hole density.

The evaluation of the elements in this circuit requires the value of the ratio n_{Ne}/n_{Nc}, in addition to the values of current gain, etc. This ratio cannot be found by comparing the equivalent circuit directly with measurements because the components in the previous circuit (Fig. 5) represent the greatest number of independent characteristics which can be determined. The ratio n_{Ne}/n_{Nc} can only be found by indirect methods, for example,[13] by modulating the thickness of each depletion layer and thus sampling the doping level adjacent to each junction. Once the ratio is determined for a given transistor, the amended equivalent circuit (Fig. 6) can be used and the excess hole densities adjacent to the junctions can be calculated from it for any circuit application. However, the same result can be achieved by using the previous circuit, calculating ϕ_e and ϕ_c for given conditions, and finally using the ratio n_{Ne}/n_{Nc} to transform these values to excess hole densities.

It therefore appears that there are no advantages in the analogy between potential and excess hole density, and in the more complicated active equivalent circuit based upon it. The electrical characteristics of the transistor can be adequately described in terms of the excess charge product set up at each junction, and the excess hole density is only relevant to the internal mechanism of the device. As the excess hole density varies over the face of each junction, in inverse proportion to the variation in electron density, any value obtained must be some kind of average, according to the measurement technique for n_{Ne} and n_{Nc}.

13. Conclusions

A more general conclusion can be drawn from the fact that the only significant quantities in the equivalent circuit (Fig. 5) are the ratios of the passive components, the physical constant q/kT, and the semiconductor constant $\phi_0 = qn_N p_N$ which is independent of impurity concentration. This conclusion is simply that variations in doping cannot be detected by measurements at d.c. or at low frequencies, where Q_B follows ϕ_e and ϕ_c with negligible lag.

The action of the drift field is of course to reduce the transit time of carriers travelling from emitter to collector and thus to improve the current gain and cut-off frequency in the normal direction at the expense of those in the inverse direction. These effects are reflected in the component values in the equivalent circuit, but they could equally well be obtained, at least in principle, by altering the geometrical arrangement of the diffusion transistor.

At high frequencies, where the equivalent circuit fails due to the distributed nature of the capacitances in the base, a one-dimensional analysis[6] shows that the shape of the frequency response becomes dependent on the drift field, and therefore its magnitude can again be inferred indirectly if the design of the transistor is known to be essentially one-dimensional.

For the analysis of transistor behaviour at low frequencies, it is concluded that the analogy between excess hole density $q(p-p_N)$ and voltage is not particularly useful except in the idealized case where n_N is constant. Some of the equivalent circuits based on this analogy, which have been developed with simplifying assumptions such as one-dimensional geometry, constant field, or even zero field, are valid without these assumptions. In every case, a convenient description is provided by the analogy between potential and excess charge product $q(p-p_N)n_N$.

14. Acknowledgment

15. References

1. J. G. Linvill and J. F. Gibbons, "Transistors and Active Circuits", p. 93 (McGraw-Hill, New York, 1961).

2. Linvill and Gibbons, p. 110.

3. Linvill and Gibbons, p. 14.

4. Linvill and Gibbons, p. 59.

5. Linvill and Gibbons, p. 115.

6. Linvill and Gibbons, p. 118.

7. S. S. Hakim, "Junction Transistor Circuit Analysis", p. 85 (Iliffe, London, 1962).

8. Hakim., p. 70.

9. F. J. Biondi (Editor), "Transistor Technology", Vol. II, p. 380 (Van Nostrand, New York, 1958).

10. Biondi, p. 242.

11. Biondi, p. 412.

12. N. H. Fletcher, "Some aspects of the design of power transistors", *Proc. Inst. Radio Engrs*, **43**, pp. 551–9, May 1955.

13. M. B. Das and A. R. Boothroyd, "Impurity-density distribution in the base region of drift transistors", *I.R.E. Trans. on Electron Devices*, **ED-8**, pp. 475–81, November 1961.

A Low-voltage Glow Discharge Tube

By

R. F. HALL, Ph.D., D.I.C.†

AND

B. J. STOCKER, M.Sc., A.R.C.S.†

Summary: A new low work function surface, Au-Cs-O, is described which can be used as the cathode of a glow discharge tube. Using neon as the gas filling, a maintaining potential of 30–35 V is obtained; with helium the potential is 22–26 V.

The variation of breakdown and maintaining potentials with gas pressure and electrode separation are shown for parallel-plate discharge gaps in neon, using this cathode surface. The gas similarity laws are not obeyed. Ways in which the current-maintaining voltage characteristics differ from those for conventional discharges are also discussed.

Possible applications are considered, especially the use of the Au-Cs-O surface in a trigger tube or surge diverter.

1. Introduction

In 1931 Druyvestyn and Warmoltz[1] obtained a maintaining potential (V_m) of 37 V in a neon glow discharge. The cathode consisted of a layer of caesium oxide deposited on a nickel base, on to which caesium was evaporated, giving a surface with a photoelectric work function of about 1 eV. However, the layer was not stable and V_m rose rapidly with time. Similar destruction of the usual Ag-O-Cs photocell cathode also takes place in the glow discharge, although low initial maintaining voltages may be obtained.

A significant reduction in the photoelectric fatigue of the Ag-O-Cs photocell cathode has been effected by incorporating a uniform distribution of gold particles in the layer, thus increasing the conductivity.[2] Incorporation of gold particles in a caesium oxide matrix, by a method to be described in the next section, enables a low work-function cathode layer with a very high conductivity to be formed. This cathode layer, deposited on a metal base, and used in a diode filled with neon at about 10 torr, may have a very long life under glow discharge conditions. Typical initial maintaining potentials are in the region 30–35 V. With helium, maintaining potentials in the range 22–26 V may initially be obtained, but the life of the cathode is short. The relatively long life for a photosensitive surface, with either gas, is undoubtedly due to the high conductivity of the semiconducting layer, with consequently less electron trapping and scattering by lattice defects.

Properties of the new cathode and tentative explanations of the low maintaining potentials have been discussed in a recent paper.[3] It was concluded that the usual secondary emission processes at the cathode were field enhanced, or augmented by field emission. It is probably no coincidence that this, and other cathode surfaces giving a low V_m, are good electron-electron secondary emitters.

2. Preparation of the Cathode

A thin layer of gold is deposited on a metal base (usually molybdenum) by evaporation. Caesium is then distilled to this layer at 100°C until it is completely transformed into CsAu.[4] Oxidation of this compound results in the formation of caesium oxide(s) and free gold, accompanied by a decrease in work function. Oxidation, also at 100°C, may be carried out while maintaining a glow discharge in neon, until V_m reaches a minimum (corresponding to minimum work function of the cathode surface). At this stage the CsAu is not completely oxidized. The compound formed on the cathode base will be denoted by Au-Cs-O.

Deposition of gold on the metal substrate by plating or other methods has so far resulted in tubes with low initial V_m but with short lives. The reason for this is not known for certain, although it has been found that the structure of the layer and the surface condition of the substrate affect the life of a tube. For example, a porous gold layer (obtained by evaporation in an inert gas at a pressure of a few torr) and a thin tarnish layer on the metal substrate both help to increase the life of the cathode.

3. Design of the Tube

Although tubes produced in these laboratories have not been designed with any specific application in mind, at present there appears to be no reason why the design of a tube with a low-voltage cathode should differ substantially from that of a conventional tube. Several points must be borne in mind however:

(1) The stability of this cathode cannot be expected to be as good as that of a pure metal cathode which has been cleaned by sputtering.

(2) The design and positioning of a trigger must be carefully considered in relation to possible non-uniformity of the cathode surface.

(3) Increase of mean current density at the cathode above about 2 mA/cm² leads to rapid deterioration of

† Mullard Research Laboratories, Redhill, Surrey.

the cathode surface. However, quite high peak current densities at low duty cycles can be tolerated, of the order of tens of mA/cm².

(4) Because of the dependence of breakdown and maintaining potentials on the field at the cathode, the similarity relationships are not strictly obeyed (see next section).

Tubes need not be larger than conventional stabilizer or trigger tubes even though they include a means of evaporating the gold, and a caesium source. Those constructed by the authors in the early development stage[5] used a B9A base, and satisfactory results were obtained with both cylindrical cathodes about 10 cm² in area and flat plates 1 cm square.

Parallel-plate geometry is physically the simplest electrode arrangement in a gas discharge tube, and the characteristics of such a tube with a low-voltage cathode will now be considered.

4. Parallel-plate Diode Characteristics

The electrodes used for these measurements were molybdenum disks 0·5 mm thick and 15 mm in diameter. A layer of Au-Cs-O was formed on the electrode to be used as cathode. The anode was free to slide perpendicularly to its plane, so that the gap could be varied by an external magnet. The design of the tube was similar to that described by Weston.[6] The electrodes could not be sputtered to maintain gas purity, but the Au-Cs-O film itself acts as an efficient getter for oxygen without appreciably altering the properties of the cathode. Evidence for this is observed during the final stages of oxidation of the CsAu—as the minimum value of V_m is reached quite large amounts of oxygen are readily taken up,
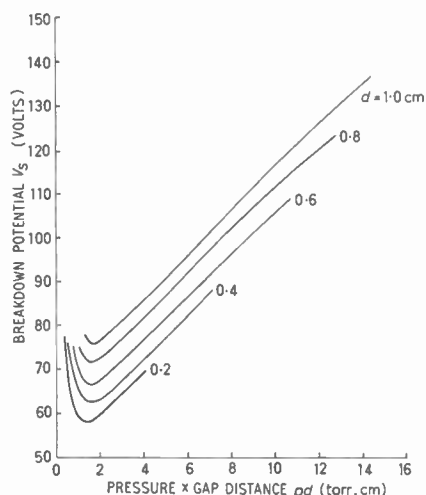
Fig. 1. Paschen curves for pure neon with the Au-Cs-O cathode, for various values of gap distance d.
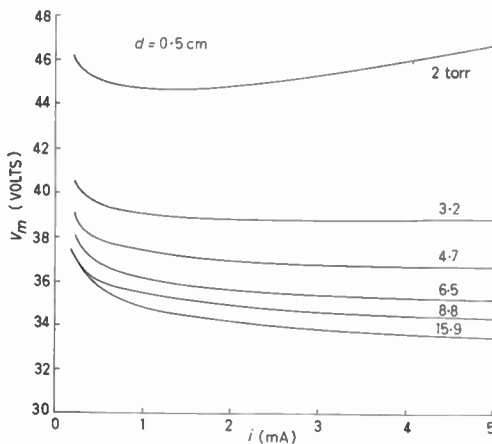
Fig. 2. Variation of maintaining voltage V_m with current i, for various values of pressure p.

although V_m changes very little. Also the inclusion of a getter in a life-tested diode makes very little difference to the initial value of V_m or its rate of rise during life. Since Au-Cs-O is photoelectrically sensitive in the visible region, breakdown statistical delay could be eliminated by means of an external source of light of low intensity. Measurements were made with pure neon and pure helium. Only results obtained for neon will be discussed in detail, since helium is unsuitable for a practical tube. All characteristics shown are for neon.

Paschen curves for the low-voltage diode are shown in Fig. 1. The breakdown potential V_s is plotted against the product of pressure p and gap distance d for various values of d. The marked variation of V_s with d for given pd indicates a possible dependence on the weak field in the gap at breakdown. Otherwise the Paschen minimum for each curve occurs at about the same value of pd as for pure metal cathodes.[6] Values of V_s for helium show a similar dependence on d.

The current–voltage characteristics for various pressures, for $d = 0·5$ cm, are shown in Fig. 2. The characteristic is almost independent of pressure in the range 11–16 torr. For pressures above 5 torr the interesting feature of the curves is the monotonic decrease of V_m with increasing current i. This trend appears to continue into the abnormal glow region if the current is raised rapidly. However, operation at relatively high current densities for any length of time results in a slow increase of V_m due to deterioration of the cathode. As the current is increased in the normal glow region, the area of cathode covered by the glow does not increase in proportion due to 'glow edge effects' which are often observed on a contaminated metal cathode. The current density correspondingly

314

increases even though the cathode is not completely covered. Under these conditions, V_m would be expected to pass through a minimum corresponding to maximum ionization efficiency. However, due to the increase of the field at the Au-Cs-O cathode with increasing current density, the ionization efficiency continues to increase because of enhanced emission from the cathode. This results in a monotonically decreasing V_m. At lower pressures V_m would be expected to pass through a minimum because the field at the cathode is not so intense, the tube voltage occurring over a greater cathode fall distance.
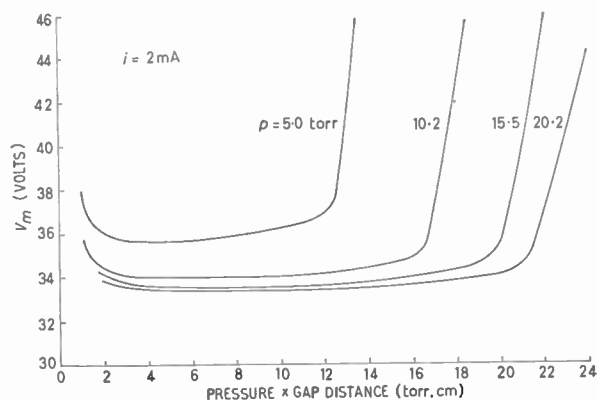
Fig. 3. Maintaining voltage V_m as a function of pd for various values of pressure p.

Curves of V_m against pd for various values of p are shown in Fig. 3. Since the $V_m - i$ characteristics do not show a minimum above 5 torr, an arbitrary current of 2 mA was chosen. Above this value an upward drift of V_m became noticeable due to deterioration of the cathode. Values of V_m are rather lower than those for the corresponding conditions in Fig. 2 since the $V_m - i$ characteristics were measured last of all, and by this time the cathode had deteriorated slightly under the relatively severe experimental conditions. The curve for 15·5 torr may be compared with that given by Weston for a pure metal cathode at 15 torr. The latter curve shows that V_m is substantially constant, at 120 V, between pd values 10–35 torr.cm, and rises only slowly above this to 135 V at $pd = 50$ torr.cm. The flat parts of the curves correspond to the combined length of the negative glow and Faraday dark space regions in the normal glow discharge. The relatively short regions for the low-voltage discharge would be expected, due to the lower energy of the electrons reaching the negative glow from the cathode fall space.[7] The relatively sharp increase in V_m as d is increased is believed to be associated with the formation of a 'ball-of-fire' type of anode glow, rather than a thin sheath of glow covering the anode observed for pure metal electrodes. Deviation from the

similarity laws is noted for both series of characteristics, so that no conclusions can be drawn from this. Deviation from the similarity laws for helium is not significant above about 10 torr.

5. Life of the Tubes

D.c. life tests with many tubes, carried out at current densities of 1 mA/cm², have indicated that V_m rises by a few volts in several thousand hours. The life of the Au-Cs-O cathode rapidly becomes shorter above mean current densities of 1 mA/cm². Deterioration probably takes place because of chemical decomposition of the Au-Cs-O film by ion bombardment and also by the resulting production of heat.

6. Applications

Specific applications must be considered in relation to the possible life of the tube under the given circuit conditions. This is a more important consideration than for conventional glow discharge tubes, since for a long life the tubes must be operated at relatively low mean current densities.

The first application that comes to mind is, of course, as a stabilizer tube. Run at low current densities, a suitably designed diode may be adequate in many circuits as a stabilizer tube, but cannot be expected to match the excellent stability of a tube such as the 85A2 for example. An attempt was made to measure the temperature coefficient of V_m over the range 30–60°C. However, any variation with temperature was masked by the inherent short-term fluctuations which were of the order of tens of millivolts.

The most promising application for the Au-Cs-O cathode at present is in a trigger tube. For example, a tube has been operated with a trigger-cathode breakdown potential of about 70 V, and a main-gap breakdown potential of more than 100 V. This tube could then be operated as a trigger tube from an 85 V line.

Due to its capacity to withstand high peak currents at low duty cycles, the low-voltage tube should find an application as a surge diverter, the peak current that can be tolerated depending on the surface area of the cathode.

At present, use of the low-voltage cathode in other types of gas discharge tubes such as decade counters, or numerical indicator tubes, appears less likely. This is partly due to the short life to be expected at the relatively high current densities which are often required in this type of tube. Any marked improvement in life may make these applications possible.

Finally, some application may make use of the fact that the cathode is sensitive to visible light. A strong source of ambient light can lower the breakdown voltage appreciably, while V_m is unaffected.

Experiments to determine electrical performance data for possible commercial applications have not yet been carried out, apart from the diode characteristics in Section 4. However, de-ionization times have been found to be of the same order as for conventional tubes, and no delay effects have been encountered which would seriously limit the speed of response of a tube in an a.c. circuit.

7. Conclusions

A new material (Au-Cs-O) with low work function and high conductivity has been developed as the cathode of a low-voltage glow discharge tube. Processing of the cathode is similar to that for a photocell.

Specific applications for which the tube may be developed in the near future include trigger tubes and surge diverters. Other applications may require development of a cathode with greater stability and a longer life under conditions of high mean current density. Several other secondary emitting surfaces have been found to give low maintaining voltages when used as the cathode of a glow discharge, a discussion of which lies outside the scope of this paper. An investigation into their properties is proceeding, and there are good prospects of developing a cathode with a longer life.

Another factor which must be taken into account in the design of a tube is that the characteristics do not obey the gas similarity laws. This is particularly true of the breakdown voltage, the minimum value of which varied, for example, from 58 V to 76 V for a parallel-plate diode as the gap was varied between 0·2 and 1·0 cm at constant pd. For the same diode the main-taining potential was about 34 V at 2 mA, and varied only slightly with pressure above about 10 torr. An increase of current still further decreased V_m, but deterioration of the cathode, above about 1 mA/cm^2, resulted in a slow rise of V_m with time.

The life of any low-voltage cathode is very dependent on current density, but is generally satisfactory for the applications suggested above if the mean current is limited to 1 mA/cm^2. Otherwise the low-voltage tube has most of the advantages associated with cold-cathode glow discharge tubes, such as absence of a heater, long shelf life, low temperature coefficient and visual indication.

8. References

1. N. J. Druyvestyn and N. Warmoltz, "Photometry of neon light", Z. Phys., **68**, pp. 378–94, 1931.

2. P. V. Timofeev and Y. I. Lun'kova, "Emission of electrons from oxygen-caesium cathodes containing particles of gold in the intermediate layer", Zh. Tekh. Fiz., **10**, pp. 12–19, 1940.

3. R. F. Hall and B. J. Stocker, "An inert-gas glow discharge with low maintaining potential (22–35 V)", Proceedings of the 6th International Conference on Ionization Phenomena in Gases, Vol. 2, pp. 57–61, 1963.

4. A. H. Sommer, "Alloys of gold with alkali metals", Nature, **152**, p. 215, 1943.

5. R. F. Hall and B. J. Stocker, British Patent No. 937391 (1963).

6. G. F. Weston, "Glow discharge characteristics of helium-neon mixtures", Brit. J. Appl. Phys., **10**, pp. 523–6, 1959.

7. J. R. Acton and J. D. Swift, "Cold Cathode Discharge Tubes", p. 179 (Heywood, London, 1963).

A Note on Multiplicative Receiving Systems for Radar

By

R. BLOMMENDAAL†

Summary: A multiplicative system is examined which consists of a 21-element array with a tapered amplitude distribution and a monopole. Radar applications are considered in the following respects:

(1) The influence of the distance between the phase-centres on beam-width and side-lobe level.

(2) Multiple target response for several relative strengths, phases and angular separations.

The results are evaluated using an analyser for complex Fourier series and compared with similar situations in linear arrays. An estimate is given of the unwanted phenomena to be expected. Further experimental research seems to be justified and the difficulties of such experiments are outlined.

1. Introduction

This study was undertaken to determine the advantages of a multiplicative receiving antenna when used in a radar system. It is well known that non-linear systems can give considerably narrower beam-widths than linear systems of the same size. Therefore it seemed attractive to use a non-linear antenna when very narrow beams are necessary, especially since antenna production costs do not rise proportionally with their size. Many recent publications have treated the combination of an array with uniform illumination and an omnidirectional antenna in a multiplicative system. Not much attention has been given, however, to tapered illuminations and their resulting side-lobe levels and beamwidths when used in a multiplicative array. Because antennae with a tapered amplitude distribution have broader beams than the corresponding uniform illuminations, the multiplication process can be very profitable. Because a 0·5 deg slotted waveguide antenna with a $k + $cosine-squared aperture distribution was available for future experiments, this type of illumination was examined from the point of view of beamwidth and side-lobe level. An analyser for complex Fourier series was available to compute the series:

$$E(\phi) = \sum_{n=-20}^{+20} a_n e^{jn\phi}$$

as a continuous function of ϕ for any set of complex a_n. A 21-element array was chosen for the computations, so that an equivalent antenna of double length would just suit the 41 harmonics of the computer.

† Nederlandsch Radar Proefstation, Noordwijk, Holland.

The main disadvantage of a non-linear array is its distortion when more than one target is present, which is caused by the resulting cross-products. In radar the relative strengths, phases and angular separations of the targets cannot be predicted and any combination of these can be expected. A general expression for the case of two targets is given and used to study two examples in general. This expression is transformed into a series representation that is suitable for use with the Fourier series analyser. No limit was set for the number of patterns that could be examined. For a range of target combinations the responses have been evaluated and a representative collection is presented here.

2. Beamwidth and Side-lobe Level

The system that will be analysed is illustrated in Fig. 1. It consists of two antennae with (field-) patterns $A(\phi)$ and $B(\phi)$, where ϕ stands for:

$$\phi = \frac{2\pi p}{\lambda} \sin \theta \qquad \quad \ldots\ldots(1)$$

The phase centres are at distances a and b from the scanning centre. If the signals from the two antennae are substantially coherent and fed into a suitable multiplier, the output of the multiplier can be shown to be:

$$E(\phi) = A(\phi).B(\phi)\cos\left\{\left(\frac{2\pi a}{\lambda} + \frac{2\pi b}{\lambda}\right)\sin\theta\right\}$$

$$= A(\phi).B(\phi)\cos\left(\frac{a+b}{p}\phi\right) \qquad \ldots\ldots(2)$$
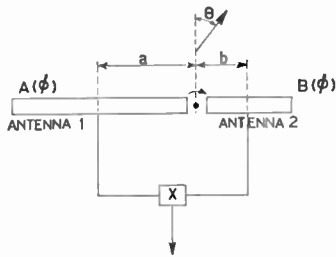
for the case of a single point target.

Fig. 1. Multiplicative system.

When antenna 2 is a 21-element array with element spacing d, the radiation pattern can be given by:

$$B(\phi) = \sum_{n=-10}^{+10} a_n e^{jn\phi}; \qquad \phi = \frac{2\pi d}{\lambda}\sin\theta \quad ......(3)$$

a_n is a complex number that indicates the phase and amplitude of the nth element. With an omnidirectional antenna for antenna 1 ($A(\phi) = 1$), the pattern $B(\phi)$ is seen to be multiplied by a cosine function, the period of which can be adjusted by varying the distance between the phase-centres. The coefficients a_n have been chosen according to a $0.4 + $ cosine-squared law. Results for $a+b = 10d$, $7d$ and $5d$ are shown in Fig. 2.

The original beamwidth in ϕ of antenna 2 is 19.4 deg and is reduced to 8.3, 11.1 and 13.5 deg respectively. In Fig. 2(a) the first side-lobe is seen to be very high (8.8 dB). If the actual multiplication process is performed, for instance, with a phase modulator and a phase-sensitive detector, the system will have a d.c.-output that can be of positive or negative sign. By rectification of the output the first high side-lobe can be suppressed and a very low side-lobe level is the result.[1] The beamwidth is reduced by a factor 0.43. Another solution is to make $a+b$ so small that the highest side-lobe has the same value as in the original pattern (see Fig. 2(c)). In this case the beamwidth reduction still is 0.70 and the side-lobe level remains 27 dB.

3. Multiple Target Response

The main disadvantage of the multiplicative system is the distortion that is caused by cross-products when contributions of more than one target are present. The relative phases and amplitude ratios of the reflections and their angular location will determine the response of the system. Because there is an infinite number of possible combinations the general analysis will be limited to the case of two targets. Even in this simple case the influence of the phase
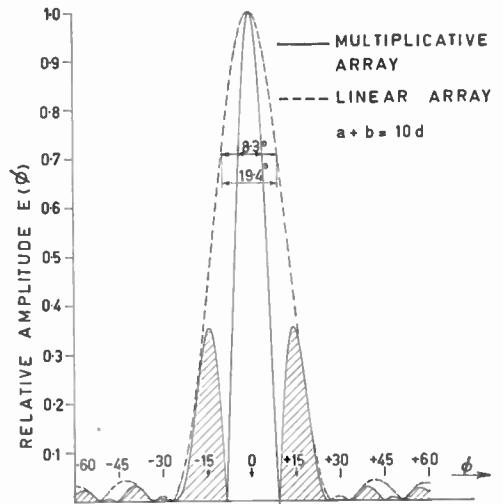
(a)
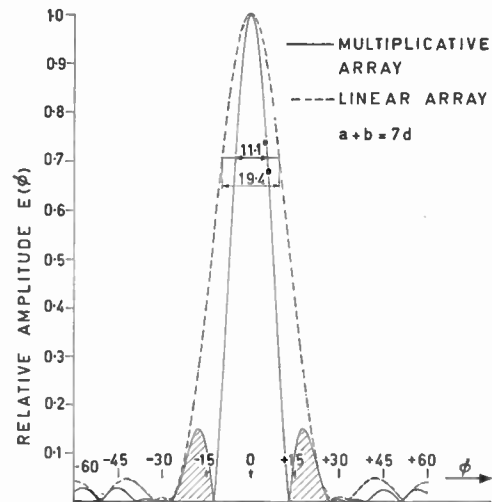
(b)
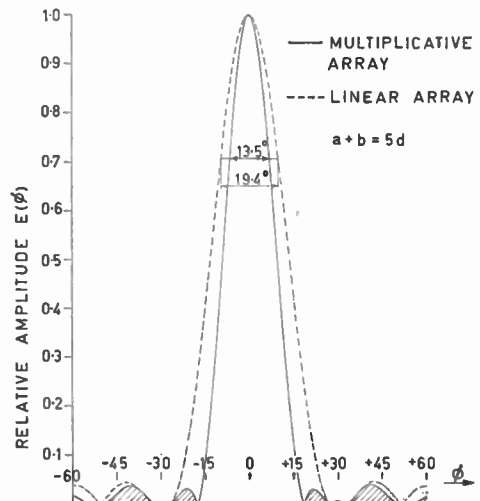
(c)

Fig. 2. Response of multiplicative system with one point target. Responses of reversed polarity are hatched.

makes conclusions difficult. We consider once more the situation of Fig. 1 and assume that two targets are present with amplitudes r_1 and r_2, phases α_1 and α_2 and bearings ϕ_1 and ϕ_2 respectively. The multiplier output will be:

$$E(\phi) = r_1^2 A(\phi-\phi_1).B(\phi-\phi_1)\cos\left\{\frac{a+b}{p}(\phi-\phi_1)\right\} +$$

$$+ r_2^2 A(\phi-\phi_2).B(\phi-\phi_2)\cos\left\{\frac{a+b}{p}(\phi-\phi_2)\right\} +$$

$$+ r_1 r_2 A(\phi-\phi_1).B(\phi-\phi_2)\times$$

$$\times \cos\left\{\frac{a}{p}(\phi-\phi_1)+\frac{b}{p}(\phi-\phi_2)-\alpha_1+\alpha_2\right\} +$$

$$+ r_1 r_2 A(\phi-\phi_2).B(\phi-\phi_1)\times$$

$$\times \cos\left\{\frac{a}{p}(\phi-\phi_2)+\frac{b}{p}(\phi-\phi_1)+\alpha_1-\alpha_2\right\} \quad (4)$$

where $\phi = (2\pi p/\lambda) \sin\theta$.

Example 1

Suppose antenna 1 can be considered to be omnidirectional and the systems scanning centre is coincident with the phase centre of antenna 2. We set $A(\phi) = 1$ and $b = 0$ and find:

$$E(\phi) = r_1 B(\phi-\phi_1)\left[r_1 \cos\left\{\frac{a}{p}(\phi-\phi_1)\right\} +\right.$$

$$+ r_2 \cos\left\{\frac{a}{p}(\phi-\phi_2)+\Delta\right\}\left] +\right.$$

$$+ r_2 B(\phi-\phi_2)\left[r_1 \cos\left\{\frac{a}{p}(\phi-\phi_1)-\Delta\right\} +\right.$$

$$+ r_2 \cos\left\{\frac{a}{p}(\phi-\phi_2)\right\}\left] \quad ......(5)\right.$$

where $\Delta = \alpha_1 - \alpha_2$.

The response in the direction $\phi = \phi_1$ will not only depend on r_1 but also on the term

$$r_2 \cos\left\{\frac{a}{p}(\phi_1-\phi_2)+\Delta\right\}$$

which can be of considerable importance when r_1 and r_2 are of approximately the same magnitude. In the extreme case that $r_1 = r_2$, zero output will result if

$$\frac{a}{p}(\phi_2-\phi_1) = \Delta + 180 \deg \qquad ......(6)$$

This condition depends on both $\phi_2 - \phi_1$ and Δ and shows that the phase difference and the angular separation can more or less be traded.

Example 2

If instead of $b = 0$ we take $a = 0$, then the scanning centre is coincident with the omnidirectional antenna:

$$E(\phi) = r_1 B(\phi-\phi_1)\left[r_1 \cos\left\{\frac{b}{p}(\phi-\phi_1)\right\} +\right.$$

$$+ r_2 \cos\left\{\frac{b}{p}(\phi-\phi_1)-\Delta\right\}\left] +\right.$$

$$+ r_2 B(\phi-\phi_2)\left[r_1 \cos\left\{\frac{b}{p}(\phi-\phi_2)+\Delta\right\} +\right.$$

$$+ r_2 \cos\left\{\frac{b}{p}(\phi-\phi_2)\right\}\left] \quad ......(7)\right.$$

where $\Delta = \alpha_1 - \alpha_2$.

For $r_1 = r_2$ the response is zero if $\Delta = 180$ deg for all separations of the two targets. This must be the case of course because two equal contributions in antiphase are incident on the omnidirectional antenna, which does not move in the field as in the case of example 1.

The extreme situations that are quoted above are only a fraction of all possible combinations of amplitudes, phases and bearings of two point targets. To get an impression of less extreme situations we look once more at the case of example 1. We specify antenna 2 as a 21-element linear array with a 0·22 +cosine-squared aperture distribution and a sidelobe level of 32 dB. We also assume:

$$A(\phi) = 1 \qquad\qquad a = 10d \qquad \phi = \frac{2\pi d}{\lambda}\sin\theta$$

$$B(\phi) = \sum_{n=-10}^{+10} a_n e^{jn\phi} \qquad b = 0 \qquad a_n = a_{-n}$$

where d = element spacing.

For two targets in the directions $\phi = 0$ and $\phi = \psi$ we can write expression (5) as (see Appendix):

$$E(\phi) = \sum_{p=-20}^{+20} b_p e^{jp\phi} \qquad\qquad ......(8)$$

$$b_p = \tfrac{1}{2}a_{p-10}[r_1 + r_2 e^{-j(10\psi-\Delta)}] \times$$

$$\times [r_1 + r_2 e^{-j\{(p-10)\psi+\Delta\}}] \qquad p > 0$$

$$b_{-p} = b_p^*$$

$$b_0 = a_{10}[r_1 + r_2 e^{-j(10\psi-\Delta)}][r_1 + r_2 e^{j(10\psi-\Delta)}] \quad ...(9)$$

From the result $b_{-p} = b_p^*$ it appears that $E(\phi)$ is a real function, as it should be. The number of harmonics has increased from 21 to 41. This is in agreement with the case where antenna 2 is illuminated uniformly and the resulting beamwidth corresponds with a uniformly illuminated antenna of double length.

All coefficients b_p have the common factor:

$$[r_1 + r_2 e^{-j(10\psi-\Delta)}]$$

that is independent of p. The modulus of this factor will affect the coefficients and therefore $E(\phi)$ directly in magnitude. To get a survey of the systems performance when two targets are present a number of

responses have been computed with the Fourier series analyser. The separation of the targets has been kept within the beam of the 21-element array in view of the proposed radar application.

A series of results for $\psi = -20$ deg and $r_1 = r_2 = 1$ is given in Fig. 3. The amplitudes of the diagrams are in correct proportion with the case of Fig. 3(b) which was used as a reference. For comparison the corresponding patterns of a 41-element linear antenna with the same aperture distribution have been added. A special case is $\Delta = 20$ deg, when the output is zero. Both shape and amplitude of the responses change with Δ. It is clear that the output must be

rectified and the responses of reversed sign have been hatched.

In some cases there is some error in the angular location of the targets' responses. Both too great and too small separations occur.

In examining the resolving power of the system one must define a resolution condition. Pedinoff and Ksienski[2] consider two targets resolved when the image clearly indicates the presence of two beams. In Fig. 3(a), (b) and (c) there are clearly two beams, while in Fig. 3(d) this is less evident. If $\Delta = 20$ deg however there is no output and for $\Delta = 80$ deg (not shown) there is only one beam. It should be noted
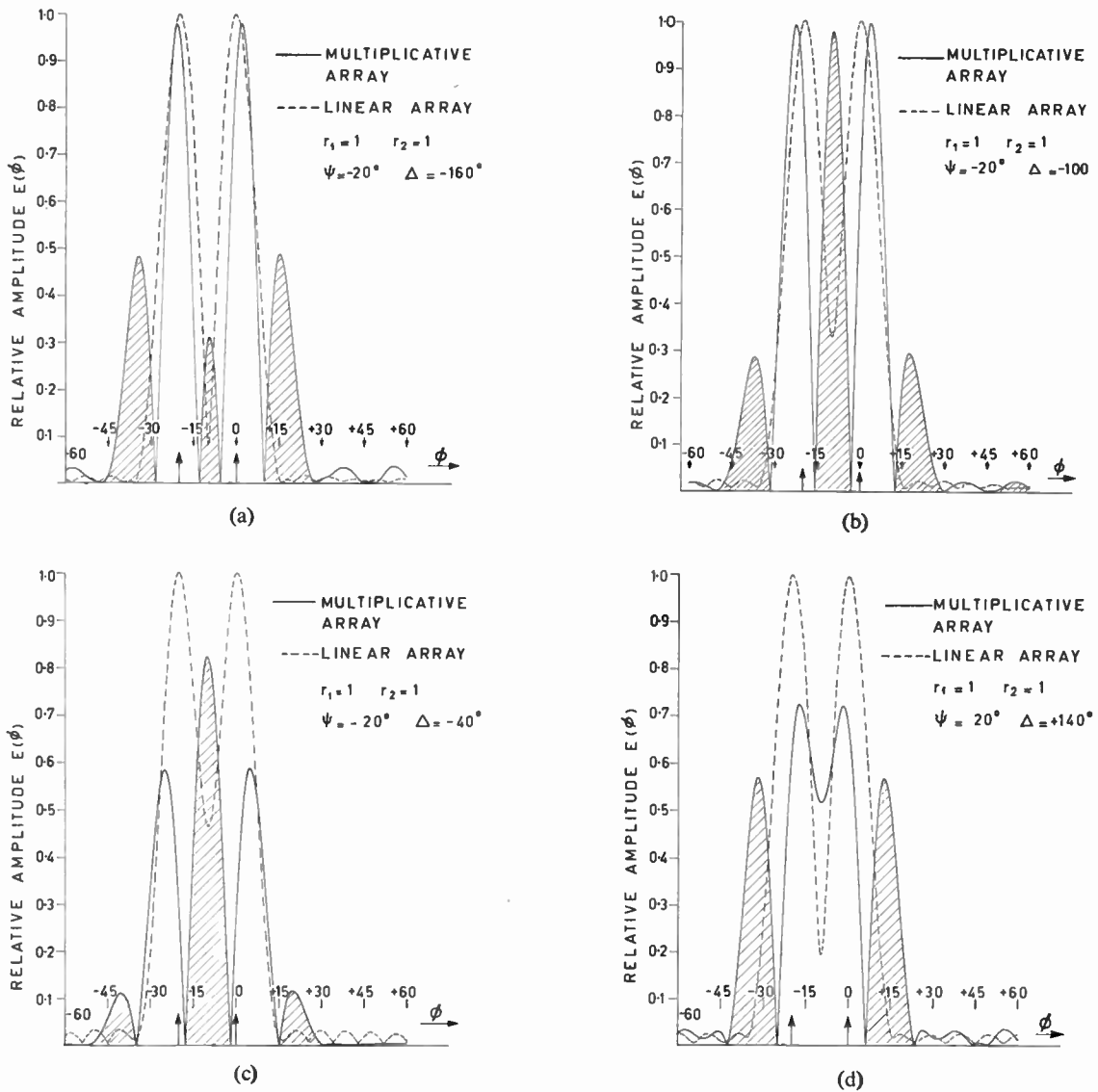
Fig. 3. Response of multiplicative system with two targets of equal strength at separation of 20 deg for several relative phases.
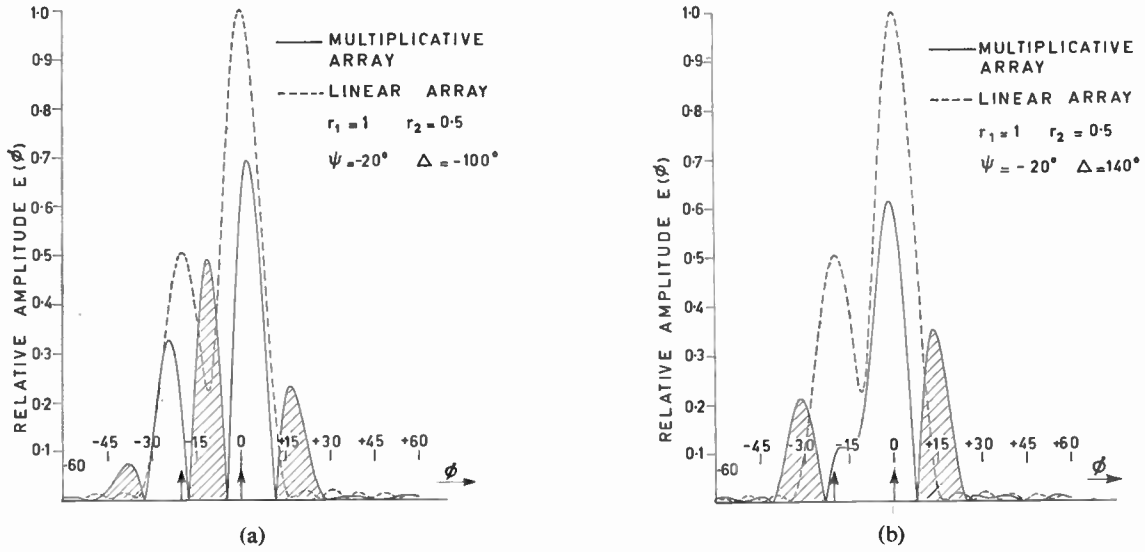
Fig. 4. Response of multiplicative system with two targets of unequal strength.

for a fair comparison that the resolving power of linear systems also depends on the phases of the reflections. For instance, two equal in-phase targets at very small angular separation will not be resolved. If they are in antiphase, however, the response will have a split beam (as in Fig. 3(a)), which resembles the odd pattern of a monopulse antenna. There is a zero between the two beams and the targets will definitely be resolved. In the diagrams of Fig. 3 the influence of Δ in the linear case comes forward in the depth of the 'dip'. In Fig. 3(a), (b) and (c) the multiplicative array has a better resolution because the beams are separated by a beam of reversed polarity.

If the targets have unequal reflections (Fig. 4(a) and (b), 5) the shape of the responses appears to be similar, but the responses of the greater target decrease also.

The range of cases for $\psi = -30$ deg has been computed, but except for the separation the responses are the same if one realizes the equivalence of separation and phase. One interesting result is given for $\psi = -10$ deg in Fig. 6. This is a very extreme situation where the positive responses are very wide apart, and might be taken for a negative response.
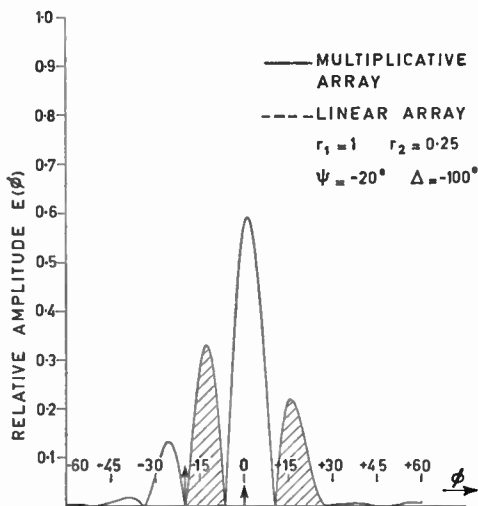
Fig. 5. Response of multiplicative system with two targets of unequal strength.
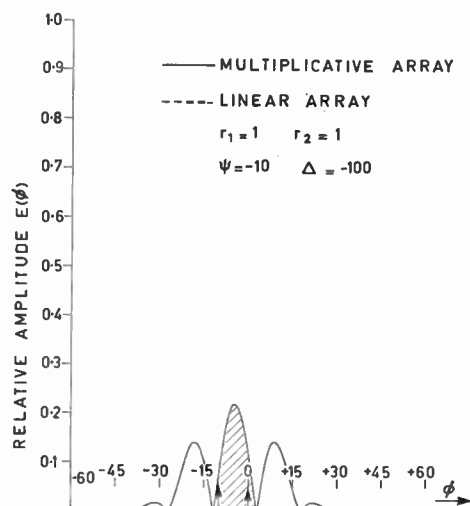
Fig. 6. Extreme case of error in the angular location of the responses.

4. Experimental Results from a Radar with a Multiplicative Receiving Array

The first radar experiments have been made with the combination of two antennae which is shown in Fig. 7. The large slotted waveguide array has the following properties:

length : 4·50 m   frequency : 9240 Mc/s
beamwidth : 0·50 deg   polarization: vertical
side-lobe level : −27 dB

To one of the sides is fitted a smaller antenna which is of the same type, but longitudinal polarization is obtained by means of a polarizing filter. The amplitude distribution is uniform and produces vertical and horizontal beamwidths of 2·9 deg and 25 deg respectively.

Both antennae are connected to a pair of identical mixers by waveguide runs of equal length (to reduce frequency-dependence). A two-channel rotary joint is used and both mixers are symmetrically coupled to the local oscillator. In the low power channel of the
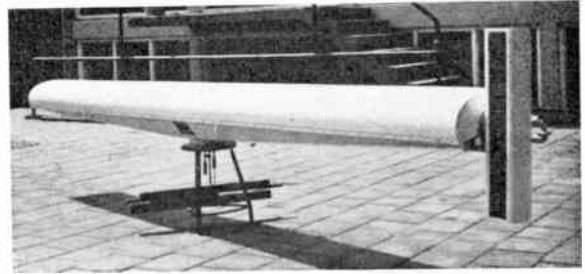
Fig. 7. Slotted waveguide antennae.

smaller antenna a waveguide phase-shifter is added for minor adjustments. An identical set of i.f.-amplifiers is used having a centre frequency of 60 Mc/s and a measured bandwidth of 22 Mc/s. A coaxial switch connects either both outputs to a phase-sensitive detector or the channel of the larger antenna to a conventional detector. Both detector outputs are connected to the video amplifier at the same time. This arrangement makes quick switching possible
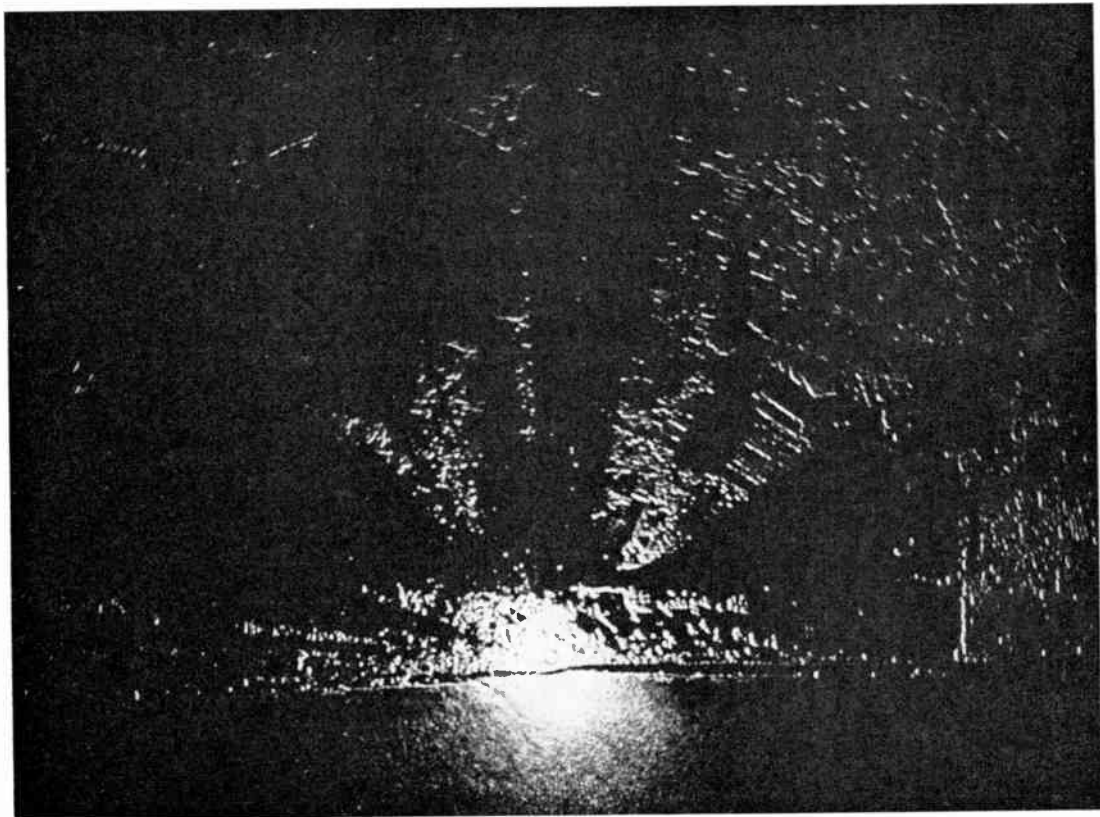
Fig. 8. P.P.I. display of Noordwijk on Sea using a linear system. Scale: 1 cm = 365 m.

between the two receiving systems. Except for the detection circuits the same equipment is used in both cases. The phase-sensitive detector which performs the multiplication process will operate well with pulses as short as 0·1 microseconds.

A normal type of harbour radar transmitter is used which has a peak power output of 3 kW and a pulse-length of 0·11 microseconds. The usual type of t/r-switch is used to couple the transmitter to the wave-guide of the larger antenna.

Beamwidth measurements have been made using the larger antenna for transmitting and both antennae for multiplicative reception. The effective beamwidth is reduced by a factor of two to 0·25 deg in agreement with the theoretical value.

Two photographs of the p.p.i.-display in both systems are shown in Figs. 8 and 9 for comparison. It is clearly seen that the multiplicative system has a better angular resolution. For instance the echoes of the overhead railway line in the top left-hand corner and the block of buildings near the coastline on the

right have gained detail. The entire picture in the multiplicative case gives the impression of consisting of points. Because no 0·25 deg linear antenna was available for direct comparison, it is not sure that the splitting up of all extended targets is correct. Although it is not directly evident from the photo-graphs, there is an indication of reduced sea-clutter in the multiplicative system. The dark region on the photograph, however, is probably due to the narrow vertical beam of the small antenna.

5. Conclusions

In conclusion we can say that no misleading responses are to be expected. The resolving power varies with the phase angle of the targets, but is in most cases better than the resolution of the 41-element array. The 21-element array could be improved when used in a multiplicative system. A very un-desirable behaviour is the zero response for certain conditions. In practice the phase difference between two targets that are not exactly at the same range
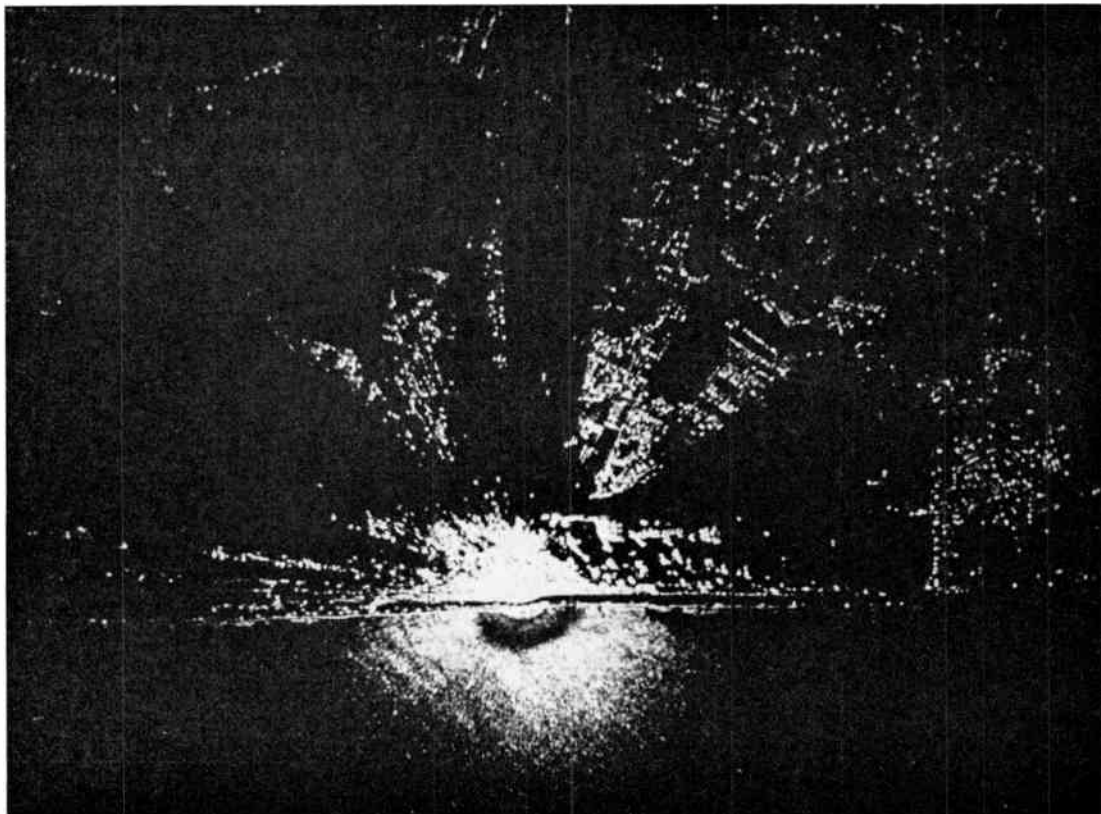
Fig. 9. As Fig. 8 but using a multiplicative system.

can possibly be changed by a small frequency shift. It is not to be expected that moving targets will be masked permanently, because their phase will change rapidly.

6. References

1. V. G. Welsby, "Multiplicative receiving arrays", Proceedings of a Symposium held at Copenhagen, June 1962, Pergamon Press.

2. M. E. Pedinoff and A. Ksienski, "Multiple target response of data-processing antennas", *I.R.E. Trans. on Antennas and Propagation*, AP–10, No. 2, pp. 112–26, March 1962.

3. A. Ksienski, "Signal processing antennas", *Microwave J.*, **4**, No. 10, pp. 77–85, October 1961.

7. Appendix

Let

$$A(\phi) = 1 \qquad a = 10d \qquad \phi = \frac{2\pi d}{\lambda} \sin \theta$$

$$B(\phi) = \sum_{n=-10}^{+10} a_n e^{jn\phi} \qquad b = 0 \qquad p = d$$

Target 1: amplitude r_1 phase α_1 bearing $\phi = 0$

Target 2: amplitude r_2 phase α_2 bearing $\phi = \psi$

Following expression (5):

$$E(\phi) = r_1^2 \cos 10\psi \sum_{n=-10}^{+10} a_n e^{jn\phi} +$$

$$+ r_2^2 \cos 10(\phi - \psi) \sum_{n=-10}^{+10} a_n e^{jn(\phi - \psi)} +$$

$$+ r_1 r_2 \cos(10\psi - \Delta) \sum_{n=-10}^{+10} a_n e^{jn(\phi - \psi)} +$$

$$+ r_1 r_2 \cos(10\phi - 10\psi + \Delta) \sum_{n=-10}^{+10} a_n e^{jn\phi}$$

$$= \tfrac{1}{2} r_1^2 \sum_{p=0}^{+20} a_{p-10}(e^{jp\phi} + e^{-jp\phi}) +$$

$$+ \tfrac{1}{2} r_1 r_2 \sum_{p=0}^{+20} a_{p-10} \{ e^{-j\{(p-10)\psi + \Delta\}} \times e^{jp\phi} +$$

$$+ e^{j\{(p-10)\psi + \Delta\}} \times e^{-jp\phi} \} +$$

$$+ \tfrac{1}{2} r_1 r_2 \sum_{p=0}^{+20} a_{p-10} \{ e^{-j(10\psi - \Delta)} \times e^{jp\phi} +$$

$$+ e^{j(10\psi - \Delta)} \times e^{-jp\phi} \} +$$

$$+ \tfrac{1}{2} r_2^2 \sum_{p=0}^{+20} a_{p-10} \{ e^{jp(\phi - \psi)} + e^{-jp(\phi - \psi)} \}$$

$$= \sum_{p=-20}^{+20} b_p e^{jp\phi}$$

The Discussion on this paper and two associated papers on Multiplicative Receiving Arrays will be published in a forthcoming issue.

A Switch-regulated Power Supply System using the Silicon Controlled Rectifier

By

J. A. CAYZER (*Graduate*)†

Summary: Switch regulation is discussed and a full analysis of the circuit using a silicon controlled rectifier given resulting in formulae for the design of this circuit.

An example is given of a complete circuit including the control amplifier for a 25 V 4 A supply unit. Performance details are included and the possible variations and applications are considered briefly.

1. Introduction

In transistor systems impedance levels are in general low, and it is advisable therefore to place power supply stabilizers as close as possible to the units which will draw their power from them.

In the equipment for which this power supply system was designed (an experimental radar data handling equipment), the circuitry was kept at a constant temperature by means of a liquid cooling system. It was desirable therefore to keep the wattage dissipated by the stabilizers to a minimum in order to avoid local hot spots and waste of coolant power.

Each cabinet contained up to sixty printed circuit cards and its own power supply stabilizers. The bulk supply system to feed these stabilizers was situated alongside the refrigerators which provided the cooling for the circuit cabinets.

The equipment contained both analogue and digital circuitry and required power at eight different potentials. To meet these requirements the bulk power supply had to provide outputs regulated to $\pm\frac{1}{2}$ V under all input and load conditions.

2. Switch Regulation

In any power supply system where a stabilized d.c. output is required from an a.c. supply of varying amplitude, much power may be dissipated by shunt or series regulation; also the size of the smoothing components is inversely proportioned to the frequency of the supply. However, with a switch regulated system the losses and the bulk of the smoothing components can be considerably reduced.

The basis of the system to be described is a variable pulse rate input to a series LC circuit which filters the input pulses of d.c. power, as shown in Fig. 1. It is arranged that the switch makes contact with A until the voltage at Y has returned to its initial value. It then makes contact with B and when current ceases to flow in L the voltage at Y will again be at its original value.

† Formerly with Decca Radar Ltd., now with Texas Instruments Ltd., Manton Lane, Bedford.

The peak-to-peak ripple content (v) of the output will be a function of the load current, the difference between the voltage at A and the output voltage, and the values of L and C. Thus

$$v = \frac{Li_R^2}{2Ce_L}$$

where i_R = load current

e_L = potential across AY.

The time for which the switch is made to A is given by

$$t_1 = \frac{2Li_R}{e_L}$$

It is clear from the above that t_1 is proportional to the load current under these conditions and therefore
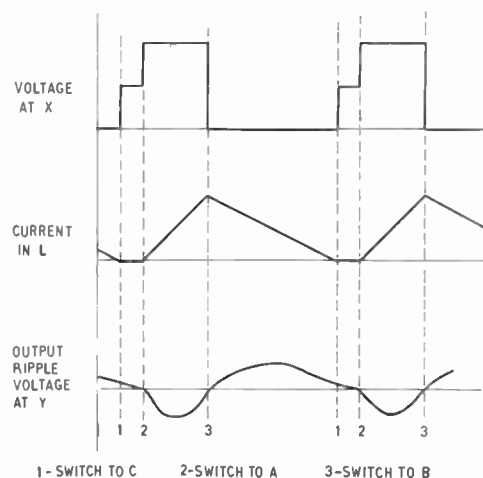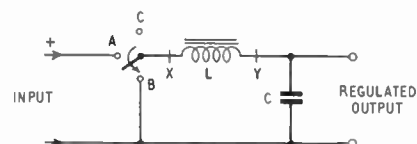

Fig. 1. Basic circuit of switch regulator and associated voltage and current waveforms.

under very light loads the 'on' pulse width of the device would be very short, and considerable losses would occur during switching.

To avoid this unnecessarily high switching speed, a definite pulse width may be used for the input pulse, as determined by some external timing device, and its frequency varied according to load conditions. This results in a higher peak-to-peak ripple value at low currents, namely

$$v = \frac{Li_p^2}{4Ce_L}$$

where i_p is the peak pulse current.

At full load, with e_L equal to the minimum value expected, it can be arranged that the following pulse occurs immediately after the recovery period of the first pulse—this is the condition of peak p.r.f. and the mean output current is exactly half the peak current (triangular current waveform).

This gives the starting point in the design procedure. An inductance L is required which will allow the minimum input voltage to build up a current in it, equal to twice the full load current, in the time selected for the input pulse. If the switch gives switching times of 1 to 2 μs, for example, then an input pulse of 100 μs will give about 3% switching time, hence reasonably small switch losses.

When a pulse width has been chosen the value of L is determined from

$$L = e_{L(min)} \frac{t_1}{i_p}$$

and the value of $i_{p(max)}$ when e_L is high may be found by transposing the same formula.

An inductance of L henrys which must not saturate with a current of $i_{p(max)}$ amperes is therefore required.

If the maximum ripple voltage which can be tolerated is known C can then be found from

$$v = \frac{Li_{p(max)}^2}{4Ce_L}$$

namely $$C = \frac{Li_{p(max)}^2}{4ve_L}$$

This gives a minimum value of C, which in practice will probably have to be doubled, as the currents are generally so high as to make the voltage across the effective series R of the capacitor an appreciable part of the ripple voltage. Series inductance of the capacitor will introduce a step function into the output voltage if a non-inductive winding is not specified.

3. Switched Regulation using Silicon Controlled Rectifiers

Lower ancillary power is required for s.c.r.'s than transistors and their peak current and voltage handling capabilities were superior to equivalently priced transistors at the time the original work was carried out. Turning off the s.c.r. was the main difficulty and a novel circuit has been designed for this.

3.1. *The Switch and its Turn-off Mechanism*

The s.c.r. provides the A contact of the switch as in Section 2 above and MR1 provides the B contact. The s.c.r. is triggered into the conducting state by a blocking oscillator and is turned off by means of the series LC circuit L_F, C_F.

The voltage waveform at X is not the same as in Fig. 1 due to the presence of these 'turn off' elements (L_F, C_F).
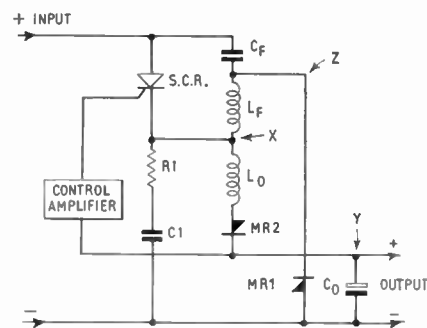

Fig. 2. Simplified circuit of a regulator using a silicon controlled rectifier.

Due to the action of MR1 and MR2 when the circuit (Fig. 2) is at rest, the voltage at X and Z is zero (Fig. 3, A and B). On firing the s.c.r., current builds up rapidly in L_F and charges C_F until Z is at twice the input voltage. The current in L_F will then reverse (Fig. 3, D), and providing L_F is small compared to L_0, the current in it will soon exceed that in L_0. The s.c.r. cannot take this reverse current and switches off as the voltage at X rapidly rises, and the voltage supplied by the capacitor at Z is split between L_F and L_0 according to the ratio of their respective inductances. Actually the s.c.r. conducts in the reverse direction for approximately 1 μs, due to hole storage, and the current in L_F when the s.c.r. actually turns off is considerably in excess of I_0, which in turn can lead to a very high voltage being generated at X. To avoid this voltage pulse the RC filter C1 R1 is connected from this point to earth. Current in L_0 continues to build up, therefore, at an ever-diminishing rate, as C_F discharges and finally ceases to rise, when the voltage across the inductors is zero. The voltage is then reversed and is caught by MR1. The current in the inductors then falls linearly to zero (as the output voltage remains approximately constant).

While the s.c.r. is reverse biased (s.c.r. recovery time, Fig. 3) the holes and electrons in the centre of
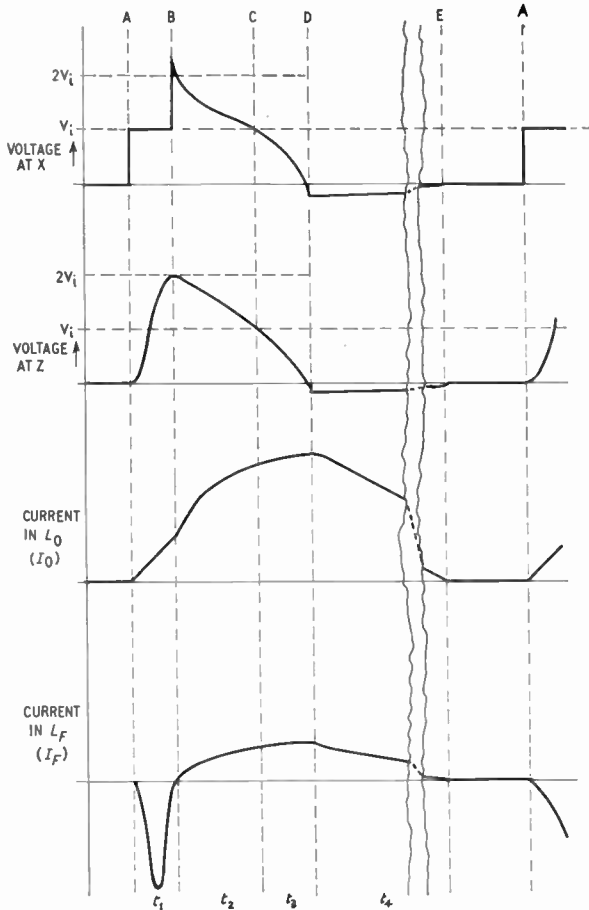
Fig. 3. Waveforms in the circuit of Fig. 2.

Note: The peak positive values of the currents in L_0 and L_F are in fact identical.

the s.c.r. must recombine before the s.c.r. can block forward voltages again. It is important, therefore, that C should be capable of maintaining a reverse voltage at X while discharging through L_0, during the recovery period.

The recovery time for an s.c.r. is typically 10–15 µs and this is one of the major factors determining the input pulse width.

3.2. *Analysis and Design Calculations*

The switching operation discussed above will be seen to consist of four distinct periods, namely

(i) Switch on. C_F charges to twice the line voltage. Current in L_0 builds up linearly with time.

(ii) Switch off. C_F discharges to line voltage. Current in L_0 increases sinusoidally (series circuit conditions C_F, $L_F + L_0$).

(iii) C_F is discharged to zero volts. MR1 catches L_0 and L_F voltage at earth. Current in L_0 reaches maximum value.

(iv) The current in L_0 and L_F reduces to zero linearly at a rate determined by the output voltage.

The time for the first period (i) is determined by the time-constant for L_F, C_F and R_F and is equal to slightly more than the half period time (time for half a period + time for the current to rise to the L_0 value)

being given by

$$t_1 \simeq \frac{\pi}{\sqrt{\dfrac{1}{L_F C_F} - \dfrac{R_F^2}{4L_F^2}}}$$

where R_F is the total effective series resistance of the circuit (capacitors, inductor, s.c.r. and wiring).

The end of the second period is determined by the time at which the voltage across the s.c.r. is zero (and hence across C_F). After this time the s.c.r. has once more to block forward voltage.

This time is a minimum when the output is short-circuited and it is the minimum time which is of interest in order to ensure that the s.c.r. will not still conduct. The equivalent circuit is shown in Fig. 4.
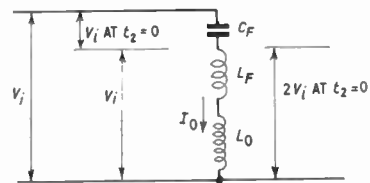
Fig. 4. Equivalent circuit during t_2 and t_3.

Now $L_F \ll L_0$. Therefore $L_0 \simeq L_0 + L_F$. The current in L_0 when $t = t_1$ is given by

$$I_{01} = \frac{V_i}{L_0} t_1 \text{ amperes} \qquad \ldots\ldots(1)$$

assuming that V_i is constant. For the second period we have

$$\left(sL_0 + \frac{1}{sC_F}\right)\bar{I}_0 = \left(\frac{\overline{V}_i}{L_0} t_1\right)L_0 + \frac{2\overline{V}_i}{s}$$

$$\bar{I}_0 = \frac{\overline{V}_i\left(t_1 + \dfrac{2}{s}\right)}{\left(sL_0 + \dfrac{1}{sC_F}\right)}$$

Therefore

$$I_0 = \frac{V_i t_1}{L_0}\cos\omega_2 t + \frac{2V_i}{\omega_2 L_0}\sin\omega_2 t \qquad \ldots\ldots(2)$$

where

$$\omega_2 = \frac{1}{\sqrt{(L_0 C_F)}}$$

or

$$I_0 = \frac{V_i}{L_0} k \sin(\omega_2 t + \alpha) \qquad \ldots\ldots(3)$$

where

$$k = \sqrt{t_1^2 + \frac{4}{\omega_2^2}}$$

and

$$\alpha = \tan^{-1} \frac{t_1 \omega_2}{2}$$

Now the voltage (V_{L_0}) across L_0 is given by

$$\overline{V}_{L_0} = s L_0 \overline{I}_0$$

i.e.

$$V_{L_0} = k V_i \omega_2 \cos(\omega_2 t + \alpha) \qquad \ldots\ldots(4)$$

The end of the second period is determined by the time at which $V_{L_0} = V_i$.

Hence t_2 is given by (from (4))

$$t_2 = \frac{\cos^{-1}\left(\frac{1}{k\omega_2}\right) - \alpha}{\omega_2} \qquad \ldots\ldots(5)$$

The end of the third period is determined by the time at which $V_{L_0} = 0$, i.e. from (4)

$$0 = k V_i \omega_2 \cos(\omega_2 t_3 + \alpha)$$

as

$$k \neq 0, \quad V_i \neq 0 \quad \text{and} \quad \omega_2 \neq 0$$

Therefore

$$\cos(\omega_2 t_3 + \alpha) = 0$$

and

$$t_3 = \frac{\pi/2 - \alpha}{\omega_2} \qquad \ldots\ldots(6)$$
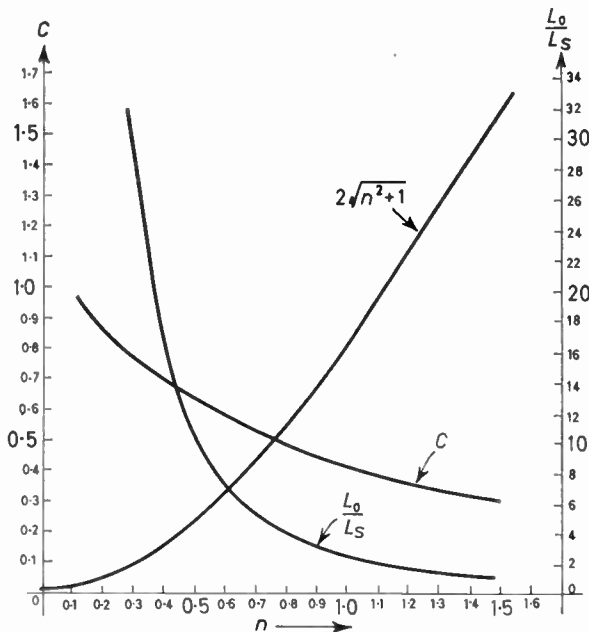
Fig. 5. Graph showing the relations between L_0, L_s, C, and n between $2\sqrt{n^2+1}$ and n. Note: for the value of $2\sqrt{n^2+1}$ add 2 to the value read against the C axis.

From (3) the current in L_0, (I_0) at $t = t_3$, (I_{0t_3}) is given by

$$I_{0t_3} = \frac{V_i}{L_0} k \sin\left(\frac{\omega_2(\pi/2 - \alpha)}{\omega_2} + \alpha\right)$$

Therefore

$$I_{0t_3} = \frac{V_i k}{L_0} \qquad \ldots\ldots(7)$$

Now

$$k = \left(t_1^2 + \frac{4}{\omega_2^2}\right)^{\frac{1}{2}}$$

and

$$\alpha = \tan^{-1} \frac{t_1 \omega_2}{2}$$

Let

$$n = \frac{t_1 \omega_2}{2} \qquad \ldots\ldots(8)$$

Therefore

$$\alpha = \tan^{-1} n \qquad \ldots\ldots(9)$$

and

$$k = \frac{2}{\omega_2}(n^2 + 1)^{\frac{1}{2}} \qquad \ldots\ldots(10)$$

From (5)

$$t_2 = \frac{1}{\omega_2}\left(\cos^{-1} \frac{1}{2\sqrt{n^2+1}} - \tan^{-1} n\right)$$

Therefore

$$t_2 = \sqrt{L_0 C_F}(C)$$

where

$$C = \cos^{-1} \frac{1}{2\sqrt{n^2+1}} - \tan^{-1} n \qquad \ldots\ldots(11)$$

and from (7)

$$I_{0t_3} = \frac{V_i k}{L_0} = \frac{V_i}{L_0}\left(\frac{2\sqrt{n^2+1}}{\omega_2}\right)$$

$$= \frac{V_i}{L_0}(2\sqrt{(n^2+1)(L_0 C_F)})$$

Therefore $I_{0t_3} = 2\sqrt{n^2+1}\sqrt{\frac{C_F}{L_0}}(V_i) \qquad \ldots\ldots(12)$

Now

$$n = \frac{t_1 \omega_2}{2} = \frac{\pi\sqrt{L_s C_F}}{2\sqrt{L_0 C_F}} = \frac{\pi}{2}\sqrt{\frac{L_s}{L_0}} \qquad \ldots\ldots(13)$$

where L_s is defined by

$$\frac{1}{L_s C_F} = \frac{1}{L_F C_F} - \left(\frac{R_F}{2L_F}\right)^2 \qquad \ldots\ldots(14)$$

and R_F is the series resistance of L_F and C_F.

It has been stated above that t_2 is dependent upon the s.c.r., hence $L_0 C_F$ may be determined from (11) and a choice of n. The other parameter which must be known for the design to proceed is the required load current. If it is assumed that the output is short-circuited as for the above analysis, then the current flowing in L_0 will decay at a rate determined by the voltage due to the diodes in the circuit—in the practical case these diodes form a semi-voltage stable load on the circuit—hence the decay will be linear and the current flowing at I_{0t_3} will be twice the mean load current during this period. As this is the longest period it will be approximately twice the mean load current for the whole cycle, i.e.

$$I_{0t_3} = 2I_{sc}$$

where I_{sc} is the load current under short-circuit conditions.

The maximum load current under normal output voltage conditions will be less than this and is modified by a factor $(V_i - V_0)/V_i$, where V_i and V_0 are the input and output voltages respectively.

3.3 Design Procedure

(i) Define maximum load current.

(ii) Define output voltage and ripple content.

(iii) Decide input voltage maximum and minimum (at least double the highest output voltage required).

(iv) Choose an s.c.r. Mean current will be $I_L(V_L/V_i)$, where I_L = maximum load current; V_L = load voltage; V_i = minimum input voltage.

(v) Determine maximum t_2 required for the s.c.r. Hence this is the minimum value permitted and is used for the calculations.

(vi) Find $L_0 + C_F$ from:

$$t_2 = C L_0 C_F \qquad I_{0t_3} = 2I_L \frac{V_i}{V_i - V_0}$$

$$I_{0t_3} = 2\sqrt{n^2 + 1}(V_i)\sqrt{\frac{C_F}{L_0}} \text{ (use minimum } V_i)$$

$C, 2\sqrt{n^2 + 1}$ may be found from the graph for different values of n. (Fig. 5.)

(vii) Determine L_s from the graph.

(viii) Calculate maximum I_{0t_3} for maximum V_i—this gives maximum magnetizing force to be allowed for in L_0.

(ix) Calculate peak I_F for maximum V_i to give maximum magnetizing force to be allowed for in $L_F (I_{F(peak)} = V_{i(max)} C_F/L_s)$.

(x) Check (vi) to (x) for different values of n (values of n from 0·3 to 0·5 are usually best and it is best to use these two values initially).

(xi) Determine C_0 from ripple requirement.

$$C_0 = \frac{L_0 I_p^2}{4 V_i v}$$

where I_p = maximum value of I_{0t_3}

V_i = maximum input voltage

v = peak-to-peak ripple.

(The value of C_0 will almost certainly have to be increased in practice because of its effective series resistance to about two to four times C_0 as found above.)

3.4. Circuit Precautions

If the s.c.r. receives a turn-on pulse during any part of the cycle when a non-zero current is flowing in L_0 other than when it is on or is reverse biased, it is liable to turn on and remain on indefinitely; hence the output would reach the input voltage level unless checked.

On switch-on under overload conditions, with a simple control amplifier operating the s.c.r. by means of a blocking oscillator, the blocking oscillator will free run and may cause the s.c.r. to be triggered at an inconvenient time. Steps must be taken, therefore, to ensure that the cycle must be complete and the current in L_0 is zero before the s.c.r. is triggered again. This gives inherent protection against overloads.

The purpose of MR2 in Fig. 2 is to ensure that current does not 'ring' back into the switching circuit from the output capacitor and so leave an undefined voltage at X, which would alter the turn-off conditions and could again lead to the output being raised to the input potential.

From the above it will be appreciated that it is necessary to protect the output circuits from the possibility of these high voltages.

4. An Example of a Complete Circuit

The circuit of Fig. 6 shows a typical arrangement. For an output of 25 V ± 0·5 V at 4 A, the capacitor C_f would be 1 μF, and the turn-off inductor 11 μH; L_0 should be 330 μH. For reliable operation the turn-off time of the s.c.r. should be less than 20 μs. The circuit will operate satisfactorily on any input voltage between 75 V and 125 V. The theoretical value of the output capacitance required is only 150 μF but at least 500 μF will be required on account of the series resistance.

The peak current in the s.c.r. is around 40 A and this current has to be allowed for in L_f. The peak current to be catered for in L_0 is 15 A.

The output voltage is determined by the difference in the breakdown voltages of the two Zener diodes, MR7 and MR8; VR1 permits fine adjustment. VT3
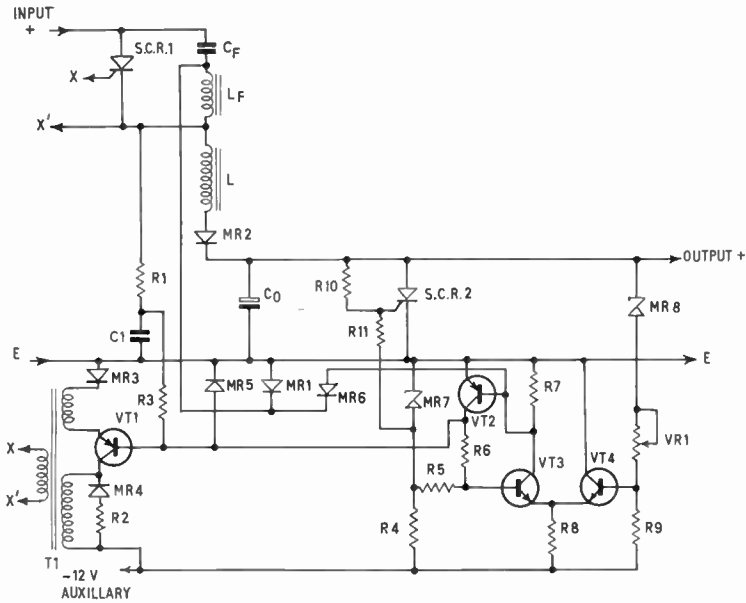
Fig. 6. Complete circuit of a switch regulated power supply using s.c.r.s.

and VT4 together with VT2 form a trigger circuit giving positive information to VT1 as to the state of the output voltage, R3 informs the control circuit when the voltage on C1 is high (i.e. during the first part of the cycle), and MR6 performs a similar function during the second part of the cycle when MR1 is conducting. Thus the blocking oscillator (VT1 and T1), can only fire the s.c.r. at or after a completed cycle. SCR2 provides protection to the load if by any chance SCR1 fails to turn off.

5. Performance

The degree of regulation obtainable with this type of unit depends on the gain of the control circuit and to some extent on the value of C_0. A simple prototype gave the following results:

Input voltage	Load current	Output voltage
95 V	Zero	26·0 V
65 V	4 A (maximum)	25·7 V

The efficiency of the units changes little with variation of input voltage ($\pm 1\%$ for $\pm 20\%$), which is an advantage particularly when considering temperature control or cooling requirements.

6. Conclusion

The principles contained in the design of this power supply could easily be adapted to the design of a number of power conversion schemes, for instance, one may generate a high voltage from a low one by rearranging the circuit. Another interesting point to note is that the system can easily be designed to cope with very widely varied input voltages, hence it could easily operate from a three-phase supply without a reservoir capacitor. The output of the supply is essentially constant current up to the desired voltage and has therefore an ideal motor drive characteristic. The main advantages are, however, to be found in the small size, low weight and high efficiency compared with other systems.

7. Acknowledgment

The author would like to record his thanks to the Directors of Decca Radar Limited, for permission to publish this paper, which was based on work carried out as part of an experimental environment-stabilized data handling and display system.

Phased Array Radar Systems

By

K. F. MOLZ †

Summary: Phased array systems provide a number of features which make them particularly suitable for long range, high target density, radar applications such as satellite tracking and space surveillance. These include long-range detection capability, high transmitted power, inertialess electronic beam steering, multiple receiving-beam capability, high reliability and extreme flexibility.

Basic principles of phased array system operation are presented and a 90-element linear array feasibility model and a large scale u.h.f. planar array are described. Several potential applications of phased array systems for space surveillance are discussed.

1. Introduction

Phased array systems provide a number of features which make them particularly suitable for long range, high target density, radar applications such as satellite tracking and space surveillance. Several examples demonstrating potential operational capabilities and functioning of phased array systems will be presented. These examples will show the increased performance to be derived from the use of phased array systems. Important features include:

(a) High scan and slowing rates achieved by inertialess electronic beam steering.

(b) High transmitted power obtained from large numbers of moderate output transmitters.

(c) Extreme flexibility which provides high traffic capability and multi-pulse operation.

(d) Associated computer functions which permit determination of target trajectories and orbits and maintenance of ephemerides on large numbers of satellites.

Before discussing the performance capabilities of phased array systems and several potential operational space surveillance applications, let us briefly review the history of phased array development and also review the characteristics and fundamental techniques of phased array radars.

2. History of Phased Arrays

In 1958, the author's organization undertook a techniques study and development to prove the feasibility of using electronically steerable array radar (ESAR) in various long-range, high-traffic capacity aerospace sensor systems. The electronic steering

† The Bendix Corporation, Baltimore, Maryland.

feature and the distributed nature of the system promise performance capabilities unobtainable with other types of radar.

Following the initial six-months study period a feasibility model of the linear array for u.h.f. was designed and constructed on the roof of the Company's engineering building at Towson, Maryland. This model consists of one row of 90 radiating elements and associated transmitters, receivers, modulators, programmer, etc.

Operation of the feasibility model substantiated the validity of the basic approach. However, the programme was re-directed by the Advanced Research Projects Agency and the Air Force toward L-band operation and it was decided to construct a full scale ESAR model. The model, which is housed in a five-storey building, has been in operation since November 1960, eleven months after design was started. The radar is shown in Fig. 1. Figure 2 shows a log periodic
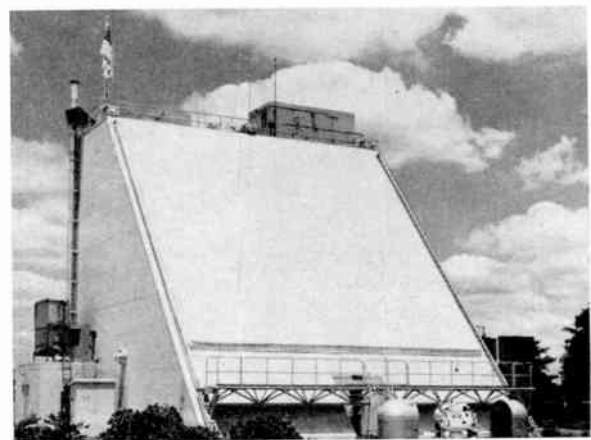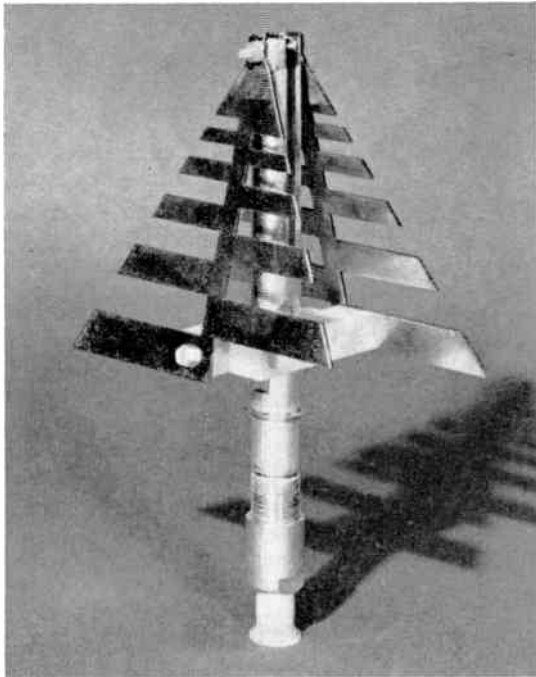

Fig. 1. L-band model.

Fig. 2. Log-periodic antenna.

antenna element and Fig. 3 an antenna panel prior to encapsulation.

This model is a two-dimensional, low-power, L-band, electronically steered, phased array radar. The array consists of 8100 individual radiating elements of which 760 are active. Space tapering techniques are employed to produce a 1·5 deg pencil beam. The system employs phase-phase steering which provides two-dimensional control of transmit and receive beams.

An intensive analytical programme has paralleled the experimental and operational activities on ESAR. A number of mathematical models of the system have been considered in detail in order to derive an adequate testing and evaluation programme for the L-band model and also to determine system parameters for a variety of applications. Studies have included detection theory and prediction theory covering a variety of target models and 'threat' situations, investigation of optimum scan programmes for various system configurations and operational environments, determination of fundamental system limitations, development of operational philosophies, and cost minimization analysis.

In addition to these fundamental analyses, a considerable effort has been devoted to analysis and design of individual components and particular techniques, especially high voltage supplies, modulation

techniques, r.f. power amplifiers, the radiating elements, beam steering and beam forming systems, and space taper techniques.

The significance of these programmes can be summarized as follows:

(a) Design, production, installation, and operation of a full-scale planar array.

(b) Demonstration of two-dimensional phase-phase beam steering techniques.

(c) Development of phased array components at 400 Mc/s and 1200 Mc/s.

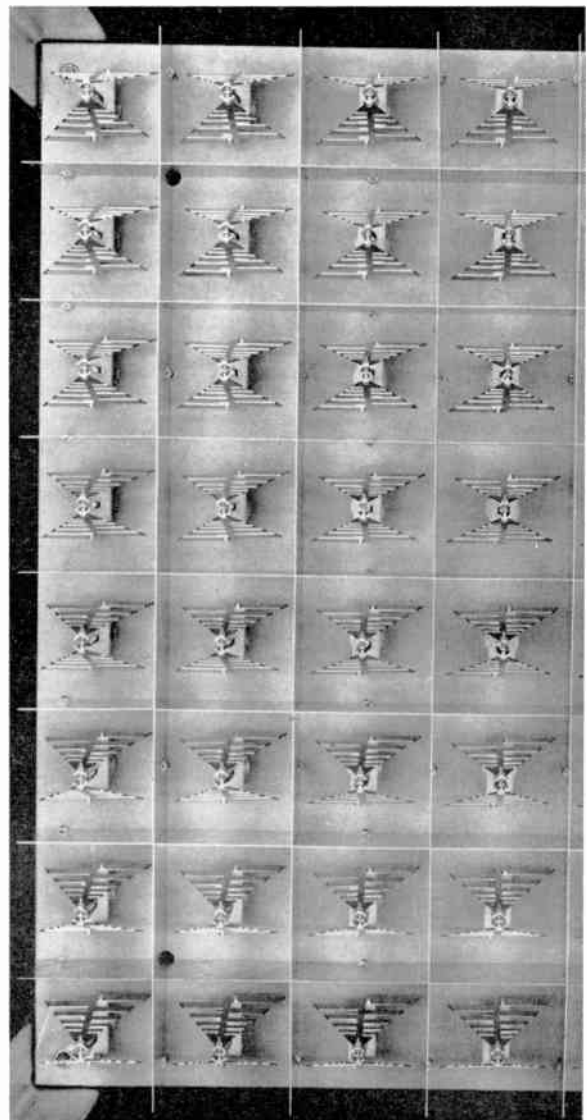(d) Design and development of instrumentation for phased array system support and evaluation.

Fig. 3. Part of antenna panel prior to 'potting'.

(e) Accumulation of reliability data and development of maintenance techniques.

(f) Development of advanced components to permit higher power operation.

(g) Analytical studies to broaden the theoretical background of phased array techniques and to determine operational applications.

(h) Accumulation of accurate cost data for the design, production, installation, and maintenance of phased array components.

(i) Establishment of a strong project team versed in both the technical discipline applicable to phased arrays and the management skill for execution of large scale programmes.

3. Characteristics of Phased Array Systems

A phased array is a set of two or more antennas radiating or receiving electromagnetic energy with a known and controllable relationship between the time phase of the energy in each element. It is this coherence which gives the phased array some of its most valuable properties. Figure 4 illustrates the most elementary array, which consists of two radiators fed from a common signal source and a phase or delay control on each element. As can be seen, the signals will add vectorially in space, and the direction of the maximum will be controlled by the time delay or phase of the signals on each element. The shape of the beam is controlled by the number and spacing of the elements and the relative amplitude and phase of the currents. The strength of the signal can be increased by putting an amplifier ahead of each antenna.

3.1. *Multi-element Planar Array*

The multi-element planar array consists of many thousands of individual antenna elements located on a flat metallic ground-plane with about one-half wavelength spacings between adjacent elements. Although it is also possible to locate the antenna elements of a phased array on other types of surfaces, such as spheres, cones, or cylinders, and while such arrangements are of interest, they will not be considered further.

3.2. *Individual Transmitters and Receivers*

By individual transmitters and receivers we mean that a transmitter and a receiver are connected to a fairly large percentage of the elements, if not to all. This use of many small transmitters makes possible radiated powers that are thousands of times greater than those attainable from conventional antenna-transmitter systems.

3.3. *Electronic Phase Control*

By electronic phase control we mean that the phase of the signal in each channel is controlled elec-

tronically. This electronic control of phase permits changing beam position in microseconds compared with seconds with mechanically scanned antennas.

4. Phased Array Features

Some of the features of phased arrays which make them particularly suitable for the space direction and tracking problem are obvious from the properties discussed above; however, some other features which may not be immediately apparent will now be described.
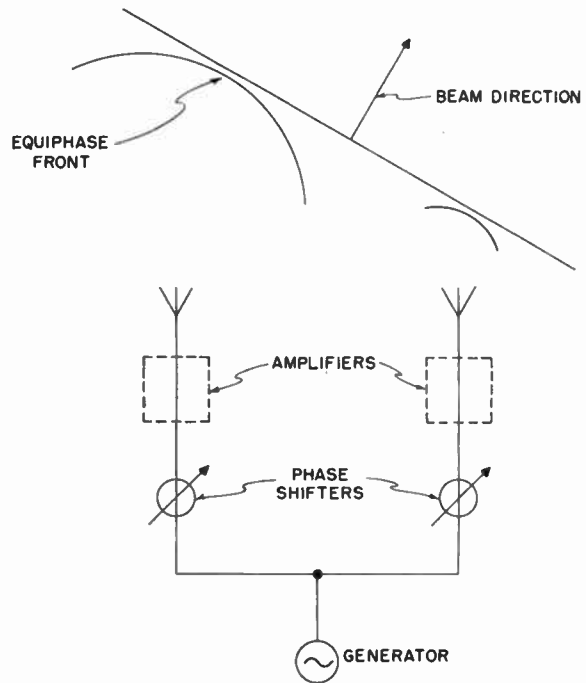

Fig. 4. Elementary phased array.

4.1. *Reliability*

With many thousands of units operating in parallel, the failure of a few channels will not seriously detract from the radar's performance. In addition, the tolerances needed on the phase and amplitude control of the signals in each channel become greater as the number of active elements is increased. With adequate monitoring circuits, defective parts can be located and replaced without switching the set off. Hence, if the channel failure rate is kept reasonably low, the set can operate continuously, as long as prime power is supplied. This assumes that those few units which are common to all channels are multiplexed.

4.2. *Utilization of Beam Mobility*

The beam mobility afforded by the fast electronic scan feature allows the design of a single radar set which can perform a number of functions. With a
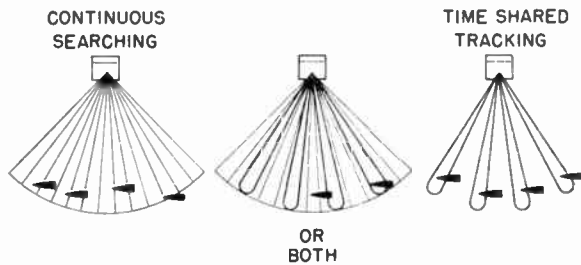
Fig. 5. Time-sharing feature.

given set of installed components, the radar set has the ability to transmit a number of pulses per unit of time depending on the peak and average power limitations of the tubes. These pulses can be utilized in a variety of ways as illustrated by Fig. 5. They can all be transmitted in a sequence of directions for establishing a search pattern, or they can be transmitted in specific directions for tracking targets previously detected, or they can be used in any combination of search and track desired. The way in which the radar can track a number of targets at once is shown in Fig. 6. This is accomplished by transmitting a group of pulses on each pulse repetition interval which might be spaced 20 or 25 microseconds apart. The group of pulses can be transmitted all in one direction for long-range targets or in different directions for short-range targets. They would be transmitted in the direction of the targets previously acquired in a search mode.

The receiving operation can be best understood by again referring to Fig. 6. Immediately after the last pulse is transmitted, the received beam is positioned in the direction of the earliest expected return (target 1). After this return is received, the beam is moved to the position of the next return and remains there until target 2 is received. This action is continued until all target returns are received.
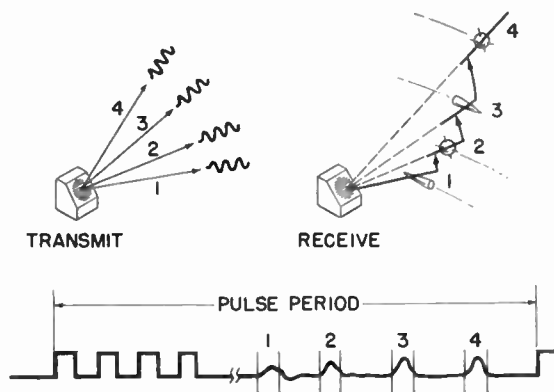
Fig. 6. Multiple-target tracking.

The advantage here is evident; one phased array radar can perform the function of a number of trackers. Further, it may not be necessary to track every target with a data interval equal to the pulse repetition interval; hence, each pulse can be time-shared between 1, 2, 5, 10, or more targets.

In the foregoing example, the beam mobility gives a phased array a time efficiency that a mechanical radar cannot have due to its inertia.

The speed and operating complexity of an array are such that human operators or mechanical controls are inefficient and computers must be used for radar control. This gives an extra degree of flexibility, since a change of operating modes can be accomplished merely by changing programs and no mechanical or circuit changes are required. In fact, by providing sophisticated control circuitry, we can have a truly 'adaptive' radar whose operating modes change automatically as the tactical situation changes.
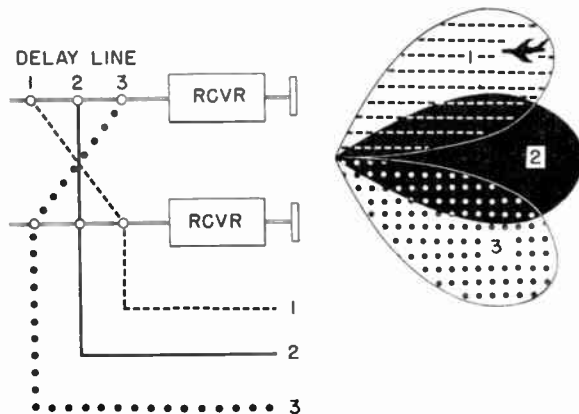

Fig. 7. Multiple-beam formation.

4.3. Multiple Beam-forming Ability

By a technique known as post-amplifier beam-forming, any number of simultaneous receiving beams, each using the full aperture and each pointing in a different direction, can be formed. These beams can be steered as a cluster by the phasing control. A simplified explanation of this operation for the case of two antenna elements is shown in Fig. 7. Note that each receiver feeds a delay line that is tapped as shown. Signals arriving in beam 2 add up in phase only on the number 2 output line. A signal arriving in beam 1 reaches the top antenna a short time before it reaches the lower antenna, and this delay is exactly compensated by the delay line for the beam 1 output.

The post-amplifier beam-forming technique extends the utility of phased arrays—the most obvious is its

use for monopulse tracking. Historically, the first use of this type of operation was in the MUSA system in the early 1930s.

It is also possible to have a system in which receiving beams are formed looking in all directions at once and that does not need to be steered. Only those beam outputs would be used which were directed toward the power being transmitted. A linear receiver array using this principle has been built for an air-traffic control application and is now being tested.

4.4. Coverage

While it is generally considered necessary for a radar set to have hemispherical coverage, there are many applications where this may not be necessary. Planar phased arrays, by their nature, are generally constrained to cover a solid angle of about 90 × 90 deg.

The coverage of a phased array is determined by the antenna element spacing and the antenna element pattern. The limitation imposed by the spacing factor is due to the formation of grating lobes. These grating lobes are spurious beams that appear as the array is scanned away from boresight and are due to the periodic nature of the radiated energy. Consider a linear array which is fed with signals whose phase shift between elements is 180 deg ($\lambda/2$). These signals will add up in two directions in space to form beams pointing in either direction along the axis of the array.

For the spacing generally employed in planar arrays (about $0 \cdot 57\lambda$), it is possible to scan about 45 deg along the principal diagonals and further along any of the others. However, as the spacing is reduced towards $0 \cdot 5\lambda$, it becomes possible to scan more than 80 deg before grating lobes appear along any diagonal when relatively narrow beams are used, although the beam broadens considerably, but predictably, at large scan angles.

The second and more important factor is the coverage which can be obtained from one antenna element when embedded in an array, since the gain of the array at any pointing angle is the product of the array gain and the effective element gain at the angle of interest. The effective element gain, or coverage, is far different from that measured on an isolated element due to the effects of mutual coupling. In the past, one of the more difficult array problems has been the effects of mutual impedance on antenna pattern, but they are now well understood and can be controlled.

This angular coverage can be placed wherever desired by appropriately tilting the plane of the array. To cover a full hemisphere will, in general, require three or more faces. This feature, however, is not the severe limitation that one would expect at first sight, since a search radar uses a fixed amount of average power to cover a given volume of space to detect targets. This power requirement depends primarily on the range and the target size and not on the type of radar used. Hence, three phased arrays would need only one-third of the power per array required by another type of radar that covered the whole area by itself. Studies have shown that the controlling factor in the cost (in high-power, long-range radars) is the average power required and that type of radar (mechanical, phased array, etc.) has much less effect. In fact, there are some instances where it is less costly to use a phased array because of power economy. This is particularly true when very high powers are needed because the voltages required for single-tube power generation are extremely high and they are difficult to generate and control.

4.5. Complexity and Economics

When first considered, the thought of the many thousands of transmitters and receivers which must be built, installed, and checked is, to say the least, horrifying. Further reflection shows, however, that this is an asset instead of a liability. A phased array is essentially a number of identical and relatively simple radar sets operating in parallel. The size of each channel of a typical array approximates to that of a two-way mobile radio and the circuit complexity compares with that of a television set, but like television sets, they can be produced in quantities of several hundred thousand per year in a relatively small production facility and at a reasonable cost. In addition, a uniformity and reliability is obtained which is difficult to match in small-quantity, job-lot operations. This basic simplicity and redundancy is the factor that makes phased array radars practical in many applications.

4.6. Beam-steering Methods

A large number of beam-steering methods have been proposed for arrays and several have been investigated thoroughly. Consequently, the designer has a choice of methods with which to work. The best method for any given array depends on many factors.

A phased array radar forms a beam by energizing many antenna elements in exact phase relationship to each other. If these are all in phase, a planar phase front is created and so a beam is formed perpendicular to the array. When a given incremental phase difference is introduced between adjacent antenna elements, a plane phase-front is created at an angle to the plane of the elements, and a beam is formed perpendicular to the phase-front, but angled with respect to the planar face of the array.
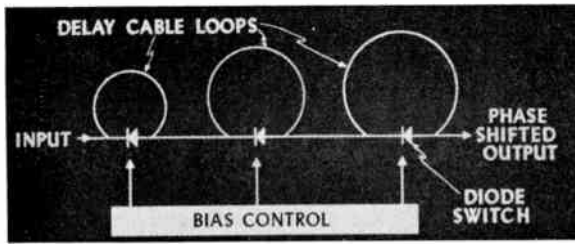
Fig. 8. Switched delay circuit.

Two basic methods which can be employed to generate the incremental phase differences between successive antenna elements and thus steer the beam are real time delay and simulated time delay by controlled phase shifts, introduced by mixers. In the real-time delay or delay-line method (Fig. 8), switching occurs at the signal frequency and requires multi-output computation to determine the required phases. Phase control steering (Fig. 9) is the preferred method for several reasons. Since phase relationships are preserved in the process of mixing, the phase increments can be applied at i.f., and can be applied separately for steering in each of two dimensions and then combined, by mixing, to give the exact phase required by each element. Only one digital input is required for each dimension. Tapped delay lines provide the basis for the precise analogue computation, even though the signals themselves are not delayed.

Once the signals exist at each antenna element, they can be summed in a number of different ways. Radio-frequency beam-forming by directional couplers or hybrids or by post-detection beam-forming networks may be used.

It is possible to create many simultaneous beams on 'receive'. Each uses the full aperture and each can point in a different direction. Signals arriving in each beam give an output only in the corresponding output channel because this is the only channel that
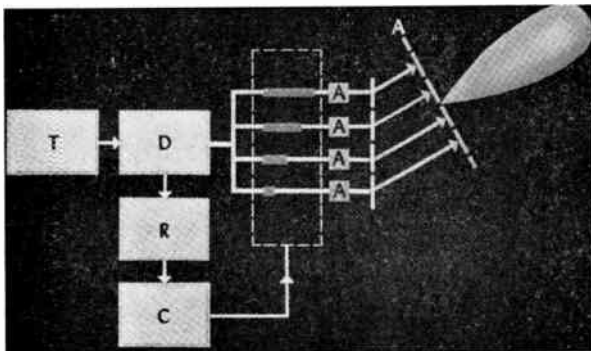

Fig. 9. Steering by variable phase.

permits the signals to add up in phase. Each of these beams can be steered independently, but it is generally more useful to use a network such as this to apply small fixed squints to the beams, and to steer the whole group as a unit in either a monopulse quad, which can give finer angular accuracies and the information needed to track a target, or as a cluster which can be used to re-acquire a target that has been lost.

4.7. *Space Taper*

Side-lobes are a consideration in any radar, and these are controlled in a conventional system by the proper illumination of the reflector so that the radiation from the centre is stronger than from the edges. This pattern is called a taper. In the same way, the radiation from the centre of a phased array should be stronger than from the edges.
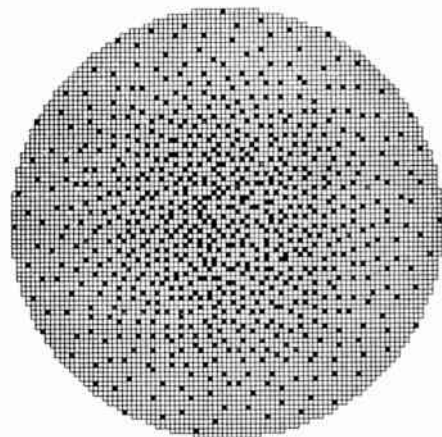

Fig. 10. Space-tapered aperture.

This could be accomplished by using more powerful transmitters in the centre, but the great advantage of a large number of identical modules would then be lost. Instead, all modules are kept identical, but some of the antenna elements are left unused, and the active radiating elements are spaced much closer at the centre than at the edges (Fig. 10). Fewer than half of all of the elements need be active, and by the technique of space tapering as shown, side-lobe levels can be controlled. It is worth noting that the width of the beam depends on the total aperture, and the only loss due to the unused elements is a power loss.

By using only the transmitters in the centre, the beam can be broadened, or if the power of more units is desirable, the phasing can be 'spoiled' in such a way as to broaden the beam.

5. Phased Array Radar Surveillance Systems and Networks

5.1. *Satellite Tracking Array*

Having discussed some of the properties of a phased array, it is now possible to discuss a tentative application. The space detection and tracking problem requires many orders of magnitude more of performance than that needed for detecting and tracking aircraft targets. Detecting targets at 4000 miles rather than at 200 requires an increase in performance of 160 000 times if the targets are the same size, because the detection range of a search radar is proportional to the fourth root of the product of the average power and receiving aperture. To get this increase, we can only enlarge the aperture by a relatively small factor (due to size limitations), and for any further increase in performance, we must raise the transmitted power. This is where the phased
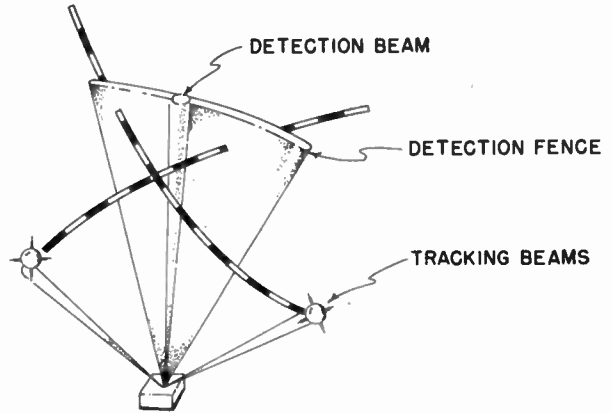
Fig. 12. Operating modes.

array comes into its own, because it can get this power increase easily. Figure 11 shows the kind of peak power which can be achieved at u.h.f. using tubes available today. One can readily see that the power increases needed are available.
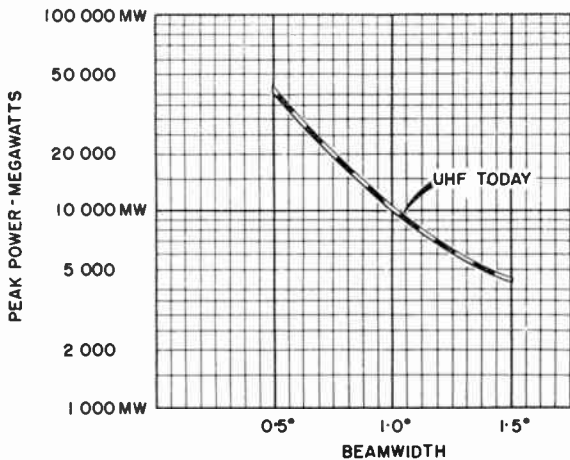
Figure 12 shows how a space tracking radar could perform the dual functions of detecting and tracking. Using the time-shared principle described earlier, a planar array lying on its back would use a portion of the transmitted pulses to form a detection search fence. As satellites cross the fence and are detected, the other pulses would be used to track them. If they are old satellites, the information would be used to update catalogue files. If they are new, as could be rapidly determined, the tracking process would supply enough data to predict their orbit. Using two radar sets and a short-range gap filler with their planes facing up, practically all objects crossing the United States could be detected and tracked (see Fig. 13).

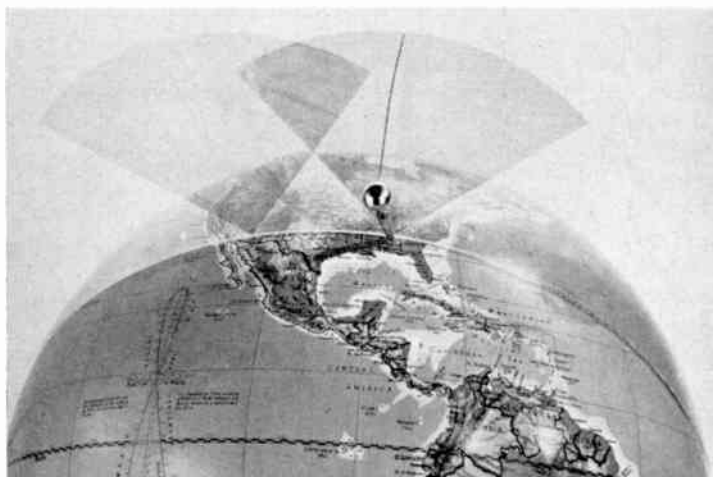Fig. 11. Total peak radiated power.
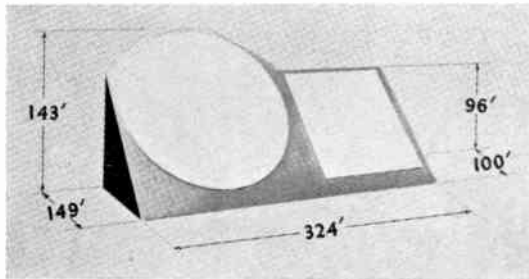
Fig. 13. Radar deployment with gap filler.

Fig. 14. SPADAT radar 'A' system.

One problem that exists in satellite detection is due to the varying altitudes of the targets. The scan period must be short compared with the time it takes a satellite to cross the fence. The following table shows the approximate time it takes a satellite to cross a 1·5 deg beam as a function of the altitude of a vehicle crossing at right angles to the scan direction.

Table 1

Time required for satellite to cross a 1·5 deg beam for several altitudes

Altitude (n.m.)	Time in fan (seconds)
100	0·65
200	1·3
500	3·7
1000	8·2
2000	17·0

If the period of scan is made small enough to get even one look at a satellite as it crosses at an altitude of 100 miles, the scan rate becomes completely impractical. However, this problem can be easily eliminated by making use of the fact that the return signal from a target at 200 miles is 40dB higher than that from a target at 2000 miles. By providing extra receiving beams at right angles to the scan direction or 'fence', the transmitting beam side-lobes can be used to illuminate the low altitude targets. These are about 30dB down, and there is more than enough energy for the purpose. Use of this technique allows the scan period to be set for the object at the highest altitude. A scan period or frame time of 5 seconds to cover 90 deg will give a minimum of three looks per pass on targets 2000 miles distant.

The equipment for such a radar could be housed in a structure about 125 ft cube. With an operating frequency in the u.h.f. band, about 3200 transmitters and receivers would be needed. The average power from each transmitter would be only 30 W. The antenna elements would be located on the roof of the structure in a circular aperture about 115 ft in diameter.

5.2. SPADAT Radar

A brief description of the SPADAT system is presented, since this equipment represents the first phased array space surveillance radar equipment. The system configuration and size are shown in Fig. 14.

The radar is a minimum system capable of expansion to greatly improved performance, representing greater range and altitude coverage, increased number of satellite ephemerides which can be maintained, increased capacity or rate for handling new targets, and improved angular and range resolution. Such a performance can be achieved by using higher-powered transmitters, increased numbers (higher percentage space taper) of transmitters, increased active transmitter antenna aperture, shorter (and higher power) pulse lengths for a given range and use of multiple-faced systems.

The SPADAT system can recognize objects as 'old' (previously detected) satellites, new satellites, or ballistic missiles by processing a series of tracking data points to determine the orbit or trajectory of the particular target. Correspondence of tracking data or detection location and time to the ephemeris data of a previous target identifies it as an 'old' satellite. The phased array radar provides surveillance by maintaining a low angle detection screen such that all satellites or missiles up to the maximum altitude coverage must pass through the screen. The width or thickness of the screen is great enough to ensure a sufficient number of pulse hits on the target as it travels through the screen. If the object is a new target or if more information is required on the target, a tracking beam tracks the target as it continues beyond the surveillance screen.
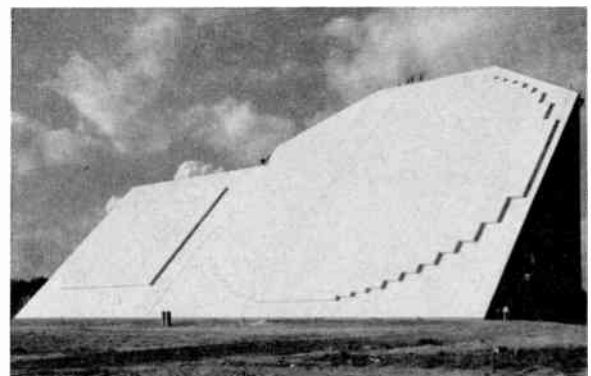

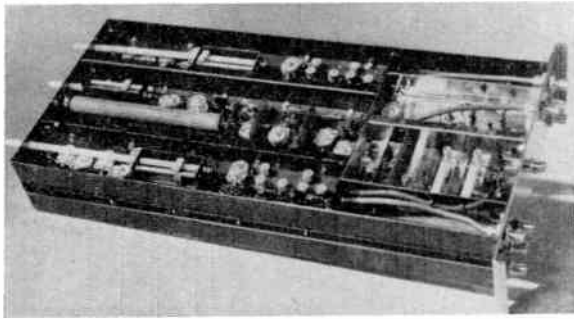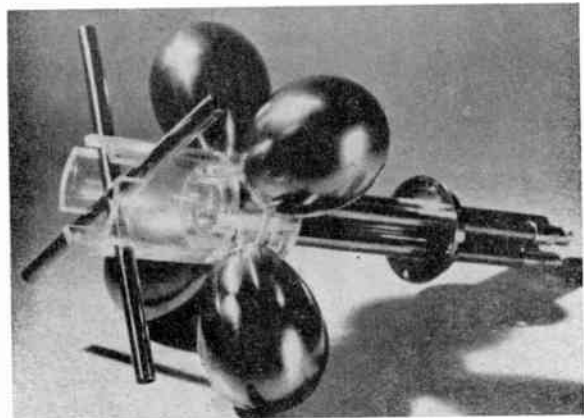Fig. 15. SPADAT system at Eglin, Florida.

Fig. 16. Receiver module.

Fig. 17. Antenna element.

Figure 15 shows the SPADAT structure, while Fig. 16 shows one of the receiver modules and Fig. 17 an antenna element.

6. Conclusion

From the space surveillance systems and networks that have been described, it is evident that phased array radar systems will make a critically important contribution to the space surveillance program and ballistic missile and satellite defence. The SPADAT program now in progress provides the basic design and fabrication techniques and operating experience in the immediate future upon which more comprehensive systems must be based.

DISCUSSION

Under the chairmanship of Mr. W. K. Grimley, O.B.E.

Dr. D. H. Davies: Has the 'mean time between failures' for this equipment been calculated? Accepting that individual element failure does not constitute system failure, what is the likely rate of failures servicing action?

Mr. K. F. Molz (*in reply*): M.t.b.f. has little meaning in a repairable ground system. Availability which is the ratio of operating time divided by the sum of operating and downtime is a much more meaningful term. The availability of the equipment is guaranteed to be 0·7 or better for the whole system. This low number reflects mainly the low availability on the two digital computers which are not redundant. The transmitter and receiver faces, themselves, have better than 0·999 availability.

We expect, roughly, one receiver failure per 10 000 hours of operation (per receiver) and a transmitter failure every 1000 hours (the transmitter output tube is guaranteed for 3500 hours). Some phase readjustment during and beyond these times can be expected but this will not require that the module be removed for repair.

Mr. P. Bradsell: Would the author discuss the receiver protection problem in phased array radars, i.e. protection from damage by the local transmitter?

Mr. Molz (*in reply*): The problem here is similar to conventional radars. In the AN/FPS-85 we have separated the transmit array to obtain the isolation afforded by separate antennas but this is an approach not unknown in conventional radars. In addition, the receiver has a protective diode (solid-state duplexer). Provisions are also made to prevent accidental over-steering of the transmit beam into the receive face.

Dr. J. H. Blythe: How is the effect of mutual impedance between antenna elements taken into account in the design of ESAR? Are screens used between the crossed-dipole elements of the AN/FPS-85?

Mr. Molz (*in reply*): Antenna element design was optimized for low mutual coupling. The element was tested in the presence of neighbouring elements and satisfactory pattern was obtained under those test conditions.

Screens are not used between the AN/FPS-85 elements —rather the pattern is controlled by the director.

Mr. B. W. Osborne: What noise factor is obtained, in view of the use of the diode t/r switches behind the receiving elements?

Mr. Molz (*in reply*): Less than 4 dB on the AN/FPS-85 system.

Dr. K. Milne: When a partially filled aperture is used the peak gain is the same as that obtained when all the

elements are contiguous. The beamwidth is narrower than that of the smaller aperture array with contiguous elements and the excess energy must appear as an increase in side-lobe levels. If the spacings in the partially filled aperture are periodic, the excess energy appears as well-defined grating lobes of high intensity. The effect of random spacings is to smear out the greatest lobes so that they appear at all angles at a fairly low intensity.

Mr. Molz (*in reply*): Yes, this is generally the case.

Mr. J. Croney: Leaving out elements randomly does keep the main beamwidth appropriate to the aperture unchanged but it does lose gain, the lost gain going into smeared out side-lobes. Is it not possible to get the beamwidth required therefore from a smaller aperture using some of the multiplicative techniques we have heard about during this Symposium?

Mr. Molz (*in reply*): In a search and track radar there are many other simultaneous requirements in addition to small beamwidths needed for overall system performance. In particular, presence of multiple targets argues against multiplicative techniques and for space taper. Also the signal/noise loss using multiplicative techniques is over and beyond that due to thinning the receivers.

Mr. A. G. Halliday†: Surely doubling the antenna dimensions and keeping the number of elements fixed will cause some increase in grating lobes?

Mr. Molz (*in reply*): No, this is not true even if the aperture is made very large. The side-lobes depend on the 'granularity' of the space taper pattern and how smoothly the distribution can approach the design illumination function. This depends on the number of active elements. As the aperture size is increased the more sparsely located active elements can be positioned in a more truly random pattern and side-lobes maintained at a low level. This, of course, assumes near optimum space taper design in all cases.

---

† Mr. Halliday died in a boating accident some two months after the Symposium.

Optimum Directional Pattern Synthesis of Circular Arrays

By

G. ZIEHM, Dr. Ing.†

Summary: An attempt is made to deduce rules for designing circular arrays which produce desired directional patterns. The paper considers circular arrays consisting of equally spaced radiators distributed on the whole circumference or on a coherent part of it.

First the general relation between the desired far-field amplitude distribution and the complex input amplitudes of the individual radiators is deduced. The results can be written as a linear system of equations. Its order depends only on the ratio of diameter to wavelength. This form permits a clear and simple numerical treatment of the task by means of a digital computer. Additionally, practical hints are given for the realization of the array feed function. Thereby it is found that a problem of the directional pattern synthesis can be reduced to a synthesis of linear networks.

In a second step the actual optimizing problem is dealt with. It is intended to find, for a given ratio of circular array diameter to wavelength, the design for which the maximum array gain is obtainable with simultaneous maximum attenuation of side-lobes. The result shows that there exists a theoretical limit for the obtainable attenuation of the dominating side-lobe, a limitation which is independent of the ratio of circular array diameter to wavelength. In the investigated case this boundary is at $-26\cdot43$ dB. The array gain increases in a nearly linear manner with the indicated ratio and there seems to be no theoretical limit to any desired increase of the gain.

All considerations are valid for transmission and for reception. The correctness of the treatment has been confirmed by a laboratory experiment.

1. Introduction

In the past two decades, numerous papers have been published on linear arrays.[1-7] One of the most important papers is that by C. L. Dolph[8] in 1946: the amplitude distribution which bears his name gives an optimum shape of the directional pattern. Unfortunately, no comparable papers exist on circular arrays, though these are more useful than linear arrays for many practical applications.

The considerations made in the following sections are valid for acoustical as well as for electromagnetic circular arrays.[9-11] Both applications use this type of source in a great number of designs and for various purposes (Figs. 1, 2 and 3). A schematic figure of a circular array arranged in a spherical co-ordinate system is shown in Fig. 4.

It can be seen that the array comprises a number of identical radiators at equidistant points which either occupy the complete circle or some coherent part of it.

If the individual radiators themselves have a directional pattern their main axes follow circular symmetry in some way—for example, they emit in the direction of the group radius.

Compared with linear arrays circular ones have the advantage that steering of the directional pattern, over any desired range of directions, does not cause any change in the form of the pattern provided that steering is carried out by purely electrical modification of the individual radiators, without any mechanical changes. The reason for this favourable property lies simply in the rotational symmetry of circular arrays.

The electrical modification to the individual radiators is generally made by adjustment of the amplitude and phase of the feeds. This is referred to as 'staggering' or 'weighting' the feed distribution. Most frequently used is simple phase staggering by means of delay lines with which—because of retardation—existing phase differences between the individual radiators are compensated. In such cases therefore one speaks of so-called 'compensated' arrays. In general,
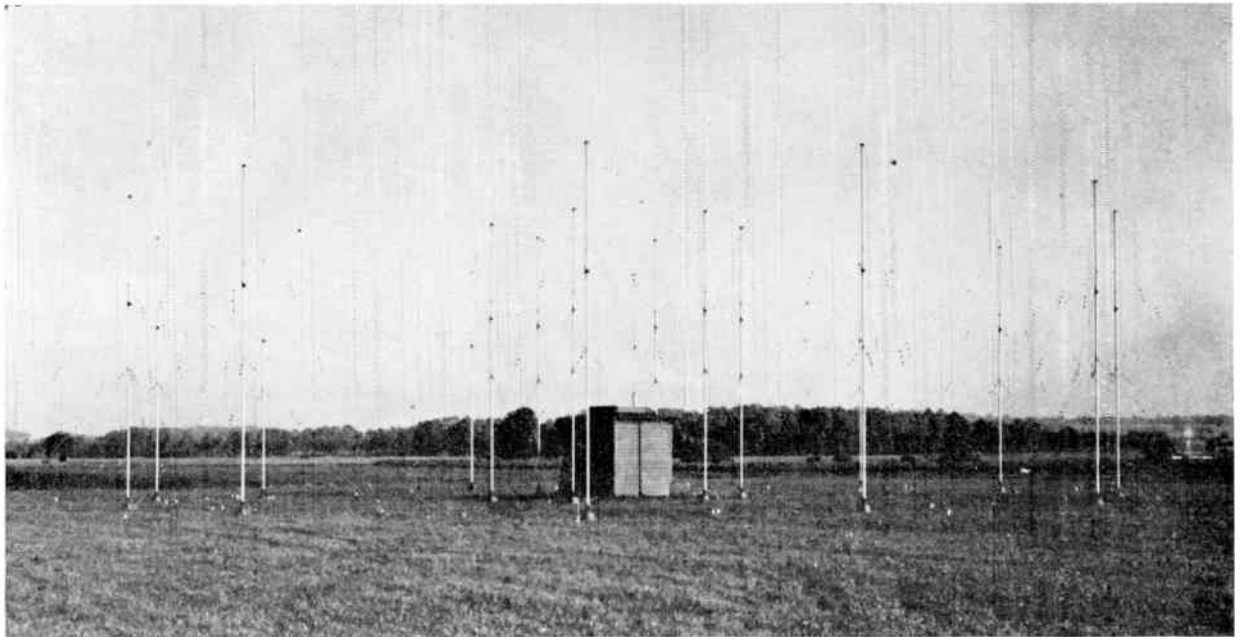
---

† Atlas-Werke A.G., Bremen.

Fig. 1. Two concentric circular antenna arrays for short-wave radio-frequencies. Operating range 1·5–30 Mc/s.
(*By courtesy of Telefunken A.G.*)

the word 'distribution' will be used in this paper for electrical modification of the radiators. The distribution represents the functional dependency of the complex amplitude of each individual radiator on its location within the group. By a suitable choice of the distribution function, any desired directional pattern can be produced.

If the adjustment of the distribution is made exclusively by linear passive means, the theorem of reciprocity is valid. Then it does not matter whether any array is used for transmission or reception. In the latter case it is rather illogical to talk of 'radiators'. Because of its shortness, however, the term is used here even when receiving systems are being discussed; also the word 'radiator' has the advantage that it is in the same way familiar to sonar and microwave engineers.
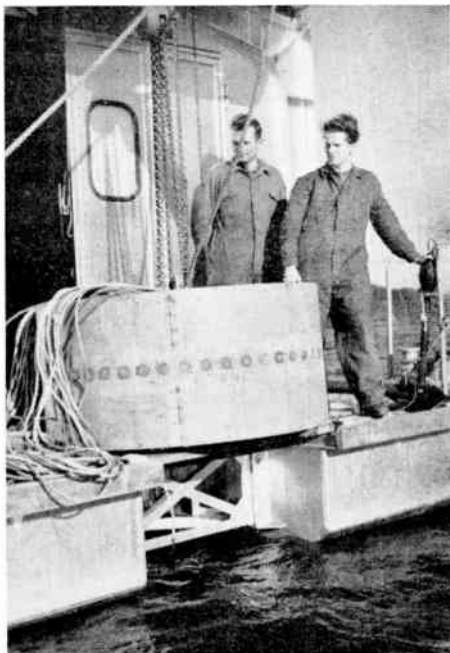
(a) Front view.
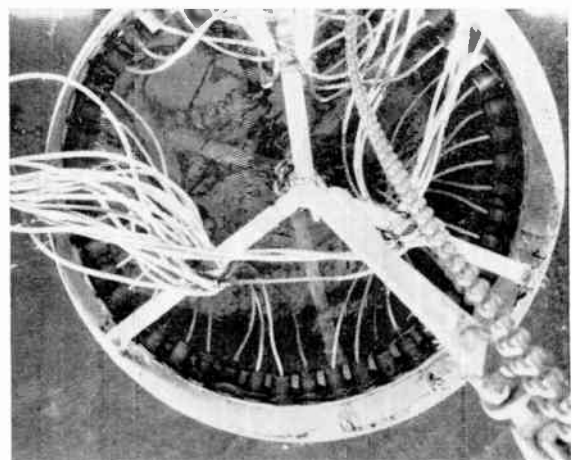
(b) top view.

Fig. 2. A circular array for sonar purposes (development stage). Operating range 30 c/s–30 kc/s.

Fig. 3. A circular array for practical purposes. The photograph shows the hull of the German submarine U–505, captured by the U.S. Navy in World War II. The submarine is now exhibited in the Museum of Science and Industry at Chicago (Illinois). The array was developed by the Atlas-Werke A.G., Bremen.

2. The Directional Pattern of Circular Arrays

In order to denote individual radiators of a circular array according to Fig. 4 the ordinal number v is introduced. If the complex input amplitude of the vth radiator has the value A_v, it produces an amplitude contribution B_v in the far field. This far field amplitude may denote the pressure, particle velocity, electric or magnetic field strength, according to the kind of radiator. If the distance r_v between the far field point F and the individual radiator is very large in comparison with its dimensions and if, furthermore, the medium in which the propagation of waves takes place is unlimited, isotropic and homogeneous, then B_v can be indicated as the product of three factors:

$$B_v = A_v \frac{\exp[-jkr_v]}{r_v} K(\theta_v, \phi_v) \qquad \ldots\ldots(1)$$

The second factor represents the simplest solution of the wave equation in spherical co-ordinates; k is the wave number. The origin of the spherical system (r_v, θ_v, ϕ_v) coincides with that point at the source which is regarded as the reference point. It is designated as phase centre, centre of gravity, origin or a similar term.

The third factor K depends only on θ_v and ϕ_v. It represents the so-called intrinsic directional function of the vth radiator. Generally, K is complex. It is normalized so that $K = 1$ in the direction of the main beam.

In order to include the effects of all individual radiators, the sum must be taken of all B_v. For this purpose it is practical to choose as the common

reference point of all radiators a point on the centre line perpendicular through the plane of the circular array (Fig. 4). The far field contribution of the vth radiator for this reference point M is:

$$\begin{aligned} B_v = A_v \frac{\exp[-jkr_0]}{r_0} \exp\{jkR[\sin\theta_0\sin\Theta_v \\ \times \cos(\phi_0-\Phi_v) + \cos\theta_0\cos\Theta_v]\} \times \\ \times K[(\theta_0;\phi_0);(\Theta_v;\Phi_v);(\alpha_x;\alpha_y;\alpha_z);R] \ldots\ldots(2) \end{aligned}$$

The transition from eqn. (1) to eqn. (2) is obtained by transformation of co-ordinates which although elementary in form, can only be realized in a complicated manner.

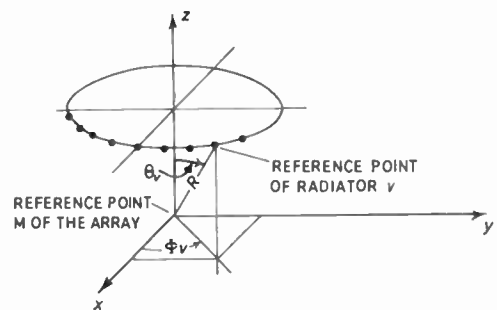$(R;\Theta_v;\Phi_v)$ = co-ordinates from reference point of vth radiator in system x, y, z (Fig. 4)

$(r_0;\theta_0;\phi_0)$ = co-ordinates of far field point F in system x, y, z (Fig. 4)
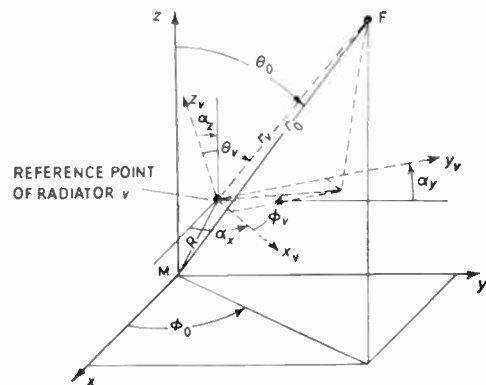
r_0 = distance

θ_0 = angle of incidence

ϕ_0 = azimuth angle

$(\alpha_x;\alpha_y;\alpha_z)$ = angles between corresponding axes of systems (x, y, z) and (x_v, y_v, z_v) (Fig. 4).

(a) Circular array with co-ordinate system $(x; y; z)$.

(b) Reference point of the vth radiator with its own system $(x_v; y_v; z_v)$ in system $(x; y; z)$.
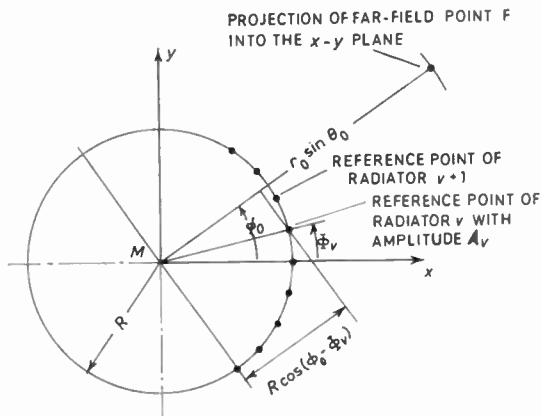
Fig. 4.

Fig. 5. Coherent part of a circular array with a finite number of radiators in the x-y plane.
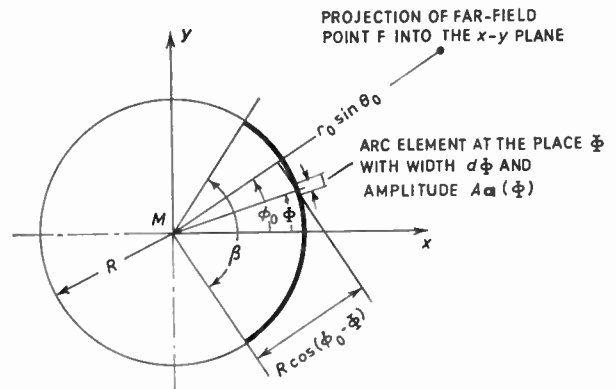

Fig. 6. Coherent part of a circular array with continuous distribution of radiators in the x-y plane.

The directional function K_G of the circular array then follows from:

$$K_G = \sum_v A_v \exp\{jkR[\sin\theta_0 \sin\Theta_v \cos(\phi_0 - \Phi_v) + \cos\theta_0 \cos\Theta_v]\} K[(\theta_0; \phi_0); \times (\Theta_v; \Phi_v); (\alpha_x; \alpha_y; \alpha_z); R] \quad \ldots\ldots(3)$$

Normalization of K_G is only possible if certain assumptions are made about A_v and K. As, however, it is just this complex amplitude A_v which is to be found, information must first be available about the normalized directional function K and about the individual radiators.

It is assumed:

(1) The reference points of the individual radiators, according to Fig. 4, are situated in the x–y-plane, i.e. $\Theta_v = 90°$. (Fig. 5 gives a picture of this special case.)

(2) The z_v-axis of all radiators is parallel to the z-axis of the group, i.e. $\alpha_z = 0°$; $\alpha_x = \alpha_y$.

As

$$\left.\begin{array}{c} r_v \\ r_0 \end{array}\right\} \gg R,$$

hence

$$\theta_v = \theta_0$$

(3) The directional function of all radiators is independent of the co-ordinate ϕ_v. In connection with assumption (2) follows:

$$K = K(\theta_0)$$

As

$$\left.\begin{array}{c} r_v \\ r_0 \end{array}\right\} \gg R,$$

hence

$$\phi_v = \phi_0$$

Equation (3) can then be simplified as follows:

$$K_G = K(\theta_0)\sum_v A_v \exp[jkR\sin\theta_0 \cos(\phi_0 - \Phi_v)] \quad \ldots\ldots(4)$$

This relation is generally known. The assumptions (1) to (3) which it contains are, however, not absolutely necessary for the synthesis procedure described in Section 3. It could also be applied—with suitable modifications—to eqn. (3) directly, except that the calculation would then be far more difficult to follow.

With the 'compensated' circular arrays mentioned above, the phase of A_v, by using delay lines, is chosen so that for a required direction the function K_G is proportional to the arithmetical sum of all A_v. The advantage of this method lies in the fact that the 'compensation' is independent of frequency over a relatively wide band.

For better understanding of the synthesis procedure and to simplify the resulting calculations we shall suppose that the distance between the radiators is made smaller and smaller and the number of radiators larger and larger.

Figure 6 shows the process which is denoted mathematically by certain changes.

(1) $\Phi_v \to \Phi$ represents a continuously-changing angle, as opposed to Φ_v which only had discrete values. As reference direction for Φ it is convenient to choose an axis of symmetry of the circular array (Fig. 6).

(2) $A_v \to dA(\Phi) \cdot dA(\Phi)$ is the complex input amplitude of a sector element of width $d\Phi$ occupied by radiators. $dA(\Phi)$ can therefore be understood as the product of the dimensionless complex distribution function $a(\Phi)$ and the element of arc $d\Phi$. For reasons of dimensions, the third factor A is added which, however, is unessential for the calculation of the directional function.

$$dA(\Phi) = Aa(\Phi)d\Phi$$

For the non-normalized directional function of a circular array continuously occupied by sources, there finally results:

$$K_G = K(\theta_0) \int_{-\beta/2}^{+\beta/2} a(\Phi) \exp[jkR\sin\theta_0 \cos(\phi_0 - \Phi)]\,d\Phi$$

$$\cdots\cdots(5)$$

This form of representation is not yet sufficient for further considerations. As is usual in the theoretical treatment of radiation fields, we require a representation of the directional function in terms of different types of waves or 'spatial harmonics' as these are called.

The Jakobi-Anger relation offers the possibility of expressing eqn. (5) in spatial harmonics.[12-16]

If now we consider, as an example, only the special case of an even distribution function, namely:

$$a(\Phi) = a(-\Phi),$$

eqn. (5) can be brought into the following form:

$$K_G = K(\theta_0)2\pi \sum_{m=0}^{\infty} \left\{ \begin{matrix} j^m J_m(kR\sin\theta_0) \\ \left[\frac{\varepsilon_m}{\pi} \int_0^{\beta/2} a(\Phi)\cos m\Phi\,d\Phi \right] \end{matrix} \right\} \cos m\phi_0$$

$$\cdots\cdots(6)$$

$$\varepsilon_m = \begin{cases} 1 \text{ when } m = 0 \\ 2 \text{ when } m \neq 0 \end{cases}$$

$J_m(kR\sin\theta_0) =$ Bessel function of order m and argument $kR\sin\theta_0$.

The introduction of π is to give the term in square brackets an obvious meaning.

3. The Synthesis of Directional Functions as Desired

Equation (6) represents K_G as a function of ϕ_0 in the form of a Fourier series. Because of choosing an even distribution function $a(\Phi)$, K_G also has become an even function of the type:

$$K_G(\theta_0; \phi_0) = \sum_{m=0}^{\infty} H_m(\beta; k; R; \theta_0) \cos m\phi_0 \quad \cdots\cdots(7)$$

The complex amplitude H_m of the mth spatial harmonic is in detail:

$$H_m = 2\pi K(\theta_0)j^m J_m(kR\sin\theta_0) \times$$

$$\times \left[\frac{\varepsilon_m}{\pi} \int_0^{\beta/2} a(\Phi)\cos m\Phi\,d\Phi \right] \quad \cdots\cdots(8)$$

By suitable selection of $a(\Phi)$ there is now the possibility of bringing the amplitude H_m to a nominal value when certain values β, k, R, θ_0 and $K(\theta_0)$ are given.

This nominal value results from the shape of the directional function required. If, for example, a directional function is to be produced according to Fig. 7, the amplitude and phase of H_m will have to correspond to the spectral representation shown in Fig. 8(a). It can be seen that the phase of H_m can only
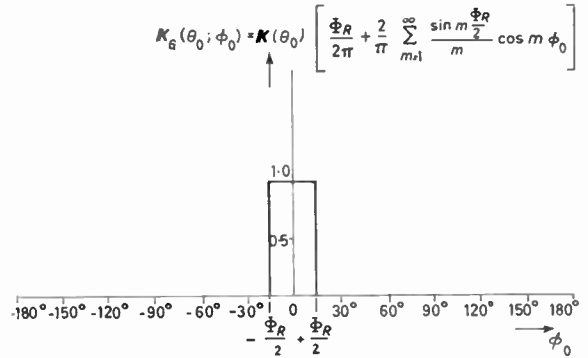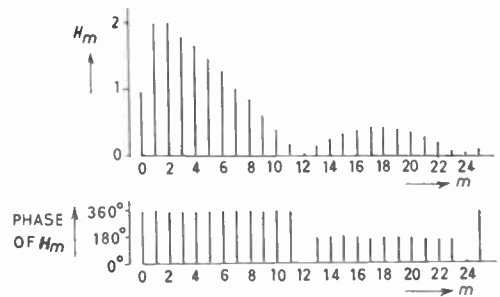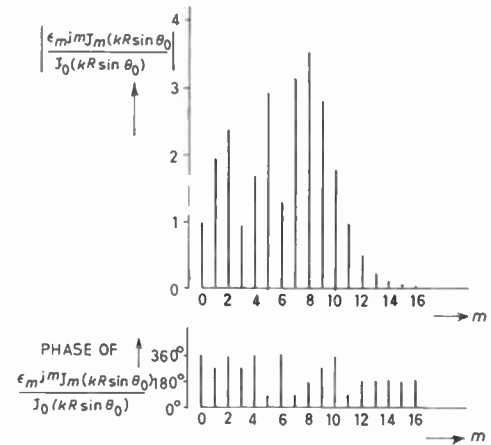

Fig. 7. The rectangular shaped directional pattern.

(a) Spectrum of the rectangular series shown in Fig. 7 with $\Phi_R/2 = 15°$.

(b) Spectrum of the complex factor $\varepsilon_m j^m J_m(kR\sin\theta_0)$ for $\dfrac{\text{circular array diameter} \times \sin\theta_0}{\text{wave length}} = 3$

corresponding to $kR\sin\theta_0 = 9\cdot43$ normalized to $J_0(9\cdot43) = -0\cdot1819$.

Fig. 8.

assume the values $0°$ and $180°$. The envelope over the amplitude spectrum must show the well known $|\sin x/x|$-character. Next, in Fig. 8(b), are given the amplitude and phase spectra of the complex factor

$\varepsilon_m j^m J_m(kR \sin \theta_0)$ by which the integral given by eqn. (8) must be multiplied in order to obtain (apart from the constant factor), the desired value H_m. In order to be able to compare the spectrum lines in a comprehensible manner with Fig. 8(a), the line with $m = 0$ is brought to the value 1 and all other harmonics are related to this value. It can be seen that the amplitude spectrum shows a completely different shape from the desired one. First of all, with the given argument $kR \sin \theta_0$ only a finite number of amplitudes is available. Also the phase spectrum shows lines with the values $0°$, $90°$, $180°$ and $270°$. How is the function $a(\Phi)$ to be chosen so that the product according to eqn. (8) just gives the nominal value H_m? The expression in square brackets with the definite integral in eqn. (8) represents in a comprehensive manner the Fourier coefficients of the even function $a(\Phi)$. It would therefore be obvious to try the general solution of a trigonometrical sum for $a(\Phi)$:

$$a(\Phi) = \sum_{\mu=0}^{\infty} a_\mu \cos \mu \Phi \qquad \ldots\ldots(9)$$

With this arrangement the definite integral of eqn. (8) can be solved and the relation between the nominal value H_m and a_μ is reduced to:

$$H_m = K(\theta_0)j^m J_m(kR \sin \theta_0)\varepsilon_m \beta/2 \sum_{\mu=0}^{\infty} a_\mu P_{m\mu} \quad \ldots\ldots(10a)$$

with the real coefficients

$$P_{m\mu} = \frac{\sin(m+\mu)\beta/2}{(m+\mu)\beta/2} + \frac{\sin(m-\mu)\beta/2}{(m-\mu)\beta/2} \quad \ldots\ldots(10b)$$

If the following temporary abbreviation is introduced

$$Y_m = \frac{H_m}{\beta/2\varepsilon_m j^m J_m(kR \sin \theta_0)K(\theta_0)}$$

eqn. (10a) can be written in the general form

$$Y_m = \sum_{\mu=0}^{\infty} P_{m\mu} a_\mu \qquad \ldots\ldots(11)$$

This relation represents a linear system containing an infinite number of equations. In reality, however, the maximum value of m is—as mentioned—always finite for a certain argument $kR \sin \theta_0$ because of the behaviour of the Bessel functions.

For $0 < m < r$, the value of a_μ can be computed from the linear system up to the rth equation. Thus $0 < \mu < r$.

The finite linear system of equations then has the explicit form:

$$Y_0 = P_{00}a_0 + P_{01}a_1 + \ldots P_{0\mu}a_\mu + \ldots P_{0r}a_r$$
$$Y_1 = P_{10}a_0 + P_{11}a_1 + \ldots P_{1\mu}a_\mu + \ldots P_{1r}a_r$$
$$Y_m = P_{m0}a_0 + P_{m1}a_1 + \ldots P_{m\mu}a_\mu + \ldots P_{mr}a_r \qquad \ldots\ldots(12)$$
$$Y_r = P_{r0}a_0 + P_{r1}a_1 + \ldots P_{r\mu}a_\mu + \ldots P_{rr}a_r$$

According to Cramer's rule the solutions a_μ result from:

$$a_\mu = \frac{2}{\beta}(-1)^\mu \frac{1}{\Delta} \sum_{m=0}^{r} \frac{j^m}{\varepsilon_m} \frac{H_m}{K(\theta_0)J_m(kR \sin \theta_0)} D_{m\mu} \quad \ldots\ldots(13)$$

Δ = determinant of the linear system of equations

$D_{m\mu}$ = minor which results by omitting the mth line and the μth column.

Even an extensive linear system of equations can easily be solved on a digital computer, so that a quick numerical determination of a_μ is possible. If a_μ is known the required distribution function results from eqn. (9) with an upper limit of the sum $\mu = r$.

It appears advantageous to clarify the steps of the solution by a particularly simple example. With the complete circular array, $\beta/2 = \pm\pi$. The coefficients $P_{m\mu}$ according to eqn. (10b) have for this special case the following values:

$$P_{m\mu} = \begin{cases} 0 \text{ when } m \neq \mu \\ 1 \text{ when } m = \mu \end{cases}$$

In the linear system of eqns. (12), only the terms of the main diagonal are different from 0. $a_\mu = Y_{m=\mu}$ is obtained, or in detail:

$$a_\mu = \frac{H_\mu}{\pi\varepsilon_\mu j^\mu K(\theta_0)J_\mu(kR \sin \theta_0)} \qquad \ldots\ldots(14)$$

$$\varepsilon_\mu = \begin{cases} 1 \text{ when } \mu = 0 \\ 2 \text{ when } \mu \neq 0 \end{cases}$$

If by using this result the distribution function $a(\Phi)$ is formed with eqn. (9) it can easily be demonstrated that in the special case of a complete circular array the values $a(\Phi)$ and $a(180° - \Phi)$ are complex conjugates of each other.

For this reason and because of the specially simple calculation of a_μ, the complete circular array is most suitable for an experimental check of the synthesis procedure described.

In order to reduce the mathematical effort, the considerations of this Section were limited to circular arrays with continuous distribution of radiators. The calculation could be extended to arrays in which there is a uniform finite distance between the individual radiators. Though this case is more important in practice than the case examined here, we shall refrain from a detailed presentation now because of the more extensive calculation involved. It may be mentioned only that, analogous to eqn. (12), a linear system of equations can be derived for the relation between the required directional function and the necessary amplitude distribution. The coefficients $P_{m\mu}$ of this system of equations follow, of course, a law of formation which differs from eqn. (10b).

4. An Arrangement to Realize the Distribution Function

In dealing with this subject, only the principal points of view will be discussed. The question as to optimum power radiation in case of transmitting—or maximum

$$I_v \simeq [U_T - R_T \sum_v I_v] G_v$$
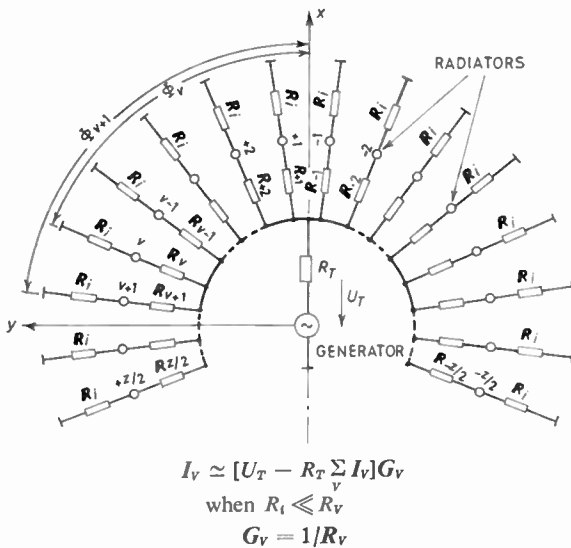when $R_i \ll R_v$
$$G_v = 1/R_v$$

Fig. 9. An arrangement for adjusting the distribution function.

sensitivity in case of receiving—cannot be answered in more detail. The reciprocity theorem is valid, and so it is the same whether an arrangement for the case of transmitting or another one for the case of receiving is indicated. Here only the case is considered in which the circular array radiates electromagnetic or acoustical waves.

In order to approach the synthesis procedure, it seemed to be advantageous for didactic reasons to suppose a circular array with a continuous distribution of radiators. For the present technical problem, which is to design an arrangement in which the individual radiators have to be fed with a determined nominal amplitude, the idea of an infinite number of radiators is difficult to grasp. It has already been mentioned, however, that the principle of synthesis can be used also in an analogous way with a finite number of radiators. The solution of the linear system of equations in analogy with eqn. (12), in this case does not give the coefficients a_μ of the continuous

distribution function $a(\Phi)$, but the complex amplitude A_μ of the individual harmonics into which the distribution can be split up. In analogy to eqn. (9) there exists for the distribution the following equation:

$$A_v = \sum_{\mu=0}^{r} A_\mu \cos \mu \Phi_v \qquad \ldots\ldots(15)$$

The only difference, compared with eqn. (9), is in the non-continuity of the angle Φ_v, contrary to the continuously-changing angle Φ.

A_μ and A_v can mean currents or voltages. In Fig. 9 the individual radiators are symbolized by the circles with the ordinal numbers $-z/2 < v < +z/2$. The internal resistances of the radiators are equal and have the value R_i. The generator U_T supplies the individual radiators through its internal resistance R_T and the different coupling impedances R_v. The current of the vth source is calculated as:

$$I_v = \frac{U_T - R_T \sum_v I_v}{R_v \left[1 + \dfrac{R_i}{R_v}\right]} \qquad \ldots\ldots(16)$$

It can be seen that the internal resistance of the generators is only of importance for the absolute value of I_v. The relation of individual I_v values to each other—which is of importance here—is, however, independent of R_T.

As questions of sensitivity are not to be discussed here, it can be assumed that $R_i \ll R_v$. Then eqn. (16) simplifies as follows:

$$I_v = [U_T - R_T \sum_v I_v] G_v \qquad \ldots\ldots(17)$$

where $G_v = 1/R_v$ and denotes the coupling admittance between generator and vth individual radiator. The coupling admittance can be conceived as the resulting admittance of the connection in parallel of individual $G_\mu \cos \mu \Phi_v$:

$$G_v = \sum_{\mu=0}^{r} G_\mu \cos \mu \Phi_v \qquad \ldots\ldots(18)$$

The individual G_μ are thus proportional to solutions A_μ of the linear system of equations. Figure 10 shows
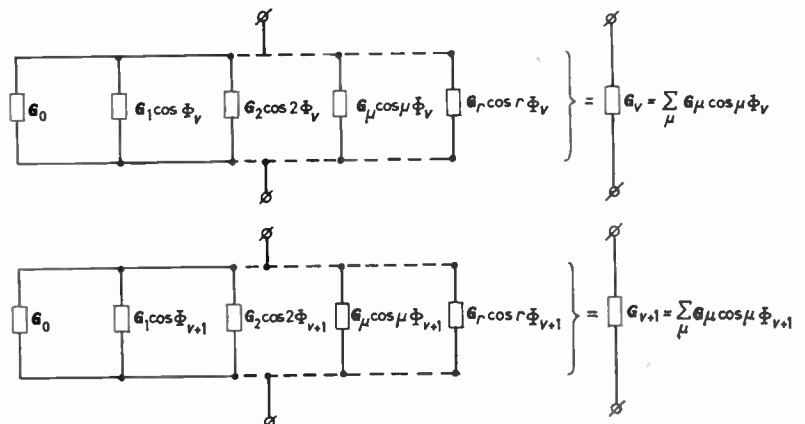

Fig. 10.
The realization of coupling admittance for the radiator v and $v + 1$.

as example the form of the coupling admittances for the individual radiators v and $v+1$.

Further analogous variations of this procedure for the adjustment of the distribution can be thought of. No essential new points of view will occur thereby.

5. Practical Limits of the Synthesis Procedure

Figure 7 represented the shape of an ideal directional pattern and its spatial harmonics in spectral representation were given. The spectrum of the directional pattern of a circular array with uniform distribution of amplitudes was indicated in Fig. 8(b) for an arbitrarily chosen case. Sections 3 and 4 gave theoretical and practical rules according to which the spectrum with the controlling complex factor

$$\varepsilon_m j^m J_m(kR \sin \theta_0)$$

can be influenced in such a way that the spectrum of a desired directional function results. The applicability of this synthesis procedure has practical limits. In order to outline them, it is necessary to enter briefly into particulars of the Bessel functions (see Fig. 11).

(1) The Bessel functions leave the abscissa for the first time like parabolas of mth order. They increase to a maximum and then continue as slightly attenuated oscillations about the abscissa.

(2) Generally, for a given argument $kR \sin \theta_0$, part of the Bessel function has already passed through one or more extreme values. The number of these extreme values becomes larger as the order m of the function is reduced.

(3) All other Bessel functions of the same argument as in (2), of orders up to $m = \infty$, will not yet have reached the first maximum. The value function corresponding to the argument thus becomes smaller as the order m of the function is increased.
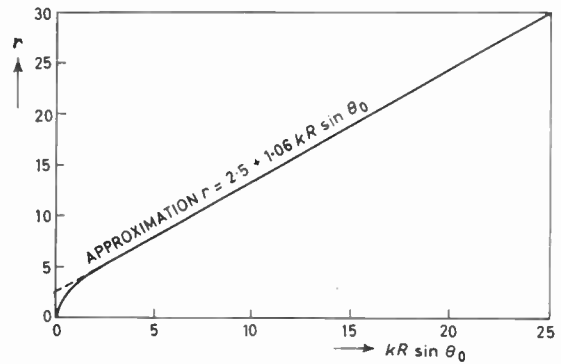
Fig. 12. The spatial harmonic of maximum order r, which has to be considered for the synthesis as a function of $kR \sin \theta_0$.

Theoretically, there are always all spatial harmonics available with a circular array and these could be brought to the nominal value according to the described procedure, the nominal value corresponding to some ideal directional pattern. Practical considerations, however, set a limit to the procedure. Spatial harmonics of small amplitude cannot be controlled in the desired way but only those which at least once reach a certain threshold. A precise determination of the threshold value will not be given here as, for this, many points of view have to be considered carefully. According to considerations made here a pre-requisite for the applicability of the synthesis is the condition:

$$J_m(kR \sin \theta_0) \geqslant 0.02 \qquad \ldots\ldots(19)$$

According to this requirement, two points have to be considered:

(1) The argument $kR \sin \theta_0$ must not be placed on the zero-position of a Bessel function, unless the corresponding spatial harmonic in the spectrum of the nominal directional function must also become zero.
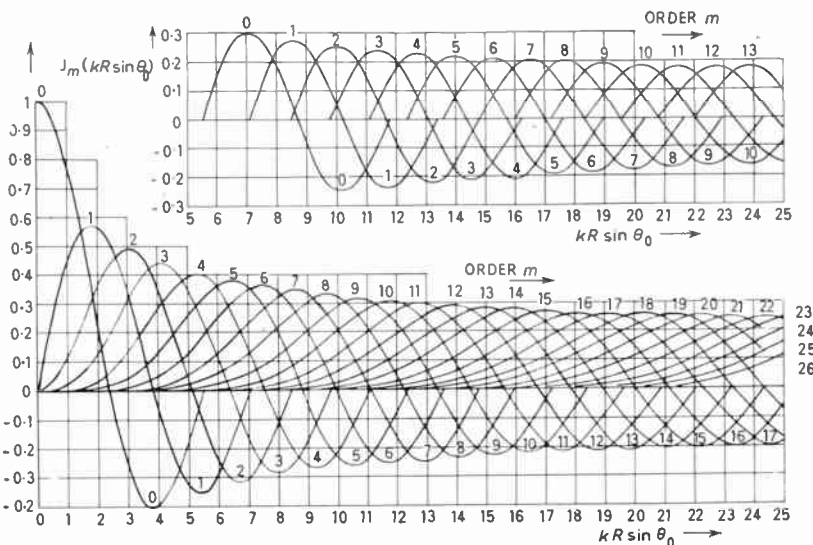
Fig. 11.
The Bessel functions $J_m(kR \sin \theta_0)$.

(2) From those Bessel functions of the argument $kR \sin \theta_0$ which have not yet reached their first maximum, from the condition laid down in eqn. (19), the spatial harmonics of order $m = r$ are given unambiguously. In Fig. 12, r is marked above $kR \sin \theta_0$. The curve was drawn up to $r = 15$ according to values taken from various tables of functions and then simply extended linearly. Above $r = 15$ small deviations from the representation in the drawing are thus possible.

When $kR \sin \theta_0 \geqq 3$, Fig. 12 shows that the order of the highest harmonic is given, to a good approximation, by

$$r = 2 \cdot 5 + 1 \cdot 06 \, kR \sin \theta_0 \qquad \ldots (20)$$

6. Optimum Directional Pattern

From the conclusions reached in the previous Section, it is seen that, for a given array diameter, frequency and angle of elevation, only a definite number of spatial harmonics are available for the synthesis of a desired directional function.

Their amplitudes and phase angles can be influenced at will according to the methods of Sections 3 and 4. Now it must be established which combination of these influenced spatial harmonics gives a 'good' directional pattern. The term 'good' is to be understood here as having a minimum half-power-beamwidth with a maximum attenuation of the side-lobes. It appears likely that it will not be simple to answer this question because of the large number of parameters which at first can be chosen freely.

The method of solution appears to be as follows:

(1) The desired normalized directional pattern is given—for example, the ideal function according to Fig. 7—and its spectrum is determined by calculating the Fourier coefficients.

(2) The deviations which will occur between the given function and the function which is really obtainable, are investigated for the case in which not all harmonics from 0 to ∞ are available but only those of orders of $0 < m < r$ (r follows from Fig. 12).

Among the great number of possible functions which can be given, only the ideal rectangular curve will be considered here. This deliberate limitation is made for two reasons. First, the mathematical effort involved in treating the rectangular series with a non-infinite number of harmonics is less than with other functions. Secondly, the rectangular curve already supplies important rules for design which are applic-

able also to other given functions, the mathematical treatment of which, however, is more difficult.

6.1. *The Rectangular Series with a Non-Infinite Number of Harmonics*

Starting with a rectangular series according to Fig. 7, the law for the formation of nominal values of amplitudes H_m in eqn. (8) is as follows:

$$H_0 = \frac{\Phi_R}{2\pi} ; \qquad H_{\substack{m \\ m \neq 0}} = \frac{2}{\pi} \frac{\sin m \dfrac{\Phi_R}{2}}{m} \qquad \ldots (21)$$

After realizing the synthesis procedure, for the circular array, the directional pattern—in a non-normalized form—is obtained:

$$K_G(\theta_0; \phi_0) = K(\theta_0) \left[\frac{\Phi_R}{2\pi} + \frac{2}{\pi} \sum_{m=1}^{r} \frac{\sin m \dfrac{\Phi_R}{2}}{m} \cos m\phi_0 \right] \qquad \ldots (22)$$

In the following consideration, attention is only drawn to the term in square brackets which will be abbreviated by $K_G(\phi_0)$.

The question is: what value must be assumed for the beamwidth ϕ_R of the ideal rectangular curve—with a given r—so that the directional pattern obtained will show the above-mentioned optimum properties?

The following calculations follow partly the ideas of the American mathematician J. W. Gibbs, who made investigations on Fourier series with non-infinite numbers of harmonics.[17-22] First, for determining the position of maxima of eqn. (22), its derivative is formed. The differentiated Fourier series can be transformed to a closed form by using the summation formula of a geometric series. Thus the following is obtained:

$$2\pi K_G'(\phi_0) = \frac{\sin \left[\dfrac{2r+1}{2} \left(\phi_0 + \dfrac{\Phi_R}{2} \right) \right]}{\sin \left[\dfrac{1}{2} \left(\phi_0 + \dfrac{\Phi_R}{2} \right) \right]} - \frac{\sin \left[\dfrac{2r+1}{2} \left(\phi_0 - \dfrac{\Phi_R}{2} \right) \right]}{\sin \left[\dfrac{1}{2} \left(\phi_0 - \dfrac{\Phi_R}{2} \right) \right]} \qquad \ldots (23)$$

or after application of the addition-theorems:

$$2\pi K_G'(\phi_0) = C \sin \left[\frac{2r+1}{2} \phi_0 - \tan^{-1} \left(D \tan \frac{\phi_0}{2} \right) \right] \qquad \ldots (24a)$$

$$\text{with: } C = \frac{\pm 2}{\cos \dfrac{\Phi_R}{2} - \cos \phi_0} \times \sqrt{1 - \cos \phi_0 \cos \frac{\Phi_R}{2} - \left(\cos \frac{\Phi_R}{2} - \cos \phi_0 \right) \times \cos \left[(2r+1) \frac{\Phi_R}{2} \right]} \qquad \ldots (24b)$$

$$D = \cot \frac{\Phi_R}{4} \tan \left(\frac{2r+1}{2} \frac{\Phi_R}{2} \right) \qquad \text{......(24c)}$$

Equation (24a) is specially suitable for the location of the maximum positions of K_G. These are generally found from the implicit transcendental relation:

$$\frac{2r+1}{2} \phi_0 - n\,180° = \tan^{-1}\left(D \tan \frac{\phi_0}{2} \right) \quad \text{......(25)}$$

A special case is that in which the azimuth angle $\phi_0 = \pm \Phi_R/2$. In this case it can be seen that the sine-function in eqn. (24a) has the value zero, whereas the amplitude C becomes infinite because of the denominator before the root. The natural value of K_G' follows from eqn. (23). Likewise, K_G itself has no maximum value at the position $\phi_0 = \pm \Phi_R/2$. For a further analytical treatment of the problem it is necessary to obtain K_G in closed form. An exact integration of eqn. (23) meets with considerable difficulties. Therefore an approximate solution is obtained. This solution is valid under the condition that $|\phi_0 \pm \Phi_R/2| \leqq 60°$ and in its normalized form is as follows:

$$K_G(\phi_0)$$
$$= \frac{\text{Si}\left[\frac{2r+1}{2}\left(\phi_0 + \frac{\Phi_R}{2}\right)\right] - \text{Si}\left[\frac{2r+1}{2}\left(\phi_0 - \frac{\Phi_R}{2}\right)\right]}{2\,\text{Si}\left(\frac{2r+1}{2}\frac{\Phi_R}{2}\right)}$$
$$\text{......(26)}$$

(Si = sine-integral of the indicated argument).

Figures 13 and 14 show that comparison between the exactly (eqn. (22)) and the approximately (eqn. (26)) calculated curve of K_G for two values of r arbitrarily chosen at constant Φ_R. The conformity is very good in the interesting range of K_G. In the same figures, furthermore, the graphical solutions of eqn. (25) are inserted. The points of intersection of the $\tan^{-1}$ function (right-hand side) with straight lines (left-hand side) correspond for their ϕ_0-position to the maximum positions of K_G, except position $\phi_0 = \pm \Phi_R/2$. Equation (25) is of great importance for the correct design of circular arrays. This equation, therefore, is termed the 'characteristic equation' in the following Section.

6.2. Unambiguous Directional Patterns

A comparison of directional pattern K_G in Figs. 13 and 14 shows a different behaviour near $\phi_0 = 0°$.

Fig. 13. The directional pattern K_G—exactly and approximately calculated—and the graphical solution of the 'characteristic equation' for $\Phi_R/2 = 15°$; $r = 12$.

○ points calculated with eqn. (26)

— points calculated with the Fourier series up to $r = 12$.

Another observation is interesting. The graphical solutions of the respective characteristic equations give the following relations between parameter n and the position of maxima of K_G.

With Fig. 13:

n	0	1	2	3...
maximum position of K_G	main maximum	$\phi_0 = \Phi_R/2$	maximum value following first zero	other maxima

With Fig. 14:

n	0	1	2	3	4
maximum position of K_G	maximum position at $\phi_0 = 0$	maximum near $\phi_0 = 0$	$\phi_0 = \Phi_R/2$	maximum value following first zero	other maxima

It is evident that the choice of Φ_R has to be made in conjunction with r. A directional pattern of the shape shown in Fig. 14 is undesirable in practice as it possesses several extremes of approximately unit amplitude. Unambiguous directional patterns are preferred of the type shown in Fig. 13. Detailed investigations yield the following condition for non-ambiguity.
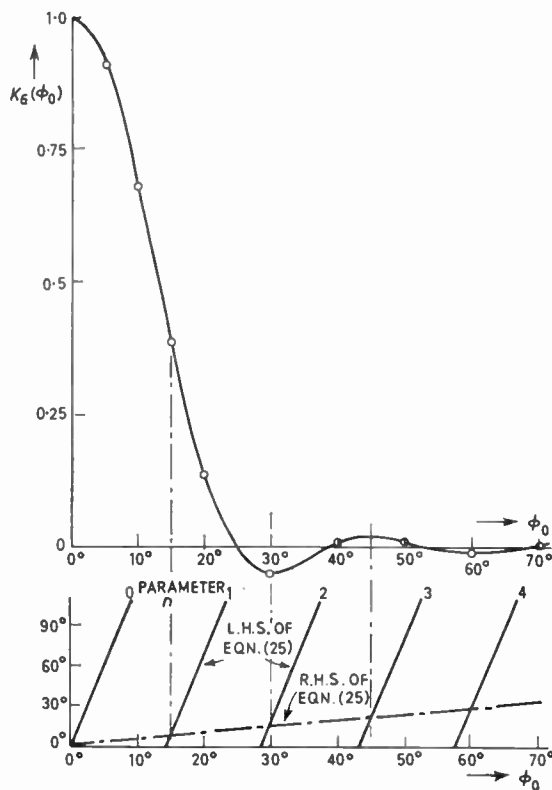
The normalized directional pattern K_G only has a single maximum of amplitude 1, if the characteristic equation of parameter $n = 0$ or $n = 1$ shows a solution $\phi_0 = \Phi_R/2$.

If a solution of the characteristic equation of parameter $n \geq 2$ is at $\phi_0 = \Phi_R/2$, this means that Φ_R has been made excessively large, and various maxima of approximately unit amplitude are obtained. In order to find the maximum admissible value of Φ_R according to the condition of non-ambiguity, the solutions represented in Figs. 13 and 14 of the characteristic equation have to be considered again. If in the argument of the $\tan^{-1}$ function the factor $D = \pm \infty$, the right-hand side of this equation becomes independent of ϕ_0 and has the constant value $\pm 90°$. In both special cases, the points of intersection with the straight line for $n = 1$ can be indicated easily. Because of the positive rise of the straight line the maximum admissible value of $\Phi_R = \hat{\Phi}_R$ follows from the point of intersection for which $D = \infty$.

$$\frac{\hat{\Phi}_R}{2} = \frac{540°}{2r+1} \qquad \ldots\ldots(27)$$

According to practical requirements only the range $\Phi_R < \hat{\Phi}_R$ will be examined in detail for the determination of the optimum directional pattern.
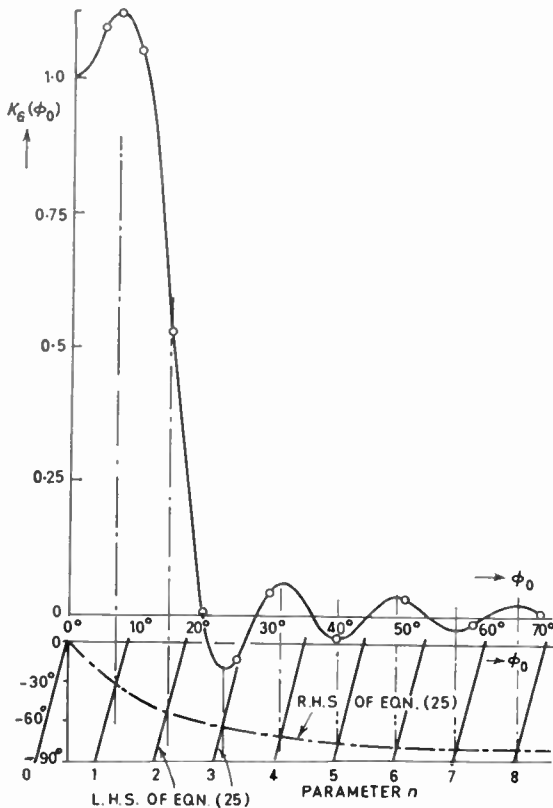
6.3. *The Dominating Side-lobe*

In Fig. 13 a general property of the rectangular series with a non-infinite number of harmonics can be seen. Among the many undesired maxima, the largest is that which follows the first zero of the function. All further considerations thus may be limited to this side-lobe alone. The azimuth position $\phi_0 = \phi_N$ is determined for $\Phi_R < \hat{\Phi}_R$ from the solution of the characteristic equation of parameter $n = 1$ or $n = 2$, depending whether the solution with $\phi_0 = \Phi_R/2$ was situated at $n = 0$ or $n = 1$. The amplitude of the dominating side-lobe follows with good approximation from eqn. (26) for $\phi_0 = \phi_N$.

Because of the implicit transcendental form of eqn. (25) it is unfortunately possible only in special cases to give the amplitude explicitly. A list of five outstanding cases is shown in Table 1. The result is surprising. It means that the attenuation of the dominating side-lobe for $D = +\infty$; 0; $-\infty$ for a circular array is independent of the ratio of diameter to wavelength. For the special cases $D = \pm 1$ there is only a very slight dependence on this ratio as under practical conditions $r \gg 1$.

With the attenuation obtained for $D = 0$, i.e. for a choice of the angle

$$\frac{\Phi_R}{2} = \frac{360°}{2r+1} \qquad \ldots\ldots(28)$$

the side-lobe of -26.43 dB represents at the same time the theoretical limit of the attenuation obtainable under the given conditions. The proof of this is complicated and will not be given here. Reference is made instead to Fig. 15 which makes clear the necessity for correct design.

6.4. *Half-power Beamwidth*

It is well known that by this term we understand the difference between the two azimuth angles which correspond to the functional values $K_G = 1/\sqrt{2}$. According to the usual nomenclature, the half-power beamwidth is designed here by $\Delta\Phi$. As it is necessary to know the half-power beamwidth for the determination of optimum properties, a closed mathematical expression must be deduced for it if possible. An exact determination meets with great difficulties. An observation, however, can be utilized which can also be seen in Fig. 13.

Fig. 14. The directional pattern K_G—exactly and approximately calculated—and the graphical solution of the 'characteristic equation' for $\Phi_R/2 = 15°$; $r = 20$.

○ points calculated with eqn. (26)

— points calculated with the Fourier series up to $r = 20$.

Table 1

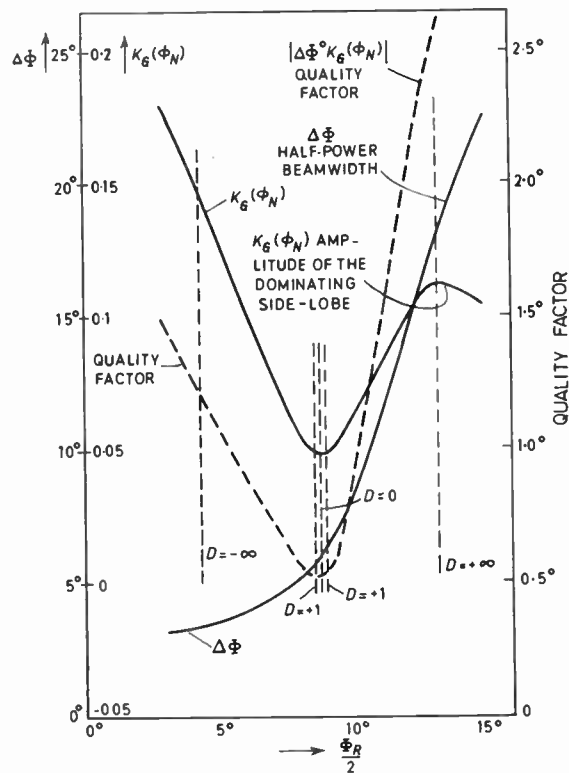| D	$\dfrac{\Phi_R}{2}$	ϕ_N	$K_G(\phi_N)$	$\Delta\Phi°$	$\left\|\Delta\Phi° K_G(\phi_N)\right\|$
$+\infty$	$\dfrac{540°}{2r+1}$	$\dfrac{900°}{2r+1}$	$\dfrac{\mathrm{Si}(4\pi)-\mathrm{Si}(\pi)}{2\mathrm{Si}(\tfrac{3}{2}\pi)} = -0.111 \equiv -19.094\,\mathrm{dB}$	$\dfrac{945°}{2r+1}$	$\dfrac{105°}{2r+1}$
$+1$	$\dfrac{180°}{r}$	$\dfrac{360°}{r}$	$\dfrac{\mathrm{Si}\left[(1+\tfrac{1}{2}r)3\pi\right]-\mathrm{Si}\left[(1+\tfrac{1}{2}r)\pi\right]}{2\mathrm{Si}\left[(1+\tfrac{1}{2}r)\pi\right]}$		
0	$\dfrac{360°}{2r+1}$	$\dfrac{720°}{2r+1}$	$\dfrac{\mathrm{Si}(3\pi)-\mathrm{Si}(\pi)}{2\mathrm{Si}(\pi)} = -0.048 \equiv -26.43\,\mathrm{dB}$	$\dfrac{444°}{2r+1}$	$\dfrac{21.2°}{2r+1}$
-1	$\dfrac{180°}{r+1}$	$\dfrac{360°}{r+1}$	$\dfrac{\mathrm{Si}\left[\left(1+\dfrac{1}{2(r+1)}\right)3\pi\right]-\mathrm{Si}\left[\left(1+\dfrac{1}{2(r+1)}\right)\pi\right]}{2\mathrm{Si}\left[\left(1+\dfrac{1}{2(r+1)}\right)\pi\right]}$		
$-\infty$	$\dfrac{180°}{2r+1}$	$\dfrac{540°}{2r+1}$	$\dfrac{\mathrm{Si}(2\pi)-\mathrm{Si}(\pi)}{2\mathrm{Si}(\pi/2)} = -0.158 \equiv -16.027\,\mathrm{dB}$	$\dfrac{339°}{2r+1}$	$\dfrac{53.5°}{2r+1}$

Fig. 15. Amplitude of the dominating side-lobe, half-power beamwidth and quality of the synthetically obtained directional pattern of a circular array with $r = 20$ (corresponding to $kR\sin\theta_0 = 16$).

Between the azimuth angles

$$\phi_0 = \frac{\Phi_R}{2} \quad \text{and} \quad \phi_0 = \frac{\Delta\Phi}{2}$$

the directional pattern is nearly linear. A corresponding Taylor development of eqn. (26) at the point $\phi_0 = \frac{\Phi_R}{2}$ gives:

$$\frac{\Delta\Phi}{2} \simeq \frac{\text{Si}\left(\frac{2r+1}{2}\Phi_R\right) + \frac{2r+1}{2}\frac{\Phi_R}{2}\left[1 - \text{si}\left(\frac{2r+1}{2}\Phi_R\right)\right] - \sqrt{2}\,\text{Si}\left(\frac{2r+1}{2}\frac{\Phi_R}{2}\right)}{\frac{2r+1}{2}\left[1 - \text{si}\left(\frac{2r+1}{2}\Phi_R\right)\right]} \quad \ldots\ldots(29)$$

Si = sine-integral of the indicated argument

si = $\frac{\sin x}{x}$-function of the indicated argument.

Specially simple forms are assumed by eqn. (29) for the special cases $D = +\infty$; 0; $-\infty$. The obtained values are also indicated in Table 1. It can be seen that the $\Delta\Phi$ values are among each other nearly in the same ratio as the Φ_R values belonging to the three special cases; only they are smaller in magnitude. As regards r, in all three cases there is practically an inverse proportionality beginning from $r > 4$.

6.5. *The Quality of a Directional Pattern*

As standard for the quality of a directional pattern, the product of half-power beamwidth and amplitude of the dominating side-lobe can be taken. It is desired to obtain an amount as small as possible for this product. In Table 1 values for the three special cases are inserted. It can be seen that the optimum design must prevail if $\Phi_R/2$ is chosen slightly smaller than $\frac{360°}{2r+1}$. The calculation of the precise value can be made by means of the deduced formulae. This calculation, however, does not offer any essential surprise, as the relative change of the half-power beamwidth with Φ_R is below that of the dominating side-lobe. Therefore, no further analytical treatment of the problem will be given here. The example for $r = 20$ shown in Fig. 15 gives a clearer impression than lengthy calculations.

7. Summary of Results for the Synthesis

In Fig. 16 various curves are shown as functions of the ratio of the circular array diameter to the wavelength with known angle of elevation.

Curve (1): Beamwidth Φ_R of the rectangular pattern according to Fig. 7 for an attenuation of the dominating side-lobe of $-26\cdot43$ dB (theoretical maximum attenuation). This curve indicates how Φ_R and the values following with eqn. (21) of the Fourier coefficients H_m must be chosen in order to be able to carry out the synthesis according to Section 3.

Curve 2: Practically obtainable half-power beamwidth $\Delta\Phi$ of a circular array, if curve (1) is taken as the basis for design. The gain of the circular array is approximately inversely proportional to $\Delta\Phi$.

Curve (3): Maximum admissible beamwidth $\hat{\Phi}_R$ according to eqn. (27) of the rectangular series. If $\Phi_R > \hat{\Phi}_R$ is chosen, ambiguous directional patterns are obtained as was shown in Fig. 14 for example. For $\Phi_R = \hat{\Phi}_R$, the attenuation of the dominating side-lobe is $-19\cdot09$ dB.

Curve (4): Beamwidth Φ_R of the rectangular pattern of the special case $D = -\infty$ in Table 1 (attenuation of dominating side-lobe $-16\cdot03$ dB).

If the chosen value Φ_R is between curves (3) and (4), the position of the dominating side-lobe results from the characteristic eqn. (25) with $n = 2$.

8. Experimental Verification of the Synthesis Method

The method described in Section 3 can be checked by a simple experiment. The experimental arrangement described by Leisterer[23] comprises a so-called strip-collector which is fed through a loaded transmission line terminated by its characteristic impedance and a contact plate which rotates on the strip collector, and is very well suited to laboratory experiments. Anywhere on the contact plate contacts can be fixed which correspond to the radiators of an array. Figure
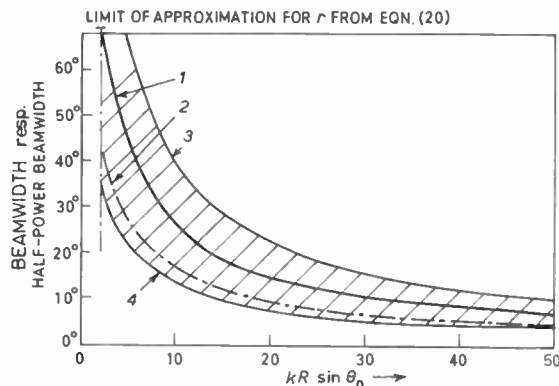

Fig. 16. Half-power beamwidths obtainable with various methods of synthesis, as a function of the ratio of circular array diameter to wavelength, for a given angle of incidence θ_0.

17 shows the measuring arrangement with a linear array of three individual radiators. The strip collector simulates the plane progressing wave field. The output voltages of the sources are added with the desired phase positions and amplitudes and are indicated.

Because of the reasons given at the end of Section 3, the test was made with a complete circular array consisting of 40 radiators (angular distance from radiator to radiator is 9°). Frequency and diameter of the circular array were chosen in a way that the argument of Bessel functions was $kR \sin \theta_0 = 6$. According to eqn. (20) in these circumstances Bessel functions up to the order $r = 9$ are to be considered.

For the half-power beamwidth of the rectangular series a value—not quite of the optimum condition of eqn. (28)—of $\Phi_R/2 = 20°$ was taken as a base (optimum value: $\Phi_R/2 = 19°$). Equations (14) and (9) give the distribution function $a(\Phi)$ of a circular array continuously filled with radiators. Though with 40 radiators it cannot really be called a continuous circular array, for the sake of simplicity the complex amplitudes of the radiators—according to the value of the function $a(\Phi)$ at the radiator-position—were chosen. To avoid errors, a calculation of the amplitude of the directional pattern was again made with the adjusted distribution. In Fig. 18 the calculation and result of the measurement have been compared. It can be seen that this proves the correctness of the method. Whereas there is a good agreement between calculation and measurement in the main beam, in the other part of the directional pattern there are variations. As reasons
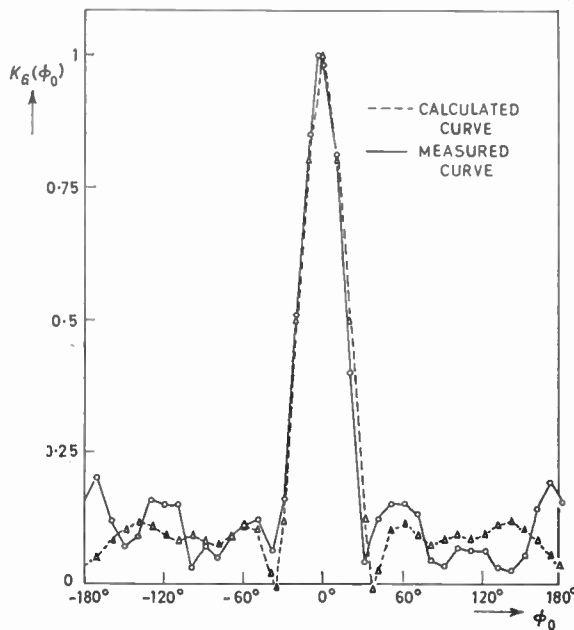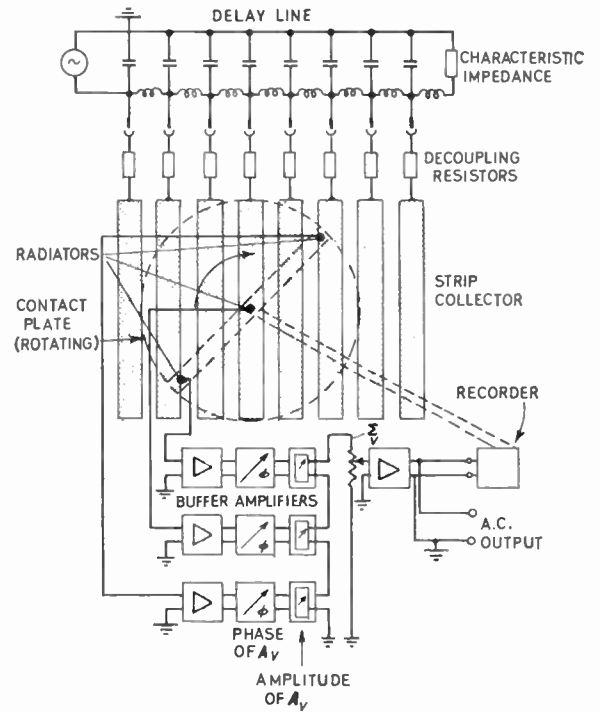

Fig. 17. Laboratory arrangement for measuring the directional pattern of an array[23].

one should mention the unsteady change in the phase with the simulation of the wave field, and tolerances in the adjustment of the distribution function.

9. Conclusion

The present paper gives a contribution to the general problem of the synthesis of directional patterns. Although synthesis problems have, for nearly four decades, been subject to theoretical and practical treatment, until today no really final papers exist. This proves the difficulty of the task. It is only very recently that a paper was published[24] which in this connection should find special attention.

10. References and Bibliography

1. N. Davids, E. G. Thurston and R. E. Mueser, "The design of optimum directional acoustic arrays", *J. Acoust. Soc. Amer.*, **24**, p. 50, January 1952.

2. R. F. Harrington, "Side-lobe reduction by non-uniform element spacing", *Proc. I.R.E.*, **43**, p. 187, March 1961.

3. K. Baur, "Antennenzeilen mit gedämpften Nebenzipfeln", *Elektron. Rdsch.*, **74**, p. 217, June 1960.

4. S. A. Schelkunoff, "Electromagnetic Waves", p. 340 (van Nostrand, New York, 1943).

5. H. Zuhrt, "Elektromagnetische Strahlungsfelden", p. 277 (Springer-Verlag, Berlin, 1953).

6. S. A. Schelkunoff and H. T. Friis, "Antennas Theory and Practice", p. 188 (John Wiley, New York, 1952).

Fig. 18. Calculated and measured curves of the directional pattern.

7. D. G. Tucker, "Some aspects of the design of strip arrays", *Acustica*, **6**, p. 403, 1956.

8. C. L. Dolph, "A current distribution for broadside arrays which optimizes the relationship between beamwidth and side-lobe-level", *Proc. I.R.E.*, **34**, p. 335, 1946.

9. Meinke/Grundlach, "Taschenbuch der Hochfrequenztechnik, 2. Aufl.", p. 490, 572 (Springer-Verlag, Berlin, 1962).

10. O. Brosze and H. Stenzel, "Leitfaden zur Berechnung von Schallvorgängen 2. Aufl.", p. 22 (Springer-Verlag, Berlin, 1958).

11. K. Menges, "Über Richtcharakteristiken von ebener Strahlerflächen, Strahlerstrecken mit ungleichmäßiger Amplitudenverteilung und der Halbkreislinie" (S. Hirzel Verlag, Leipzig, 1941).

12. N. W. McLachlan, "Bessel Functions for Engineers" (Oxford, Clarendon Press, 1958).

13. G. N. Watson, "A Treatise on the Theory of Bessel Functions (Cambridge University Press, 1958).

14. F. Jahnke-Emde, "Tables of Higher Functions" (Leipzig, 1952).

15. W. Magnus and F. Oberhettinger, "Formeln und Sätze für die speziellen Funktionen der mathematischen Physik", (Springer-Verlag, Berlin, 1948).

16. W. Rehwald, "Elementare Einführung in die Bessel-, Neumann- und Hankel-Funktionen" (S. Hirzel Verlag, Stuttgart, 1959).

17. J. W. Gibbs, "Fourier's series", *Nature (London)*, **59**, p. 200 and pp. 606, 1898 and 1899.

18. J. W. Gibbs, "Fourier's series", Papers, London, New York, Bombay, 1906, Vol. 2, p. 258.

19. C. Runge, "Theorie und Praxis der Reihen", p. 170 (S. Hirzel Verlag, Leipzig, 1904).

20. H. Weyl, "Die Gibb'sche Erscheinung in der Theorie der Kugelfunktionen", *Rend. Circ. Mat. Palermo*, **29**, p. 308, 1910.

21. H. Weyl, "Über die Gibb'sche Erscheinung und verwandte Konvergenzphänomene", *Rend. Circ. Mat. Palermo*, **30**, p. 377, 1910.

22. K. W. Wagner, "Einführung in die Lehre von den Schwingungen und Wellen", p. 38 (Dietrich'sche Verlagsbuchhandlung, W. Klem, Wiesbaden, 1947).

23. R. Leisterer, "Ein Verfahren zur raschen Ermittlung der Richtcharakteristik von Strahlergruppen", Proc. 3rd International Congress on Acoustics, 1959, p. 685.

24. J. Harris and H. E. Shanks, "A method for synthesis of optimum directional patterns from nonplanar apertures", *I.R.E. Trans. on Antennas and Propagation*, AP-10, No. 3, p. 228, 1962.

25. H. Rindfleisch, "Die Grossbasis-Peilanlage Wullenwever" *Nachrtech. Z.*, **9**, p. 119, March 1956.

DISCUSSION

Under the chairmanship of Mr. W. K. Grimley, O.B.E.

Mr. J. Croney: These circular arrays at first sight seem attractive as a design for electronic scanning radar antennas. In this case, however, one would be also very interested in the way in which the main beam and side-lobe levels changed relatively not only in azimuth but in elevation. The optimizing process should therefore consider deterioration of the pattern in elevation, and this deterioration seems likely to be worse than for an array where all the radiators are in one line. Has any consideration been given to this?

Dr. Ziehm (*in reply*): Consideration has not been given to this point of view in the present paper in an explicit manner. But it can be seen from Fig. 4(a) and (b) and equation (3) that we are doing some investigations to realize a good quality of the radiation pattern with respect to the angle of elevation.

It should be mentioned that a similar method of synthesis to that described in the paper could be used for the case with all radiators in a line.

Radio Engineering Overseas . . .

The following abstracts are taken from Commonwealth, European and Asian journals received by the Institution's Library. Abstracts of papers published in American journals are not included because they are available in many other publications. Members who wish to consult any of the papers quoted should apply to the Librarian, giving full bibliographical details, i.e. title, author, journal and date, of the paper required. All papers are in the language of the country of origin of the journal unless otherwise stated. Translations cannot be supplied.

PROTECTION AGAINST INTERFERENCE IN SECAM COLOUR TELEVISION

An investigation has been carried out by a Czech engineer of the protection ratios against periodic signal interference in the r.f. channel of the SECAM colour television system. An experimental colour television equipment with a SECAM coding and decoding apparatus has been used for the investigation and the influence of interference in the reproduced picture judged by a set of 10 viewers. The protection ratio necessary against interference with a sine signal in the r.f. channel of the SECAM colour television system is 50 dB in the carrier wave region of the black-and-white picture and 44 dB in the region of the colour carrier wave.

"Protection ratios with periodic signal interference in the r.f. channel of the SECAM colour television system", J. Pazderák. *Slaboproudý Obzor (Prague)*, **25**, No. 4, pp. 219–23, April 1964.

AN ECONOMIC DESIGN OF SHIFT REGISTER

A recent German paper describes a shift register which uses one ferrite core per bit for storage and one capacitor for intermediate storage. It is shown how the stability can be considerably improved by the interposition of an inductance between core and capacitor and how this inductance and the data of the other circuit components must be chosen for a specific application. For this calculation a simple method is used which also allows calculation of the stability of the shift register.

"Calculation of the stability of shift registers with ferrite cores for storage", Herbert Stegmeier and Rolf Zscherpel. *Archiv der Elektrischen Übertragung*, **18**, No. 4, pp. 249–54, April 1964.

A LARGE MULTI-FREQUENCY ANTENNA

A huge parabolic antenna for over-the-horizon communications has been built in Japan recently. The antenna consists of a rectangular-aperture mirror of 16 m in height and 25 m width, together with a multiband primary-radiator system with dual polarization. The primary-radiator system is composed of three types of radiators. The first one is a square-aperture horn with both vertical and horizontal polarization in 2000 Mc/s band, the second is a sectoral horn array with vertical polarization in 800 Mc/s band, and the third is a half-wave-dipole array with horizontal polarization in 700 Mc/s band. The performance of the antenna satisfies the requirements for the telephone—and television—relay links in a long path of 341·5 km between Kagoshima and Amami-ōshima, which uses an intervening island as a diffracting obstacle. The antenna gains are approximately 39·5 dB in the 700 Mc/s band, 42·5 dB in the 800 Mc/s band and 50·5 dB in the 2000 Mc/s band.

"A huge parabolic antenna with dual polarization in multi-band for o/h communications", Takashi Kirsuregawa, Seibei Tachikawa and Yoshihiro Takeichi. *Mitsubishi Denki Laboratory Reports*, **5**, No. 2, pp. 223–37, April 1964.

INFORMATION CAPACITY OF HUMAN VISION

The channel capacity of vision under specified television viewing conditions has been calculated by an Australian engineer using his own theory on visual thresholds. A foveal capacity of $2·5 \times 10^5$ bits/second is found and arguments are advanced which indicate that this estimate is reasonable. The significance of the result to visual communications is discussed. The estimated visual capacity is about two orders less than the channel capacity required for the transmission of standard C.C.I.R. television signals. This would indicate that the television system is far from being informationally matched to its ultimate receiver. Mis-match of two types is indicated; the television system is capable of presenting data which cannot be seen by the viewer and some data which the viewer could see are not presented. Thus the study reaffirms known conclusions that there is scope in television communication research both for reducing the required channel capacity and for improving the apparent quality.

"On the channel capacity of the human sense of vision", Z. L. Budrikis. *Proceedings, The Institution of Radio and Electronics Engineers Australia*, **25**, pp. 228–34, April 1964.

LASER RESONATORS

Fabry-Pérot resonators not only play an important role in the realization of lasers, but also from their behaviour valuable conclusions may be drawn concerning the properties of optical transmission lines with periodic focusing elements. A German engineer has described a perturbation method of analysis which takes into consideration systematic imperfections that occur in practical mirrors and lenses. The efficiency of this method is shown by treating the resonators with mirrors or arbitrary radii of curvature, as perturbed confocal resonators with known properties. The diffraction losses of arbitrary resonators have been determined with satisfactory accuracy which improves with increasing diffraction losses.

"Calculation of Fabry-Pérot laser resonators by perturbation analysis", D. Gloge. *Archiv der Elektrischen Übertragung*, **18**, No. 3, pp. 197–203, March 1964.