FOUNDED 1925
INCORPORATED BY
ROYAL CHARTER 1961

"To promote the advancement of radio, electronics and kindred subjects by the exchange of information in these branches of engineering."

THE RADIO AND ELECTRONIC ENGINEER

The Journal of the Institution of Electronic and Radio Engineers

VOLUME 33 No. 4 APRIL 1967

Computers in Britain

SPEAKING recently at the annual luncheon of the Electronic Engineering Association, the chairman of the Association, Mr. R. J. Clayton, stated that Great Britain was probably the only country other than America with an independent computer industry designing, producing and delivering full ranges of computers which are sold in world markets. Much early original research on electronic computers was carried out in British universities in the years immediately after the war and applied to scientific problems, while the use of computers for business purposes was also pioneered in this country some thirteen years ago—papers describing these machines were read at an Institution Convention in 1954. The 'third generation' computers now in production, using integrated circuits, represent an advance as spectacular in a short time as can be found in any branch of technology. The day when computers can in effect be said to design their successors is near since computers are already used for circuit design.

The greatest man-power problem in the utilization of computers is in the work of the systems analyst and programmer, or—to use current jargon—the design of 'software'. As in the information field, where research is often undertaken in ignorance of the fact that it had already been done and documented, a considerable waste of effort can result by developing programs which have been already prepared elsewhere. The British computer industry has therefore taken a significant step in joining with the Ministry of Technology in setting-up the National Computing Centre which will make the future use of computers considerably more effective.

The National Computing Centre will not carry out computation as a service—its first main purpose is to make available information and advice about computer applications and programs in the fields of interest of its members who already represent a cross-section of the whole of British industry and commerce. Its second role is as an educational force in informing management of the potentialities for computers in their organizations, and in providing and encouraging training in systems analysis, programming principles and computer applications. Thirdly, the Centre will do research into programming and operating methods. To help it in its work, the Centre will initially have, at its headquarters in Manchester, two of the latest and most powerful computers produced by British firms. Although primarily catering for users in Great Britain, the Centre will co-operate with similar organizations overseas, particularly within the Commonwealth. In computing, as in most fields of scientific endeavour, free interchange between nations offers great advantages.

It is seldom that an inventor can see, or even foresee, the eventual fruits of his work and often many years elapse before it can be implemented fully. The father of the computer, for such is Charles Babbage generally agreed to be, would probably have been more astounded than most inventors at the immense industry which has resulted from his concepts and work. In view of his acknowledged contribution to the foundation of such an important British industry, it is strange that apparently the only identification of his name with this technology in Britain is in the Award which the I.E.R.E. makes annually for the outstanding paper on electronic computers published in its *Journal*.

F. W. S.

INSTITUTION NOTICES

Obituary

Council has learned with regret of the deaths of the following senior members of the Institution: Mr. V. J. Cooper, B.Sc.(Eng.), C.Eng. (Member), who was the first Chairman of the Television Group Committee.

Mr. G. R. Scott-Farnie, C.B.E., C.Eng. (Member), who was elected a member of Council in December 1966 and served on the Radar and Navigational Aids Group Committee.

Full obituary notices are published in the March–April issue of the Institution's *Proceedings*.

Conference on 'The Integration of Design and Production in the Electronics Industry'

Registration forms for this Conference, which will be held at the University of Nottingham from 10th to 13th July 1967, are now available from the Institution. The application form included at the back of this issue may be used to request forms and further information.

The registration charge for the Conference is £10 to members of the sponsoring Institutions (I.E.R.E., I.E.E., I.Prod.E.) and £13 to non-members. The charge for accommodation in the University's Halls of residence is £10 for three nights and includes the Conference Dinner on the evening of Monday, 10th July.

The programme for the Conference will be published in the May issue and synopses of papers will appear in the June issue of *The Radio and Electronic Engineer*.

Conference on 'R.F. Measurements and Standards'

The venue and date for the Joint I.E.R.E.-I.E.E. Conference on 'R.F. Measurements and Standards', which will be held near London in November 1967, will be confirmed very soon, and further details about the arrangements will be announced in due course.

Meanwhile, the Organizing Committee can still consider offers of papers for the Conference, which will cover the following aspects of measurements and standards in the frequency range 100 kHz to 3 GHz:

power; impedance; attenuation; noise;

voltage and current, including field strength.

(Frequency measurement as such will not be included.)

Offers of papers, which should be accompanied by synopses, should be sent to the Joint Conference Secretariat, I.E.R.E., 8–9 Bedford Square, London, W.C.1., by 31st May 1967.

Dinner of Council and Committees

The Eleventh Dinner of the Institution's Council and Committees will be held on Thursday, 15th June 1967 at the Savoy Hotel, London. The guest of honour will be the Immediate Past President, Colonel G. W. Raby, C.B.E.

Further details may be obtained on application to the Secretary, at 9 Bedford Square, London, W.C.1.

IFAC Pulse Symposium

An International Symposium on 'Pulse Rate and Pulse Number Signals in Automatic Control' will be held in Budapest from 9th to 11th April 1968. It is sponsored by the IFAC Components Committee, in co-operation with the Hungarian Academy of Sciences and the Scientific Society for Measurement and Automation.

The Organizing Committee have invited contributions to the Symposium in the field of automatic control systems and their components, involving the use of pulse-rate and pulse-number signals as applied to the following techniques:

digital differential analysers; incremental and digital-operational computing; designing procedures and associated hardware; possibilities and limitations of the various principles in automatic control applications; relative merits of the components and methods applied.

Short summaries of proposed papers should be submitted before 31st May 1967.

Further information is available from, and synopses of papers should be sent to: Organizing Committee, International Pulse Symposium, POB 63, Budapest 112, Hungary.

CORRECTIONS

The following corrections should be made to the paper 'The PAL Colour Television System', which was published in the March 1967 issue of *The Radio and Electronic Engineer:*

Page 148, Fig. 3: The duration of the colour burst should read 10 ± 1 cycle.

Page 151, The second expression in eqn. (12) should read: $2jS \sin \alpha$.

Page 157, Fig. 16: The first two bursts on line 2 (' 1. field') are inverted and the third burst should be omitted. On line 5 ('4. field'), the first three bursts should be inverted. On line 6 ('1. field') the third burst should be omitted.

Page 158, Appendix, Para. 4: The subcarrier burst duration should read: 10 cycles ± 1 cycle.

U.D.C. 621.396.677

Broadbanding of Concentric Planar Ring Arrays by Space Tapering

By

RUBEN DAS†

Summary: The paper discusses the application of concentric ring arrays as a new approach to space tapering for planar arrays. It describes the design of an array of non-uniformly spaced concentric rings of non-uniform density containing discrete elements. The array can be operated over a large frequency range and wide scan-angles. The ring array was designed to simulate the performance of a Taylor-weighted circular aperture. An array having a diameter of 34 wavelengths was designed with ten concentric rings containing a total of 759 elements and a thinning factor of 0·21. This array has a side-lobe level of −20 dB or better over a range of sin θ up to 2·5 at mid-band. Taking a dielectric filled square-guide as the radiating element, the mutual coupling was calculated. The element spacing is 0·498 λ_{mid} for the innermost and 1·028 λ_{mid} for the outermost ring. The mutual coupling does not produce any appreciable deleterious effect on the array-factor. For the scan angle of $\theta_0 = 60°$ and $\phi_0 = 0°$, the array has a side-lobe level better than −20 dB for a frequency range of 0·75 λ_{mid} to 1·25 λ_{mid}.

List of Symbols

a	Radius of aperture, broad dimension of waveguide, function defined in eqn. (16)
A(dB)	array-factor in dB
A_K	total number of elements in the Kth ring
b	narrow dimension of waveguide
C, C_{KL}	coupling coefficient defined in Section 3.1
D	diameter of aperture
E	array factor
E_1, E_0	electric field as defined in Appendix
E_θ, E_ϕ	electric field, elevation and azimuthal components respectively
J_0	Bessel function of zero order
k	phase constant in the free space $= \frac{2\pi}{\lambda_0}$
K	number of ring
l	length along the circumference Fig. 16
L	number of element
r	variable radius of ring, Fig. 16
R	distance of field point
R_K	radius of the Kth ring
W_r	power received by the waveguide
W_t	power that can be transmitted by the radiator
x	co-ordinate of the abscissa, Fig. 16
α, α_{KL}	angle defined in Fig. 20, phase angle of elements
β	phase constant $= \frac{2\pi}{\lambda}$
β_{10}	phase constant in the guide $= \frac{2\pi}{\lambda_g}$
Γ	reflection coefficient of the radiating guide
σ	beam broadening factor defined in Taylor's paper[11]
δ	argument of Bessel function, function defined in eqn. (17)
Ψ	angle defined in Fig. 16
θ	elevation angle
ϕ	azimuthal angle
θ_0	elevation scan angle
ϕ_0	azimuthal scan angle
λ, λ_0	free-space wavelength
λ_g	wavelength in the guide
μ	relative permeability of the medium
ε	relative dielectric constant of the medium

† Formerly at Syracuse University Research Station, Syracuse, New York; now with Loral Electronic Systems, The Bronx, New York.

1. Introduction

A number of authors[1-15] have studied various aspects of ring arrays. Many of these investigations [1, 3, 5-9] were concerned with the azimuthal radiation patterns. Experiments of Neff and Tillman[1] with two rings of diameters of 0·318λ and 0·955λ containing altogether 13 dipoles with one at the centre gave a side-lobe level of −16 dB, rather than −20 dB

as hoped. Tillman *et al.*[6] made a rigorous analysis of currents, impedances and azimuthal patterns of a single-ring array containing equally-spaced axial dipoles. A ring of radius $0{\cdot}318\lambda$ with six elements produced a -20 dB side-lobe level. Knudsen[8] considered the case of equally-spaced elements having equal amplitudes, placed along a circle of radius from $0{\cdot}125\lambda$ to 2λ, the number of elements varied from 5 to 9 and was primarily concerned with the radiation patterns of a ring containing odd or even number of elements as compared to that of a continuous ring. Knudsen[7] also calculated the characteristics of arrays of small radius, i.e. $k \times a \ll 1$, and having identical tangential or radial dipoles but not identically located.[7] Mack *et al.*[3] made a theoretical analysis and carried out refined experimental measurements of the circuit and azimuthal radiation properties of circular arrays of thin dipoles over a wide range of element lengths, diameters, spacings, etc. Azimuthal radiation patterns were measured for a total number of equally-spaced elements of 2, 3, 4 or 5 in the array and a range of element spacings from $0{\cdot}25\lambda$ to λ. The linear radiators had a half length of $0{\cdot}25\lambda$ to $0{\cdot}625\lambda$. The beam width increased with increasing length of radiators. James[9] considered the polar response of an array of similar radiators placed uniformly around a circle. The side-lobes were reduced by using either directional elements or by non-uniform weighting of the elements or both. James[9] recommended an element spacing of $\lambda/2$ or less for suppressing the grating lobes.

Stearn and Stewart[5] made an analysis of azimuthal radiation patterns of circular arrays with diameters of 2 to 9λ containing 3 to 18 concentric circles, the current amplitudes of which are obtained from Chebyshev radiation pattern function. The elements of each circle have equal current amplitude. A side-lobe level of -20, -30 or -40 dB could be designed.

Nelson[10] analysed the elevation and azimuthal patterns of ring arrays having radii up to 1000λ using the J_0 approximation. The side-lobe levels of -8 dB, $-12{\cdot}3$ dB and $-15{\cdot}5$ dB were obtained with a single, two and three rings respectively within $1{\cdot}4°$ in elevation or azimuth. Schroeder[4] made analysis and synthesis of ring arrays to reduce both azimuthal and elevation side-lobes. The radiation patterns of two five-ring arrays of diameter 20λ and having 755 and 189 elements with $0{\cdot}25\lambda$ and λ element spacing respectively, show side-lobes of -16 dB and -14 dB at several places. The 'star' configuration, which is complicated from the point of view of scanning, shows a greater uniformity of the side-lobe levels. For a ten-ring star array of diameter 10λ the worst side-lobe is -17 dB at mid-band.

The grating lobes of an array consisting of one ring are of considerably smaller magnitude compared to those of a linear array. This study[16, 17] was undertaken to find out the feasibility of using an array of concentric rings containing discrete elements over a large frequency range and wide scan angles with a low side-lobe level. Effects of both elevation and azimuthal scanning on radiation patterns are investigated. No amplitude taper was studied as it is considered to be comparatively complicated in practice. The principle of space tapering is used to reduce the side-lobe level. (Lo and Lee[59] gave a review of the work on space-tapered arrays.)

2. Design

It is, initially, assumed that:

(i) all the antenna elements are isotropic, each having the same weight,

(ii) a ring of discrete antenna elements performs like a continuous ring, and

(iii) there is no mutual coupling between the elements.

The design goal was a 2° pencil-beam with a side-lobe level better than -20 dB over the operating region. The starting point was a Taylor-weighted circular aperture to give a pencil-beam of 2° and a side-lobe level of -30 dB. Use was made of the Tables derived by Hansen.[12] Referring to these, from Table I,

$$2a/(\sigma\lambda) = \frac{60{\cdot}55}{2} = 30{\cdot}275$$

From Tables I and II, a suitable value of $\bar{n}$ is 5. So $2a/\lambda = (30{\cdot}275)\ (1{\cdot}1180) \simeq 33{\cdot}8$ and an aperture diameter of 34 wavelengths is required. The aperture distribution for a Taylor-weighted circular aperture is plotted from Table III and Fig. 1. It is desired to realize this distribution with the help of a limited number of rings containing discrete antennae to give us approximately the result that could be obtained with a continuous circular aperture. The Taylor-weighted distribution function of a continuous circular aperture is shown in Fig. 1. The distribution function of a ring is on a circular line whereas the Taylor distribution function of a circular aperture is confined to a plane. One way of comparing the two distribution functions will be to project them both on a straight line. As both the circular aperture and the ring are circularly symmetrical, only one quadrant of the space will be considered. The Taylor distribution function is projected on a line in the following way. For each radius in the quadrant, the ordinate is divided into several small equal segments, the larger the number of segments, the more accurate will be the projection. The distance of each of the segments is measured from the origin, i.e. the centre of the circle, and the distribution function for each of these distances is found

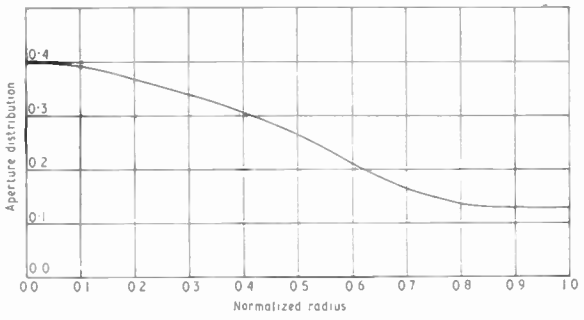


Fig. 1. Circular aperture Taylor-weighted distribution for −30 dB side-lobe, $\bar{n} = 5$.

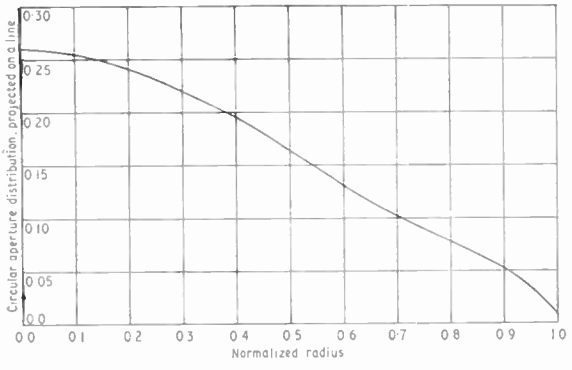


Fig. 2. Circular aperture Taylor-weighted distribution for −30 dB side-lobe, projected on a line, $\bar{n} = 5$.

from Fig. 1. When all these distribution functions are added and suitably normalized, the distribution function projected on a line is obtained for this particular radius. The Taylor distribution function, projected on a line, is plotted in Fig. 2. The distribution function of a continuous ring, when projected on a line is given in the Appendix (Section 7).

$$\frac{dl}{dx} = \frac{r^2}{\sqrt{r^2 - x^2}} \qquad \text{.....(1)}$$

and

$$l = r \sin^{-1} \frac{x}{r} \qquad \text{.....(2)}$$

Equation (1) is plotted in Fig. 3. The normalized area under this curve is $l = \pi r/2$. For half a wavelength separation between the elements of a ring, the total number of elements in the Kth ring is

$$A_K = (2)(2\pi R_K) = 4\pi R_K$$

where $K = 1, 2, 3 \ldots$.

The total number of elements spaced half a wavelength apart in the outermost ring is (2) (34π) $\simeq$ 214. The ordinate of Fig. 3 at $r = 0$ is equal to the weight per unit length of the ring.

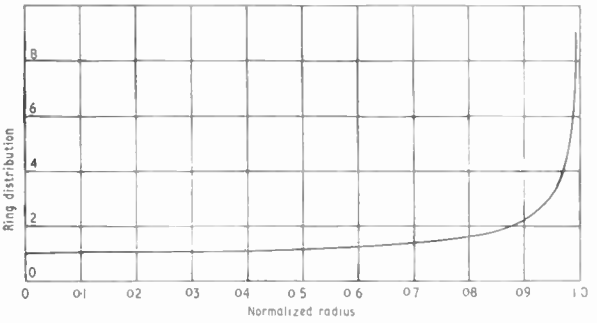


Fig. 3. Continuous ring distribution projected on a line.

2.1 *Rings of Uniform Density*

The area under the curve of Fig. 2 is equal to the total number of elements in the array. The object was to simulate the above-mentioned area with rings of antennae. Figure 2 is re-plotted in Fig. 4(a), the ordinate scale being suitably labelled for convenient plotting of a number of curves each given by eqn. (1). The first (outermost) ring has a radius of 17 wavelengths. The plot of its distribution function is given by bc P...def (Fig. 4). The ordinate of this curve at $r = 0$ is $216/(\pi/2) = 137$. The area under this curve is equal to the area under the lower portion of the curve of Fig. 4(a) except for a positive error given by P...deg. To cancel this, a negative error PQc of equal magnitude is added and thus the position of the second ring is obtained at Q. The process was repeated assuming antenna spacings to be half a wavelength in all the rings. Each time the distribution function of a ring overshoots that of the circular aperture giving a positive error, such as MN, exactly equal negative errors like XL, are added fixing the positions of the rings at e, Q, R, S, T, U, V and W. The corresponding radii, expressed in wavelengths, are 17, 11·1, 9·15, 7·3, 5·79, 4·46, 3·15 and 1·87.

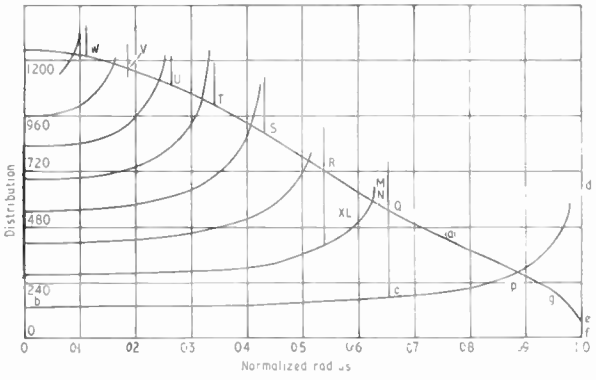


Fig. 4. Simulation of circular aperture with rings of uniform density.

2.2 *Rings of Non-uniform Density*

This procedure gave large errors especially for the first ring. The side-lobe level increases with the error. In addition, non-uniformly spaced arrays result in better side-lobe levels[18] and increased bandwidth.[19] With rings of non-uniform density, the weight and as such the ordinates (Fig. 4) for the distribution function at $r = 0$ will have different values for different rings.

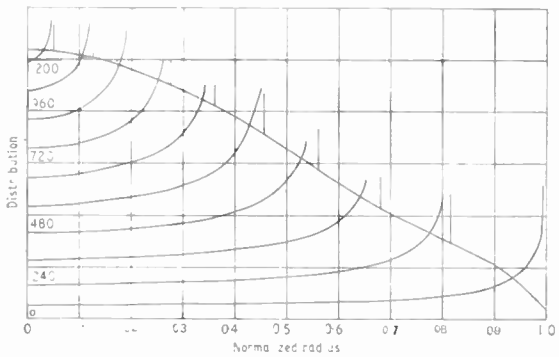


Fig. 5. Simulation of circular aperture with rings of non-uniform density.

The spacings of antennas were selected to be $0{\cdot}5\lambda$ for the innermost and λ for the outermost rings with intermediate values for the inner rings. Because of the reduced density of the outer rings, a larger number of rings were required to keep the number of elements approximately the same as before. The design is shown in Fig. 5. The radii of the rings, as determined from Fig. 5, are 17, 13·85, 11·55, 9·51, 7·74, 6·13, 4·67, 3·4, 2·13 and 0·85 and the corresponding number of elements are 107, 122, 117, 109, 93, 75, 58, 43, 27 and 11 respectively.

The array-factor of a ring array is given by (see Appendix 2):

$$E = \sum_{K=1}^{10} \sum_{L=1}^{A_K} \exp\left[j\beta R_K \times \{\sin\theta\cos(\phi-\alpha_{KL}) - \sin\theta_0\cos(\phi_0-\alpha_{KL})\}\right] \quad \ldots\ldots(3)$$

or

$$E = \sum_{K=1}^{10} \sum_{m=-\infty}^{\infty} A_K (j)^{-mA_K} \exp(-jmA_K\delta) \times J_{-mA_K}(\beta R_K \delta) \quad \ldots\ldots(4)$$

The first term ($m = 0$) of eqn. (4) gives the pattern of a continuous ring. The remaining terms are correction terms of higher order Bessel functions. If the number of elements is large, the first correction term ($m = 1$) alone will be important. The higher-order Bessel functions introduce lobes like grating lobes of a linear array. The first grating lobe of a ring can be self-compensated in the following way.

A higher-order Bessel function has its first maximum when the order is nearly equal to its argument, i.e. when the argument is equal to the number of elements in that ring. For large arguments $J_0(\delta)$ can be replaced

Table 1

Ring number	Ring radius in wavelength	Number of elements	Spacing of elements in wavelengths	Spacing between radii in wavelengths
1	17	104	1·027	3·20
2	13·85	122	0·713	2·30
3	11·55	116	0·625	2·04
4	9·51	110	0·543	1·81
5	7·70	91	0·531	1·57
6	6·13	73	0·527	1·39
7	4·74	59	0·505	1·38
8	3·36	42	0·503	1·13
9	2·23	28	0·500	1·12
10	1·11	14	0·498	

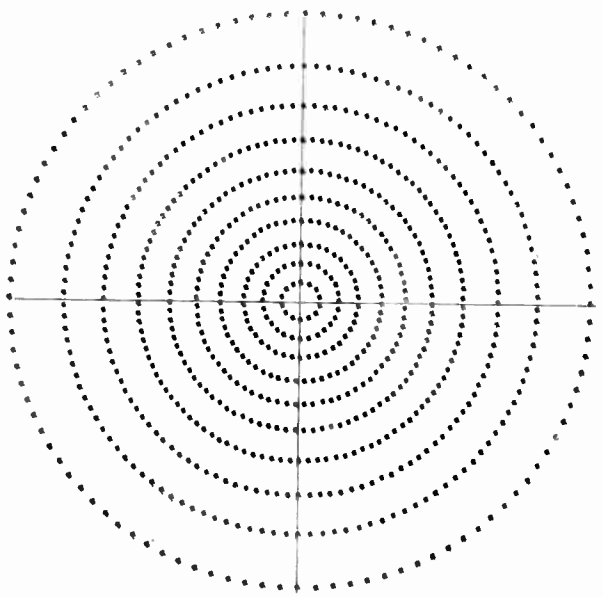


Fig. 6. Physical plot of the array containing ten concentric rings.

No. of elements:	104	122	116	110	91	73
radii:	17·0	13·85	11·55	9·51	7·70	6·13
No. of elements (cont.):	59	42	28	14		
No. of radii (cont.):	4·74	3·36	2·23	1·11		

The number of elements and the corresponding value for the radius given above are applicable to all figures from Fig. 6 to 10 and from Fig. 14 to 17.

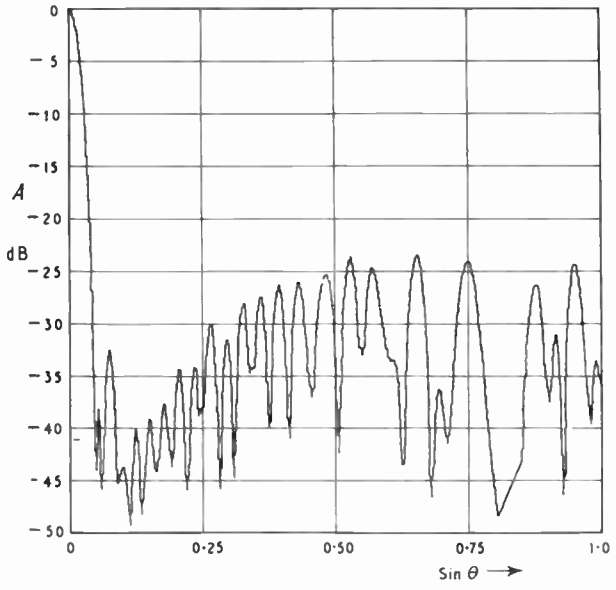


Fig. 7. Array-factor of a 10-ring array as a function of sin θ (J_0 approximation), beam pointing boresight.

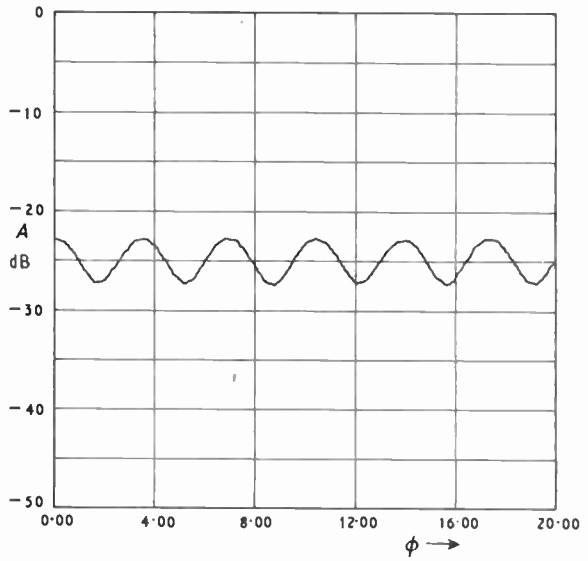


Fig. 9. Array-factor of a 10-ring array, beam pointing boresight.

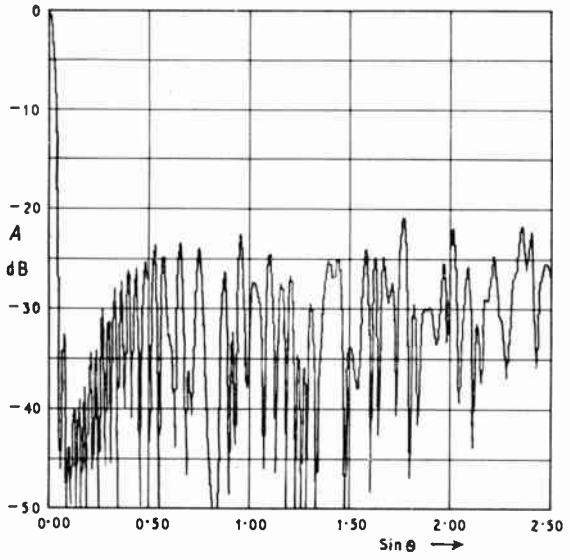


Fig. 8. Array-factor of a 10-ring array as a function of sin θ (element summing method), beam pointing boresight.

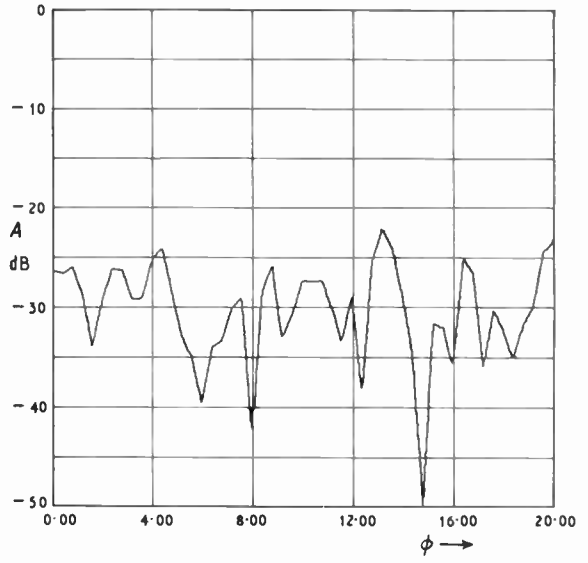


Fig. 10. Array-factor of a 10-ring array, beam pointing boresight.

by cos δ. If the first maximum of the higher-order Bessel function coincides with a negative peak of $J_0(\delta)$, then the first grating lobe, due to the ring under consideration, could be minimized. This means that the number of elements in a ring, for optimum grating lobe conditions, be an odd multiple of π. This argument is valid mainly for the inner rings as the peak of a Bessel function of a large order is fairly broad. The modified designed values of Fig. 5, as outlined above, are given in Table 1. Figure 6 shows a physical plot of the array.

2.3 *Results of the Array of Ideal Elements*

The array factor is calculated using only the first term of eqn. (4) and is plotted in Fig. 7 as a function of sin θ in steps of 0·005. The array factor is also calculated by the element summing method (eqn. (3)) and the results are shown in Figs. 8–10. From Figs. 7 and 8, it is found that for values of sin θ up to 0·72 the ring array of discrete elements behaves like a continuous aperture. The array-factor of the array as a function of ϕ in steps of 0·4° was studied for a

number[16] of values of $\sin\theta$, some of which are presented in Figs. 9 and 10. The side-lobe level is better than -20 dB over the region investigated.

3. Effect of Mutual Coupling

The effect of the mutual coupling between individual antennas as well as in an array have been studied by a number of investigators.[20-48] A large number of these investigations are experimental[43, 46] with theoretical studies limited to simple cases like dipoles,[28] rectangular slots[33] or uniformly-spaced radiators[40, 41] in a large array. The elements in the array under consideration are not uniformly spaced.

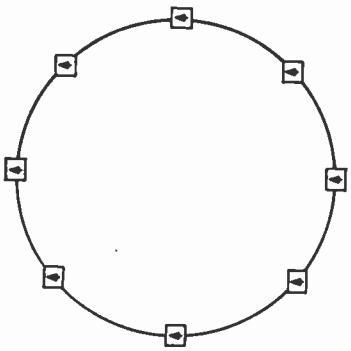


Fig. 11. Ring array with square waveguide radiators (Y-polarization).

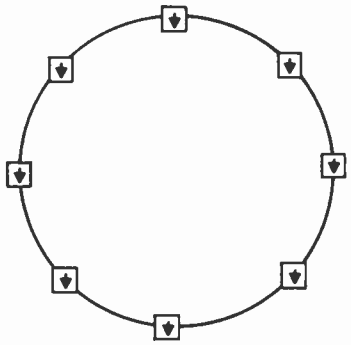


Fig. 12. Ring array with square waveguide radiators (X-polarization).

To obtain an idea of the effect of the mutual coupling on the array-factor, some realistic type of element was necessary. The element is required to operate over a large frequency range and a broad beamwidth. Moreover, they are required to have polarization diversity. Rectangular waveguide radiators[51] and dielectric-filled circular waveguide radiators[49, 50] fulfill some of the criteria. Metter[51] mentions that 'An X-band horn antenna has been developed that can be used whenever there is a broad beamwidth requirement with either horizontal, vertical or elliptical polarization'. For simplicity of analysis of the mutual coupling and the element pattern, the following assumptions will be made:

(a) the radiator is a square waveguide completely filled with a dielectric,

(b) a ground plane is located at the mouth of the radiating guides,

(c) there is no mismatch between the radiator and the radiating space, i.e. the reflection coefficient Γ is zero,

(d) the field at the aperture of all the radiators is plane polarized, let us call this Y-polarized (Fig. 11),

(e) mutual coupling is present only if the incident field at an aperture due to the radiation from another radiator is Y-polarized (Fig. 11). X-polarized (Fig. 12) field is completely absorbed inside the guide,

(f) any radiator is situated in the far field region of all the other radiators.

3.1 *Array-factor with Mutual Coupling*

When scanned to a direction θ_0, ϕ_0, the coupling coefficient C is given by (see Appendix, Sect. 7.3)

$$C = B \exp\left(\frac{-j2\pi R}{\lambda_0}\right) \times \exp\left[\frac{-j2\pi R_M}{\lambda_0} \sin\theta_0 \cos(\phi_0 - \alpha_{MN})\right]$$
$$= P_{MN} - jQ_{MN}, \text{ say} \qquad \ldots\ldots(5)$$

The total coupling to the Lth element by all other elements of the ring is

$$C_{KL} = \sum (P_{MN} - jQ_{MN}) \qquad \ldots\ldots(6)$$

The summation extends over all the elements of the array except the Lth element of the Kth ring.

In the absence of the mutual coupling, the weight of each element was taken as unity. The weight of each element in the presence of the mutual coupling is $(1 + C_{KL})$. The element pattern will modify the far-field array-factor given by eqn. (3). The array factor of a ring array, taking the mutual coupling and the element pattern into account, is given by

$$E = [E(\theta, \phi, \theta_0, \phi_0)][E_{\theta,\phi}] \qquad \ldots\ldots(7)$$

where

$$E(\theta, \phi, \theta_0, \phi_0) = \sum_{K=1}^{10} \sum_{L=1}^{A_K} (1 + C_{KL}) \times \exp\left[j\beta R_K \{\sin\theta \cos(\phi - \alpha_{KL}) - \sin\theta_0 \cos(\phi_0 - \alpha_{KL})\}\right] \qquad \ldots\ldots(8)$$

and $E_{\theta,\phi}$ = element pattern.

(a) E-plane array-factor $\phi = \frac{\pi}{2}$

$$E = \left[E\left(\theta, \phi = \frac{\pi}{2}, \theta_0, \phi_0\right)\right]\left[1 + \frac{\lambda_0}{\lambda_g}\cos\theta\right] \times \times \left[\frac{\sin\left(\frac{\pi a}{\lambda_0}\sin\theta\right)}{\left(\frac{\pi a}{\lambda_0}\sin\theta\right)}\right] \quad \ldots\ldots(9)$$

(b) H-plane array-factor $\phi = 0$,

$$E = [E(\theta, \phi = 0, \theta_0, \phi_0)]\left[\cos\theta + \frac{\lambda_0}{\lambda_g}\right] \times \times \left[\frac{\cos\left(\frac{\pi a}{\lambda_0}\sin\theta\right)}{\left(\frac{\pi a}{\lambda_0}\sin\theta\right)^2 - \frac{\pi^2}{4}}\right] \quad \ldots\ldots(10)$$

The array-factor in decibels is given by

$$A(\text{dB}) = 10\log_{10}\frac{E^2}{E_0^2} \quad \ldots\ldots(11)$$

where E_0 is the value of E for $\theta = \phi = 0$

3.2 *Results*

3.2.1 Mutual coupling

With the designed values of the ring array given in Table 1, the coupling coefficient for each element was calculated using eqn. (6) for Y-polarization,

$$\lambda = \lambda_{\text{mid}}/1{\cdot}33, \theta_0 = 60^\circ$$

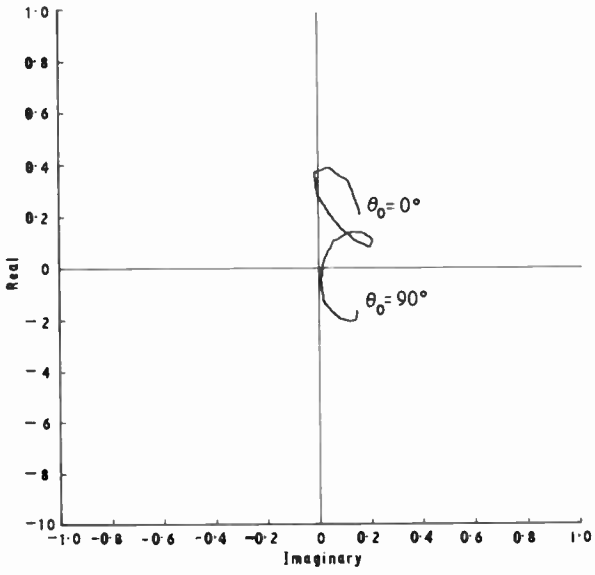


Fig. 13. Coupling coefficients as a function of scan angle θ_0 for element no. 10 of ring no. 10, $\phi_0 = 0°$, $\lambda = \lambda_{\text{mid}}/1{\cdot}33$.

The length of each side of the square guide is $0{\cdot}45\lambda_{\text{mid}}$. The coupling due to ten elements on either side of the element under consideration on the same, and two adjacent rings were taken into account, i.e. the effect of approximately sixty elements was considered. For the innermost and outermost rings, the coupling due to forty elements was considered. Figure 13 shows the coupling coefficient of an element as a function of the scan angle $\theta_0(\phi_0 = 0°)$.

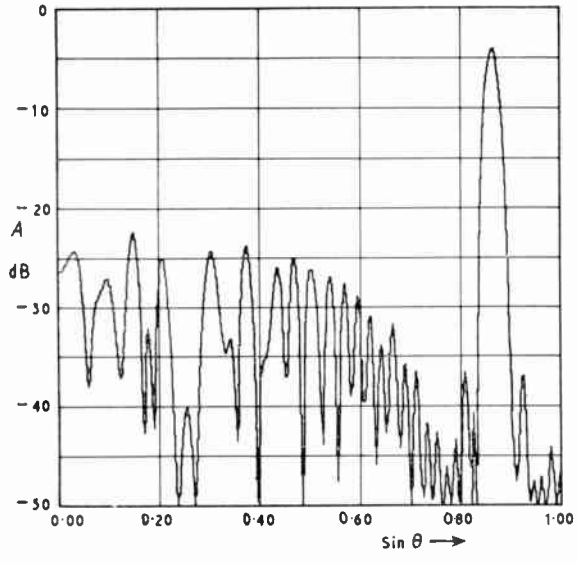


Fig. 14. Array-factor of a 10-ring array with mutual coupling and element pattern taken into account. Beam pointing $\theta_0 = 60°$, $\phi_0 = 0°$, $\lambda = \lambda_{\text{mid}}/1{\cdot}33$, $\phi = 0{\cdot}000$.

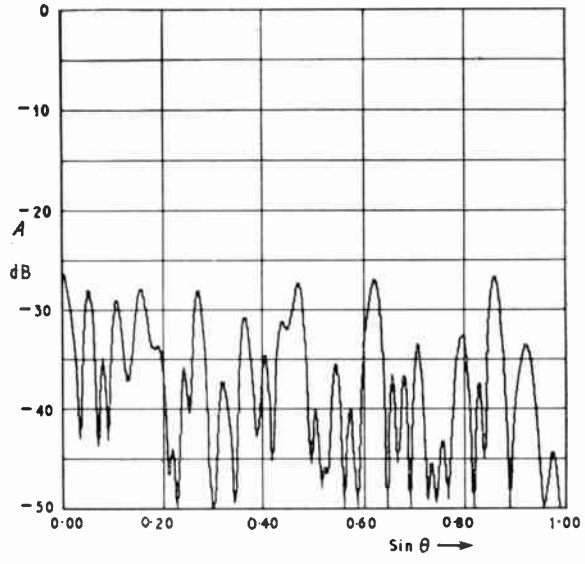


Fig. 15. Array-factor of a 10-ring array with mutual coupling and element pattern taken into account. Beam pointing $\theta_0 = 60°$, $\phi_0 = 0°$, $\lambda = \lambda_{\text{mid}}/1{\cdot}33$, $\phi = 90{\cdot}000$.

3.2.2 Array pattern with mutual coupling

Figures 14 and 15 show the array-factor with mutual coupling and element pattern taking into account for $\lambda = \lambda_{mid}/1{\cdot}33$, $\theta_0 = 60°$. Figures 16 and 17 show the array-factor for $\lambda = 1{\cdot}25\lambda_{mid}$, $\theta_0 = 60°$.

4. Conclusions

The array, with isotropic radiators, has a side-lobe level of -20 dB or better over a range of $\sin\theta$ up to 2·5 at mid-band (Fig. 8). It is interesting to compare Figs. 7 and 8 and to find that the array-factor of rings with discrete elements is the same as that of continuous rings for $\sin\theta$ up to 0·72 or $\theta = 46°$. Several azimuthal patterns were studied,[16] and the side-lobes were found to be no worse than -20 dB over the region investigated.

The far-field expressions (eqns. (3) and (4)) for a rectangular waveguide radiator neglect higher-mode fields in the aperture and the tangential component of the magnetic field on the outside surface of the waveguide. The values of the mutual coupling coefficients, as a result, are approximate. By and large, the coupling coefficients are small. One of the highest values is 0·4 for the innermost ring (Fig. 13).

The side-lobe levels (Figs. 14–17) of the array pattern, with mutual coupling taking into account for beam pointed at $\theta_0 = 60°$, $\phi_0 = 0°$ are better than -20 dB for elevation angles from 0° to 90° and for azimuthal angles of 0° and 90° for a frequency range of $0{\cdot}75\lambda_{mid}$ to $1{\cdot}25\lambda_{mid}$. There is no evidence of any serious deleterious effect on the array pattern due to the mutual coupling. The reduction of the gain of the main-beam (Figs. 14 and 16) is practically due to the reduction of the gain of the element at that value of the elevation angle. The results reported here were computed with the help of an IBM 1620 computer.

A total number of $(4\pi)\times(17^2) = 3630$ elements of half a wavelength spacing can be accommodated in the area having a radius of 17 wavelengths. The total number of elements used is (Table 1) 759. So the array is thinned by a factor of $759/3630 = 0{\cdot}21$.

Willey[60] designed a 900-element space-tapered circular planar array having 19·9% space taper and with a maximum side-lobe level of -24 dB for bore sight pattern, showing agreement with the calculated value of $-24{\cdot}5$ dB. He mentioned, qualitatively, the degradation of the side-lobe level with scanning.

It will be profitable to make a more detailed study to design a ring array with a predicted side-lobe level.

5. Acknowledgment

It is a pleasure to acknowledge valuable and profitable discussions with Messrs. Paul W. Howells, Carl W. Gerst, Sidney P. Applebaum and Hugh A. Hair.

The work was supported by Advanced Research Project Agency and Rome Air Development Centre under Contract No. AF 30(602)–3523.

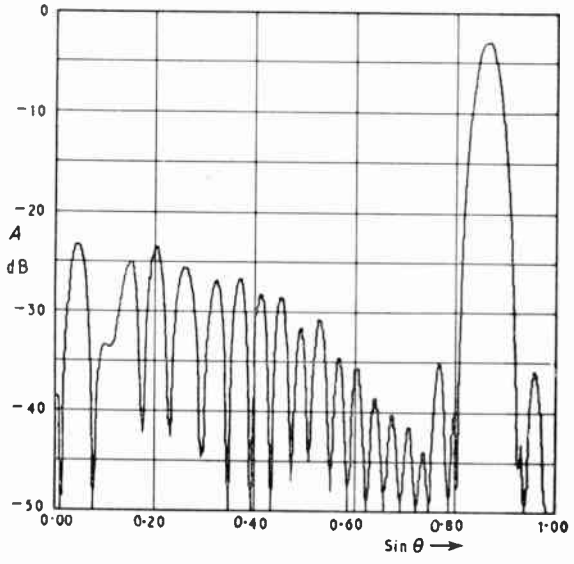


Fig. 16. Array pattern of a 10-ring array with mutual coupling and element pattern taken into account. Beam pointing $\theta_0 = 60°$, $\phi_0 = 0°$, $\lambda = 1{\cdot}25\lambda_{mid}$, $\phi = 0{\cdot}000$.

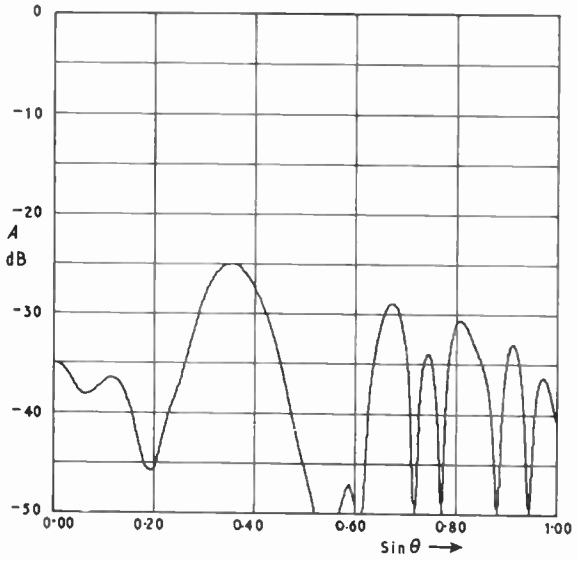


Fig. 17. Array pattern of a 10-ring array with mutual coupling and element pattern taken into account. Beam pointing $\theta_0 = 60°$, $\phi_0 = 0°$, $\lambda = 1{\cdot}25\lambda_{mid}$, $\phi = 90{\cdot}000$.

6. References

1. H. P. Neff and J. D. Tillman, 'An Electronically Scanned Circular Antenna Array', Inst. Radio Engrs: International Convention Record, Part 1, pp. 41–47, 1960.
2. J. D. Tillman, W. T. Patton, C. E. Blakely and F. V. Schultz, 'The use of a ring array as a skip range antenna', *Proc. Inst. Radio Engrs*, **43**, pp. 1655–1660, November 1955.
3. Richard B. Mack, 'A Study of Circular Arrays'; T.R. No. 381, 'Experimental Equipment'; T.R. No. 382, 'Self and Mutual Admittances'; T.R. No. 386, 'Plan for Practical Application to Arrays of Twenty or Fewer Elements', May 1 1963, Cruft Laboratory, Harvard University, Massachusetts.
4. K. G. Schroeder, 'Analysis and Synthesis of Ring Array Elements', TIS R62 EMH55, December 1962, HMED, General Electric, Syracuse, New York.
5. C. O. Stearns and A. C. Stewart, 'An investigation of concentric ring antennas with low sidelobes', *Trans. Inst. Elect. Electronics Engrs on Antennas and Propagation*, **AP-13**, No. 6, November 1965.
6. C. E. Hickman, H. P. Neff and J. D. Tillman, 'The theory of a single-ring circular antenna array', *Trans. Amer. Inst. Elect. Engrs*, **80**, Part 1, pp. 110–115, May 1961.
7. H. L. Knudsen, 'Radiation from ring quasi-arrays', *Trans. I.R.E.*, **AP-4**, No. 3, pp. 452–472, July 1956.
8. H. L. Knudsen, 'The necessary number of elements in a directional ring aerial', *J. Appl. Phys.* **22**, pp. 1299–1306, November 1951.
9. P. W. James, 'Polar patterns of phase-corrected circular arrays', *Proc. Instn Elect. Engrs*, **112**, No. 10, pp. 1839–1848, October 1965.
10. W. D. Nelson, 'Ring arrays', *Proc. of I.R.E. Convention on Military Electronics*, pp. 153–156, June 1960.
11. T. T. Taylor, 'Design of circular apertures for narrow beamwidth and low sidelobes', *Trans. I.R.E.*, **AP-8**, pp. 17–22, January 1960.
12. R. C. Hansen, 'Tables of Taylor distributions for circular aperture antennas', *Trans. I.R.E.*, **AP-8**, pp. 23–26, January 1960.
13. B. S. McCartney, 'Proposals for an electronically scanned circular array', *Proc. I.E.E.*, **110**, No. 7, pp. 1220–1222, July 1963 (Letters).
14. R. G. Fenby, 'Limitations on directional patterns of phase-compensated circular arrays', *The Radio and Electronic Engineer*, **30**, pp. 206–222, October 1965.
15. Ta-Shing Chu, 'On the use of a uniform circular arrays to obtain omni-directional patterns', *Trans. I.R.E.*, **AP-7**, pp. 436–438, October 1959.
16. Advanced ECCM Techniques, Quarterly Technical Reports, 1st July 1965 through 30th September 1965 and 1st October 1965 through 31st December 1965. Syracuse University Research Corporation, Syracuse, New York. (Classified.)
17. Ruben Das, 'Concentric ring array', *Trans. I.E.E.E.*, **AP-14**, No. 3, pp. 398–400, May 1966 (Letters).
18. C. H. Tang, 'On a design method for non-uniformly spaced arrays', *Trans. I.E.E.E.*, **AP-13**, pp. 642–643, July 1965.
19. A. Ishimaru and Y. S. Chen, 'Thinning and broadbanding antenna arrays by unequal spacings', *Trans. I.E.E.E.*, **AP-13**, pp. 34–42, January 1965.
20. S. J. Rabinowitz, 'The Conductance of a Slot in an Array Antenna', Technical Report No. 192(U), Lincoln Laboratory, M.I.T., DDC 209908, 31st December 1958.
21. J. L. Allen, 'Gain and impedance variation in scanned dipole arrays', *Trans. I.R.E.*, **AP-10**, pp. 566–572, 1962.
22. S. Edelberg and A. A. Oliner, 'Mutual coupling effects in large antenna arrays II: compensation effects', *Trans. I.R.E.*, **AP-8**, pp. 360–367, 1960.
23. L. Stark, 'Radiation Impedance of a Dipole in an Infinite Array', Formal Technical Document, Hughes Aircraft Company, Fullerton, California, May 1960.
24. H. A. Wheeler, 'The radiation resistance of an antenna in an infinite array or waveguide', *Proc. I.R.E.*, 36, No. 4, pp. 478–487, 1958.
25. P. W. Hannan, 'The Element-Gain Paradox for a Phased Array Antenna', Report 1161, Wheeler Laboratories, Great Neck, New York, 25th June 1963.
26. W. E. Rupp, 'Coupled Energy as a Controlling Factor in the Radiation Patterns of Broad-Side Arrays', Abstract of the Eleventh Annual Symposium on U.S.A.F. Antenna Research and Development Program. University of Illinois, 16–20th October 1961.
27. L. I. Parad, 'Mutual Effects between Circularly Polarized Elements', 12th Annual Symposium U.S.A.F. Antenna Research and Development Program, October 1962.
28. P. S. Carter, Jr., 'Mutual impedance effects in large beam scanning array', *Trans. I.R.E.*, **AP-8**, pp. 276–285, May 1960.
29. L. A. Kurtz and R. S. Eliott, 'The effect of mutual impedance in electronically scanned antenna arrays', *Proc. Electronic Scanning Symposium*, April–May 1958, ASTIA Document 152409.
30. L. A. Kurtz *et al.*, 'Mutual coupling effects in scanning dipole arrays', *Trans. I.R.E.*, **AP-9**, pp. 433–443, September 1961.
31. H. C. Baker *et al.*, 'Digital computation of the mutual impedance between thin dipoles', *Trans. I.R.E.*, **AP-10**, pp. 172–178, March 1962.
32. R. M. Kalafus, Y. K. Kwon and J. A. M. Lyon, 'Coupling Effects with Slot and Spiral Antennas', Abstract of the Thirteenth Annual Symposium on U.S.A.F. Antenna Research and Development Program, October 1963.
33. D. K. Adams, R. B. Harris, Y. K. Kwon and J. A. M. Lyon, 'The Interference Levels of Rectangular Slot Antennas in a Common Ground Plane by Simplified Analyses and Tests', Twelfth Annual Symposium on U.S.A.F. Antenna Research and Development, Monticello, Illinois, 1962.
34. A. R. Stratoti, 'An investigation of the complex mutual impedance between short helical array elements', *Trans. I.R.E.*, **AP-7**, pp. 279–280, July 1959.
35. P. S. Carter, 'Circuit relations in radiating systems and applications to antenna problems', *Proc. I.R.E.*, **20**, pp. 1004–1041, June 1932.
36. C. R. Cox, 'Mutual impedance between vertical antennas of unequal heights', *Proc. I.R.E.*, **35**, pp. 1367–1370, November 1947.
37. W. E. Rupp, 'Coupled Energy as a Controlling Factor in the Radiation Patterns of Broad-Side Arrays', Abstracts of the Eleventh Annual Symposium on the U.S.A.F. Antenna Research and Development Program, pp. 1–13, 1961.
38. John D. Dyson, 'The Conical Log-Spiral Antenna in Simple Arrays', Abstracts of the Eleventh Annual Symposium on the U.S.A.F. Antenna Research and Development Program, 1961.

39. H. E. King, 'Mutual impedance of unequal length antennas in echelon', *Trans. I.R.E.*, **AP-5**, pp. 306–313, July 1957.
40. S. Edelberg and A. A. Oliner, 'Mutual coupling effects in large antenna arrays: Part I, Slot arrays', *Trans. I.R.E.*, **AP-8**, pp. 286–297, May 1960.
41. Harold A. Wheeler, 'Simple relations derived from a phased-array antenna made of an infinite current sheet', *Trans. I.E.E.E.*, **AP-13**, pp. 506–514, July 1965.
42. P. W. Hannan, D. S. Lerner and G. H. Knittel, 'Wide-Angle Impedance Matching Calculated for a Phased Array Antenna', G-AP Symposium Digest, pp. 228–233, 1963.
43. P. W. Hannan, P. J. Mier and M. A. Balfour, 'Simulation of phased array antenna impedance in waveguide', *Trans. I.E.E.E.*, **AP-11**, pp. 715–716, November 1963 (Letters).
44. P. W. Hannan, D. S. Lerner and G. H. Knittel, 'Impedance matching a phased array antenna over wide scan angles by connecting circuits', *Trans. I.E.E.E.*, **AP-13**, pp. 342–353, May 1965.
45. M. A. Balfour, 'Phased array simulators in waveguide for a triangular arrangement of elements', *Trans. I.E.E.E.*, **AP-13**, pp. 475–476, May 1965 (Letters).
46. T. Debski and P. W. Hannan, 'Complex mutual coupling measured in a large phased array antenna', *Microwave J.*, **8**, No. 6, pp. 93–96, June 1965.
47. N. R. Brennecke and W. N. Moule, 'Use of Fences to Optimize Operating Impedance of Phased Arrays, using an Improved Measuring Technique', PTGAP International Symposium Digest, pp. 134–142, 1964.
48. N. Amitay, J. S. Cook, R. G. Pecina and C. P. Wu, 'On Mutual Coupling and Matching Conditions in Large Planar Phased Arrays', PTGAP International Symposium Digest, pp. 150–156, 1964.
49. P. J. Mier, M. A. Balfour and H. A. Wheeler, 'Circular Waveguide Loaded with Dielectric Discs for Increased Usable Bandwidth', G-MTT International Symposium Program and Digest, pp. 33–36, 1964.
50. P. J. Mier and H. A. Wheeler, 'Dielectric-lined circular waveguide with increased usable bandwidth', *Trans. I.E.E.E. on Microwave Theory and Techniques*, **MTT-12**, pp. 171–175, March 1964.
51. Raymond E. Metter, 'X-band horn antenna has broad beam-width', *Electronics*, **33**, No. 10, pp. 50–53, 4th March 1960.
52. C. W. Gerst, 'Broadbanding Direction Finding Techniques for Reconnaissance Satellites', General Electric No. R 63 ELS-33, 24th May 1963, Syracuse, New York.
53. L. J. Chu, 'Calculation of the radiation properties of hollow pipes and horns', *J. Appl. Phys.* **11**, pp. 603–610, 1964.
54. Samuel Silver, 'Microwave Antenna Theory and Design', pp. 341–345, Boston Technical Publishers (Massachusetts, 1964).
55. S. Silver, *ibid.*, p. 199.
56. D. Midgley, 'A theory of receiving aerials applied to the radiation of an electromagnetic horn', *Proc. Instn Elect. Engrs*, **108**, pp. 645–650, November 1961. (I.E.E. Paper No. 3718 E.)
57. S. Silver, *ibid.*, p. 346.
58. G. Ziehm, 'Optimum directional pattern synthesis of circular arrays', *The Radio and Electronic Engineer*, **28**, pp. 341–355, November 1964.
59. Y. T. Lo and S. W. Lee, 'A study of space tapered arrays', *Trans. I.E.E.E.*, **AP-14**, pp. 22–30, January 1966.
60. R. E. Willey, 'Space tapering of linear and planar arrays', *Trans. I.R.E.*, **AP-10**, pp. 369–377, July 1962.

7. Appendix

7.1 *Distribution Function of a Ring Projected on a Line*

Consider a small portion Δl of a continuous ring (Fig. 18) at P so that the arc is equal to the tangent.

$$\Delta l^2 = \Delta x^2 + \Delta y^2$$
$$= \Delta x^2 + \Delta l^2 \cos^2 \psi$$

or

$$\frac{dl}{dx} = \frac{r}{\sqrt{r^2 - x^2}} = \frac{1}{\sqrt{1 - \left(\frac{x}{r}\right)^2}} \quad \ldots\ldots(12)$$

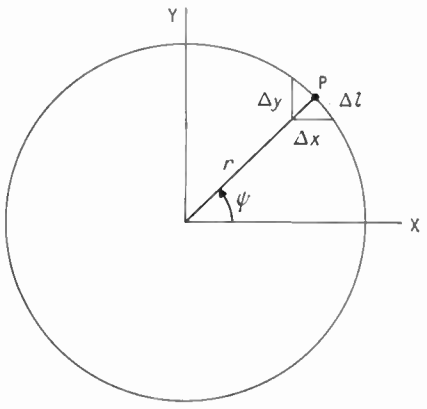


Fig. 18. Distribution function of a continuous ring projected on a line.

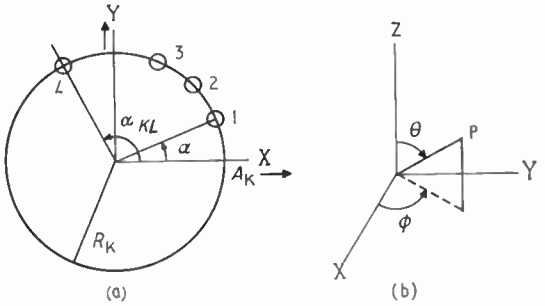


Fig. 19. Ring array and the co-ordinate system.

The array-factor of a ring array had been derived by a number of investigators.[1, 10, 52] Consider equally-spaced elements around the circumference of the Kth ring having a radius of R_K (Fig. 19).

$$\alpha = \frac{2\pi}{A_K} \quad \text{and} \quad \alpha_{KL} = \frac{2\pi L}{A_K} \quad \ldots\ldots(13)$$

where A_K = total number of elements in the Kth ring. The phase centre of the array is at the centre of the rings.

Let the beam be pointed in an arbitrary direction (θ_0, ϕ_0). This will multiply the signal in the Lth element by

$$\exp\left[-j\beta R_K \sin\theta_0 \cos(\phi_0-\alpha_{KL})\right]$$

where

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

The phase of the signal received at the Lth element from the direction θ, ϕ is

$$\exp\left[j\beta R_K \sin\theta \cos(\phi-\alpha_{KL})\right]$$

The array factor of the ring, therefore, is

$$E(\theta,\phi,\theta_0,\phi_0) = \sum_{L=1}^{A_K} \exp\left[j\beta R_K\{\sin\theta\cos(\phi-\alpha_{KL}) - \sin\theta_0\cos(\phi_0-\alpha_{KL})\}\right] \quad \text{.....(14)}$$

This could also be expressed in the form[52]

$$E = \sum_{m=-\infty}^{\infty} A_K (j)^{-mA_K} e^{-jmA_K\delta} J_{-mA_K}(\beta R_K a) \quad \text{.....(15)}$$

where

$$a = \left[(\sin\theta\cos\phi-\sin\theta_0\cos\phi_0)^2 + (\sin\theta\sin\phi-\sin\theta_0\sin\phi_0)^2\right]^{\frac{1}{2}} \quad \text{.....(16)}$$

and

$$\cos\delta = \frac{1}{a}(\sin\theta\cos\phi-\sin\theta_0\cos\phi_0) \quad \text{.....(17)}$$

For a number of rings, we shall have to sum over all the rings. Equations (14) and (15) reduce to

$$E = \sum_{K=1}^{10}\sum_{L=1}^{A_K} \exp\left[j\beta R_K\{\sin\theta\cos(\phi-\alpha_{KL}) - \sin\theta_0\cos(\phi_0-\alpha_{KL})\}\right] \quad \text{.....(18)}$$

and

$$E = \sum_{K=1}^{10}\sum_{m=-\infty}^{\infty} A_K (j)^{-mA_K} \times e^{-jmA_K\delta} J_{-mA_K}(\beta R_K a) \quad \text{.....(19)}$$

7.2 *Calculation of the Mutual Coupling*

The far-field expressions for a rectangular waveguide radiator have been calculated by Chu[53] and are given by Silver:[54]

$$E_\theta = -\left(\frac{\mu}{\varepsilon}\right)^{\frac{1}{2}} \frac{\pi a^2 b}{2\lambda_0^2 R}\sin\alpha \times \left[1+\frac{\beta_{10}}{k}\cos\theta+\Gamma\left(1-\frac{\beta_{10}}{k}\cos\theta\right)\right] \times \left[\frac{\cos\left(\frac{\pi a}{\lambda_0}\cos\alpha\sin\theta\right)}{\left(\frac{\pi a}{\lambda_0}\sin\theta\cos\alpha\right)^2-\left(\frac{\pi a}{2}\right)^2}\right] \times \left[\frac{\sin\left(\frac{\pi b}{\lambda_0}\sin\theta\sin\alpha\right)}{\left(\frac{\pi b}{\lambda_0}\sin\theta\sin\alpha\right)}\right] \exp(-jkR) \quad \text{.....(20)}$$

and

$$E_\phi = -\left(\frac{\mu}{\varepsilon}\right)^{\frac{1}{2}} \frac{\pi a^2 b}{2\lambda_0^2 R}\cos\alpha \times \left[\cos\theta+\frac{\beta_{10}}{k}+\Gamma\left(\cos\theta-\frac{\beta_{10}}{k}\right)\right] \times \left[\frac{\cos\left(\frac{\pi a}{\lambda_0}\sin\theta\cos\alpha\right)}{\left(\frac{\pi a}{\lambda_0}\sin\theta\cos\alpha\right)^2-\left(\frac{\pi}{2}\right)^2}\right] \times \left[\frac{\sin\left(\frac{\pi b}{\lambda_0}\sin\theta\sin\alpha\right)}{\left(\frac{\pi b}{\lambda_0}\sin\theta\sin\alpha\right)}\right] \exp(-jkR) \quad \text{.....(21)}$$

The origin is taken at the centre of the aperture. R is the distance of the field point and

$$\beta_{10} = \frac{2\pi}{\lambda_g} \quad \text{and} \quad k = \frac{2\pi}{\lambda_0}$$

(λ_0 = free space wavelength.)

As the radiation at the plane of the aperture $\theta = 90°$, Fig. 20) is of interest, the E_α-component alone will couple to the guide. The coupled field is $E_0 = E_\alpha \cos\alpha$. For the square waveguide with no reflection E_0 is, from eqn. (21),

$$E_0 = -\sqrt{\frac{\mu}{\varepsilon}}\frac{\pi a^3}{2\lambda_0^2 R}(\cos^2\alpha)\left[\frac{\cos\left(\frac{\pi a}{\lambda_0}\cos\alpha\right)}{\left(\frac{\pi a}{\lambda_0}\cos\alpha\right)^2-\frac{\pi^2}{4}}\right] \times \left[\frac{\sin\left(\frac{\pi a}{\lambda_0}\sin\alpha\right)}{\frac{\pi a}{\lambda_0}\sin\alpha}\right]\frac{\lambda_0}{\lambda_g}\exp\left(-j\frac{2\pi R}{\lambda_0}\right) \quad \text{.....(22)}$$

The far-field region is generally defined[55] as a space located at a distance of D^2/λ from the radiating aperture, where D is the diameter of the aperture and λ is the operating wavelength. The nearest neighbour

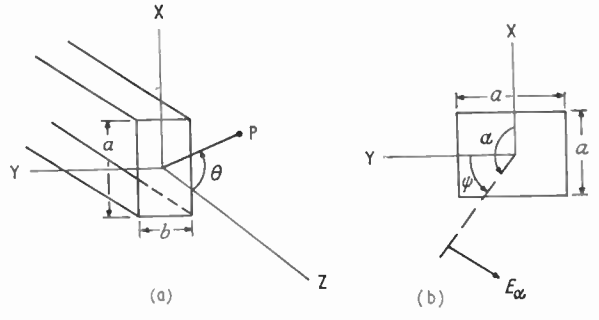


Fig. 20. Co-ordinate system used for the square waveguide.

radiator in this array will be, for smaller spacings, in the near field. All the other radiators will be in the far-field region. It is assumed, for simplicity, that any radiator is situated in the far-field region of all other radiators of the array.

7.3 *Principle of Field Fitting*[56]

Consider the Lth element of the Kth ring (Fig. 21) in the field of the Nth transmitting element of the Mth ring. The incident plane wave at L can be represented by the Fourier series:

$$\frac{4E_0}{\pi}\left[\cos\frac{\pi x}{a}+\tfrac{1}{3}\cos\frac{3\pi x}{a}+\tfrac{1}{5}\cos\frac{5\pi x}{a}+\ldots\right] \quad \ldots\ldots(23)$$

where E_0 is given by eqn. (22).

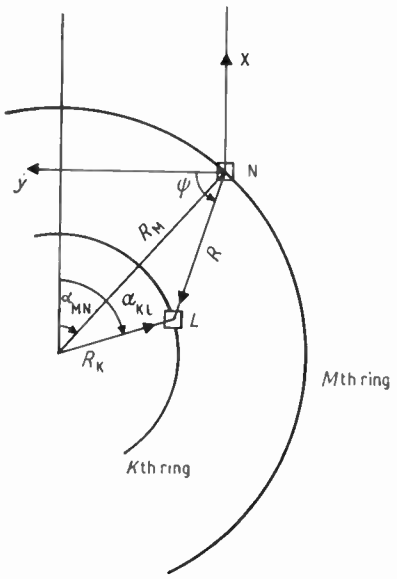


Fig. 21. Calculation of R, the distance between the two radiating elements.

The fundamental component that will couple to the guide is

$$E_1=\frac{4E_0}{\pi}\cos\frac{\pi x}{a} \quad \ldots\ldots(24)$$

Power received by the guide

$$W_r=\frac{1}{2Z_H}\int_{-a/2}^{a/2}\int_{-a/2}^{a/2}E_1^2\,dx\,dy \quad \ldots\ldots(25)$$

where Z_H is the wave impedance for the dominant-mode propagation. Therefore,

$$W_r=\frac{1}{f\lambda_g\mu}\frac{4}{\pi^2}a^2E_0^2 \quad \ldots\ldots(26)$$

The total power that can be transmitted[57] by the radiator is

$$W_t=\frac{\omega\mu a^4 2\pi}{4\pi^2\lambda_g} \quad \ldots\ldots(27)$$

The coupling C is defined[32] as

$$C=\left(\frac{W_r}{W_t}\right)^{\frac{1}{2}}=\frac{2}{\pi f\mu a}E_0 \quad \ldots\ldots(28)$$

For reducing the size, the waveguide is completely filled with a material having a dielectric constant of 4 (e.g. fused quartz has a dielectric constant of 3·78) and a low loss-tangent.

Substituting the value of E_0, the expressions for the coupling coefficient C is

$$C=-\tfrac{1}{2}\left(\frac{a}{\lambda_0}\right)^2\left(\frac{\lambda_0}{R}\right)(\cos^2\alpha)\frac{\lambda_0}{\lambda_g}\left[\frac{\cos\left(\frac{\pi a}{\lambda_0}\cos\alpha\right)}{\left(\frac{\pi a}{\lambda_0}\cos\alpha\right)^2-\frac{\pi^2}{4}}\right]\times$$

$$\times\left[\frac{\sin\left(\frac{\pi a}{\lambda_0}\sin\alpha\right)}{\frac{\pi a}{\lambda_0}\sin\alpha}\right]\exp\left(\frac{-j2\pi R}{\lambda_0}\right) \quad \ldots\ldots(29)$$

$$=B\exp\left(-j2\pi R/\lambda_0\right)$$

When scanned to a direction θ_0, ϕ_0, the coupling coefficient C is

$$C=B\exp\left(\frac{-j2\pi R}{\lambda_0}\right)\times$$

$$\times\exp\left[\frac{-j2\pi R_M}{\lambda_0}\sin\theta_0\cos\left(\phi_0-\alpha_{MN}\right)\right] \quad \ldots\ldots(30)$$

7.4 *Calculation of R, the Distance between the Two Radiating Elements*

Let i_x = unit vector in the positive x-direction
i_y = unit vector in the positive y-direction
R_K = radius of the Kth ring
R_M = radius of the Mth ring
L = element number of the Kth ring

It can be shown that

$$\bar{R}=\bar{R}_K-\bar{R}_M$$
$$=i_x[R_K\cos(\alpha_{KL})-R_M\cos(\alpha_{MN})]$$
$$-i_y[R_K\sin(\alpha_{KL})-R_M\sin(\alpha_{MN})] \quad \ldots\ldots(31)$$

Manuscript first received by the Institution on 21st June 1966 and in revised form on 24th October 1966. (Paper No. 1108.)

U.D.C.621.382.323

Field-effect Studies on Indium Antimonide Films

By

C. JUHASZ, B.Sc.†

AND

Professor
J. C. ANDERSON, D.Sc.(Eng.),
Ph.D., C.Eng., M.I.E.E., A.M.I.E.R.E.†

Reprinted from the Proceedings of the Joint I.E.R.E.–I.E.E. Conference on 'Applications of Thin Films in Electronic Engineering' held in London on 11th–14th July 1966.

Summary: Field effect studies have been carried out on thin film InSb inert insulator solid–solid interfaces. The films were prepared by the protected recrystallization method, namely vacuum evaporation and subsequent annealing of the resulting insulator-semiconductor-insulator sandwiches. The high annealing temperatures permissible without danger of re-evaporation of the more volatile component gave films with high mobility of approximately 10^3–10^4 cm²/V-s and relatively low carrier concentrations of approximately 10^{16}–10^{17}/cm³. The good bonding of the insulator on to an inherently clean semiconductor surface resulted in a low fast surface state density of approximately 5×10^{11}/cm², which would permit the use of these films for field effect transistor applications. Thin films are very well suited for field effect investigations. Surface Hall effect and surface conductivity against surface potential gave the trapped and mobile charge density with the relevant Hall and field effect mobility directly. A model of fast surface states based on these results is proposed.

1. Introduction

The insulated gate field effect transistor, in which the conductivity of a thin semiconductor film is modulated by a gate electrode, is a promising form for a thin film transistor, because device operation involves majority carrier processes which do not require high perfection single crystal films. The realization of such a device depends primarily on a sufficiently low density of fast surface states in order to obtain usable conductivity modulation. Such a thin film transistor (t.f.t.) can be fabricated by standard thin film preparative techniques and is therefore compatible with thin film circuits.

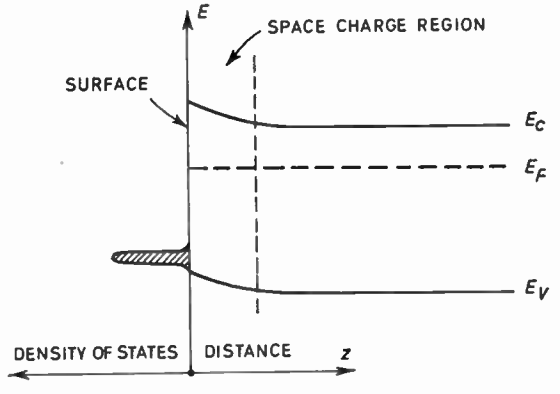


Fig. 1. An energy band diagram for the surface of a semiconductor. The horizontal axis to the right represents distance from the surface into the bulk and to the left it represents density of states. The vertical axis represents energy.

Working t.f.t.'s of this type were originally envisaged by Shockley[1] and successfully made by Weimer.[2] This was rapidly followed by transistor action demonstrated in a range of evaporated materials,[3] namely CdS, CdSe and Te, and recently extended to PbS,[4] InSb,[5] PbTe[6] and Si.[7] A notable success is that of the CdSe t.f.t. in an all-evaporated integrated circuit scan generator.[8] Most of these materials have field effect mobilities in the region of 10–10^2 cm²/V-s. This paper describes a preparative technique and experimental results for InSb films, which offer the advantage of higher field effect mobilities in the region of 10^3 cm²/V-s.

2. Semiconductor Surface Properties

The surface properties of a semiconductor crystal in contact with vacuum, can be summarized as dangling bonds on the surface giving rise to a high density, $\sim 10^{14}$/cm², of acceptor-like states called 'intrinsic states'. In an energy band diagram this can be represented as bending of the bands at the surface, as shown in Fig. 1. The intrinsic states prohibit usable conductivity modulation, since the maximum charge density which can be induced with fields in the region of the usual electric breakdown strength, is in the region of 10^{13}/cm².

In the case of a solid/solid interface, between an insulator and semiconductor, the intrinsic states will be modified by the taking-up of the dangling bonds. There may, however, be additional 'extrinsic states'

† Department of Electrical Engineering, Imperial College of Science and Technology, London.

introduced which are characteristic of the particular insulator/semiconductor combination, and may be acceptor or donor type. These states are referred to as 'fast states' in the literature, and their time-constant is around 10 μs for InSb at liquid nitrogen temperatures.[9]

If the dielectric material has a number of impurity levels in its band gap, the applied electric field may terminate on these, resulting in a large density of slow states which also give rise to degradation of the conductivity modulation. The time-constant of these so-called 'slow states' is generally in the region of a few seconds or higher. Their effect will not be considered here.

The gate electrode and the semiconductor film form the plates of a capacitor. A negative charge on the gate, $-Q_g$, induces a positive charge, $+Q_g$, on the semiconductor and vice versa. A part, Q_{ss}, of this induced charge may become trapped in surface states on the semiconductor surface and is thus immobilized[1] and a part, Q_{sc}, may enter the space charge region of the semiconductor below the surface and is mobile, contributing to the conductivity parallel to the surface.

Thus we may write

$$Q_g = Q_{ss} + Q_{sc} \qquad \text{.....(1)}$$

The applied charge Q_g is known from the geometrical capacitance between gate and semiconductor.

The trapped charge Q_{ss} obeys Fermi–Dirac statistics and, from the occupation of the surface states of density N_{ti} and energy E_{ti}, is given by

$$Q_{ss} = \sum_{i=1}^{\infty} \frac{qN_{ti}}{1+g\exp(E_{ti}-E_F)/kT} \qquad \text{.....(2)}$$

where E_{ti} is the trap energy and g is a degeneracy factor. It can be seen that Q_{ss} depends on the band bending at the surface.

The mobile charge in the space charge region, Q_{sc}, is a unique function of band bending and by measurement of galvanomagnetic properties (conductivity, Hall coefficient, and magneto-resistance), Q_{sc} can be calculated from experimental results.

This then allows N_{ti} and E_{ti} to be deduced by curve fitting. It must be noted, however, that theoretically N_{ti} and E_{ti} can only be obtained uniquely from field-effect measurements if the amount of band bending is very large.

Space charge calculations for the surface of a non-degenerate semiconductor have been well summarized by Many, Goldstein and Grover[10] with comprehensive references to the original work. In the case of InSb, however, degenerate statistics must be used for the conduction band since the bulk Fermi level is less than $3kT$ from the conduction band edge at room temperature. The degenerate semiconductor has been considered by Seiwatz and Greene[11] and their model has been extended[12] to permit calculation of the mobile carrier densities in the space charge region in the presence of charge applied to the gate.

If n_b is the bulk carrier density and n the density of n type carriers at any distance, z, below the surface, then we define the surface excess of carriers relative to the flat band case as

$$\Delta N = \int_0^{\infty} (n-n_b)\,dz \text{ per unit area of surface} \qquad \text{.....(3)}$$

For holes, a similar definition applies.

The surface excess of carriers contributes to the surface galvanomagnetic properties, and the contribution will depend on the effective surface mobilities μ_{ns} and μ_{ps}, as defined by eqns. (4) and (5). The difference between conductivity and Hall mobilities is generally less than 13% and therefore these are assumed to be identical.[13]

The effective surface mobilities should reduce with increasing band bending[14, 15, 16] due to surface scattering.

The excess surface conductivity contribution is defined as

$$\Delta\sigma = (\Delta N\mu_{ns} + \Delta P\mu_{ps})q \qquad \text{.....(4)}$$

and, similarly, using the Hall coefficient R, it can be shown that

$$\Delta(R\sigma^2) = (-\Delta N\mu_{ns}^2 + \Delta P\mu_{ps}^2)q \qquad \text{.....(5)}$$

For a two-carrier system in the limit of small magnetic fields[17]

$$R\sigma^2 = R_1\sigma_1^2 + R_2\sigma_2^2 \qquad \text{.....(6)}$$

where

$$\sigma = \sigma_1 + \sigma_2 \qquad \text{.....(7)}$$

and suffixes 1 and 2 refer to carriers of differing mobilities and densities.

Thus, for a surface layer in parallel with the bulk material

$$\sigma = \sigma_b + \Delta\sigma \qquad \text{.....(8)}$$

and similarly

$$R\sigma^2 = R_b\sigma_b^2 + \Delta(R\sigma^2) \qquad \text{.....(9)}$$

Equation (9) was originally obtained by Petritz from a circuit analogy and more rigorously proved by the solution of the Boltzmann transport equation.[18]

In the case of one excess carrier in the space charge region, eqns. (8) and (9) have a very simple physical interpretation.

If

$$\Delta N|\mu_{ns}| \gg \Delta P\mu_{ps}$$

then

$$\left|\frac{\Delta(R\sigma^2)}{\Delta\sigma}\right| = \mu_{ns} \text{ and } \left|\frac{(\Delta\sigma)^2}{\Delta(R\sigma^2)}\right| = -q\Delta N \qquad \text{.....(10)}$$

and this also applies for the case of holes.

Since $\Delta(R\sigma^2)$ and $\Delta\sigma$ can be measured experimentally the effective surface mobility and excess mobile charge are obtained directly, without resort to theoretical estimates based on various surface scattering and charge trapping models.

In the case of the description of t.f.t. operation it is useful to define a field-effect mobility as

$$\mu_{FE} = \frac{\Delta\sigma}{\Delta Q_g} \qquad \text{.....(11)}$$

since Weimer[3] has shown that the trans-conductance of the t.f.t. is adequately given, below the knee in the characteristic, by

$$g_m = \frac{\mu_{FE} C_g V_D}{L^2} \qquad \text{.....(12)}$$

where C_g is the gate semiconductor capacitance, V_D the source-drain voltage, and L the source-drain distance.

In the present experiments $\Delta\sigma$, $\Delta(R\sigma^2)$ and μ_{FE} are measured.

3. Experimental

Evaporations were generally carried out in an oil-diffusion-pumped system with a cold trap and liquid-nitrogen-cooled Meissner spiral in the bell jar, operating below 5×10^{-7} torr during evaporation. In addition some films were prepared in an all-baked stainless steel system at approximately 10^{-9} torr but these were not significantly different from those prepared at 10^{-7} torr. Full details of the plant and experimental procedure have been given elsewhere.[19]

Films were deposited by evaporation of insulator, SiO or MgF_2, from a molybdenum boat on to a substrate, usually mica, heated to 100°C. This was followed by flash evaporation of coarse-grained InSb powder and then by a final protective insulator layer. The evaporation was immediately followed by annealing at approximately 500°C, for approximately 30 minutes, followed by slow cooling. The annealing serves to recombine any free In and Sb to the compound and, to be effective, must be at a temperature close to the melting point because of the relatively small diffusion coefficients of In and Sb in the compound. In addition some recrystallization takes place during the anneal. The final layer of insulator prevents re-evaporation of Sb, from the film, during annealing but must not permit diffusion of either component through it. SiO, MgF_2 and a number of other dielectric materials have been found satisfactory in this respect.[19]

It was found necessary to ensure that substrate, sources and powders were thoroughly outgassed before evaporation. Failure to observe this precaution led to unsatisfactory results. The semiconductor powder was prepared from single-crystal, n-type InSb with an impurity concentration of $10^{14}/cm^3$ and mobility of 4×10^5 cm^2/V-s at 77°K. It was freshly crushed, before use, in a quartz mortar. The insulator materials were pure SiO (Kemet) and zone-refined MgF_2 (B.D.H.).

The field and Hall effect samples were made by scratching the film sandwich to the dimensions shown in Fig. 2. Ohmic contacts were soldered on to the films using spec.-pure In as a solder, which penetrated the top insulator layer without difficulty. For field-effect measurements the field-plate was provided by a layer of In on the back of the mica substrate, the latter being of about 10 μm thickness.

4. Results

Over 400 films were prepared and their bulk properties measured in air using standard d.c. Hall effect techniques; in some cases the measurements were extended down to 77°K. A.c. field-effect mobility was measured in a large number of films and, for those of the SiO and MgF_2-covered films which showed a conductivity minimum, a full analysis, at room temperature, of the $\Delta\sigma$ and $\Delta(R\sigma^2)$ variations was made. The surface state densities quoted, therefore, are applicable only to the 'best' films prepared.

A thorough investigation of the bulk properties and growth of films was made and has been reported elsewhere.[19] The effects of annealing temperature, time, vacuum conditions, evaporation source, substrate material and a range of different insulators was investigated with InSb films in the thickness range 3000–5000 Å. This led to the conclusion that the most important parameter is annealing temperature which must be near to 500°C for optimum properties. The

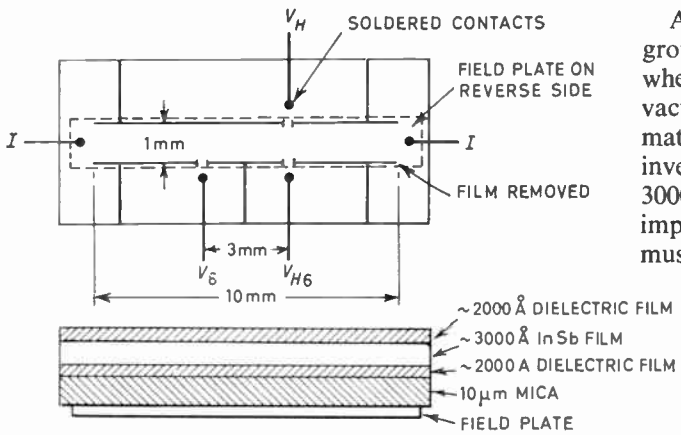


Fig. 2. Construction of thin-film sandwich specimens for galvanomagnetic and field-effect measurements.

films obtained were n-type with an apparent impurity concentration in the range 10^{16} to $10^{17}/cm^3$. Films prepared on amorphous or evaporated-insulator-covered substrates exhibited mobilities of about 6×10^3 cm^2/V-s at room temperature whilst those prepared on single-crystal substrates gave approximately 10^4 cm^2/V-s, the highest value obtained being 15 310 cm^2/V-s.

All mobilities were found to decrease with decreasing temperature, a typical result, for a film of InSb 4200 Å thick in an MgF_2 sandwich, being shown in Fig. 3. In some films a much more rapid decrease was observed. Similar results have been obtained by other workers.[20-22]

Extensive electron microscopy was carried out on a large number of films in an attempt to correlate the Hall mobility with crystal structure. In general the larger crystallite sizes corresponded with the higher mobilities. In Fig. 4(a) is shown a film of 510 Å thickness in which the mean crystallite diameter is 600 Å and the mobility was 1050 cm^2/V-s. In Fig. 4(b) the film thickness is 2500 Å, mean crystallite diameter 3000 Å and mobility 6170 cm^2/V-s. These data suggest that crystallite boundaries play a part in determining the mobility, but it is stressed that there are several other possible scattering mechanisms which may also be operative and will also limit the mobility.

Hall mobility was also found to decrease with decreasing film thickness, in agreement with other workers.[23] At a thickness of 500 Å mobilities of approximately 1000 cm^2/V-s were obtained.

Field-effect experiments under large a.c. signal conditions were carried out on SiO and MgF_2-covered films at room temperature. Both enhancement and depletion layers were observed with SiO-covered films depending on preparation conditions but reproducibility was poor. It appears that close control of preparation will be necessary if these films are to be used in t.f.t.'s. The surface properties of MgF_2-covered films were more reproducible giving a weak depletion layer at the surface in dry air corresponding to an upward band bending. The band bending was found to be sensitive to ambient conditions, especially water vapour, which drove the surface towards n-type. Typical variations of $\Delta\sigma$ and $\Delta(R\sigma^2)$ with applied charge are shown in Fig. 5. From this it can be seen that the conductivity minimum is reached.

Analysis of these measurements indicated the presence of two types of fast surface states, one almost coincident with the conduction band edge, and the other below mid-gap probably closer to the valence band edge. The density of fast states near the conduction band edge is around $5 \times 10^{11}/cm^2$, whilst the density of those below mid-gap could not be determined from the present experimental results as it was not possible to induce a sufficiently strong inversion layer on the films.

The field effect mobilities, predicting the usefulness for t.f.t. operation, were generally higher than 10^3 cm^2/V-s sufficiently away from the conductivity minimum on the 'n'-side. The surface Hall mobilities were always higher than the field effect mobilities, indicating that field ionization effects as proposed by Abraham and Poehler[24] were not operative in these films.

5. Discussion

The increase in impurity concentration in thin semiconductor films compared with that of the starting material is a well-known phenomenon. This could certainly be due to the incorporation of impurities during evaporation. However, the present field-effect studies show that the surface state densities observed can account for a part of this increase. This is supported by the fact that little difference was found between films evaporated at 10^{-7} and at 10^{-9} torr, and between films evaporated from quartz, vitreous carbon and refractory metal boats.

The decrease of mobility with decreasing temperature indicates the presence of some sort of barrier which limits transport of carriers. At higher temperatures some carriers are able to get over or through the barriers but as the temperature is lowered, fewer and fewer carriers are able to do so. Barlow *et al.*[25] have suggested oxide layers between crystallites as the possible barriers but, since electron microscopy and diffraction fail to reveal any separate phase in the present films, this is not thought to be applicable in the present case. A more probable effect is the scattering of carriers by ionized flaws surrounded by a space-charge region. This would apply to line and point defects but also, and more significantly, to grain boundaries. Interfacial surface states must occur at grain boundaries and will be accompanied by space-charge layers. The dependence of mobility on crystallite size suggests that grain boundaries represent the major mobility-limiting barrier in these films.

The decrease of mobility with film thickness is also mainly a result of the same mechanism since crystallite size reduces as the films become thinner. However, surface scattering also limits mobility.

Analysis of the field-effect data also supports the model of a potential barrier limited mobility, since the surface Hall mobility increases with field plate voltage and then decreases at higher voltages. The initial increase is due to the barriers effectively being lowered in potential so that they are not so effective in limiting mobility. At higher gate potentials surface scattering becomes more effective and mobility is then reduced. Similar observations have been made on CdS films by Waxman *et al.*[26] although the origin of the potential

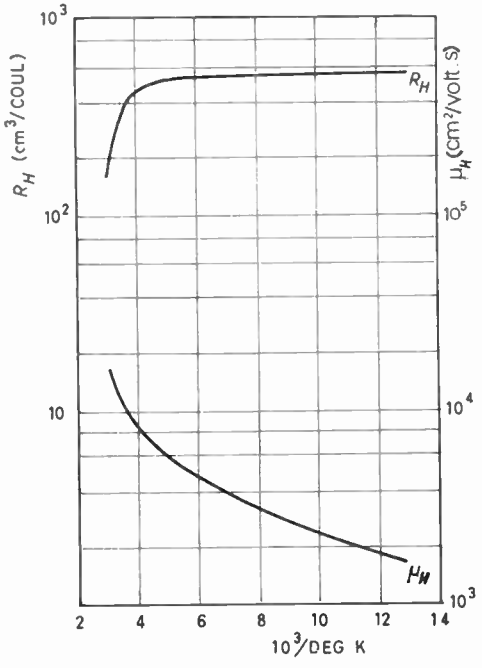


Fig. 3. Hall mobility, μ_H, and Hall constant, R_H, as a function of reciprocal temperature for a typical MgF_2/Insb/MgF_2 sandwich.

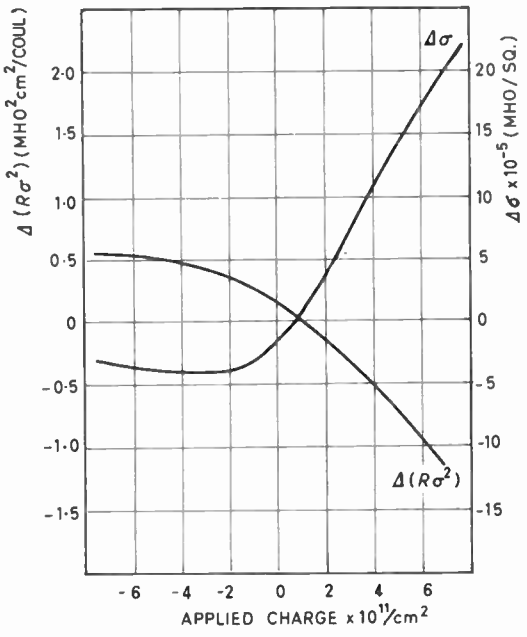
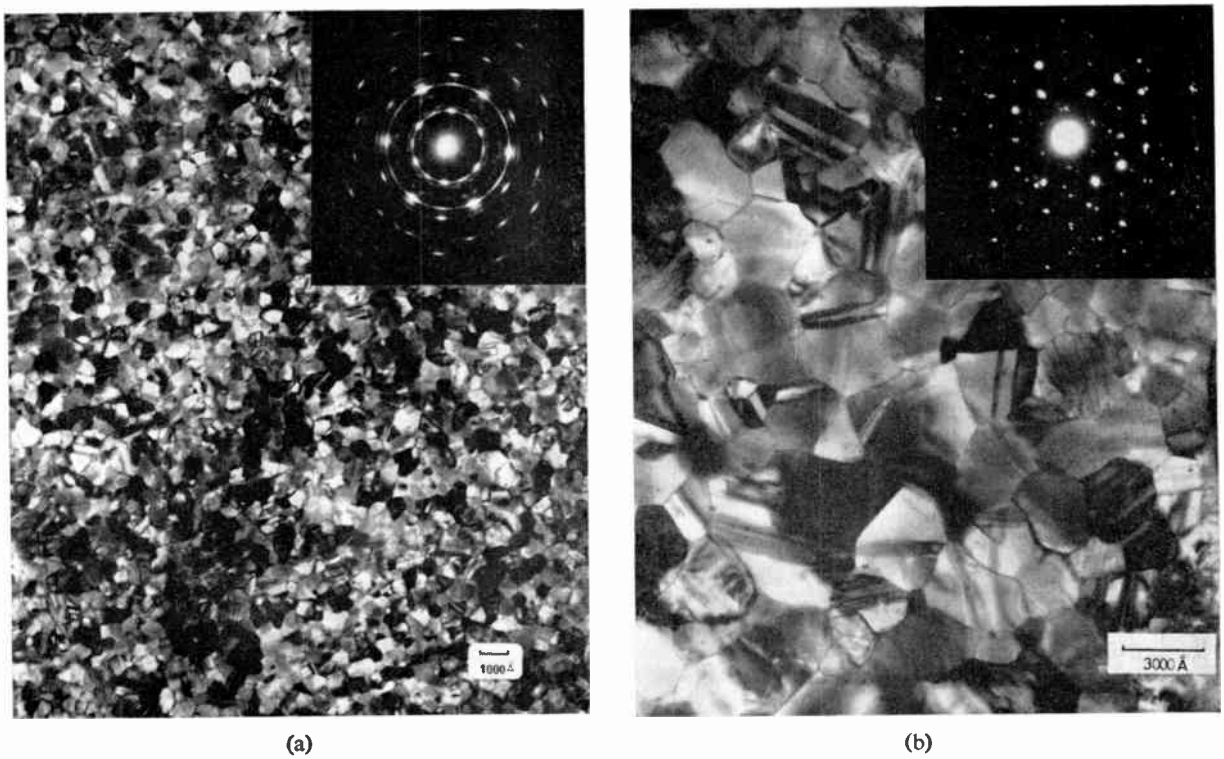


Fig. 5. $\Delta(R\sigma^2)$ and $\Delta\sigma$ as a function of charge applied to the field plate for an MgF_2/InSb/MgF_2 film which shows the conductivity minimum.



(a) (b)

Fig. 4. Electron micrographs and diffraction patterns for (a) 510 Å Insb film; (b) 2500 Å InSb film.

barrier is different in the CdS case. It should be noted that the reduction in mobility at high gate potentials, in the case of the present InSb films, is not so great as would be predicted from the theory of perfectly diffuse surface scattering. This indicates that a proportion of the carriers undergo specular reflection at the surface, so that serious degradation of the mobility at high gate fields should not occur.

6. Conclusions

The most significant feature of these results is the fact that adequate conductivity modulation and field-effect mobilities can be obtained entirely by thin film preparative techniques and that high-perfection single crystals are not required. The mobilities obtained in thin InSb films are orders of magnitude larger, than those in evaporated Ge and Si. The field-effect type of measurement used allows a given semiconductor/insulator combination to be assessed from the point of view of its performance as an insulated gate field-effect transistor, as well as yielding basic information on surface states at interfaces. The 'protected recrystallization' or solid encapsulation preparative technique gives a well-defined, reproducible interface and, in the case of $InSb/MgF_2$, provides the basis for a t.f.t. using an InSb film.

With the field-effect mobilities, of approximately 10^3 $cm^2/V\text{-}s$ which have been achieved, it should be possible to operate a t.f.t., based on this system, at 10^9 Hz. Alternatively, source-drain distances could be increased, for a given lower frequency, to ease fabrication problems.

7. Acknowledgment

One of the authors (C. J.) is indebted to Dr. Jay N. Zemel for many useful discussions about field-effect measurements in the early stages of the work.

8. References

1. W. Shockley and G. L. Pearson, 'Modulation of conductance of thin films of semiconductors by surface charges', *Phys. Rev.*, **74**, pp. 232–3, 15th July 1948.
2. P. K. Weimer, 'The TFT—a new thin film transistor', *Proc. Inst. Radio Engrs*, **50**, pp. 1462–9, June 1962.
3. P. K. Weimer, 'Physics of Thin Films', Vol. 2. (Academic Press, New York, 1964.) This gives a comprehensive bibliography of the earlier work.
4. W. B. Pennebaker, 'PbS thin film transistors', *Solid-State Electronics*, **18**, No. 5, pp. 509–15, May 1965.
5. V. L. Frantz, 'Indium antimonide thin-film transistors', *Proc. Inst. Elect. Electronics Engrs*, **53**, p. 760, July 1965. (Letters).
6. J. F. Skalski, 'A PbTe single-crystal thin-film transistor', *Proc. Inst. Elect. Electronics Engrs*, **53**, p. 1792, November 1965. (Letters).
7. C. A. Salmon and L. Young, 'Evaporated silicon thin-film transistors', *Proc. Inst. Elect. Electronics Engrs*, **53**, pp. 2156–7, December 1965. (Letters).
8. P. K. Weimer, G. Sadasiv, L. Meray-Horvath and W. S. Homa, 'A 180-stage integrated thin-film scan generator', *Proc. Inst. Elect. Electronics Engrs*, **54**, pp. 354–60, March 1966.
9. J. L. Davies, 'Surface states on the (iii) surface of indium antimonide', *Surface Science*, **2**, pp. 33–39, 1964.
10. A. Many, Y. Goldstein and N. B. Grover, 'Semiconductor Surfaces'. (North Holland, Amsterdam, 1965.)
11. R. Seiwatz and M. Greene, 'Space charge calculation for semiconductors', *J. Appl. Phys.*, **29**, pp. 1034–40, July 1958.
12. C. Juhasz and J. C. Anderson, unpublished.
13. J. N. Zemel, 'Surface transport theory', *Phys. Rev.*, **112**, pp. 762–5, 1st November 1958.
14. J. R. Schrieffer, 'Effective carrier mobility in surface-space charge layers', *Phys. Rev.*, **97**, pp. 641–6, 1st February 1955.
15. Y. Goldstein, N. B. Grover, A. Many and R. F. Greene, 'Improved representation of calculated surface mobilities in semiconductors. II. Majority carriers', *J. Appl. Phys.*, **32**, pp. 2540–1, December 1961.
16. N. B. Grover, Y. Goldstein and A. Many, 'Improved representation of calculated surface mobilities in semiconductors. I. Minority carriers', *J. Appl. Phys.*, **32**, pp. 2538–9, December 1961.
17. E. H. Putley, 'The Hall Effect and Related Phenome na' (Butterworth, London, 1960.)
18. R. L. Petritz, 'Theory of an experiment for measuring the mobility and density of carriers in the space-charge region of a semiconductor surface', *Phys. Rev.*, **110**, pp. 1254–62, 15th June 1958.
19. C. Juhasz and J. C. Anderson, 'Preparation of high mobility thin films of indium antimonide', *Proceedings of the 3rd International Vacuum Congress*, Stuttgart, 1965.
20. E. B. Dale and G. Senecal, 'Annealing effects in evaporated InSb films', *J. Appl. Phys.*, **33**, pp. 2526–31, August 1962.
21. S. A. Semiletov and P. S. Agalarzade, 'Structure and electrical properties of thin films of InSb', English translation in *Soviet Phys. Cryst.*, **9**, No. 4, pp. 409–13, January–February 1965.
22. K. G. Günther and H. Freller, 'Properties of evaporated layers of InSb and InAs', *Z. Naturforsch.*, **16a**, pp. 279–83, March 1961. (In German.)
23. W. J. Williamson, 'High mobility thin films of indium antimonide vacuum deposited on to a cold substrate', *Solid State Electronics*, **9**, pp. 213–24, March 1966.
24. D. Abraham and T. O. Poehler, 'A physical description of CdSe tft operation', *Intl J. Electronics*, **19**, pp. 165–179, 1965.
25. H. E. M. Barlow, R. Koike and R. Udea, 'Negative magnetoresistance effect in evaporated indium-antimonide films', *Proc. Instn Elect. Engrs*, **112**, pp. 1849–55, October 1965.
26. A. Waxman, V. E. Henrich, F. V. Shallcross, H. Borkan and P. K. Weimer, 'Electron mobility studies in surface space-charge layers in vapor-deposited CdS films', *J. Appl. Phys.*, **36**, pp. 168–75, January 1965.

Manuscript received by the Institution on 10th May 1966. (Paper No. 1109.)

U.D.C. 621.37/9.002

The Preparation and Application of Tantalum Thin Film Passive Components

By

R. NAYLOR,
M.Sc., C.Eng., M.I.E.E.†

AND

R. FAIRBANK,
B.Sc.(Tech.)†

Reprinted from the Proceedings of the Joint I.E.R.E.–I.E.E. Conference on 'Applications of Thin Films in Electronic Engineering' held in London on 11th–14th July 1966.

Summary: The paper commences with a comparison of thin film and diffused passive components, then describes the methods of fabricating resistors and capacitors by using sputtered tantalum and also inductors using nichrome/gold. The reasons for choosing tantalum as the resistor material and sputtering as the deposition method are discussed.

The work described in the paper has concentrated on making resistor and capacitor dimensions as small as possible without seriously affecting their properties. Resistors less than 0·001 in wide and capacitors with 1 pF per 0·001 in square of electrode area can be made, so that small, and thus economical, circuits can be realized.

The performance of thin film resistors, capacitors and inductors is given and one section is devoted to the behaviour of resistors when dissipating high powers (over 1 kW/in²) at elevated temperatures.

Methods of bringing together active devices and the thin film components are considered, emphasizing that since the active and passive devices can be made small, then the method of combining the two should take advantage of these small dimensions. The two methods discussed are the 'flip-chip' techniques for adding active devices to a substrate with thin film components and the use of an active substrate, in which the thin film components are fabricated on oxidized silicon which contains the active devices. The use of 'flip-chip' techniques for mounting integrated circuits on thin film interconnection patterns is also discussed.

Examples are given which demonstrate how thin film resistors, capacitors and inductors with active or passive substrates can be used in electronic circuits.

1. Introduction

The use of thin film passive devices in microelectronics is now well established, but it is unfortunate that, in general, they have been considered as rivals to integrated circuits. It is far better to consider the thin film and integrated circuit techniques together and to choose the one which suits a particular application better. It may be that a combination of the two techniques, producing the so-called hybrid circuit, provides the best method for a given requirement.

It is accepted that ultimately the integrated circuit provides the most economic method of fabricating large scale quantities of small pieces of electronic circuitry, provided that the performance of the diffused passive components is adequate. The numerous different ranges of logical circuits fall into this category.

In some applications, e.g. linear circuits, the limitations of diffused passive components are severe, and it is in these cases that the superior properties of thin film components can be used.

One disadvantage of integrated circuits, which does not apply to thin film circuits, is component interaction due to inadequate isolation, and is sometimes more important at high frequencies. This generally requires an extra stage of design during the development and this is important in small-scale production. Another limitation which sometimes occurs in silicon integrated circuit design is that the doping levels of the semiconductor required for diffused resistors and transistors are not compatible. In such cases, the use of thin film components enables the transistors to have an optimum design.

2. The Preparation of Passive Components

The main requirements for the manufacture of thin film passive components are a method of depositing the materials with a uniform and controllable thick-

† Automation Systems Division, Ferranti Ltd., Wythenshawe, Manchester.

ness, accurate pattern definition and for the deposited films to be stable in their particular environment and over the expected operating life of the device.

2.1 *Resistors*

The two main limitations in the use of diffused resistors are their relatively poor tolerance on resistance, say $\pm 20\%$ and temperature coefficient of resistance, say between $0{\cdot}1\%$ and $0{\cdot}5\%$ per deg C. Resistors of considerably better performance, such as $\pm 5\%$ in tolerance and $0{\cdot}01\%$ per deg C temperature coefficient of resistance, are attainable when using thin films of tantalum as the resistor material and sputtering as the method of deposition. The tolerances given here are for large scale quantity production. It has been quoted that tantalum resistors can be made to an initial tolerance between $\pm 0{\cdot}1\%$ and $\pm 0{\cdot}02\%$ depending on resistance value, by individual trimming.[1]

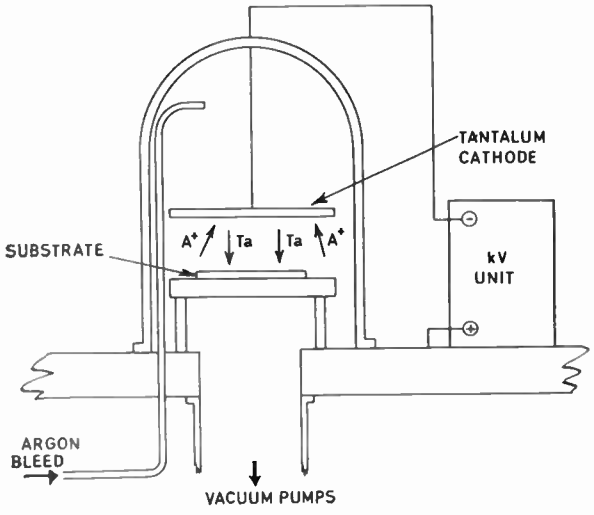


Fig. 1. Simplified sputtering equipment.

2.1.1 Resistor materials

The two most frequently reported materials for resistor fabrication are nichrome or tantalum. The latter was chosen for the work described in this paper since it is a refractory material (melting point about 3000°C) and is thus expected to be very stable at operating temperatures. Tantalum is not an alloy but a single element and thus there are no worries about composition changes during deposition. Another point in favour of using tantalum for resistors is that, when it is nitrogen doped, it has a higher resistivity ($\sim 250\ \mu\Omega$ cm) than nichrome (~ 80–$130\ \mu\Omega$ cm). Thus, for a given sheet resistivity, tantalum films are thicker and so easier to produce.

2.1.2 Deposition techniques

There are two principal vacuum methods of depositing thin films, one by thermal evaporation using either conventional heating or electron beams, and the other by sputtering.

In sputtering[2] a discharge is struck in a gas such as argon (see Fig. 1) so that the argon ions are accelerated towards the tantalum cathode and, on striking the cathode, they collide with and transfer their energy to the tantalum atoms. These atoms are then displaced from the cathode and deposited as a film on the substrate.

Sputtering was chosen as the method of deposition because of its following advantages:

(a) Since the deposition source can be made as large as the vacuum system will allow, the method of deposition is planar in principle. This means that it is easier to obtain a high degree of uniformity in film thickness across the substrate.

(b) The atmosphere in the bell-jar during sputtering can be far more reproducible for a given pumping capability, than for evaporation. The procedure prior to sputtering is to pump the bell-jar down to about 1×10^{-5} torr, then to admit a nitrogen–argon mixture until the pressure rises to about 5×10^{-2} torr, which is the pressure required for sputtering. The nitrogen doping reduces the effects of impurities outgassed during processing and enables reproducible resistor films to be made with good properties.[3] It should be mentioned that any impurities present before the required deposition commences are removed by getter sputtering, i.e. a closed shutter is used to prevent the sputtered tantalum atoms which trap the impurities, from being deposited on the substrate. Thus, the effects of gaseous impurities can be considerably reduced by using sputtering as the method of deposition. In evaporation, the residual atmosphere during deposition is of unknown composition and thus would be expected to give more variable results.

(c) The high melting point of tantalum creates difficulties if it is evaporated, since at these temperatures it reacts to some degree with the crucibles, etc. Sputtering is essentially a low temperature process, thus overcoming these difficulties and enabling films to be deposited with low power dissipation. Typical sputtering conditions are:

power supplies: 80 mA, 3·5 kV, source resistance 25 kΩ;

pressure: 0·05 torr;

atmosphere: nitrogen doped argon;

substrate temperature: glass substrate 450°C oxidized silicon substrate 500°C;

sputtering rate: ~ 75 Å/min.

The sheet resistivity of the tantalum film is monitored during deposition by using a stencil type of mask to deposit a strip of tantalum between conducting films. When the resistance of this monitor reaches the required value, the sputtering is stopped.

After the tantalum film is removed from the vacuum equipment, the sheet resistivity can be trimmed by methods which convert some of the film to insulating oxide. Two methods are available for oxidizing the films. One is to anodize the tantalum and the other to convert the metal to oxide thermally by baking in air. It should be mentioned that these layers provide a protective cover for the resistors.

2.1.3 Processing of resistor patterns

So far, the paper has described a method for producing thin films of tantalum on an insulating substrate, but it is still necessary to determine the shape, or aspect ratio, of this material to form resistors. Before resistor fabrication, the sheet resistivity of the film is decided from a study of the range of resistor values required in the circuit. The individual resistor values are obtained from this by using the appropriate aspect ratio.

For economic reasons, the area occupied by each resistor is made as small as possible. The limit for a given tolerance is caused either by power dissipation, or processing limitations. The processing limitations are not normally severe when using smooth and flat substrates such as glass, oxidized polished silicon or possibly glazed ceramic.

When processing resistors on a glass substrate, the limit on size reduction tends to be due to power considerations, and a limit of 20 W/in^2 has been chosen as an acceptable value. If glazed ceramics are considered, then the power limitation can be relaxed up to something like 100 watts for each square inch of resistor area.

If oxidized silicon is used as a substrate, then, provided the silicon is mounted on a good heat sink, it is possible to run the resistors with power dissipations of 500 to 1000 W/in^2 of resistor area.

The method chosen to define the resistor shapes uses photoresist and etching techniques, since these provide more accurate edge definition than can be obtained with stencil masks. The substrate is covered with a sputtered layer of copper and holes are etched in the copper, using photoresist and photographic masks, such that the holes correspond to the required resistor patterns. The substrate then receives the required thickness of the sputtered layer of tantalum and when placed in a copper etch, the copper and unwanted tantalum are removed, leaving the required tantalum shapes. In-contact copper masking is used as it provides an excellent method of determining the desired high accuracy of edge definition.

The remaining process to complete a resistor is to deposit the conducting materials which determine the length of the resistor and provide the necessary interconnections. Gold with nichrome underlay, or aluminium, have been used. In general, nichrome/gold is used for circuits on glass substrates when it is required to connect to other devices by soldered wires, or aluminium for circuits on oxidized silicon when it is required to connect to active devices diffused into the silicon.

The resistors have now reached the stage at which they can have their values checked, and after this initial check, the resistors are trimmed and passivated until the resistor values are within the required tolerance.

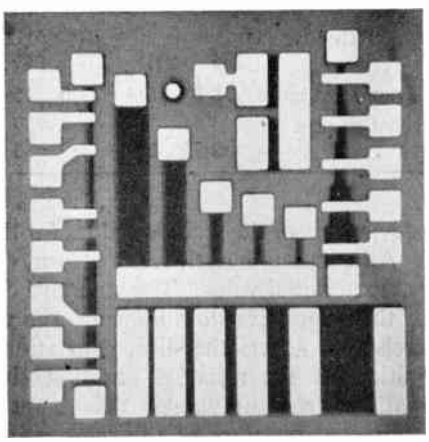


Fig. 2. Test resistors.

Figure 2 shows a series of test resistor patterns fabricated on a slice of oxidized silicon as substrate. The conductor material used to determine the resistor length in this particular example is aluminium. The area occupied by each pattern is 0·045 in × 0·045 in and the narrowest resistor is 0·0005 in wide.

2.1.4 Resistor performance

It is not easy to give a simple set of figures to indicate the resistor values, tolerances or ageing characteristics, since these figures depend considerably on film thickness and resistor area, etc.

It is possible to sputter resistor films from about 10 to 300 Ω/square. The sheet resistance of the sputtered tantalum can be controlled to about ±5%. This initial tolerance is adequate since it can be further improved by trimming at a later stage.

Resistors of 0·003 in wide and 200 Ω/square can be fabricated in slice form with tolerances of between ±5 and ±10% on absolute value and ±2% on ratio between adjacent resistors of similar width. These

tolerances deteriorate with thinner films and narrower resistors.

The temperature coefficient of resistance of the nitrogen doped resistors is approximately 80 parts in 10^6 per deg C.

Measurements were taken on a group of resistors as shown in Fig. 2. The tantalum film was sputtered to give a sheet resistivity of 25 Ω/square, aluminium was added as conductor material, and finally the slice was trimmed and passivated by baking in air. The results shown in Table 1 give the mean values together with

Table 1

Mean resistance values of different width resistors

Resistor width (10^{-3} in)	3	2	1	¾	½
Mean value of resistance in ohms	295	291	287	269	262
Standard deviation in ohms	8·0	10·4	14·0	15·0	19·3
Standard deviation as a percentage	2·75	3·6	4·9	5·5	7·4

standard deviations of the resistor values for the width indicated. All resistors had the same aspect ratio and thus the same nominal value. The standard deviation $\pm 2\frac{1}{4}\%$ for the largest resistors indicates the variation of sheet resistivity across the slice. Variations in the edge definition of the resistors are best shown by studying ratio of resistor values. These ratios were taken on adjacent resistors of the same aspect ratio but different dimensions. The results are given in Table 2.

Table 2

Resistance ratio for different width resistors

Resistance ratio for different width resistors	$\frac{R_3}{R_2}$	$\frac{R_3}{R_1}$	$\frac{R_3}{R_{\frac{3}{4}}}$	$\frac{R_3}{R_{\frac{1}{2}}}$
Mean value of ratio	1·014	1·029	1·088	1·130
Standard deviation of ratio	0·017	0·032	0·039	0·057
Standard deviation as a percentage	1·7	3·1	3·6	5·0

Some experiments were carried out to determine the power dissipation required to burn-out resistors. The resistors were made to the standard test pattern (see Fig. 2) on an oxidized silicon substrate and, during the experiment, the slice of silicon was placed on a large block of metal but no precautions were taken to ensure a low thermal resistance between the two. The power applied to the resistors was slowly increased until they burnt-out. All the resistors had the same aspect ratios, but different areas. Figure 3 is a graph showing the resistor area versus power required to burn-out the resistors.

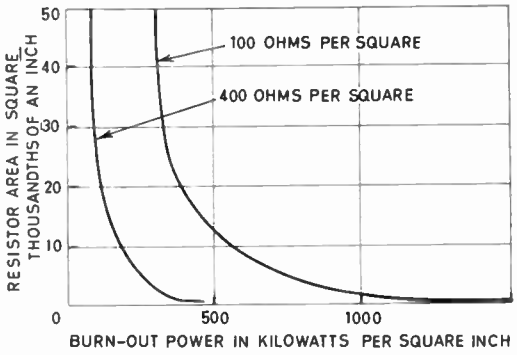


Fig. 3. Power required to burn-out resistors.

2.1.5 Ageing experiments

Some preliminary ageing experiments were started to indicate the order of magnitude of the ageing of nitrogen doped tantalum resistors.

For the tests, a set of resistors was made to the pattern required for a core store amplifier (see Section 2.3). The resistors were made from 70 ohms per square tantalum on oxidized silicon as substrate and were passivated together on the slice to 200 ohms per square by anodizing. After the silicon slice had been cut into individual dice, they were alloyed into reduced height eight-lead TO-5 headers, and the electrical connections were made by using bonded gold wires. For this test a silicon die was mounted in each can, and connected so that only the two 900-ohm resistors were dissipating power.

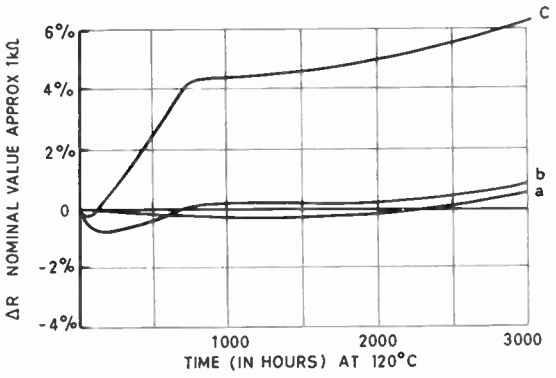


Fig. 4. Ageing of tantalum resistors.

(a) Total power dissipation = 75 mW or 2 kW/in².
(b) Total power dissipation = 200 mW or 5 kW/in².
(c) Total power dissipation = 300 mW or 8 kW/in².

During the ageing experiments, the cans were kept in an oven at 120°C and the resistors dissipated different powers at these temperatures. The resistors were allowed to cool before measurements were taken.

The graph of Fig. 4 shows the results of the experiments in the first 3000 hours. The changes in resistor values when dissipating 2 kW per square inch, i.e. 75 mW in a TO-5 can at 120°C, are less than 1%. Further work is necessary to determine the reasons for these changes.

2.2 *Capacitors*

The second passive component to be considered is the capacitor. These can be made in integrated circuits by a semiconductor p–n junction under reverse bias, or by using the protective layer of silicon oxide as dielectric. As with diffused resistors, these capacitors have adequate performance for some, but not all, applications.

The two main advantages of thin film capacitors over diffused capacitors are that they can be made with up to 1 pF per 0·001 in square and still have a high breakdown voltage, and that the capacitors are not voltage dependent.

2.2.1 Capacitor materials

Tantalum pentoxide in thin film form is a very attractive material to use as a dielectric for capacitors since its dielectric constant is 21, compared with 2 to 4 for oxides of silicon, so that capacitors can be made with a relatively small electrode area for a given breakdown voltage.

A further point supporting the use of tantalum pentoxide as dielectric for capacitors is that it can be formed by anodization.[4] The anodizing process is self-healing, in that the anodizing current concentrates on flaws so that they tend to be sealed, and the process comes to an end when the dielectric thickness everywhere corresponds to the anodizing voltage. As an example, when an anodizing voltage of 80 V is applied, the dielectric builds up to about 1200 Å, yielding capacitors of about 1 pF per 0·001 in square and 1200 pF capacitors with this thickness of dielectric have a breakdown voltage of about 40 to 50 V.

Thin film capacitors are more difficult to make than thin film resistors so that, at the moment, it is more economical to fabricate the capacitors as separate devices. However, one important feature has come to light, namely, that they can be made with a very good high frequency performance by using a silicon substrate as a backing to the tantalum electrode, so that the resistance of this lower electrode is adequately low. The high frequency loss factor of a thin film capacitor is nearly always determined by the losses in the electrodes rather than dielectric losses.[5] Hence in this case if care is taken to ensure that the sheet resistivity of the other electrode is kept low, then capacitors with good high frequency performance can be made.

2.2.2 Processing techniques

So far the main requirement for capacitors has been that they should be small in area for use in hybrid circuits in which capacitor chips are connected to active and/or resistor chips by means of the thermocompression bonding of gold wires.

The method used to fabricate the capacitors to have the good high frequency performance is to sputter approximately 2000 Å of tantalum on to a polished silicon substrate, taking care to ensure that the surface is smooth. This provides the lower electrode with the necessary low resistance. Then, by using standard anodizing techniques, the surface of the tantalum is anodized to give the dielectric of the required thickness. Figure 5 shows a section through a typical capacitor chip and Fig. 6 shows an assortment of capacitors, with values ranging from 32 pF to 1000 pF.

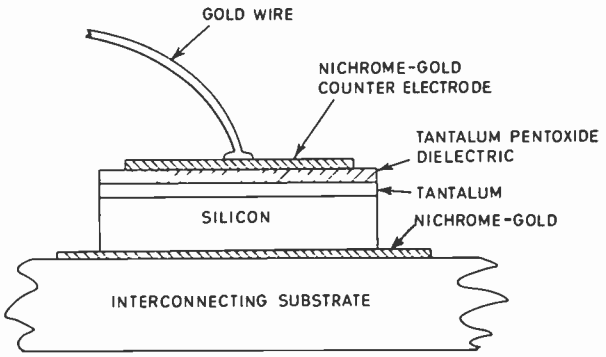


Fig. 5. Section showing construction of thin film tantalum capacitor.

A layer of nichrome-gold is then evaporated on to the dielectric to form the upper electrode and this is then etched to define the area of the capacitor. The capacitor value is also determined by the dielectric thickness, which in turn is determined by the forming voltage.

After dicing, the capacitors are mounted for use by alloying the silicon on to a conductor on an interconnecting substrate for one connection, and by bonding a gold wire to the upper electrode for the other connection.

2.2.3 Performance of capacitors

The general properties of anodized tantalum capacitors have been adequately covered in the literature.[4]

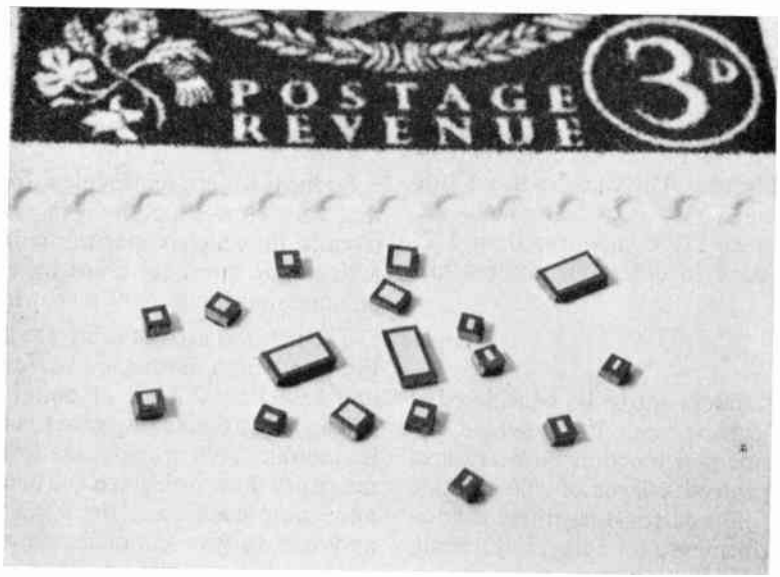


Fig. 6. An assortment of tantalum pentoxide capacitor with values from 32 pF to 1000 pF.

Typical capacitors made by the methods described in the preceding Section have breakdown voltages between 50 and 75% of the formation voltage, and working voltages of up to 25% of the formation voltage. Thus, films anodized to 80 V provide capacitors with about 1 pF per 0·001 in square suitable for 20 V working.

The leakage currents of anodized tantalum capacitors are very low. Capacitors of 1200 pF with a bias of 15 V have leakage currents of the order of 10^{-9} A at room temperature, rising to about 5×10^{-7} A at 150°C.

The loss factor of capacitors made by these methods is such that a 50 pF capacitor has an equivalent series resistance of about 5 ohms at about 50 MHz, i.e. a Q-factor of about 15 or loss factor of about 6%, thus indicating that they are suitable for use in tuned circuits at high frequencies.

One disadvantage of anodized capacitors, however, is that the leakage current is asymmetrical. This is not as serious as with ordinary electrolytic capacitors, but it appears their life and ageing characteristics are better if the capacitors are biased in the same direction as required for anodizing.

2.3 *Inductors*

Inductors are not as adaptable to thin film techniques as resistors and capacitors, since the field of the coils requires a depth comparable to the coil radius. This is important when considering substrates for inductors, since a conducting substrate located within a distance equal to the coil radius will reduce the coil inductance and increase the coil losses.

Since the making of multi-layer coils or the use of high permeability materials surrounding the coil are impractical propositions, the upper limit of inductance value is likely to be a few microhenries.

2.3.1 Inductor processing

The inductors are processed in evaporated thin films of nichrome-gold using photoresist techniques. Since the inductors are required to have a reasonable Q-factor at their operating frequency, it is often necessary to thicken the evaporated conductor by chemical plating techniques.

2.3.2 Inductor performance

An inductor occupying 0·35 inches square with conductors 0·005 in wide and with 0·005 in spacing, conveniently provides about 1 μH of inductance. When the conductor is plated up to a thickness of about 0·0005 in then the series resistance of the coils is about 4 ohms at 10 MHz, i.e. Q-factors are between 15 and 20. Thus, with the thin film capacitors described in the preceding section, it is possible to make tuned circuits with a reasonable Q for operation at tens of megahertz.

2.4 *Methods of Using Active Devices with Thin Film Circuits*

So far, only thin film passive components have been considered but, in order to complete an electronic circuit, active devices and methods of mounting them must be considered.

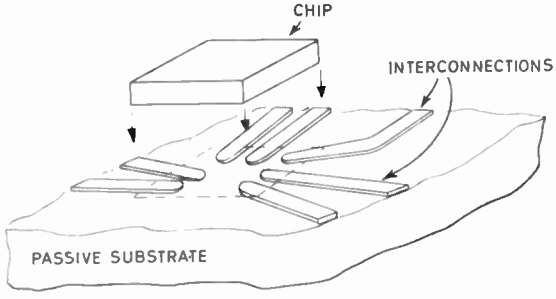
At the present time, thin film active devices are still in the research stage and they are not considered as being ready for application to circuitry.

Since the dimensions of diffused active devices in chip form are very small, it is desirable to find a method of using the chips which takes advantage of this small size. This leads to two methods of using the active devices, the so-called 'flip-chip' technique when using passive substrates, or by diffusing the transistors into the substrate when using oxidized silicon as an active substrate.

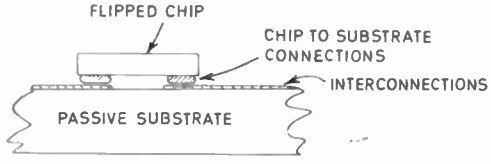
2.4.1 'Flip-chip' mounting

The 'flip-chip' method of mounting active devices, either transistors or complete circuits, on substrates provides an excellent way of mounting the chips since it is compatible with the small size of the devices. This method of mounting is not restricted to active devices since, for example, the small capacitors described above could be applied by these techniques.

Briefly, the 'flip-chip' technique consists of mounting a semiconductor chip with its active surface facing the substrate, as shown in Fig. 7(a) and (b), so that the connections are made between the two by small (0·003–0·005 in diameter) metallic 'bumps' of some form or other on either the substrate or the chip. The devices can be mounted either by thermo-compression or ultrasonic bonding.



(a) Diagrammatic representation of 'flip-chip' technique.



(b) Section through substrate and flipped chip.

Fig. 7. 'Flip-chip' method of mounting components.

Using this method, transistors can be connected to gold conductors on a substrate such that the shear force necessary to break the three bonds is in the region of 75–100 grammes.

The use of 'flip-chip' techniques also provides an attractive method of connecting together logical integrated circuits to thin film interconnecting patterns of conductors with cross-overs. This is an extremely attractive alternative to using the logical elements in TO-5 headers or flat-packs, since there are fewer interconnection interfaces, thus leading to more reliable, cheaper and smaller circuit assemblies.

2.4.2 The use of active substrates

A circuit with an active substrate is one in which the complement of active devices is diffused into silicon at the positions and passive components prepared on the insulating silicon oxide as substrate (see Fig. 8).

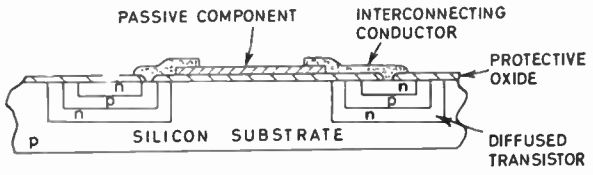


Fig. 8. Section through a thin film resistor on an active substrate.

The necessary interconnections between active and passive components are obtained by processing thin film conductors which pass through holes etched where necessary in the oxide. This type of circuit has been called a hybrid circuit.

3. Applications of Thin Film Techniques to Electronic Circuitry

The main aim in the use of thin film techniques in electronic circuitry is to provide an economical and reliable method of making the circuits.

The most important step leading to low cost devices is to make them as small as possible, provided that this does not cause any significant decrease in yield or loss of tolerances, so that large numbers of the devices can be handled in any process.

The degree to which the dimensions of resistors can be reduced and still maintain a good tolerance, is indicated in Fig. 2, in which the narrowest resistors are 0·0005 in wide. Also, capacitors with up to 1 pF per 0·001 in square can be made.

The following selection of circuits demonstrate how thin films can be used to advantage.

3.1 *An Audio Frequency* RC *Oscillator*

The particular requirement was for a low frequency oscillator (6·6 kHz) having a temperature coefficient of frequency better than 700 parts in 10^6 per degC. Since this is not easily obtained by using diffused passive components, thin film resistors and capacitors were chosen.

The circuit diagram of the oscillator is given in Fig. 9 and a photograph of the device, Fig. 10, shows the thin film resistors on the active substrate. The size of the silicon chip is 0·05 × 0·07 in.

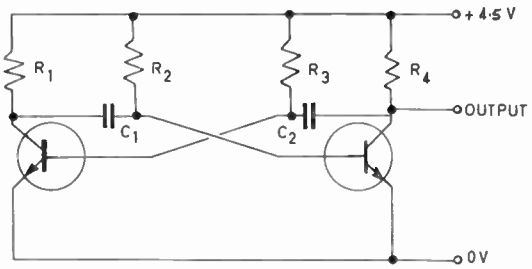


Fig. 9. Circuit diagram of an audio frequency R–C oscillator.
($R_1 = R_4 = 4\ k\Omega \pm 5\%$, $R_2 = R_3 = 90\ k\Omega$–$140\ k\Omega \pm 5\%$, $C_1 = C_2 = 1000\ pF \pm 5\%$)

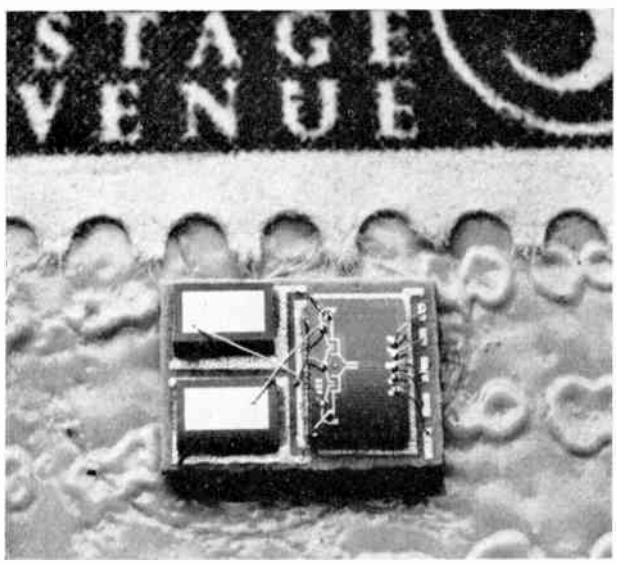


Fig. 10. An audio frequency RC oscillator.

The resistor material is arranged to be 200 Ω/square, the resistors are 0·00075 in wide and require a tolerance of ±5%. The resistor taps (see Fig. 10) were introduced during the design stage as a precaution, since this was the first circuit to be made with such narrow, high value resistors.

The anodized tantalum capacitors also require a tolerance of ±5%. The nichrome/gold counter electrodes are 0·030 in × 0·045 in in area (i.e. 0·75 pF per 0·001 in square).

The measured temperature coefficient of frequency is 430 parts in 10^6 per deg C and this can be accounted for by the temperature variation of the transistor emitter-base voltage. This dependence on transistor temperature effects could be reduced, if necessary, by designing the device to operate with higher power supplies.

In this design the dissipation is only 160 μW for each resistor when conducting, and this represents a power density of less than 1 W/in², and can be ignored.

3.2 *A Wide-band Amplifier*

A wide-band amplifier with an a.g.c. attenuator stage has been developed[6] which provides amplification of about 20 dB per stage (current gain of about 9) over the frequency range 0 to 120 MHz. The circuit has a low input impedance so that several circuits can be cascaded, provided that the necessary coupling capacitors are added. The circuit is constructed on a single silicon chip and has thin film resistors deposited on an active substrate. The circuit diagram is shown in Fig. 11 and a photograph of the device, without coupling or compensating capacitors, can be seen in Fig. 12.

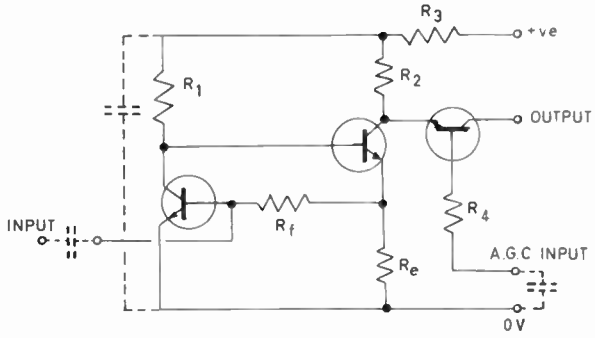


Fig. 11. Circuit diagram of a wideband amplifier.

R_1	2·7 kΩ	± 20%	R_f	880 Ω	± 20%
R_2	400 Ω	± 20%	R_e	88 Ω	± 20%
R_3	50 Ω	± 20%	Ratio $\frac{R_f}{R_e}$	= 10	± 5%
R_4	400 Ω	± 20%			

The resistors in the amplifier require a sheet resistivity of 90 Ω/square, which is a low value for diffused resistors made from a transistor base diffusion. Thus, in this particular application, the use of thin films for resistors enables the transistor and resistor design to be considered and optimized separately.

The maximum power dissipation in the resistors is about 400 W/in². Resistor tolerances of ±20% are adequate but since the gain of the amplifier is deter-

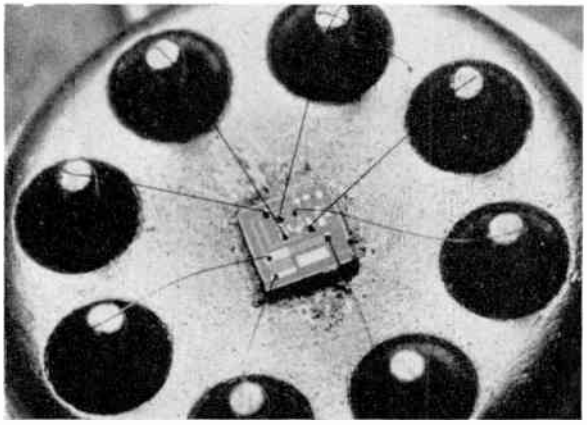


Fig. 12. A wideband amplifier.

mined by the ratio of resistors R_f and R_e, the tolerance on this ratio is kept to better than $\pm 5\%$.

3.3 *Core-store Sense Amplifier using an Active Substrate*

This circuit was developed to provide the necessary amplification for use with a 1 μs ferrite-core store, and consists of a two-stage, d.c. coupled, differential amplifier followed by a detecting circuit. The amplifier gives full output for signals of 35 mV of either polarity, and rejects noise signals of up to 12 mV differential.

Figure 13 shows the circuit diagram of the amplifier and Fig. 14 shows a photograph of an amplifier die. Again, the amplifier consists of transistors diffused into the silicon and thin film resistors deposited on the silicon oxide.

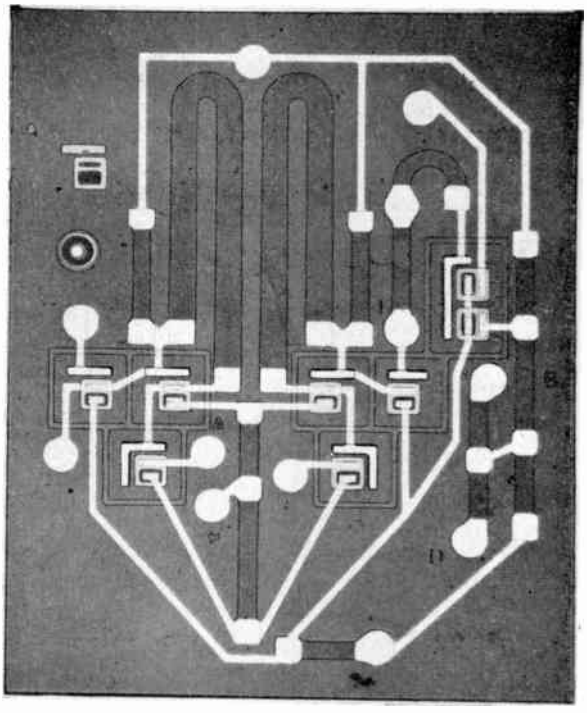
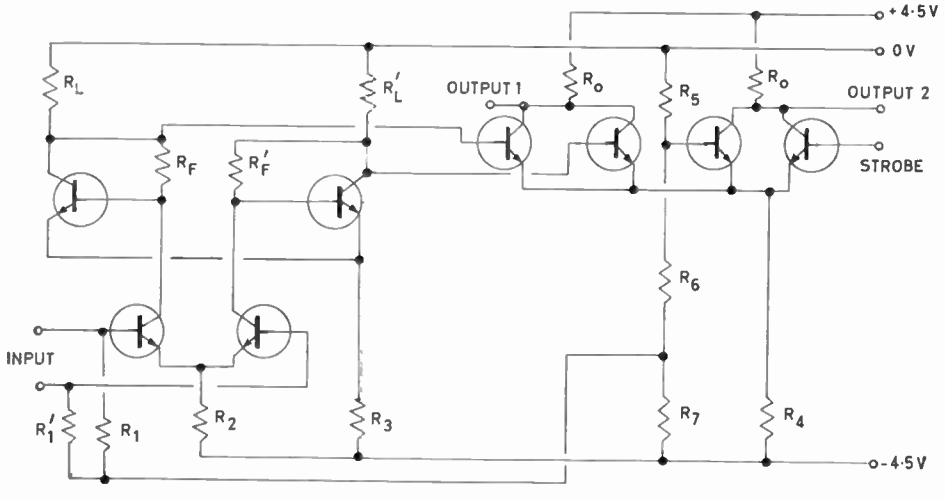


Fig. 14. A core-store sense amplifier chip.

The resistors are made with a sheet resistivity of 200 Ω/square and with tolerances of $\pm 2\%$ on ratios of R_f, $\pm 2\%$ on the ratio of R_f to R_2 and a tolerance of $\pm 5\%$ on absolute values. These figures, together with the low temperature coefficient of resistance, are more than adequate for the requirements of the amplifier.



R_1 R_1'	500 Ω
R_2	1·3 kΩ
R_3	560 Ω
R_4	560 Ω
R_5	620 Ω
R_6	1 kΩ
R_7	620 Ω
R_F R_F'	4 kΩ
R_L R_L'	1 kΩ
R_o	900 Ω

Tolerances: Absolute values $\pm 5\%$
Ratio $R_F/R_2 \pm 2\%$

Fig. 13. Circuit diagram of a core-store sense amplifier.

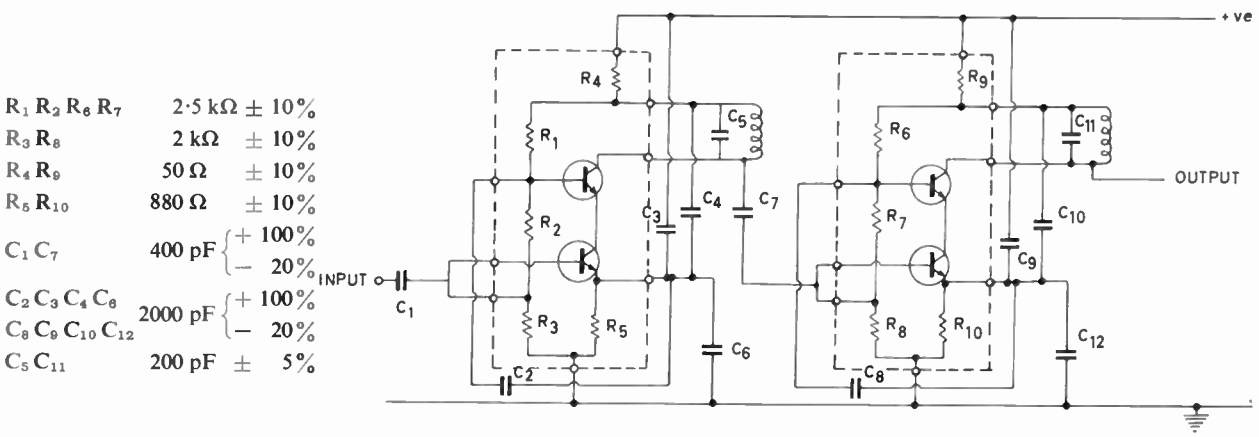


Fig. 15. Circuit diagram of a two-stage i.f. amplifier.

The power dissipation in resistor R_4 is 14 mW and this represents a power density of about 1·25 kW per square inch of resistor. The two resistors R_0 dissipate 25 mW each when their corresponding transistors are conducting, and this represents a power density of 1·5 kW per square inch of resistor. The performance of these resistors under a higher degree of loading is given in Fig. 4 and a comparison can be made as each amplifier dissipates a total of 50 mW under normal operating conditions.

3.4 *A Two-stage I.F. Amplifier*

This particular application demonstrates how thin film inductors, resistors and capacitors can be assembled together with active chips to form an i.f. amplifier. The active chips were designed some time ago for another application and thus were not adaptable to 'flip-chip' methods of mounting. However, in any future design 'flip-chip' mounting will be considered.

The circuit diagram is shown in Fig. 15 and a photograph of the device in Fig. 16. The two stages of amplification are designed to provide a centre frequency of 10·7 MHz, with a bandwidth of 3 MHz and a voltage gain of 40 dB.

3.5 *A Triple Entry* NAND *Gate*

This circuit demonstrates how well transistors can be mounted on thin film circuits by the 'flip-chip' technique to provide a very small and economic assembly, especially suited to a relatively small scale production.

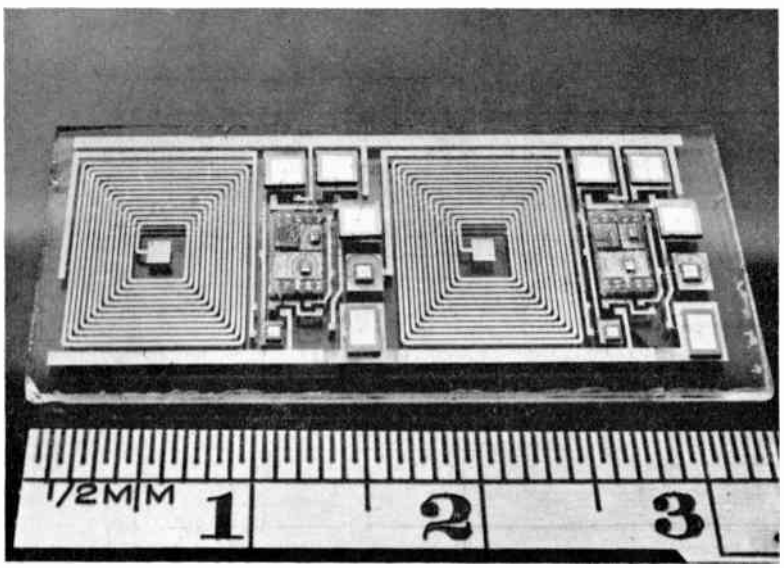


Fig. 16. A two-stage i.f. amplifier.

The circuit diagram, together with resistor values, is given in Fig. 17. The tolerances required of the resistors are ±5% on absolute values, ±5% on the ratio of the 10 kΩ to the 1·2 kΩ resistors and ±2% on the ratios of the 10 kΩ resistors. Figure 18 is a photograph of the device on a glass substrate and shows that the connections to power supplies and other units are by soldered copper wires.

The circuit was designed to provide a fan-in of three and a fan-out of four and to operate with switching speeds of the order of 1 μs.

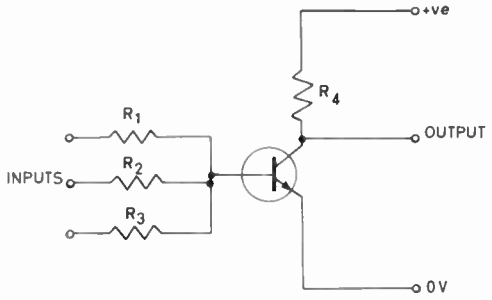


Fig. 17. Circuit diagram of a resistor–transistor logic element.

$R_1 R_2 R_3$ 10 kΩ ± 5%
R_4 1·2 kΩ ± 5%
$R_1 = R_2 = R_3$ within ± 2%
Ratios R_1 to R_4, R_2 to R_4, R_3 to R_4 ± 5%

The resistor dimensions are limited by power considerations, so that the 1·2 kΩ resistor, when fabricated from 300 ohms per square material and a power limitation of 20 W/in² has dimensions of 0·045 in × 0·180 in. With similar limitations on the 10 kΩ resistors, the dimensions are 0·005 in × 0·167 in.

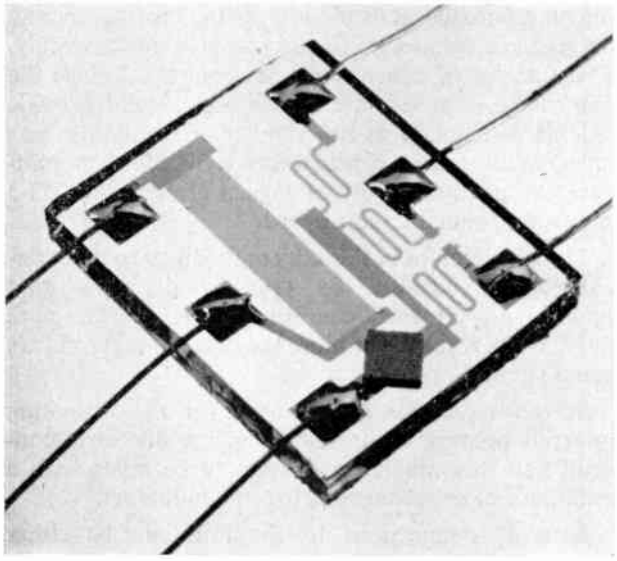


Fig. 18. A resistor–transistor logic element.

3.6 *A High-speed Serial Adder*

The serial adder was chosen to demonstrate how a group of diffused logical elements can be interconnected using 'flip-chip' techniques to mount the devices on to thin film interconnecting substrates.

Figure 19 shows the high-speed serial adder. The conductors are 0·005 in wide. Nichrome-gold with

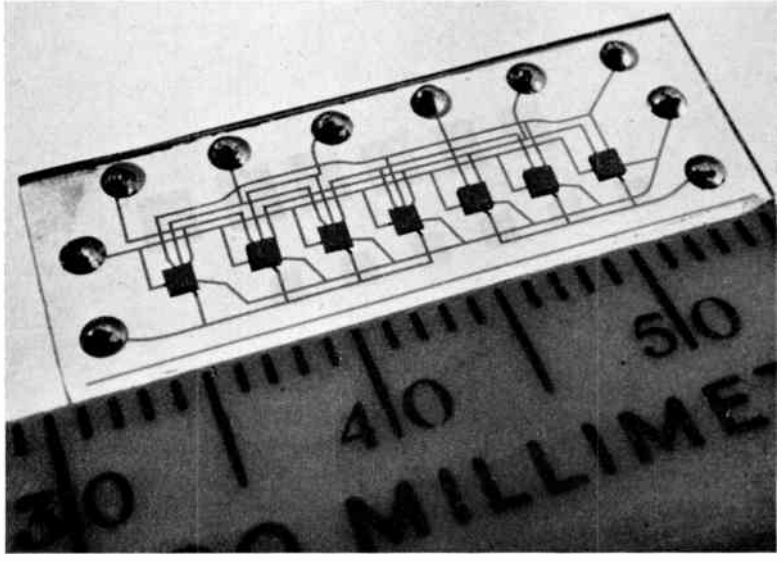


Fig. 19. A serial-adder using flipped chips.

silicon monoxide as insulation for the cross-overs and the seven integrated NOR gates complete the assembly. This method of construction is economical since the fabricating processes are simple and should provide reliable assemblies, as the number of interconnecting interfaces has been considerably reduced when compared with an assembly of logical elements in TO-5 cans on a printed circuit board.

The silicon chips in this adder are joined to the interconnecting substrate by 0·003 in diameter gold spheres, one at each connecting point. This gives satisfactory results, but more economical methods are being studied.

Silicon monoxide was chosen for the insulating material because of its relatively low dielectric constant and this enables crossovers to be made with a minimum of capacitance between conductors.

External connections to the unit can be either soldered copper wires, as in the triple entry NAND gate, or, if hermetic sealing is required, the device can be mounted in a large flat pack.

4. Conclusions

The paper shows that the use of sputtered tantalum enables thin film resistors and capacitors to be made with excellent properties, in particular, tolerances and ageing effects.

The range of circuits given to illustrate the application of thin film techniques is comprehensive and indicates that the circuits can be made small in size, and that the production yields are high.

5. References

1. D. A. McClean, N. Schwartz and E. Tidd, 'Developments in tantalum nitride resistors', *Intl. Conv. Rec. I.E.E.E.*, Part 9. (Session on Components, Materials and Production) 1964.
2. L. Holland, 'Vacuum Deposition of Thin Films', p. 401. (Chapman and Hall, London, 1956.)
3. D. A. McClean, N. Schwartz and E. D. Tidd, 'Tantalum-film technology', *Proc. I.E.E.E.*, **52**, pp. 1450–62, December 1964.
4. R. W. Berry and D. J. Sloan, 'Tantalum printed capacitors', *Proc. I.R.E.*, **47**, No. 6, pp. 1070–75, June 1959.
5. F. S. Maddocks and R. E. Thun, 'Properties of evaporated film capacitors', *J. Electrochemical Society*, **109**, No. 2, pp. 99–102, February 1962.
6. M. D. Wood, A. Bardsley and J. Walker, 'High frequency tuned amplifier using integrated circuits', *Proc. I.E.E.E.*, **52** pp. 1608–12, December 1964.

Manuscript received by the Institution on 22nd April 1966. (Paper No. 1110.)

U.D.C. 621.397.12

The Development of a Pay-Television System and Equipment

By

G. L. HAMBURGER,
Dr.-Ing., C.Eng., F.I.E.E., M.I.E.R.E.†

Presented at a Joint Meeting of the Institution's Television Group and the Royal Television Society in London on 30th November 1966.

Summary: The system described applies to wired-broadcasting networks of the multi-pair variety. The pay-television user establishes credit by the insertion of one or more coins into a special subscriber's unit. The established credit will be consumed at a faster rate for an expensive programme, and vice versa. On the other hand, the subscriber can at any time switch off, and preserve the remaining credit for some future occasion. If the credit is exhausted, the unit switches off, and no pay-television programme can be received until more money is inserted.

Means are provided in the central station from which the pay-television programmes emanate and from which the price is controlled to display numerically and to print out minute by minute how much money has been collected from viewing subscribers since the start of a particular programme. This is achieved by digital techniques combined with data transmission, and it constitutes a complete, automatic and instant accountancy system.

1. Introduction

Pay-television is any system which, in addition to or regardless of such charges as the purchase or rental of the set and aerial, service and repairs, and in most countries the licence, requires the viewer to make payments for specific programmes he wishes to see.

Such a concept is new to Great Britain but has been in operation in the U.S.A. and Canada for some time. Those systems, of which there are a number of technical variations, some of which have been put forward for use in Great Britain, have one feature in common.[8-12] Payment is made, or an obligation to pay is entered into, when the subscriber has decided to view a particular programme lasting a specific time. The charge for a sports event may be different from that demanded for a recital, but the charge is irrevocably incurred once the subscriber has made his decision to view. However, with a pay-television system having a *pro rata* basis, the viewer pays minute by minute as the programme proceeds. If he switches off for one reason or another he has in fact only paid for what he has actually viewed, and no more. This is obviously fair and economical for the use. Moreover, because of the intrinsic fairness, the programme company will also benefit, since the viewer will switch on knowing that he can always switch off having spent a very small amount—and he may well continue to see the whole programme.

The present system of pay-television has been planned as a time-limited experiment in Great Britain, with the restriction that it is permitted to operate only over wired-broadcast systems, aerial transmission being strictly excluded.

The two determining factors of the Pay-TV system at present operating in the U.K. are thus:

(1) Broadcasting over wired networks only.

(2) *Pro rata* payment by the subscriber.

Having thus laid down the operating conditions of this service, two further desirable requirements had to be met:

(3) The programme price per unit time should be controlled centrally by the company originating the programme.

(4) The actual programme consumption should be measurable so that the takings for a particular programme are at any time known at the centre of the network.

2. Outline of the Pay-Television System

The wired-broadcast system which is being used with the Pay-TV system is of the multi-pair variety. The cable consists of four pairs of conductors in an earthed flexible copper tape sheath, each pair being twisted at a different pitch to minimize crosstalk. Each pair is basically balanced against earth and carries a television programme using the lower vestigial sideband of a carrier of 3·76 MHz. The so-called viewing units are somewhat simplified television receivers and programme selection is by means of a rotary switch selecting between the four balanced pairs, thus

† Sangamo Weston Limited, Enfield, Middlesex.

obviating any necessity of tuning from channel to channel, the crosstalk between pairs being negligibly small. The viewing units are connected to the pairs of the feeder cable via a small auto-transformer so that the impedance with which the tapping point of the cable is loaded by the set is negligibly high compared with the wave impedance of the pair. A geographic area of just over one mile radius can be covered by a number of radial feeder cables emanating from a sub-station in the centre, the radii having spurs at appropriate points. The feeders are energized at their origin in the substation by suitable amplifiers. To cover larger areas further substations must be installed. The substations themselves obtain their input signals via landlines from a central station and thus can, of course, be any required distance away from the central station. Each conductor pair also carries a direct audio programme for sound broadcasting.

In order to explain the Pay-TV system, consider Fig. 1 which depicts the fundamental electrical circuits employed. In the upper part it shows how the vision channel energizes the feeder pair via a r.f. transformer with a split secondary. The viewing unit, instead of being strung across the pair via the tapping transformer (not shown), is connected to it via a subscriber's unit. This unit is essentially a series circuit consisting of an on/off switch, a credit switch, a 12·4 kΩ resistor, a moving coil of approximately 5·6 kΩ resistance, and a 10 μF capacitor. When the two switches are closed the pay-television programme can be received.

The manual on/off switch is under the control of the subscriber. This is the switch he uses to turn on the pay-television programme when he wishes to view and to turn it off if he so desires.

The position of the credit switch depends only on the credit the subscriber has established. Inserting one or more coins and turning a knob automatically closes the switch and establishes credit. As the programme proceeds this credit is being reduced by small increments. When it is completely consumed the credit switch automatically opens and no further pay-television programmes can be received unless another coin is inserted.

The 12·4 kΩ resistor together with the 5·6 kΩ moving coil keeps the impedance high for audio frequencies.

The 10 μF capacitor has a most important role. It can be seen that it is in circuit only when pay-television is being viewed. Thus every active pay-television viewer throws a 10 μF capacitor across the feeder—admittedly via 18 kΩ. Nevertheless if we were to measure the total capacitance of a feeder we would have a direct measure of the number of subscribers viewing at this moment. This is precisely what this scheme does. Owing to the necessary series resistances and also the many thousands of microfarads to be measured, standard bridge methods would be quite unsuitable for such a measurement.

The method employed is to charge all capacitors by applying to the feeder, say, a direct potential of 50V d.c. for long enough, say 10 seconds, and then to discharge the whole network through an integrator which measures the time integral of the discharge current; in other words, it measures the charge which originally resided in all capacitors of viewing subscribers. If the voltage is kept constant the charge can be considered a direct measure of the capacitance and, therefore, of the number of subscribers watching that particular programme. The factors affecting the accuracy with which the network parameters can be measured are discussed in Appendix 9.1.

2.1 *Capacitance Measurement and Subscriber Counting*

Now let us look closer at the moving coil of the subscriber's unit, which is of the kind contained in any typical d.c. milliammeter. When we applied 50V to the feeder the current would go into the 10 μF capacitor and the moving coil would rotate in one direction to its end stop. When we discharged the feeder the current would come out of the capacitor and the coil would move the other way. This to and fro rotation of the moving coil during one charge and discharge cycle—a mechanical action taking place within the subscriber's unit—can be used as an increment by which the established credit is being reduced. How this is done will be discussed a little further on.

Returning to Fig. 1, let us examine how the capacitance measurement is made. When the charge transistor TR1 is switched on, TR2 being off, the battery voltage of 50V is put across the feeder by passing through the low-pass filter and through the secondary winding of the h.f. transformer. The secondary is split from the point of view of d.c. to allow the two legs of the feeder to carry the charging potential. A capacitor closes the circuit for a.f. and r.f.

Having applied the potential long enough to ensure steady state conditions TR1 is cut off and the discharge transistor TR2 is switched on. A circuit is established through the integrator and all the charge residing in the sum total of subscriber's capacitors must pass through this integrator. This path is left open long enough to permit complete discharge.

The integrator is basically of the well-known Grassot fluxmeter type, a device of considerable antiquity in electrical technology, which was modified for the present application.[2] A fluxmeter is a moving-coil instrument with the special feature that instead of the usual restoring springs the current is fed to the moving

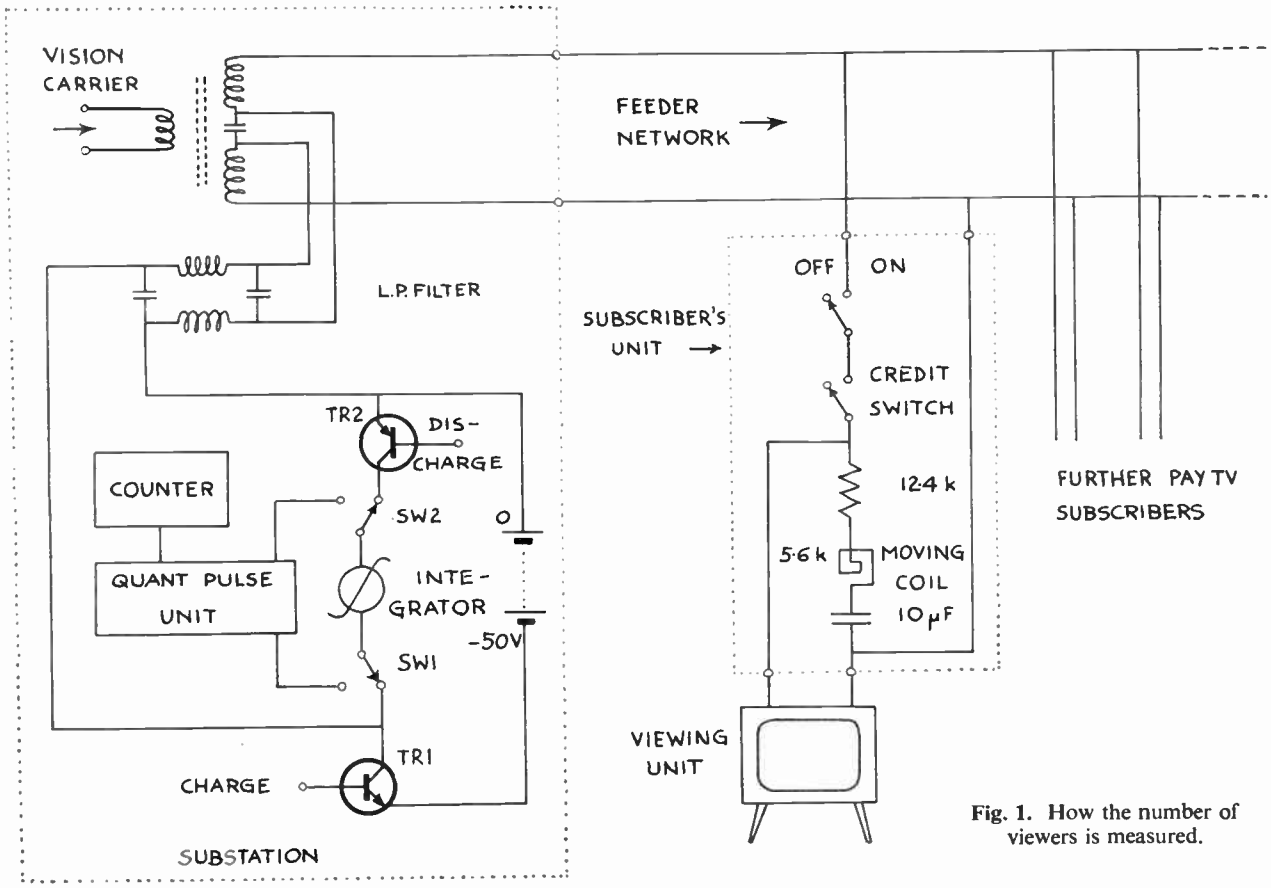


Fig. 1. How the number of viewers is measured.

coil by means of extremely fine ligaments which ideally do not exert on to the coil any torque whatever. Due to frame and circuit damping the coil resists any change of position. A constant current passing through the coil will cause the pointer to move at constant speed upscale. When the current is interrupted the pointer stops dead and remains in that position indefinitely—in theory, at least. In practice we are still faced with the fact that even the finest ligament has some residual springiness and that even the most carefully balanced movement still exhibits some degree of unbalance and, hence, torque.

The discharge current from the capacitors causes the integrator to move upscale and come to rest when they are fully discharged. One could, of course, read the value off the scale but this is not much use when bearing in mind that we have many unattended substations, the measuring results of which have to be correlated at a central point.

It was clear right from the start of this development that the only way of dealing with this problem was to convert the analogue indication into digital information. This can then be handled and processed by reliable means. Various schemes were considered such as light beams playing over circular Gray-code screens and similar arrangements, but they were all rejected due to complications, and above all, because every moving-coil instrument exhibits a certain degree of deflectional non-linearity. Consequently, any coded screen would have to incorporate this non-linearity which, of course, varies from instrument to instrument.

The scheme chosen proved to be the simplest. The unknown discharge has moved the integrator pointer upscale where it has come to rest. All we have to do is to move the pointer back to where it started from by a known discharge in the opposite direction. A constant current precisely timed until the pointer hits the starting point is one possibility. A much better scheme, however, is to push a string of identical and accurately known current pulses of the right polarity through the moving coil and stop when the pointer has reached its starting point. The number of pulses is counted and the result is a direct measure of the original discharge, and hence the number of viewing subscribers.

The lower part of Fig. 1 demonstrates this principle. The two switches SW1 and SW2 are thrown to the left. The pulses which quantize the restoring stroke

are called the 'quant pulses', and are generated at the rate of about 10 000 pulses per second in the quant pulse unit. They pass through the integrator and return the pointer to the starting point, whilst their number is being counted. Not only does this scheme offer a very simple way of digitization but it also circumvents the problem of non-linearity of the instrument deflection since the extent of the deflection does not enter the picture at all. The integrator merely acts as a temporary repository of the charge measurement only to be subsequently annulled by putting through it precisely the same charge but this time quantized and metered, and in the opposite direction. Non-linearities, therefore, cancel out completely.

Two more factors had to be resolved in this connection. The starting point of this instrument had to be defined for it has no proper restoring torque and hence no natural and controlled zero position. Secondly, the instant when the pointer on its digitizing return stroke reaches the starting point has to be accurately detected so that the return stroke and the counter can be stopped.

In order to define the starting point the pointer has to be held by some means in a well-defined position. Although the instrument is essentially torqueless there is inevitably some residual drift due to the pull of the ligaments, however feeble, to mechanical unbalances, to minute magnetic occlusions in the nominally non-ferrous moving coil system, and even to air currents. An innocuous method for zero holding is based on electrostatic attraction. Figure 2 shows how the radial tail of the electrically-earthed pointer system passes between the two prongs of a radially-positioned and stationary fork. If this fork carries a potential of a few

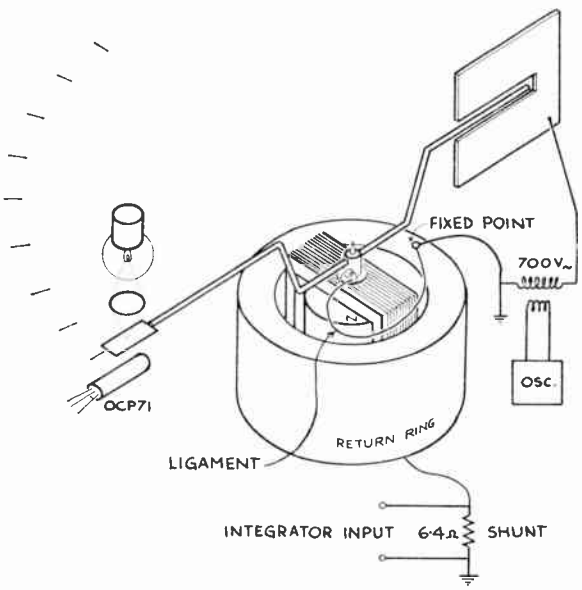


Fig. 2. The essential features of the integrator.

Fig. 3. The Pay-TV subscriber's unit.

hundred volts (a.c. is used in the present case), the electrostatic attraction keeps the pointer tail between the prongs. The potential is removed, i.e. the pointer unclamped, immediately prior to the commencement of the capacitor discharge which moves the pointer upscale.

As the pointer comes back on the digitizing stroke its passing through the zero position is detected by a photoelectric arrangement consisting of light source and photo-transistor. It has been very carefully designed to give a reliable output pulse which is reproducible to within a few minutes of arc of the pointer movement.

A great deal of care was lavished on to the design and execution of this whole capacitance measuring and digitizing arrangement. It is an unusual combination of old and new principles. It gives an overall accuracy of the order of 1 part in 1000 and is quite capable of measuring capacitance of up to tens of thousands of microfarads. Accuracy considerations are discussed in Appendix 9.3.

2.2 *Control of Subscriber's Credit*

The subscriber's unit[3] is, in its present form, a free standing unit (Fig. 3) which is inserted between the so-called drop-down lead, which itself taps the main transmission cable, and the television viewing unit. The function of the unit is to enable the pay-television subscriber to establish credit by inserting one or more, in the present case two-shilling, coins. This then keeps the path open for receiving any pay-television programme at any time—as long as some credit is left. Whether or not this credit shall be made use of depends on the subscriber, namely, it depends on whether or not he switches on or off. If the credit is

being consumed and exhausted during a programme the unit simply cuts off the picture. Although the insertion of a coin immediately restores the picture, such occurrences can be a little irritating and, therefore, a credit indicator is also fitted so that the subscriber can ensure before the programme begins that enough credit has been established.

Figure 4 shows in a symbolic and much simplified form how these functions are achieved. The coin mechanism (not shown in this figure) is of the kind where coin insertion alone is not enough to complete the function. A knob must be rotated by the subscriber before the coin can actually drop into the coin box. Only then can another coin be accepted. This system has the great advantage that any basic power or energy required for the functioning of the unit is supplied by the subscriber turning the knob. The effect is that wheel 1 in Fig. 4 rotates in the direction of the arrow by a certain amount. We can recognize in Fig. 4 a clockspring, a differential, a credit disk and indicator and the credit switch. We can also see a clock-type escapement, pallet and balance wheel. This balance wheel sits directly on the moving coil which is in the circuit as previously described. The moving coil is ligament supplied, and hence has no restoring torque.

In the state before coin insertion the pawl of the credit switch rests in the notch of the credit disk The credit switch is open, no pay-television programme can be received. Inserting the coin and turning the knob causes wheel 1 to rotate in the sense of the single arrow. Since wheel 4 of the differential is held stationary by the locked escapement it is the planet wheel 3 which moves in the single arrow direction. It takes the centre staff along which winds up the clockspring, thus storing the subscriber's knob-turning energy, and it rotates the credit disk, thus disengaging the notch from the pawl, and the credit switch closes the circuit.

Suppose now the subscriber switches on and a capacitance measurement is performed by the substation. The charge current will rotate the moving coil in one direction to its end-stop and the discharge current will cause the opposite movement to the other end-stop. The effect of this will be that the pallet lever has made one oscillation to and fro—one tick-tock—and the escapement has moved one full tooth further on in the double arrow direction. Hence wheel 4 has via 7 and 6 moved in the double arrow direction too. Since wheel 2 is now stationary (held by the coin wheel 1) the planet wheel 3 and the centre staff and credit disk have moved in the double arrow direction by a small increment, the clockspring driving the centre staff, the escapement controlling it. The important thing is that the notch in the credit disk has made a tiny step towards the waiting pawl of the credit

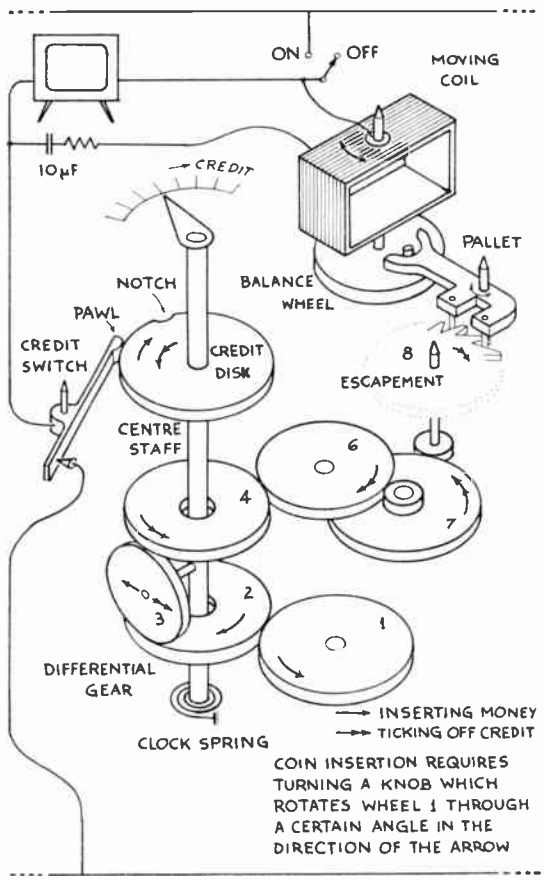


Fig. 4. How credit is consumed in the subscriber's unit.

switch. Since it has reduced the credit this step must really be called a decrement. A sufficient number of such further decrements, and the pawl will once again drop into the notch and switch off the programme, the credit being exhausted.

Considering this basic build-up from an energy point of view it can be seen that the subscriber actually supplies the energy to close the credit switch against its spring and stores further energy in winding up the clock spring when turning the coin knob (not shown). The capacitor charge and discharge current passing through the moving coil only controls the 'letting out' of the energy stored in the clock spring. It only unlocks the escapement tooth by tooth which requires control energy that is only a very small fraction of the clock spring energy thus controlled. In fact a current of far less than 400 μA is sufficient to provide this unlocking torque.

The factors in the subscriber's unit affecting accuracy and reliability are described in Appendix 9.2.

We have just seen that the effect of a charge and discharge of the feeder is not only that a capacitance measurement was performed but also that the credit

of all viewing subscribers was reduced by the same decrement. The feeder charge and discharge cycle has, therefore, a double function: measurement and credit control. Obviously the more frequently this measurement is done the faster the rate at which the credit will be reduced. Here, therefore, is a means of controlling from the substation and ultimately from a central point, how much money is to be collected per unit time for a particular programme.

However, we must recall that in order to make an accurate measurement charge and discharge times of 8 to 10 seconds each are needed. Add to these the digitizing time and the total cycle-time for one measurement is 20 to 30 seconds. This is a long time and does not leave much scope for expensive programmes unless the money value of each decrement is so high that it would be almost unusable for cheap programmes.

In order to achieve more flexibility it was decided to split the double functions of measurement and credit control of the charge and discharge cycle.[4] Bearing in mind that it is the long duration of the cycle which causes difficulty, it was decided to make a sampling measurement once every 2½ minutes only, and to take the full time of 20 to 30 seconds over it. We then assume that the popularity pattern stays constant for these 2½ minutes. The sampling measurement produces one credit decrement but in order to add further decrements for price control we can apply fairly rapid charge and discharge cycles. Obviously there is no longer any need to charge or discharge fully, in fact, one can go as fast as the mechanical inertia of the balance wheel and coil will permit. In practice 1 second was chosen for a charge and discharge cycle, and since their only purpose is to ensure that the viewer pays his programme price, these 1 second cycles have been called price pulses.[5]

For the practical system it was decided to make the value of one decrement one farthing. Therefore, since there are 24 2½ minute periods per hour one decrement per 2½ minute period would collect 24 farthings or sixpence per hour. This is then the minimum programme price per hour. Evidently 10 minutes of such a programme would cost the subscriber only 1d. A half-crown-per-hour programme would require the measuring cycle to be followed by four additional price pulses in every 2½ minute period.

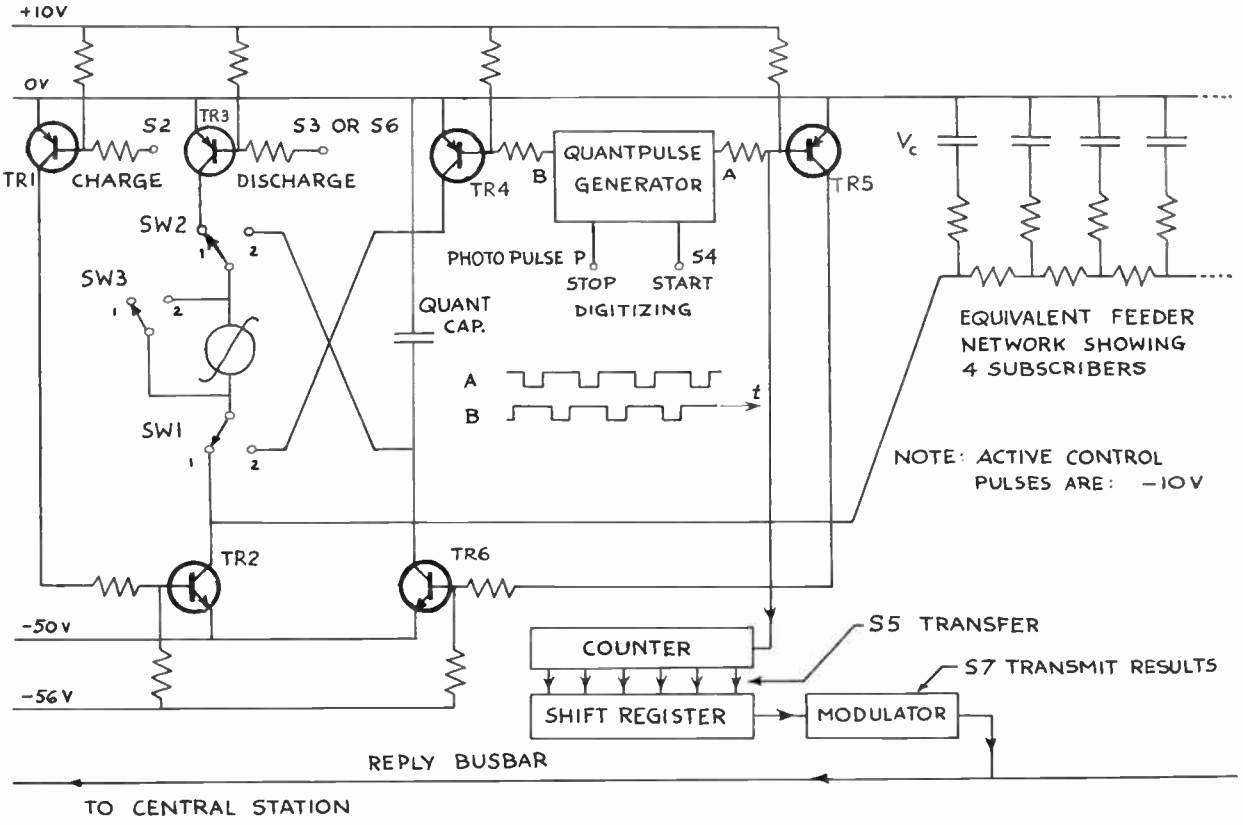
The accuracy of the integrator is discussed in Appendix 9.3.

The salient features of the Pay-TV system have now been explained and it will be useful to fill in some of the more interesting details.

3. The Capacitance Digitizer in the Substation

The measurement of viewer density is a measurement of the sum total of all the capacitors of all the subscriber's units which are switched on. In the multi-pair wired broadcast system the only place where such measurement can be done is the substation for only there is a complete d.c. connection to all the feeders radiating outward. In fact, it is at the split centre of the vision amplifier output transformer where we can make a d.c. measurement on all the feeders, as shown in Fig. 1. In a larger area there may be dozens of substations, and in order to know the viewer density of the whole area, capacitance measurements must be made in each substation and the results periodically be conveyed to the central station where they are accumulated. As already stated this calls for digitizing the capacitance measurement and conveying the digital result to the central station by means of a data transmission system.

The upper part of Fig. 5 gives more circuit details of the basic digitizing scheme in the substation, and the lower part of Fig. 5 shows a timing diagram of this scheme. A negative step signal S2 switches on transistors TR1 and TR2 causing the whole cable network to be charged to 50 volts. The switches SW1, 2 and 3 are in the rest position 1 as shown. After a period of, say, 10 seconds the step signal S2 is removed and after a small but definite delay of 5 ms another step signal S3 is applied to TR3 which, via switch SW2, the integrator, and SW1, discharges the network. After another period of 10 seconds step signal S3 is removed, switches SW1 and 2 are thrown over to position 2 and the digitizing command pulse S4 is applied to the push-pull quant pulse generator. It triggers the latter to produce two pulse trains A and B which are interleaved in time as shown in the inset timing diagram in such a way that there is always a pause between alternate pulses. It can be seen that the A pulse charges the quant capacitor by switching on TR5 and TR6, and the B pulse causes the quant capacitors to discharge through the integrator by switching on TR4. The pauses between A and B pulses ensure that one transistor is securely switched off before the other is switched on lest TR4 and TR6 would temporarily short-circuit the 50-V supply through the integrator, only to be burnt out. The digitizing discharge pulses which, as can be seen from the cross-over connection, pass through the integrator in a sense opposite to that of the previous network discharge pulse, drive the integrator movement back towards zero. The frequency of the A and B pulses is approximately 10 kHz and the number of pulses required to drive the integrator back to zero is registered in a counter. When zero is reached the pointer interrupts a light-beam shining on a photo-transistor the step-output of which, the photopulse P in Fig. 5, shuts down the quant pulse generator. The number standing in the counter is therefore a measure of the total capacitance of the cable network. Moreover the photopulse P switches on a transistor oscillator



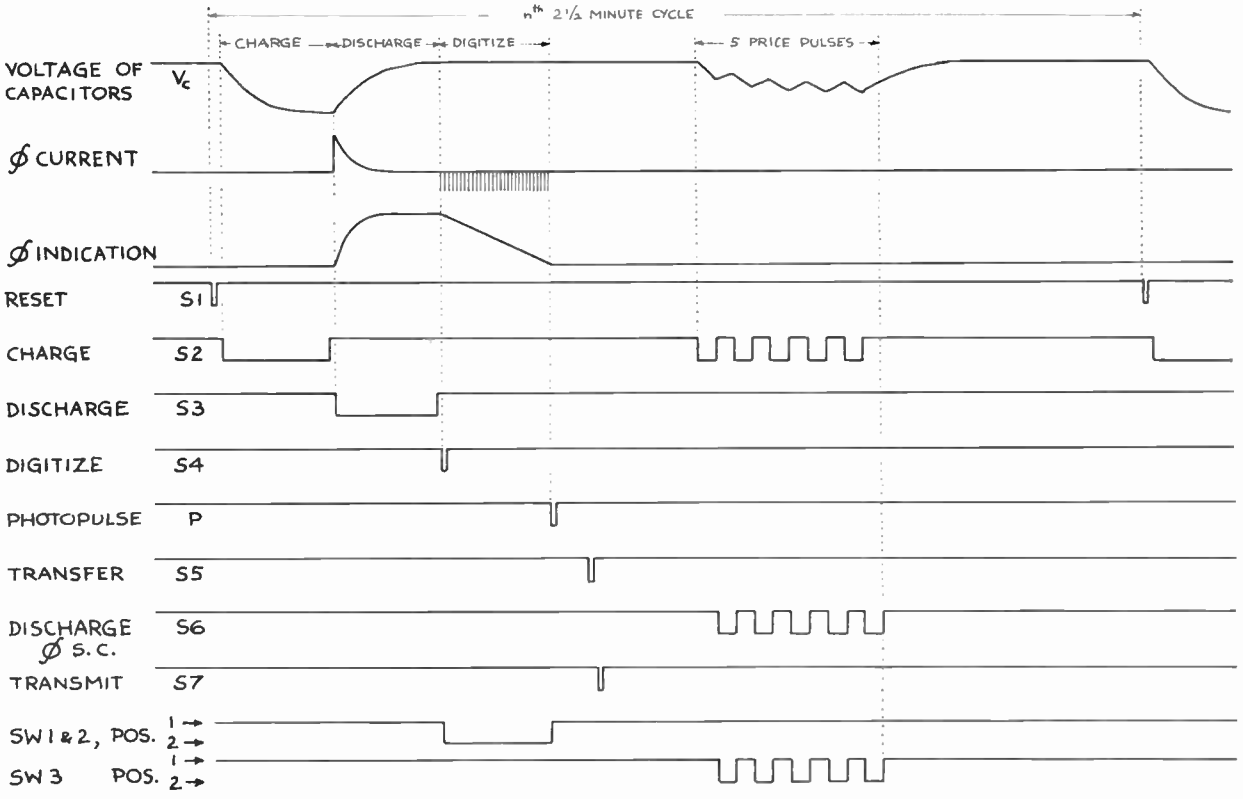


Fig. 5. Basic circuit of the capacitance measuring and price pulsing facility of the substation with its timing chart.

producing about 700 volts at a few tens of kilohertz which are applied between pointer tail of the integrator movement and a stationary fork-shaped electrode between the two prongs of which the pointer tail has come to rest. The pointer tail is at earth potential and the fork is at 700 volts a.c. The electrostatic forces keep the nominally torqueless movement in the zero position against any residual drift tendencies. Thus an accurate zero datum is maintained from which the pointer is, at the next measuring cycle, driven upscale, after the oscillator has of course been switched off, and subsequently the pointer is digitized back to the photo-electric light barrier. The latter is set up to coincide with the position of the zero holding electrode.

The number standing in the counter is transferred in parallel to a shift register by a command pulse S5. A further command pulse S7 causes the number in its coded form to be marched through a modulator over a signal line to the central station for further processing. This terminates the actual measuring sequence. The number of subscribers viewing at this moment is now known and has been conveyed to the central station. Also a farthing (¼d.) has been 'collected' from every viewing subscriber. Since there are 24 2½ minute periods in the hour this would only collect 6d. per hour. If 3 shillings had to be collected per hour five further price pulses have now to be introduced.

With switches SW1 and 2 back in the rest position 1, an S2 command signal lasting this time for ½ second, will charge the cable network, or at least partially charge it. This must be followed by a discharge, but this time not through the integrator. Therefore, switch SW3 is closed to short-circuit the integrator. This action is included in the command signal S6 which is otherwise similar to S3 as already described. S6 lasting for another ½ second is followed by S2, and S6, and S2, etc., until the requisite number of price pulses has been established. After that there is a pause until the end of the 2½ minute cycle whence another sequence will commence.

The switches SW are in fact reed relays and as a matter of design principle these reed switches are always operated under zero contact power conditions. They are only used to prepare paths under no-current conditions and the actual switching on or off is always done by the transistors. Failure of the reed switches used in this manner could, therefore, only be caused by mechanical fatigue, which is practically unknown in this case.

Care is therefore taken in the timing operations to ensure a 5 ms pause between every event necessitating reed switching so that the reed contacts have ample time to open or close and do their little contact bounce before the appropriate transistors do the actual switching.

Another most important principle should here be pointed out. The substation, essentials of which were just described, is arranged to be operated as a complete slave to the central station. This means that all signals S1 to S7 are initiated in the central station and transmitted to the substation by means of data transmission. This is vital when several substations are operated from one central station to avoid chaos should fault conditions arise anywhere in the system. Should a fault arise in one substation, for instance, if a command signal is misread, only that particular substation would fail to send back the correct information and would not interfere with the others. In any case there is a reset command signal S1 which precedes every 2½ minute cycle and resets the whole logic into the unique starting position so that every such cycle is started with a 'clean slate'.

4. Coded Command Receiver in the Substation

The purpose of this unit is to receive the various commands S1 to S7 from the controlling central station. These command signals must be transmitted over the wire broadcast network. This is done by first of all binary coding them so that S3 would, for example, be 011. The individual bits are then transmitted sequentially by carrier shift modulation. In the development model a '0'-bit was represented by a burst of a frequency f_0 lasting 800 μs. This was followed by a pause of 200 μs whence, for instance, a subsequent '1'-bit would be sent out as a 800 μs long burst of frequency f_1, and so on. Thus a bit frequency of 1 kHz is achieved which is more than adequate for this application.

For the sake of reliability and simplicity, and because there is no particular restriction on bandwidth, it was decided to adopt an asynchronous data transmission system. Using this system means that pulse duration and pulse rate need not be accurately held, and that pulse trains, i.e. words or commands, can arrive at any time without requiring prior synchronization. Figure 6 shows how this was done.

The bursts of two possible carrier frequencies are sorted out by two filters, followed by a detector and a Schmitt trigger for each channel. Waveform (1) represents the '1'-bits and waveform (2) the '0'-bits. The positive edge of (1) sets the binary B3 to '1' where it remains. The next positive edge of (1) leaves B3 at '1', whereas the following positive edge at (2) sets B3 to '0', as shown in waveform (5). Binary B3 acts as the signal source for the following shift register B2, B1 and B0. The shift pulses (4), which are required to march the pulse train of ones and noughts into the shift register are derived from the trailing edges of the '1' or the '0' signals. Therefore, the antiphase outputs of the two Schmitts are fed into an OR-gate, the output of which, waveform (3), triggers a pulse shaper with

its positive edge. The pulse shaper produces a 15 μs pulse (4) for every positive step on its input. This train of shift pulses is fed to the shift busbar of the shift register. It can be seen from the time diagram that at the end of the third frequency burst the binaries B2, 1 and 0 contain the word 011.

The contents of the shift register are decoded in a conventional diode matrix which yields a signal on one of the seven possible output lines, in our case on the third line, namely S3. S3 then controls the logic of the digitizer (Fig. 5), it sets the reed switches into the appropriate positions, it makes TR3 conductive, etc., in other words, the command S3, 'discharge through the integrator', is being obeyed.

It can be seen that during the time of receiving the signal, i.e. whilst the information is being shifted into the register, the matrix would decode various meaningless signals. In order to prevent this the matrix output is prevented from playing havoc with the digitizer logic by the closed strobe gate. The latter is only opened after the end of the word by means of the end-of-word detector.

This works on a simple RC-charging principle reminiscent of the frame sync. separator in a television receiver. The composite charge and discharge waveshape (9) is squared in another Schmitt trigger to shape (10), and its trailing edge triggers another pulse shaper to give a strobe pulse (11) at the end of the word, which momentarily opens the gate to let the S3 signal through to the control logic of the digitizer.

It might be thought that the problem of transmitting seven command signals could have been solved more simply by choosing seven carrier frequencies being separated by seven filters at the receiving end. This is true, but in practice more signals will be needed for various check functions, for print-out, etc., and above all, more signals are required in order to address each substation individually. For the seven signals S1 to S7 in Fig. 6, all but the seventh are applied to all substations simultaneously, i.e. they all charge, discharge, digitize and transfer to the shift register and, indeed, price-pulse simultaneously. But the transmission of measuring results back to the central station, signal S7, is done by one substation at a time over the reply busbar (see Fig. 5). It is, therefore, necessary that the central station can address each substation

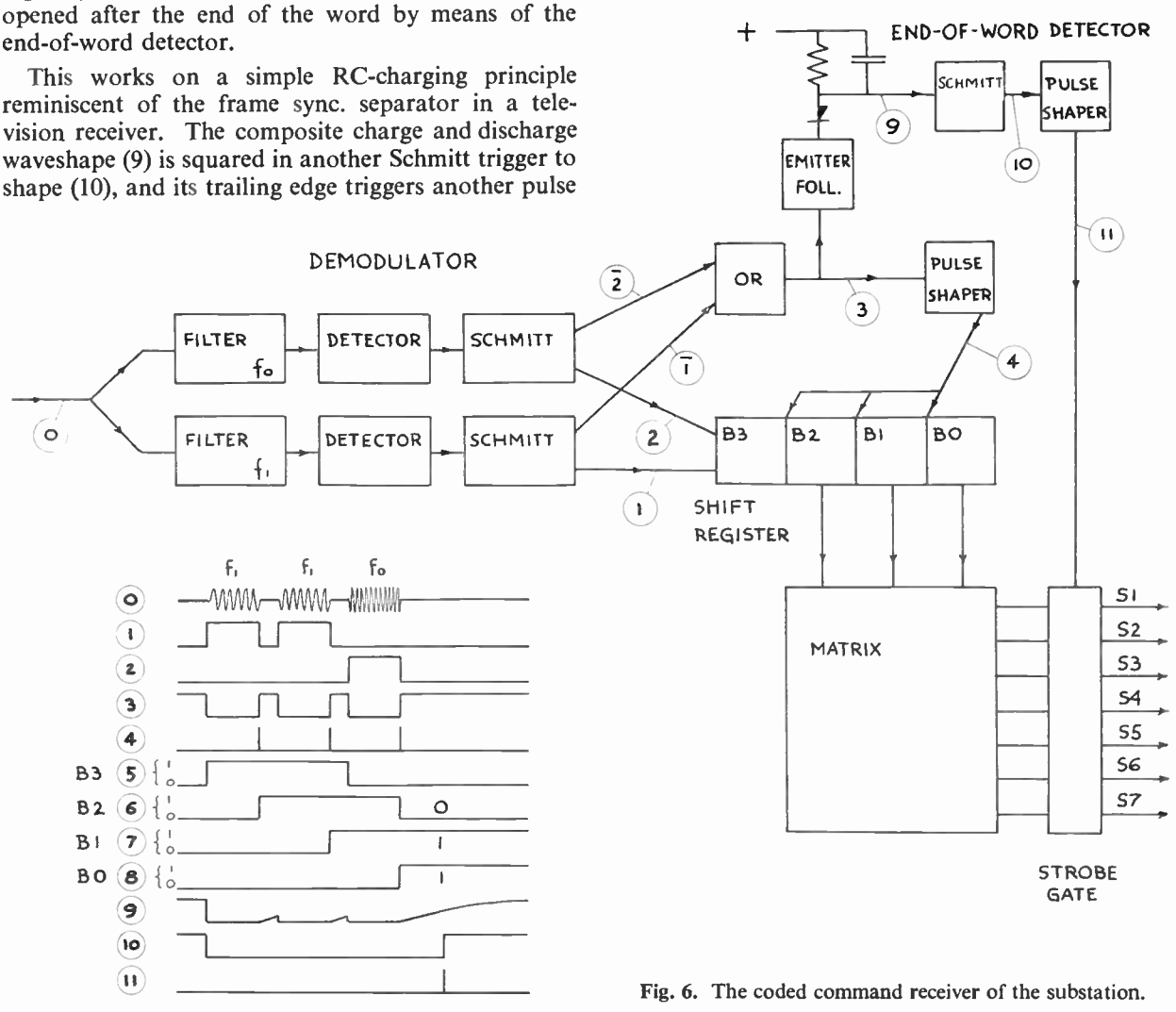


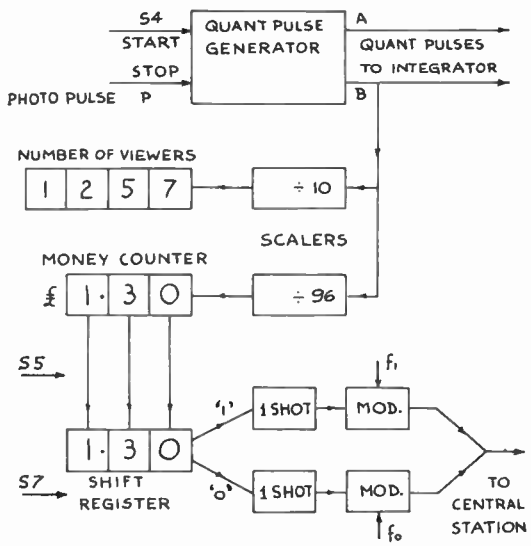
Fig. 6. The coded command receiver of the substation.

separately as well as simultaneously. This calls for further digits in the command code and it can easily be appreciated how much more flexible, how much more capable of expansion, and therefore how much simpler this digital approach is as compared with the alternative of a multiplicity of frequencies and filters with their inherent problems and requirements of constancy and accuracy. Adhering to the digital approach it can be seen that, by the addition of only one more binary bit, up to 15 command signals can be accommodated, and so on. This flexibility, coupled with the absence of multiple tuning problems led to the preference for the digital approach.

5. Money Conversion

It was claimed in the summary that this system provides 'instant accountancy' and the means of achieving this will be described in this section.

The value of one charge and discharge pulse as far as its effect on reducing the subscriber's credit is concerned was laid down as one farthing (¼d.). The capacitance with which one viewing subscriber contributes to the total is 10 μF, and finally, the quant capacitor was chosen to be 1 μF. These three statements form the first line of the table in Fig. 7. It was



NUMBER OF SUBSCRIBERS	MONEY PENNIES	£	μF	NUMBER OF QUANT PULSES
1	0·250		10	10
0·1	0·025		1	1
9·6	2·4	0·01	96	96
96	24·0	0·10	960	960
1257	314·0	1·30	12570	12570

Fig. 7. How quant pulses are converted into money values in the substation.

the aim to express all monies collected in terms of £0·01 and it can be seen from the table that this value of £0·01 or 2·4d. corresponds to 96 quant pulses. Hence if a scaler fed from quant pulses gives one output pulse for every 96 quant pulses that output pulse can be tucked away as £0·01 collected. This is shown also in the block diagram of Fig. 7. To put the fourth line of the table into words: the cable network with 96 viewing subscribers, when discharging, drives the integrator upscale. It then requires 960 quant pulses to return the integrator pointer to zero. The divide-by-96 scaler has then given out 10 output pulses at the value £0·01 each, i.e. £0·1 or 2s. has been collected, which is 96 farthings.

The money value due to one measurement now stands in the money counter. In the case of Fig. 7 £1·30 was collected from 1257 subscribers in that particular substation during the measuring phase at the beginning of one particular 2½ minute cycle. An S5 signal then transfers this information in parallel to a shift register, and an S7 signal marches its contents through a modulator via a busbar data channel back to the central station.

There, it is demodulated in the manner described in the coded command receiver shown in Fig. 6, and marched through an adder into the shift register called accumulator 1 as shown in Fig. 8. The central station then interrogates the next substation with an S7 signal. That substation now transmits the money results of its measuring phase via the busbar data channel to the central station where it is demodulated and passed, together with the contents of accumulator 1, through the adder, thereby obviously being added to the erstwhile contents of accumulator 1. There the sum of money values from two substations now resides until upon further interrogating the next substation the latter's results are added to it. This process goes on sequentially until all substations have sent their results in. In fact this whole interrogating and adding process is over in one or two seconds after which accumulator 1 contains the sum of all farthings collected from all viewing subscribers in the area covered by the central station due to the measuring phase at the beginning of one particular 2½ minute cycle.

It is now assumed that the viewing pattern does not vary during the next few seconds whilst the central station sends out the signals S2, S6, S2, S6, etc., which command all substations simultaneously to price-pulse its networks. Since every S2+S6 signal means another farthing from each subscriber track must be kept of it in the central station. This is achieved very simply by throwing switches SW1 and 2 over to position 2 and, for every price pulse, adding the recirculated, but constant, contents of accumulator 1 to those of accumulator 2. For $(n-1)$ price pulses the

contents of the latter have been increased by $(n-1)$ times that stored in accumulator 1. Therefore, one addition is missing for correct accounting, namely, the one pertaining to the measuring sequence itself. Therefore, to put matters right, the contents of accumulator 1 are once more added into accumulator 2, but without this time sending out price pulses to the subscribers. After this the switches are reset to position 1, the accumulator 1 is cleared, and is once again ready to receive the measuring results from the substations after the start of the next $2\frac{1}{2}$ minute cycle.

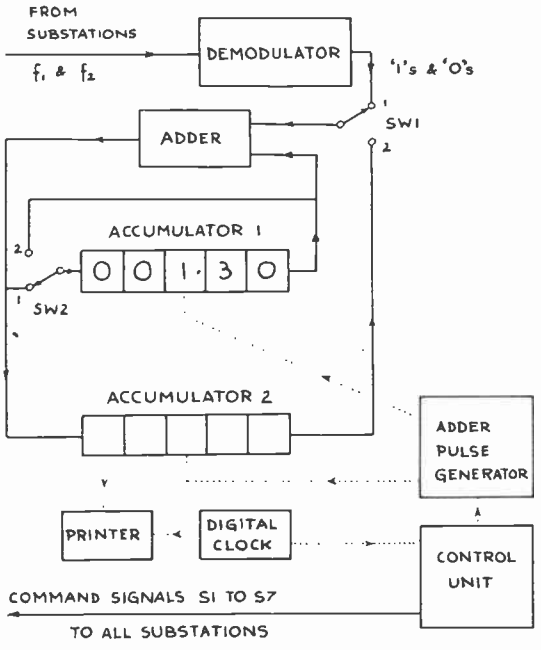


Fig. 8. How money collected from the viewing pay television subscribers is accounted for in the central station.

Provisions are made for the results of every $2\frac{1}{2}$ minute cycle to be numerically displayed and printed out together with the time derived from a digital clock so that complete records are automatically kept. Thus, at the end of every programme the sum total of all monies collected for that programme is accurately known in black and white, hence, the claim of instant accountancy.

Apart from the features already pointed out, such as the fairness of the *pro rata* payment, the importance of central programme price control and instant accountancy, the Pay-TV system has the attraction of simplicity of operation. Nearly all the complications are concentrated in a few locations, namely the central station and the unmanned substations, the latter being completely controlled by the former. The subscriber's units are passive, i.e. they require no power supplies. The tiny mechanical energy required to reduce the credit by letting out the spring, which was previously wound up by the subscriber himself inserting money, is supplied from the substation. The only 'servicing' that has to be done by the company is to empty the money boxes at appropriate intervals.

6. Concluding Note

It may perhaps be permissible to close this paper with a somewhat informal and unconventional note, which relates how this project had come about, for it represents, in the author's opinion, a good example of what can be achieved by enthusiasm and close co-operation between two Companies. In Summer 1960, Mr. K. A. Russell, Technical Director of British Relay Wireless Ltd. (B.R.W.), discussed with Mr. H. J. Lovegrove, Assistant Managing Director of Sangamo Weston Ltd. (S.W.) his basic idea for a *pro rata* pay-television system. It contained three fundamental elements of the present scheme, namely a subscriber's capacitor as a unit for subsequent popularity determination, together with its second function involving charge and discharge currents to actuate some electromechanical credit reducing device, and thirdly the basic suggestion of measuring the total capacitance by some integrating device for the purpose of viewer determination.[1] Sangamo Weston had previously designed and produced coin-operated fixed-rate collectors for B.R.W. This, together with S.W.'s experience in mass producing electromechanical devices and precision measuring instruments, and B.R.W.'s long experience in wired broadcast network operation formed a good basis for this enterprise.

The author, who was in charge of this project from its inception, could further broaden the approach by his own computing experience and thus the system logically emerged. This combined background also explains the curio of the venerable Grassot type of integrator, seemingly incongruously embedded in semiconductor circuitry and logic. However, at the time of the design some six years ago this appeared to be the best solution for an integrator with a drift time constant required to be of the order of an hour. No doubt this will one day be superseded by all solid-state circuitry as new devices and techniques become available, but for the time being it is rendering sterling service.

Although S.W.'s interests right from the start were vested essentially in the development and design of the subscriber's unit in order to mass-produce it, it was clear that this could not have been done out of context or in isolation. It was, therefore, necessary to pursue and develop fully the project in its entirety so that the specification of the subscriber's unit could be arrived at logically. Likewise the integrator was developed as an organic part of the whole system.

One substation and the central station equipment complete with numerical display and print-out facilities were actually built up in S.W.'s laboratories. A demonstration in December 1961 proved the feasibility of the scheme. This was followed by transfer of the experimental equipment to B.R.W. for field trials on their networks. Ultimately it was decided that the terminal equipment would be engineered and manufactured by Pye Ltd.

In the companion paper by Mr. K. A. Russell[7] considerations governing the choice of the system, the engineering aspects of the terminal equipment, interface problems between equipment and network, and actual reliability and operating experience of the Pay-TV system are more fully discussed.

7. Acknowledgments

The author wishes to thank Mr. Lovegrove and Mr. Russell for many helpful discussions, for much encouragement and understanding during the development, and thanks are due to the author's colleagues in all parts of the organization, Mr. A. E. Gane in particular. Finally the author wishes to express his thanks to the directors of both companies for their kind permission to publish this paper.

8. References and Bibliography

1. British Relay Wireless Ltd. British Patent Specification No. 1,002,298.
2. Sangamo Weston Ltd. British Patent Specification No. 1,049,851.
3. Sangamo Weston Ltd. British Patent Specification No. 1,046,451.
4. Sangamo Weston Ltd. British Patent Specification No. 1,046,022.
5. Sangamo Weston Ltd. British Patent Application No. 27,856/62.
6. G. L. Hamburger, Discussion on 'The design and application of modern permanent magnets', *J. Brit. Instn Radio Engrs*, **6**, p. 247, December 1946.
7. K. A. Russell, 'Pay-television: the engineering of the system and operational experience'. *J. Royal Television Soc.* To be published.
8. A. L. C. Webb and A. Ellett, 'Phonevision—an effective method for subscription television', *J. Brit. I.R.E.*, **16**, p. 205, April 1956.
9. L. S. White, 'Subscription television', *Wireless World*, **67**, p. 552, November 1961.
10. G. E. Partington, 'Pay television system', *Brit. Kinematography*, **42**, p. 16, January 1963.
11. G. M. Beedell, 'British Telemeter Home Viewing subscription television systems', *Proc. Soc. Relay Engrs*, **6**, No. 4, p. 85, March 1964.
12. P. Bass, 'Pay television by wire', *The Radio and Electronic Engineer*, **27**, No. 3, p. 209, March 1964.

9. Appendix

The Accuracy of Subscriber Determination

There are only three basic causes of inaccuracy of this measurement of the total charge of the subscriber's capacitors: the network, the subscriber's unit and the integrator. The most serious of these three is the network itself.

9.1 *Network Parameters Affecting Accuracy*

It is clear that the arch-enemy of any charge measurement is leakage. Part of the charge leaks away before it can be measured, part of it during the measurement. This was recognized at the very beginning of the project. Since its effect could have possibly nipped the whole scheme in the bud the first efforts were directed towards an assessment of accuracy as a function of leakage.

The task seemed formidable. The network cannot really be defined, and even less so its leakage. A multi-pair wire broadcast network is, within certain boundary rules, quite an arbitrary thing. Spurs go off in all directions to reach certain subscribers. Geography, withheld way-leaves, economic considerations, random time sequence of subscriber's connections, all result in this arbitrariness. The only tangible measurement one can reasonably make is the total loop resistance, r, and the total leakage, R. The loop resistance is the total copper resistance of the wire used, and is in practice measured by looking back into the end of a line or spur when its origin is short-circuited. The total leakage is measured at the feeding end by measuring across the capacitor in the split centre of the feed transformer (see Fig. 1). But of how many components this total leakage consists and where exactly they occur is quite impractical to ascertain. It was, therefore, decided to pursue an empirical path based on carefully designed model experiments and measurements, checking them with simple calculations and thus gaining physical insight into the mechanism of the error process. Simple lumped-constant networks were used to simulate the subscriber-loaded network. For the charge measurement a standard Grassot type of integrator was used reading directly against a calibrated scale, since the special integrator with digitization was hardly thought of in those early days. To make a start a very poor line with 60 ohm loop resistance and 500 ohm total leakage and without spurs was assumed. It was to be loaded with 400 subscribers. To avoid too much tedium the line was simulated by only four lumped sections by assuming a bunch of 100 subscribers in one location, A, another bunch in location, B, and similarly in C and D. The loop resistance was split up into four equal parts and the leakage remained unconnected for the time being (see Fig. 9(a)).

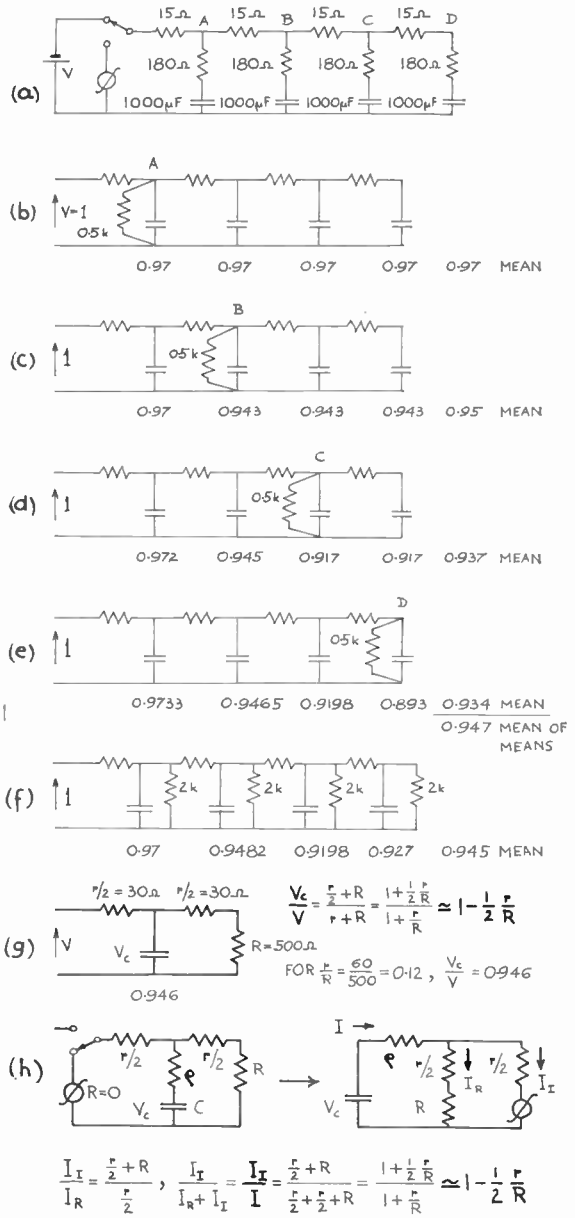


Fig. 9. Network analysis to assess errors due to leakage.

Inspection leads one to conclude that if V is applied to the input all capacitors charge up to the full voltage V and one would, upon discharging through the integrator, measure the same charge CV which would have been obtained had the experiment been done with just the four capacitors in parallel. This was confirmed by experiment and this charge was called 100%.

It was clear that the subscriber's series resistors, the 180-ohm resistors in Fig. 9(a) (being equivalent of 100 subscribers), could not in any way affect the charge condition so they were left out for the next experiments for the sake of further simplification.

The next step was to assess the charge or, what is a measure of it, the voltage of each capacitor when the lumped leakage of 500 ohm was successively placed in the four positions A to D. Figure 9(b) to (e) shows the circuit configurations. The voltages on every capacitor were calculated and noted as well as the mean voltage of all four capacitors representing the charge residing in them. It can be seen that the total charge depends very much on where, along the line, the lumped leakage has been placed; the further away from the sending end the worse the error. Then the leakage was distributed equally to all four sections and the voltages calculated. Naturally the last capacitor has the smallest charge, but the mean charge for the distributed case, i.e. 0·945, is better than in case (e) and worse than in case (b) but compares very well with the mean of the mean of cases (b) to (e)

One conclusion can be drawn right away: the knowledge of the overall leakage and loop resistance does not permit a precise statement of errors. It depends very much on the detail structure of the network.

However, in order to arrive at some rule of assessing the errors a simple equivalent circuit was stipulated. It applies the total capacitance between the centre point of the loop resistance, r, and the common return as shown in Fig 9(g). The simple arithmetic shows, for the assumed values of 60 ohm loop and 500 ohm leakage resistance, a capacitor voltage of 0·946 which is in surprisingly close agreement with 0·945 of the distributed case and the mean of 0·947 for the first four cases.

This appeared such a usable equivalent circuit that it was decided to base the next step on its validity rather than use the network of Fig. 9(a), namely to evaluate how this already reduced charge would be further diminished during the discharge measurement through the integrator due to further leakage during that measurement. The assumption here was that the integrator resistance was zero. Figure 9(h) shows the equivalent circuit in its original form, and also redrawn with the capacitor, C, as the source of energy discharging through the integrator branch and the leakage branch $(r/2+R)$ in parallel. The equation shows that the expression I_1/I, representing the percentage of the available current I which actually flows through the integrator is identical with that of Fig. 9(g). In other words: for a given ratio r/R of, say, 12%, the charge actually conveyed to the capacitors falls $\frac{1}{2}r/R$, i.e. 6% short of what it would be were there no leakage. But the integrator in turn sees of that only 94%, i.e. again 6% less. The measurement, therefore, is 12% short of what it should be. This is represented by the simple formula:

measured capacitance = actual capacitance $\times (1 - r/R)$

This simple relationship is of great value when turning the argument around and posing the question: How good must the network be to give an error of not worse than, say, 1%? Answer: the leakage resistance must be at least one hundred times greater than the loop resistance.

Unfortunately, however, one cannot fully correct for this error, because the assumption made in deriving this simple relationship, namely even distribution of loop and leakage resistance and subscriber's density on which it is based, need not necessarily hold in a practical case, nor indeed will the leakage be a fixed quantity. For an essentially overhead network the leakage will vary somewhat with the weather: in wet weather it will be worse than in dry weather.

Hence this relationship can only be used as a guide in assessing the possible errors. It should be pointed out that the value of the capacitances does nor enter into any of these considerations of errors nor, indeed, the subscriber's series resistor.

This statement must be qualified by saying that the errors that have been considered in the foregoing are those arising out of leakage. In other words we were concerned about a particular capacitance measurement without leakage (100%) and the measurement of the same capacitance in the presence of leakage

$$\left(100 - \frac{100r}{R}\,\%\right)$$

9.2 *Factors in the Subscriber's Unit Affecting the Accuracy and Reliability*

Of course, the capacitances which are the object of measurement have to be very accurate in themselves if the money conversion derived from these measurements is to be accurate. For this reason the 10 μF capacitor fitted to every subscriber's unit was specified to be within ±1%. However, since such a component cannot be obtained at an economic price it was decided from the outset that this specification had to be achieved by fitting a 9 μF ±10% capacitor, which can be economically obtained, and to pad up its capacitance to the correct value by adding in parallel suitable smaller standard components. It was felt to be worth while not to subject the final accuracy of the subscriber's capacitance to the vicissitudes of the human element in the shape of a possibly inattentive operator. For this reason an automatic padding bridge was specially developed and constructed, and incorporated into the assembly flow line. The operator has on her bench seven jigs arranged in a semicircle, each jig topped by a component bin. The bins contain 9, 2, 1, 0·5, 0·2, 0·1 and again 0·1 μF capacitors. The operator has to load each jig with the appropriate capacitor from its bin. Pressing a button will cause the digital electronic bridge to select from these seven wide-tolerance capacitors those which will combine to 10 μF ±1%. The selected capacitors, which are usually two to four in number, are ejected from the jigs within one second of pressing the button and roll towards the operator. She then proceeds to wire-wrap them on to a printed circuit board, shown in Fig. 10, except for the rather larger 9 μF capacitor which is separately held in a clamp. She then refills the empty jigs with the appropriate capacitors and the cycle is repeated for the next assembly.

Examination of Fig. 10, which shows the subscriber's unit open, reveals the switch which combines the manual on/off and the automatic credit functions. It consists of an insulated rotor which can rotate through 30° and carries two trident phosphor bronze springs. These in turn are fitted at their ends with domed rivets made of 625 gold alloy which play on gold-plated printed circuit pads. This construction combines the best of all worlds: light action, long springs, hence even and gentle contact pressure, noble metal surfaces to prevent tarnishing, and wiping contacts to brush away dust particles. All this is desirable and necessary for switching milli- or even microvolts of radio frequency and it proved in practice to be eminently reliable. The light action is required because this switch must be unlatched into the off-position, when the credit has finally run out, by an energy as tiny as 80 microwatt-seconds. This is the minimum electromagnetic energy which may be available in the moving coil 'motor' at the end of a long feeder line. This energy is still capable of unlocking the escapement against the resultant of the clock-spring minus the sum of all friction forces including the friction of the latch which releases the spring-loaded switch rotor. It is evident that great care had to be exercised, when so little power was available, in the basic and the detail design. Matters requiring special care included the tolerancing of gears, the majority of which were moulded in Delrin, and consideration of temperature and hygroscopic expansion coefficients, so that after careful tooling the components could be assembled without selection or matching, and yet guarantee proper functioning and reliability.

One more item may be worthy of mention in this connection. The conductor pair carrying microvolts or millivolts of vision carrier also carries a direct audio programme of the order of 100 volts maximum (*fortissimo!*). In early experiments on the network it was observed that, at certain audio frequencies or transients of peak amplitudes, the moving coil vibrated and sometimes even managed to move completely through its possible travel, thus unlocking the escapement by one half tooth. Such fortuitous behaviour had to be prevented and the solution obviously lay in filtering out low frequency components. An ordinary low-frequency electric filter suppressing, say 20–80 Hz,

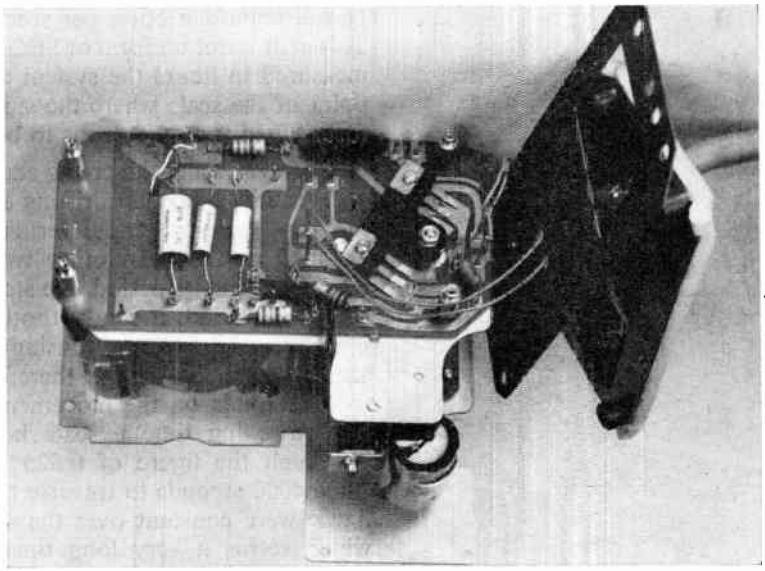


Fig. 10. Pay-TV subscriber's unit opened revealing the printed circuit carrying the padding capacitors. The rotary printed circuit switch can also be seen.

was out of the question for reasons of bulk and cost. Therefore, a mechanical filter was adopted. A simple plastic fly-wheel was stuck on to the moving coil and this addition of inertia cured the trouble.

9.3 *Factors Affecting the Accuracy of the Integrator*

As explained in the main part of the paper, the function of the integrator is to give a deflection which is proportional to the charge. It has also been pointed out that this deflection has to be maintained until digitization, and that the so-called drift of the moving-coil system has to be kept as low as possible in order to safeguard accuracy. This drift is caused by unwanted torques, which are due to mechanical unbalance, to magnetic occlusions in the nominally non-magnetic moving system being attracted into regions of higher magnetic stray fields, to air currents, but most important of all, they are due to the pull of the ligaments. Obviously the greatest care was taken in the construction to minimize each of these sources or error. The finest sapphire jewel bearings were used to keep friction to a minimum, the movement was most carefully balanced, materials specially selected to be non-magnetic, air currents excluded by a sealed case, and finally the ligaments used were the finest that can be obtained, namely a 625 gold alloy 'wire' of 0·002 in by 0·0001 in cross-section, i.e. one dimension of the rectangular ligament cross-section being of the order of the thickness of the human hair and the other being about one-twentieth of that. Evidently little more could be done to minimize these drift torques. But further improvements were wanted, and the only avenue open was to ensure that these irreducible torques would have as little an effect on the moving-coil system as possible. In fact, what resists the efforts of these stray torques is the so-called damping torque. This is due to the closed loop conductor in a magnetic field resisting being moved by setting up eddy currents. This damping torque is quantitatively expressed by B^2A^2/R where B is the magnetic flux density in the gap, A is the area circumscribed by the conductor loop of the moving coil, and R is the resistance of one turn of that loop if its thickness is equal to the total copper cross-section added to the aluminium cross-section of the coil frame. The actual subdivision of the total copper cross-section into many fine wires has only a secondary effect insofar as insulation and air spaces between the wires reduce the effective copper cross-section within a given area.

It can readily be seen from the formula that in order to increase the damping torque we want the highest flux density B, the greatest area A, and the lowest resistance R of the equivalent single-turn loop. It is interesting to develop this relationship further by establishing the law by which the damping torque increases if we were to increase the size of the whole system whilst keeping all its proportions constant and the materials the same. Flux density B will remain the same as was shown in a previous publication.[6] But the coil area A grows with the square of the linear dimension, i.e. with $[l]^2$, hence, the term A^2 grows with the fourth power, i.e. with $[l]^4$. The resistance R, being proportional to loop length over conductor cross-section, grows with $[l]^1/[l]^2$, i.e. $[l]^{-1}$. Therefore,

Fig. 11. The integrator in its cast-iron box with the drip-proof lid removed.

the damping torque B^2A^2/R grows with $[l]^0[l]^4/[l],^{-1}$ i.e. the damping torque grows with the fifth power of the linear dimension. This leads to the simple conclusion: the bigger the better, and what is more, with the fifth power too.

This was indeed implemented. The largest available coil, 1·5 inches in diameter and normally used for special relay work, was used and a special internal magnet system was designed, the internal magnet largely eliminating stray fields. A flux density B of 3500 lines/cm^2 was achieved in the gap. The overall performance showed a drift rate of better than 0·025% of full scale deflection per second on the prototype. This drift is not uniform or linear and after a long time measured in hours the system comes to rest at some point of the scale where the sum total of all the tiny fortuitous torques happens to be zero.

When assessing the errors caused by drift it should be borne in mind that drifting as such takes place all the time whilst the pointer is not electrostatically clamped at zero. In other words, drift affects the upscale discharge period, the standstill period and the digitizing period; it can be thought of as being superimposed upon the forces causing the upscale and downscale movements. Drift, therefore, acts for some 10 to 15 seconds on the movement, not merely for the brief dwelling time upscale before digitizing back. Although the figure of 0·025% per second implies some 4000 seconds to traverse the whole scale (if this figure were constant over the scale, which it is not) which seems a very long time, a duration of discharging and digitizing of only 10 seconds could give an error of 0·25% which is not necessarily insignificant as a contributor to the overall errors. Hence the concern to keep the drift as low as possible.

In order to test manufactured integrators in conditions under which they will be used on pay television, special digital test gear was built on principles already described, so that capacitance check measurements can quickly be made. It also permits measuring the drift digitally to accuracies of 1 part in 30 000. The overall accuracy of capacitance measurements in the absence of capacitor leakage was actually found to be of the order of 0·1%.

Figure 11 shows the integrator mounted on levelling screws in a solid cast-iron box. For best results the integrator is operated horizontally and mounted on a reasonably vibration-free base such as a concrete pedestal.

Manuscript received by the Institution on 4th November 1966. (Paper No. 1111/T38.)

U.D.C.621.372.832.8

An H-plane High-power TEM Ferrite Junction Circulator

By

J. HELSZAJN,
M.S.E.E., C.Eng., A.M.I.E.R.E.†

Summary: In this paper a new junction circulator is described which consists of a composite junction having a beryllia disk sandwiched between two oversized garnet disks on each side of the TEM lines. In this way the r.f. power is evenly divided among four thin garnet disks. This approach results in a very large thermal capacity. A general characterization of the junction in terms of the experimental admittance function is also given. This approach is used to make the junction one of the resonators of a two-section filter. The final circulator was tested at 150 kW peak power and 2·5 kW average power at 1300 MHz.

1. Introduction

In this paper, a high-power TEM junction circulator which was tested at 2·5 kW c.w. and 150 kW peak at a frequency of 1300 MHz with a 7% bandwidth is described. The junction circulator consists of a symmetrical distribution of ferrite material at the junction of three transmission lines. [1, 2] In this device, power entering port 1 emerges from port 2 and so on, in a cyclic manner.

The two main considerations which led to the final junction geometry were: (a) heating effects associated with the average power and (b) non-linear effects associated with the peak power. One way to eliminate the non-linear effects below ferrimagnetic resonance at high peak powers is to use low magnetization materials. However, such materials are normally associated with low Curie temperatures, a fact which is not consistent with the average power requirement of the circulator. The junction was therefore operated above ferrimagnetic resonance where first- and second-order non-linear effects are not present.[3] Since the effective permeability of the material is higher above resonance than it is below resonance, the resultant ferrite disk is geometrically reduced. This situation is particularly undesirable for a high-power circulator. To overcome this difficulty a departure was made from the classical junction geometry which uses simple disks with the well-known cut-off number $kR = 1{\cdot}84$.[4] It was felt that such a disk would be geometrically too small effectively to carry the r.f. power. Instead, a composite junction was developed which consisted of a beryllia disk sandwiched between two oversized ferrite disks on each side of the TEM line as shown in Fig. 1. In this way the r.f. power is evenly divided among four ferrite disks instead of two as in the conventional approach. This approach is particularly useful in a high average power device. The TEM lines leading up to the junction were also loaded with beryllia to further enhance the thermal capacity of the junction. A different composite junction has previously been described.[5] This geometry also differs from the classical one in that the assumption that there is no z-variation in the electromagnetic fields is no longer true.

In the experimental work, the junction geometry was characterized by an equivalent shunt parallel circuit derived from the admittance function in terms of the shunt conductance and susceptance slope parameter. This approach is useful when treating an arbitrary junction as is the case here. The classical junction representation is known to be a shunt parallel circuit[4, 6, 7] and has been characterized by a loaded-Q. In the experimental work the junction was first directly coupled to the transmission lines. The final bandwidth was then improved by making the junction one of the resonators of a two-section filter network. Bandwidth improvement in this manner has previously been mentioned.[4]

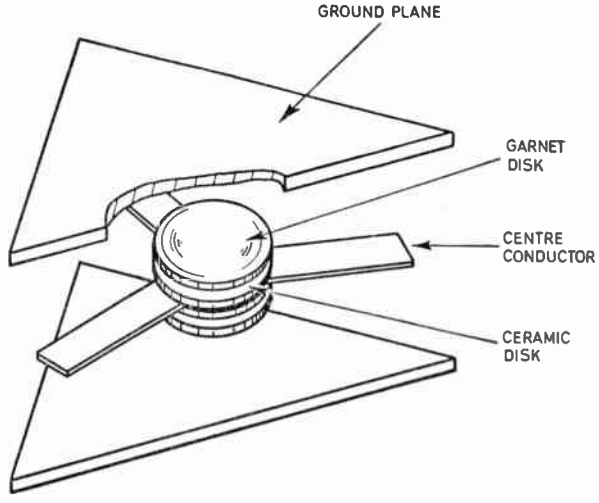


Fig. 1. Stripline junction using a garnet disk.

† Formerly with Microwave Associates Inc., Burlington, Mass., U.S.A.; now with the Department of Electrical and Electronic Engineering, University of Leeds.

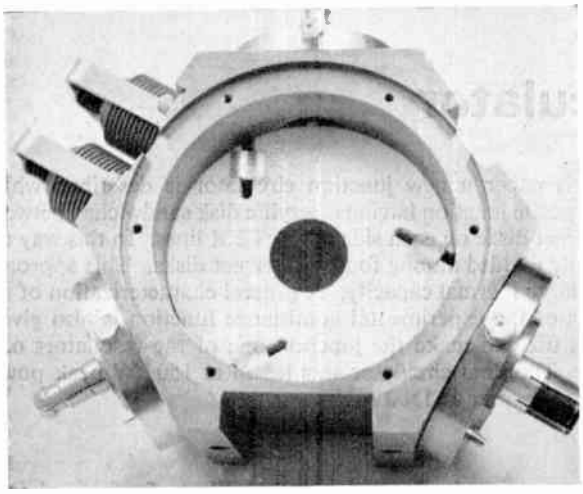


Fig. 2. High-power junction circulator.

The Appendix describes how the temperature gradient in the system can be predicted.

Figure 2 shows the circulator with one ground plane removed. (The two oil bellows were not used.)

2. Characterization of Junction Circulators

It is well known that any junction geometry which is experimentally matched is a perfect circulator. For a three-port circulator this requires two independent variables.[8] The junction can then be characterized by a parallel resonant circuit if the admittance function is known. An important quantity which relates to the junction is the loaded Q-factor (Q_L) and hence the bandwidth.

A useful formula for Q_L of a resonant circuit at the reference terminals for a system whose admittance function is known is given by:[9]

$$Q_L = \frac{\omega_0}{2G} \cdot \frac{\partial B}{\partial \omega}\bigg|_{\omega_0}$$

where B is the shunt susceptance and G is the shunt conductance. This representation is particularly useful when working with a Smith chart. From the above equation, the Q_L of the junction circulator can be determined in conjunction with the Smith chart of the experimental junction. It is worthwhile noticing that Q_L is determined by the susceptance slope parameter and the shunt conductance of the circuit.

The equivalent shunt elements can also be related to the susceptance slope parameter by:

$$b = \frac{\omega_0}{2} \frac{\partial B}{\partial \omega}\bigg|_{\omega_0} = \omega_0 C = \frac{1}{\omega_0 L}$$

It has been noted experimentally that, for a given geometry, the susceptance slope remains essentially constant over fairly large variations in shunt conductance. The equivalent circuit is shown in Fig. 3.

For an 'above-resonance' circulator with a normalized admittance level of unity, the susceptance slope parameter can be reduced in the vicinity of resonance by making the junction circulator one of the resonators of a quarter-wave coupled filter. A two-section filter is shown in Fig. 4. A more general matching network is shown in Fig. 5, where the J-inverters represent the coupling.[10]

In the experimental circulator design, the shunt resonant circuit of the filter was obtained with the help of a quarter-wave long transmission line terminated in a short-circuit. It can easily be shown that the susceptance slope for such a circuit in the vicinity of resonance is given by:

$$b = \frac{\pi}{4} Y_0$$

Hence, the susceptance slope parameter can be adjusted by varying the admittance of the short-circuited stub.

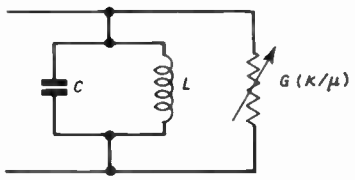


Fig. 3. Equivalent circuit of a junction circulator.

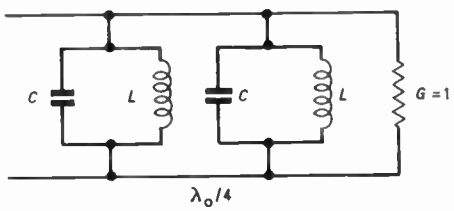


Fig. 4. Two-section filter.

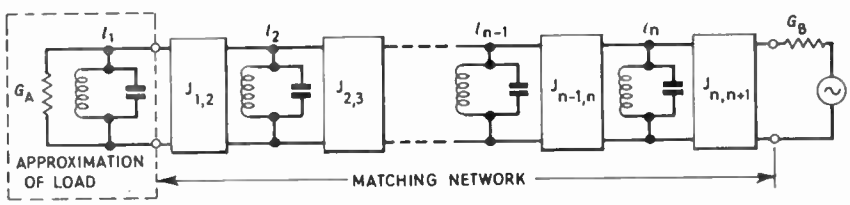


Fig. 5. Coupled-resonator matching network for shunt-resonant loads.

The insertion loss of a filter consisting of a number of n identical resonant elements spaced a quarter wave is given by:[11]

$$L = 10 \log_{10}[1+x^2 U_n^2(x)]$$

where the function $U_n(x)$ is a Chebyshev polynomial of the second kind and nth order. The polynomials corresponding to $n = 1, 2, 3, 4$ are given below with a recurrence formula from which they may be successively derived.

$$U_1(x) = 1$$
$$U_2(x) = 2x$$
$$U_3(x) = 4x^2 - 1$$
$$U_4(x) = 8x^2 - 4x$$
$$U_{n+1}(x) = 2xU_n - U_{n-1}$$

Plots of the insertion loss are shown in Fig. 6 for $n = 1, 2, 3, 4$.

Also

$$Q_L\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right) = x$$

For a narrow-band filter we can write

$$\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega} \simeq 2\,\frac{\omega - \omega_0}{\omega_0}$$

In the design of circulators we are more often concerned with the voltage standing wave ratio (v.s.w.r.). The insertion loss in terms of the v.s.w.r. is given by:

$$L = 10 \log_{10} \frac{(r+1)^2}{4r}$$

Hence:

$$\frac{(r+1)^2}{4r} = [1+x^2 U_n^2(x)]$$

This last equation can be used to determine the bandwidth for one and two section resonators from a knowledge of Q_L.

The insertion loss of the circulator at resonance is given by:

$$L = 20 \log_{10}\left[1 + \frac{Q_L}{Q_u}\right]$$

where Q_u is the unloaded Q-factor. Q_u can therefore be obtained from the knowledge of Q_L and an insertion loss measurement at the centre frequency.

3. Experimental Junction

The composite geometry was experimentally adjusted to obtain a junction resonance. The main experimental emphasis was directed towards obtaining as thin a ferrite disk as possible in order to reduce the temperature drop across the disk. A further important feature of such a junction is that the thermal capacity of the junction is also thereby increased because most of the ferrite material is replaced by the beryllia. The latter behaves as a heat sink because its thermal conductivity is large compared to that of ferrite materials. The detuning effects due to changes in the ferrite material are reduced because the ceramic is not temperature sensitive and a further advantage is that the magnetic losses are also reduced with this arrangement. This is particularly important at low frequencies.

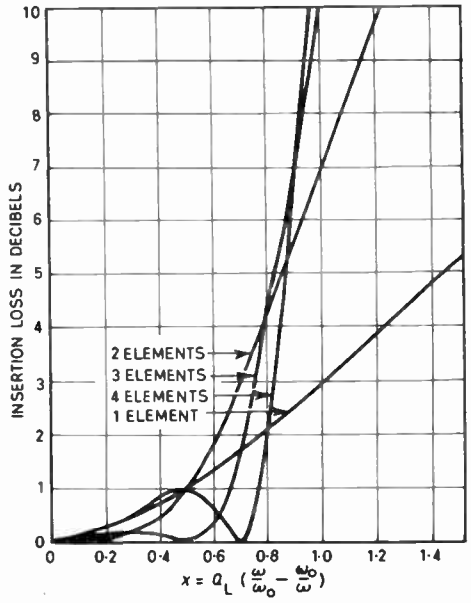


Fig. 6. Theoretical characteristics of filters employing identical elements. (After Fano and Lawson.[11])

The Smith chart of the experimental junction is shown in Fig. 7. The loaded Q-factor, Q_L, for this junction is 3·9. Figure 7 shows the bandwidth obtained by making the junction one of the resonators of a two-section quarter-wave coupled filter. The second resonator section was obtained with the help of a quarter-wave short-circuited stub. The calculated bandwidth is about 6% compared to the measured one of 7%. The insertion loss at the centre of the band was about 0·25 dB.

One of the practical problems in the development programme was the difficulty in obtaining the magnetic field required to operate above resonance with reasonably-sized magnets. This came about because of the unfavourable demagnetizing factors of the ferrite disks, and the large ground-plane spacing needed to prevent peak-power breakdown. An Alnico 9 high-energy product magnetic material was therefore used.

In the final assembly the two ground planes were water-cooled.

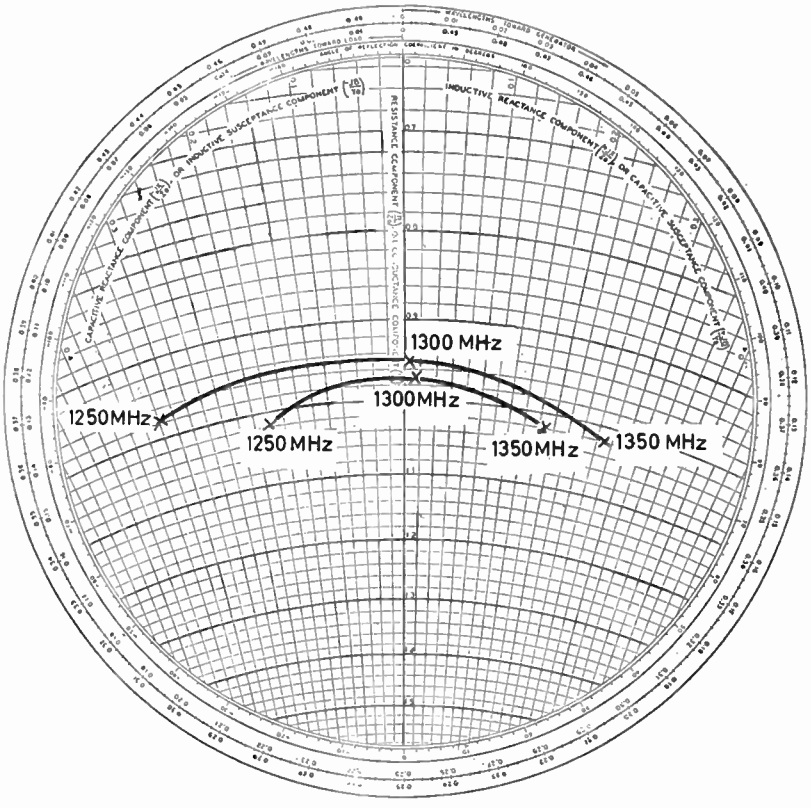


Fig. 7. Admittance plot of junction circulator with and without filter section.

4. External D.C. Magnetic Field

The external magnetic field needed to bias the junction can be predicted from Kittel's equation:

$$\omega = \gamma\left\{[H_a-(N_z-N_x)(M_s/\mu_0)]\times \times[H_a-(N_z-N_y)(M_s/\mu_0)]\right\}^{\frac{1}{2}}$$

where

$$N_x+N_y+N_z=1$$

For a thin ferrite shape

$$N_x=N_y=0, \qquad N_z=1$$

Hence

$$H_a=\frac{\omega}{\gamma}+\frac{M_s}{\mu_0}$$

For 'above-resonance' operation, the required magnetic field is n line-widths above ferrimagnetic resonance, hence:

$$H_a=\frac{\omega}{\gamma}+\frac{M_s}{\mu_0}+n\Delta H$$

where H_a = applied magnetic field in At/m
ω = angular frequency in rad/s
γ = gyromagnetic ratio, $2{\cdot}21\times10^5$ rad/s/At/m
ΔH = line-width in At/m
n = integer
$\mu_0 = 4\pi\times10^{-7}$ henry/m
M_s = saturation magnetization in weber/m^2

The material used was a modified yttrium iron garnet with $M_s/\mu_0 = 54\,000$ At/m and $\Delta H = 3600$ At/m. The applied magnetic field for $n = 3$, at 1300 MHz is 101 800 At/m.

The unloaded Q-factor, Q_u, of the junction can be determined in terms of n and ΔH. It is usually desirable to have as small a line-width as possible.

5. High-power Measurements

In the Appendix, a one-dimensional heat flow analysis has been given and this can be used to predict the maximum temperature rise in the system. The thermal conductivity of ferrites is about 5×10^{-3} cal cm^{-3} s^{-1} whereas that of beryllia is about 525×10^{-3} cal cm^{-3} s^{-1}.

In determining the power density dissipation from the insertion loss, it is assumed that the power is dissipated in a 240° wedge because one of the ports is decoupled. The insertion loss of the junction was 0·25 dB at the centre of the band. Using this number,

the maximum temperature rise was calculated to be 30 deg C when both the centre conductor and the ground planes were cooled. If it is assumed that only the ground planes are cooled, the temperature rise is 125 deg C.

This last number represents the worst possible situation because there will be significant heat flow in directions other than the axial direction. The Curie temperature of the garnet material used is 180°C.

If the beryllia is replaced by ferrite, the calculated temperature rise becomes 310 deg C. This last value indicates the improvement possible using the technique developed in this paper.

The high-power measurements fell into two parts. The junction was first tested for peak power breakdown by using a low duty-cycle magnetron. In this way areas in which peak power breakdown initially occurred were located and re-worked. The junction was then tested at the full average and peak power with the two ground planes cooled. The input water inlet was adjusted to drop the external temperature of the housing to about ambient temperature. In the initial average power test the short-circuit of the input quarter-wave stub burned out at 1·5 kW c.w. power because the foil contacts to the ground planes were not sturdy enough to carry the short-circuit wall currents. This was remedied by replacing the foil contacts by spring fingers.

6. Conclusions

In this paper a new H-plane TEM junction circulator in which the r.f. power is evenly divided among four thin ferrite disks is described. A general characterization of the junction circulator in terms of the admittance function has also been given. The characterization was then used to increase the bandwidth of the junction by making it one of the resonators of a two-section filter. The final circulator was tested at 150 kW peak power and 2·5 kW average power at 1300 MHz.

7. Acknowledgment

The author wishes to thank Dr. P. Chorney for providing the heat analysis and Microwave Associates, Inc., for permission to publish this paper.

8. References

1. H. N. Chait and T. R. Curry, 'Y-circulator', *J. Appl. Phys.* (Suppl.) **30**, pp. 152S–153S, April 1959.
2. U. Milano, J. H. Saunders and L. Davis, Jr., 'A Y-junction stripline circulator', *Trans. Inst. Radio Engrs on Microwave Theory and Techniques*, **MTT-8**, pp. 346–51, May 1960.
3. W. Suhl, 'The non-linear behavior of ferrites at high microwave signal levels', *Proc. I.R.E.*, **44**, pp. 1270–84, October 1956.
4. H. Bosma, 'On stripline Y-circulation at u.h.f.', *Trans. Inst. Elect. Electronics Engrs on Microwave Theory and Technique*, **MTT-12**, pp. 61–72, January 1964.
5. J. Helszajn, 'A ring stripline junction circulator', *The Radio and Electronic Engineer*, **32**, pp. 55–60, July 1966.
6. C. E. Fay and R. L. Comstock, 'Operation of the ferrite junction circulator', *Trans. I.E.E.E.*, **MTT-13**, pp. 15–27. January 1965.
7. H. J. Butterweck, 'Der Y-zirkulator', *Archiv der Elektrischen Übertragung*, **17**, pp. 163–76, April 1963.
8. B. A. Auld, 'The synthesis of symmetrical waveguide circulators', *Trans. I.R.E.*, **MTT-7**, pp. 238–46, April 1959.
9. 'Principles of Microwave Circuits', Ed. C. G. Montgomery, R. H. Dicke and E. M. Purcell, (McGraw-Hill, New York, 1948).
10. G. L. Matthaei, L. Young and E. M. T. Jones, 'Microwave Filters, Impedance-Matching Networks and Coupling Structures', (McGraw-Hill, New York, 1966).
11. R. M. Fano and A. W. Lawson, Jr., 'Microwave filters using quarter-wave couplings', *Proc. I.R.E.*, **35**, pp. 1318–1323, November 1947.

9. Appendix

Analysis of Heat Flow and Temperature Gradients in Disk Ferrite-Ceramic System

The model for which the analysis is described is illustrated in Fig. 8. It is assumed that the only significant heat flow is directed axially and that only a 240° wedge need be considered because one of the ports of the circulator is decoupled. The inner and outer ferrites are bounded by ground planes and a centre conductor which can act as heat sinks. To simplify the problem, the r.f. power dissipation is assumed to be uniformly distributed and evenly divided

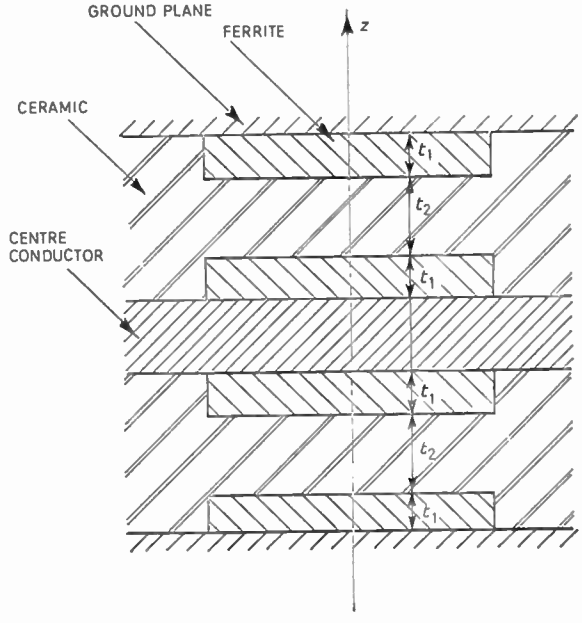


Fig. 8. Model for heat flow calculations.

among the four ferrite disks. The ferrite disks are separated on each side of the centre conductor by ceramic disks. The thermal conductivities of each of the ferrite disks and the ceramic media are given respectively by σ_1 and σ_2.

The analysis of this problem proceeds from the basic heat conduction equations:

in the inner ceramic $\nabla^2 T = 0$ .....(1)

and

in the ferrite $\nabla^2 T = -p/\sigma_1$ .....(2)

where p is the power dissipation density in each disk. Equations (1) and (2) must be solved subject to the imposed heat sinks and subject to the conditions that the temperature and heat flux across the various boundaries are continuous. The temperature distribution has been derived for two different cases. In the first case, both the ground planes and centre conductor act as heat sinks. In the second case only the ground planes act as heat sinks, the centre conductor being assumed to be thermally floating.

For the case in which both the ground planes and inner conductor are used as heat sinks (heat sink temperature is T_0) the temperature distribution from a one-dimensional solution of eqns. (1) and (2) is given by:

$$T(z) = T_0 - \frac{1}{2}\frac{p}{\sigma_1} z^2 + \frac{p}{\sigma_1} z t_1, \qquad 0 < z t_1 \qquad \text{.....(3)}$$

$$T(z) = T_0 + \frac{1}{2}\frac{p}{\sigma_2} t_1^2, \qquad t_1 < z < t_1 + t_2 \qquad \text{.....(4)}$$

$$T(z) = T_0 - \frac{1}{2}\frac{p}{\sigma_1}(2t_1 + t_2 - z)^2 + \frac{p}{\sigma_1}(2t_1 + t_2 - z)t_1, \qquad t_1 + t_2 < z < 2t_1 + t_2 \qquad \text{.....(5)}$$

The temperature distribution is uniform in the ceramic. This is a consequence of the symmetry imposed by the assumption that power dissipated in each disk is the same. The maximum temperature rise in the system is given by:

$$\Delta T = \frac{1}{2}\frac{p}{\sigma_1} t_1^2 \qquad \text{.....(6)}$$

The temperature rise is independent of the thermal conductivity of the ceramic because of the symmetry.

For the second case in which only the ground planes act as heat sinks, the temperature distribution is given by:

$$T(z) = T_0 - \frac{1}{2}\frac{p}{\sigma_1} z^2 + \Delta T, \qquad 0 < z < t_1 \qquad \text{.....(7)}$$

$$T(z) = T_0 - \frac{p t_1}{\sigma_2}[z - (t_1 + t_2)] + \frac{3}{2} p \frac{t_1^2}{\sigma_1}, \qquad t_1 < z < t_1 + t_2 \qquad \text{.....(8)}$$

$$T(z) = T_0 - \frac{1}{2}\frac{p}{\sigma_1}[z - (2t_1 + t_2)]^2 - 2\frac{p t_1}{\sigma_1}[z - (2t_1 + t_2)], \qquad t_1 + t_2 < z < 2t_1 + t_2 \qquad \text{.....(9)}$$

The region of maximum temperature occurs in the ferrite near the centre conductor. The temperature rise above the heat sink is given by ΔT

$$\Delta T = p t_1 \left[\frac{2t_1}{\sigma_1} + \frac{t_2}{\sigma_2}\right] \qquad \text{.....(10)}$$

In general, the temperature rises predicted in this analysis are conservative estimates and the actual temperatures are expected to be lower. This is because all of the heat flow paths have not been taken into account. It was assumed for convenience of analysis that the heat flow was purely axial. However, there will be heat flow in the radial direction as well as in the circumferential direction into the remaining one-third part of the circulator. These other heat flow paths will obviously result in a lower temperature rise.

Manuscript first received by the Institution on 22nd August 1966 and in final form on 10th January 1967. (Paper No. 1112.)

U.D.C.534.88

Low Frequency Sound Sources: Statement of Problem and Some Possible Solutions

By

B. S. McCARTNEY, Ph.D.†

Reprinted from the Proceedings of the I.E.R.E. Conference on 'Electronic Engineering in Oceanography' held in Southampton on 12th–15th September 1966.

Summary: The main stumbling block in a project to measure the back-scattering cross-section (target strength) of commercial fish at frequencies in the band 300 Hz to 3000 Hz has been that of providing a suitable sound source. The spectrum level requirement of this source is + 70 dB relative to 1 μbar (0·1 N/m²) at 1 m per Hz across the band, or about 100 W peak power during a 10 ms pulse at any frequency. The general problem of radiating this power level into water at these frequencies is considered. Several sources employing various principles of transduction including sparkers, boomers, explosives, hydroacoustic, electrodynamic, piezoelectric and magnetostrictive sources, are considered in terms of a common parameter—the maximum output spectrum level—as well as efficiency.

1. Introduction

There is an increasing need in oceanographic applications for low-frequency (10–3000 Hz) sound-sources. Continuous sub-bottom profiling systems have received much attention[1–3] and many types of sound source have been used for the repetitive, low-frequency (10–1000 Hz), high energy, sound pulse necessary to obtain deep penetration with reasonable resolution. Pingers on neutrally-buoyant 'Swallow' floats[4] using 2 ms pulses at 10 kHz have hitherto been detected at ranges up to about 4 miles, but recent designs[5] for longer range floats transmitting in the Sofar channel require longer pulses ($\sim$1 s) at a lower frequency, around 1000 Hz or less. Facilities at calibration stations for hydrophones and transducers should include a stable wide-band sound source covering the audio frequency spectrum. An important bio-acoustic problem is the determination of the acoustic target strength (back-scattering area) of fish and fish shoals. Haslett[6] has summarized target strength measurements for fish lengths, L, greater than 3λ, and the range of interest in the present paper is $L < 3\lambda$. For commercial fish this means frequencies from about 300–3000 Hz, corresponding to wavelengths of 5–0·5 m.

Whilst no single sound source would be expected to suit all these fields of application, the number of types of sound sources for this frequency band is considerable, each worker having his preference. This contrasts with the case at higher frequencies where resonant piezoelectric or magnetostrictive transducer arrays are now standard. The bio-acoustic problem will be considered in some detail as it has some novelty and is one with which the author is familiar. The acoustic theory of radiation by small objects is well known and available in the literature, but a little of the basic theory is presented in view of its relevance to the scattering from fish and the primary production of sound by the source. Finally some sources employing various means of transduction are briefly discussed with reference to sound spectrum level, waveform where available, and suitability for this particular problem.

2. The Measurement of the Target Strength of Fish at Audio Frequencies

If an echo-detection system is set up with an omni-directional hydrophone between the sound source and the target, and if $V_d(f)$ and $V_e(f)$ are the voltages received by this hydrophone due to the direct transmission pulse and the echo pulse respectively at a frequency f, at times t_d and t_e seconds after transmission, then the acoustic cross-section of the target is

$$\sigma(f) = 4\pi \left[\frac{c.(t_e^2 - t_d^2)}{4t_d} \cdot \frac{V_e(f)}{V_d(f)} \right]^2 \qquad \ldots\ldots(1)$$

where c is the velocity of sound.

With a wide-band transmission pulse, spectrum analysis will be necessary to give $V_e(f)$ and $V_d(f)$ before the ratio is used for $\sigma(f)$, but the spectral response of hydrophone and receiver need not be known.

The requirements for the sound source are that the source level should be sufficient that V_e can be detected above the ambient noise background and that the

† National Institute of Oceanography, Wormley, Godalming, Surrey.

pulse is sufficiently short that the trailing edge of V_d should not interfere with V_e. In practice the first requirement amounts to a source spectrum level at any frequency between 300 Hz and 3 kHz of greater than +70 dB (relative to 1 μbar) at 1 m in a 1 Hz bandwidth, which is equivalent to a power of 1 W/Hz if the radiation is omnidirectional. This is based on measuring a minimum target strength according to the locus of line AB in Fig. 1, with an echo 12 dB above a background noise level equivalent to sea state 3 noise, and with the hydrophone 10 m, and the target 30 m, from the source; these distances are reasonable for local waters where fish may be caught. By reducing the signal/noise ratio and waiting for quieter conditions results could be obtained with a source level of +60 dB (1 μbar) in 1 Hz band. The minimum value of V_e/V_d is 5×10^{-5}, so that galvanometric or oscillographic recording of V_d and V_e must be via different, calibrated amplifiers.

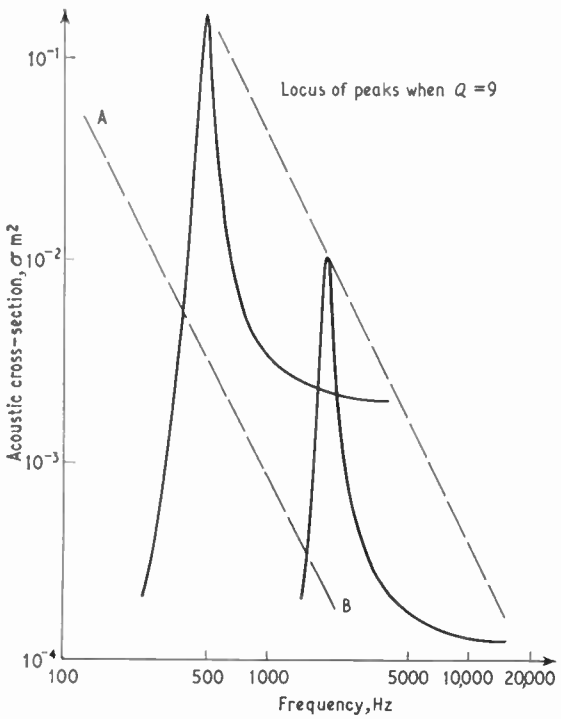


Fig. 1. Acoustic target strength spectra of resonant bubbles or swim-bladders.

The temporal limitations on the sound pulse necessitate wide-bandwidth for source and receiving circuits and place a limit to the frequency resolution of V_e and V_d and hence of σ. The frequency functions strictly represent energy density functions between limits $\pm \delta f/2$ in a measurement interval t_p where $\delta f \simeq 1/t_p$ and $t_p < (t_e - t_d)$. Further temporal restrictions may be caused by reflections from surface, bottom, anchor chains and suspension cables. A pulse length of 10 ms means barely 3 cycles at the lowest frequency and a resolution of 100 Hz. Within the 10 ms pulse a peak pressure level +90 dB (1 μbar) at 1m is required, equivalent to 100 watts during the pulse and a total energy radiation of 1 joule. If the sound source is tuned at f Hz the trailing edge of the pulse decays with a logarithmic decrement of $\pi f/Q$. Equating the residue of the direct pulse at a time t_e to the smallest expected signal V_e, for the geometry assumed above, gives the minimum value of the Q-factor of a resonant transducer according to the table below. (a_{min} is defined and discussed in Section 3.2.)

A delta function pressure pulse is difficult to generate cleanly and at the required amplitude but would be very useful since it would give the target spectrum directly from the transform of the echo waveform. The peak source level of a pulse 0·1 ms

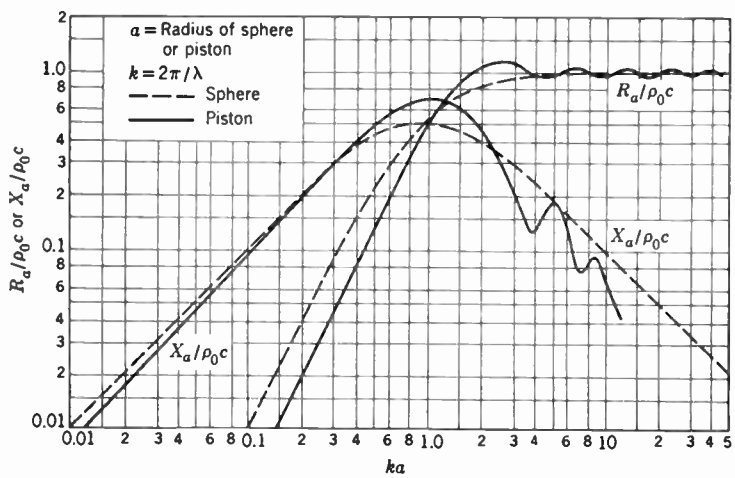


Fig. 2. Radiation impedance of a sphere and a circular piston (from Hueter and Bolt[24]).

long to cover the band would need to be +110 dB (1 μbar) at 1 m. When the source is energy or amplitude limited, there is some advantage if the pulse is as long as possible at each frequency in turn (e.g. many cycles), since this concentrates the available energy into a narrower band and a small amount of noise filtering at the receiver can be applied. More advanced matched filtering techniques are not convenient as the distortion produced by the target itself is not known at this stage.

3. Acoustic Radiation from Small Sources

3.1 *Fish*

When re-radiation of sound incident on a fish is considered, the main component responsible for scattering at wavelengths longer than the fish dimensions is the swim-bladder if present. The tissue, scales, bones and other parts scatter sound at higher frequencies[6] but can be considered acoustically transparent at low frequencies. The gas-filled swim-bladder of a fish is a compliance, forming with its mass loading the essential components of a resonant system. It is suspected that at the resonant frequency the target strength of the fish will be increased in the manner of Fig. 1, as for a spherical air bubble[7] in water. The swim-bladder is not spherical, but at the low frequency concerned may be considered so as a first approximation.[8] Thus the radiation impedance approaches the spherical wave impedance shown in Fig. 2. The low-frequency spectrum of σ has three major regions, as may be seen from Fig. 1 and eqn. (8) in the Appendix.

Thus

$$\sigma \to 4\pi a^2\left(\frac{f}{f_r}\right)^4, \quad \text{for} \quad f < f_r \qquad \ldots\ldots(2)$$

$$\sigma = 4\pi a^2 Q^2, \qquad f = f_r \qquad \ldots\ldots(3)$$

$$\sigma \to 4\pi a^2, \qquad f_g > f > f_r \qquad \ldots\ldots(4)$$

where $L = \lambda_g$, the wavelength at f_g, the limiting frequency at which the lumped parameter approximations for the swim-bladder are valid. A typical fish[6] has $a = 5{\cdot}2 \times 10^{-2} L$, and the swim-bladder has a length $0{\cdot}24\,L$, so that $f_g \simeq 50 f_r$. The value of σ given by eqn. (2) is a factor 10^8 greater than that given by the Rayleigh scattering formula for a rigid particle of the same size. This shows how dangerous it may be to assume that a reflector is rigid if there is any possibility of compliant spaces. When the fish is close to a soft or hard reflecting boundary or to other fish, the simplified equivalent circuit does not apply and the boundary conditions for the geometry involved must be applied with the wave equation to determine the overall target strength. For a shoal of N fish separated by more than a wavelength it would seem reasonable that the shoal target area $\sigma_s = N\sigma$, at least at frequencies away from the single fish resonance.[10] Near resonance or for closely-packed shoals the fish are coupled acoustically and the relationship between σ_s and σ will be more complicated.[11] Thus

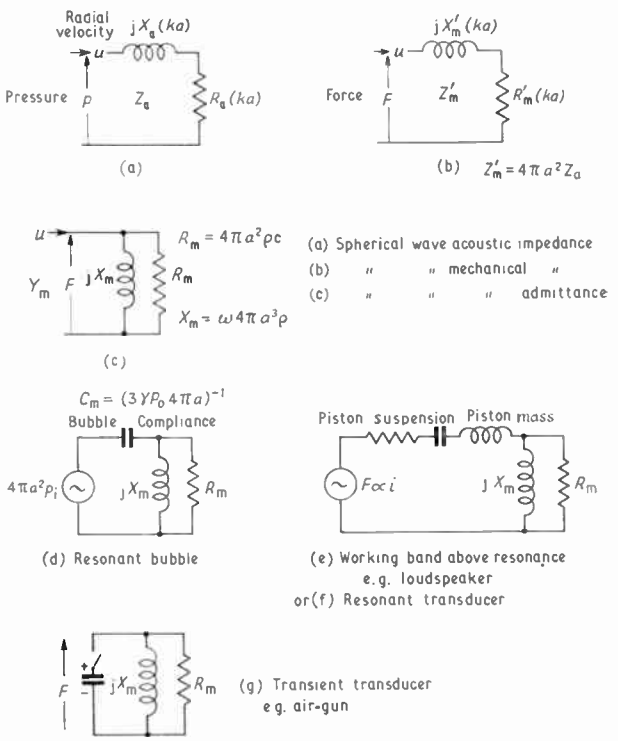


Fig. 3. Simplified equivalent circuits of transducers.

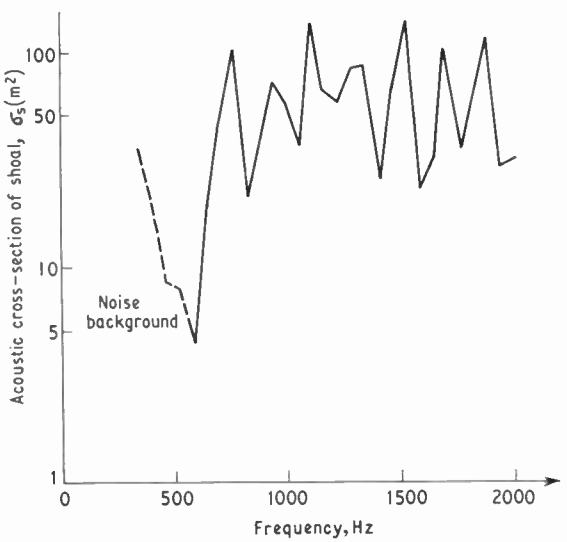


Fig. 4. Acoustic target spectrum of a shoal of fish (probably pilchards) obtained using 1000 joule Boomer.

σ_s will have a spectrum different from that of σ. An example of σ_s for what was probably a pilchard shoal is given in Fig. 4. In deep scattering layers the small (2–20 cm) myctophid and other fish which have swimbladders are well spaced and the volume scattering coefficient would thus be expected to show the resonance, and this has been observed by Hersey[12] *et al.* over the range 3–30 kHz and by Andreeva.[9]

The resonance may be broadened due to the spread in size of these fish. In a shoal of commercial fish, however, the spread of sizes is often small (about 5%).

3.2 *Linear Audio Transducers*

With a wavelength range of 5–0·5 m, the sound source will usually be a fraction of a wavelength. Radiation will then be omnidirectional and the mass loading of the medium will predominate. Though in principle the radiation load on a transducer of any geometry can be calculated it is found in practice that, for small sources, shape is a secondary factor and the radiation impedance approaches the spherical wave impedance (Fig. 2). For convenience radiation from a pulsating sphere of radius a will be considered. In order to radiate 100 watts the minimum radius to prevent cavitation near the surface in sea-water at these frequencies is 3 cm. The displacement amplitudes for different radii are shown in Fig. 5. The radial strains are very high especially for low frequencies and small radii. The range of amplitudes over the band is greater for smaller a. As the radius a increases more care must be taken to prevent flexural resonances.

It is clear that a force generator which is independent of frequency and which has a low source impedance is required to obtain a flat frequency response of pressure in the far field. Constant force can be generated by an electrodynamic (moving coil) system driven at a constant current or by a piezoelectric system driven at constant voltage. The inevitable mounting stiffness causes a resonance which must be placed below the working band. Whilst this is possible for an electrodynamic system, Young's modulus is too high and the peak strain too low to achieve this with piezoelectric materials. The simplified equivalent mechanical circuit for an underwater loudspeaker is shown in Fig. 3(e). Sims[13] has given a thorough analysis of such a design including the equalization of static pressure.

The necessity for a constant-current source inherently implies low overall efficiency, and when electrical losses in the coil are included the overall efficiency in practice drops to less than 1%. A commercially-available loudspeaker having a 10 cm diameter diaphragm has a maximum output of 63 dB (1 μbar) ± 2 dB over the band 100–7000 Hz and is limited by coil dissipation; below 100 Hz the output is 12 dB per octave less and limited by diaphragm displacement. The efficiency quoted is 0·16%. To increase the output level from this type of transducer, it will be necessary to increase the diaphragm diameter and to design a magnet and coil assembly capable of handling about 20 kW. Moving coil vibrators are available which handle this power level.

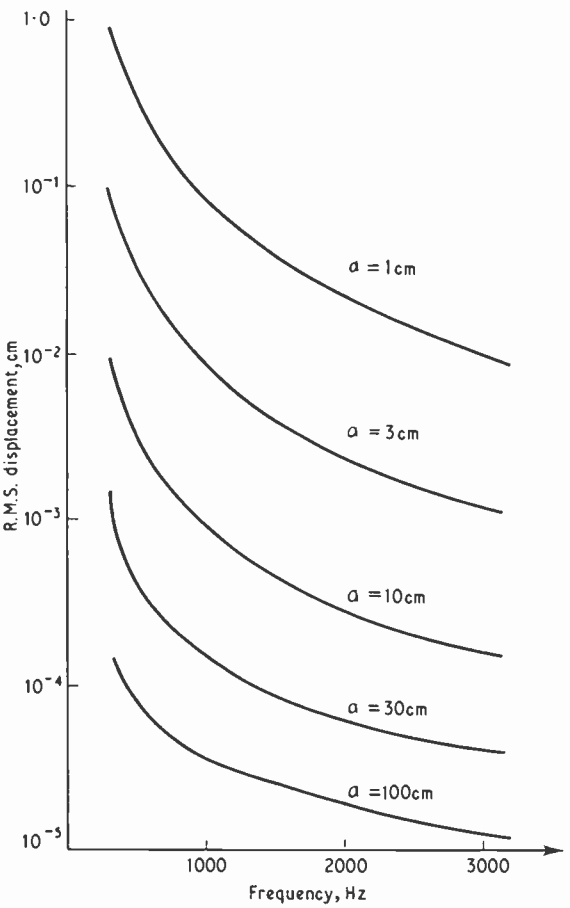


Fig. 5. Displacement amplitude of a spherical source of radius a radiating 100 watts.

The use of several tuned transducers to cover the band depends upon the working Q as seen above. The mechanical driving source impedance Fig. 3(f), is similar to the loudspeaker at resonance, and to the air bubble if the piston mass is small compared with the radiation mass; then eqn. (7) (see Appendix) relating Q and ka is valid. The minimum radius of a spherical source required to give the maximum Q is shown in Table 1. The force generator may be magnetostrictive, piezoelectric or electrodynamic as was used many years ago for the 'Fessenden oscillator'.[14] The mass of the piezoelectric or magnetostrictive stacks may be ignored as they are much shorter

Table 1

Limiting Q and minimum radius values at different frequencies

Frequency (Hz)	300	600	1000	2000	3000
Q_{max}	3·3	6·1	9·1	18	25
a_{min}(m)	0·24	$6{\cdot}5\times10^{-2}$	$2{\cdot}6\times10^{-2}$	$6{\cdot}6\times10^{-3}$	$3{\cdot}2\times10^{-3}$

than a wavelength in the material. The electromechanical coupling coefficient can approach the basic material value since the strains are uniform throughout the material. Thus to some extent electrical loading can be applied to reduce the Q, though it is an inefficient method. Generally the efficiency is high (30–80%). As the stacks operate over a limited band it is possible to design to optimize displacement, efficiency or Q.

In practice the piston mass cannot be ignored and higher Q's than eqn. (2) would suggest are obtained. Some results have been obtained using a 3·3 kHz piezoelectric transducer with a piston diameter of $\lambda/6$, a Q of 19 and a maximum output of +85 dB (1 μbar) at 1 m with an efficiency of 50%. A magnetic reluctance transducer with a diaphragm $10{\cdot}1\lambda$ diameter at 960 Hz giving a maximum output level of +75 dB (1 μbar) at 1 m, had a Q of 17 which was too high for this application. Cylindrical transducers with bender-bar driving elements are commercially available up to 1·8 kHz, with outputs in excess of +97 dB (1 μbar) at 1 m but the Q values are too high below 800 Hz and marginal otherwise.

Sims[15] added a spherical rubber enclosed air bubble to the front of an underwater loudspeaker and at the resonant frequency achieved +79 dB (1 μbar) at 1 m with an efficiency just over 3% and Q around 5 at 50 Hz. By adjusting the volume of the air the resonant frequency can be changed.

Both the wide-band and resonant types of sound source considered above are linear and reversible, being also usable as receivers. The hydroacoustic source described by Clay[16] *et al.*, is a high-power, linear source for seismic profiling, which covers the band 80–600 Hz (see Fig. 6) but which could not conveniently be used as a receiver. The source is a hydraulic analogue of a d.c. to a.c. inverter. A radiating piston is supplied by a high-pressure hydraulic pump with fluid whose flow is modulated by an electrically-controlled valve. A feedback signal proportional to piston velocity is applied at the control signal input to increase the damping at resonance and extend the bandwidth. The experience of manufacturers of vibration test equipment would suggest that extension of the hydraulic principle to frequencies over 500 Hz will be accompanied by reduced output and efficiency, due to the compliance of the hydraulic fluid.

3.3 *Irreversible Transducers*

In the last decade there has been a proliferation of sound sources for the marine seismic application in which stored energy is released impulsively, to create motion and thence an acoustic pulse. However the motion is impressed, sound radiation is ultimately determined by the radiation impedance. As changes of state or finite-amplitude effects (or both) are often involved, the empirical approach can be more fruitful.

Weston[1] has given the source spectrum level for explosives with rules for scaling according to weight and depth. The pressure waveform from an underwater explosion consists of a shock wave pulse followed by a bubble pulse and then further bubble pulse oscillations. Very high source levels can be obtained and most low-frequency propagation measurements over long distances have used explosives. A charge of 0·03 lb (13·6 g) of T.N.T. would be needed to provide the spectrum level required here, if finite amplitude effects at short range were ignored. The interval between shock and first bubble pulses, the bubble frequency and decay rate are all dependent on the depth and weight. For 0·03 lb weight the interval would be less than 40 ms, so that the first bubble pulse would interfere with the target echo and the bubble oscillation frequency would exceed 30 Hz. The main disadvantage of explosives is the need for elaborate safety precautions, storage facilities and trained operators, not always available on non-military vessels.

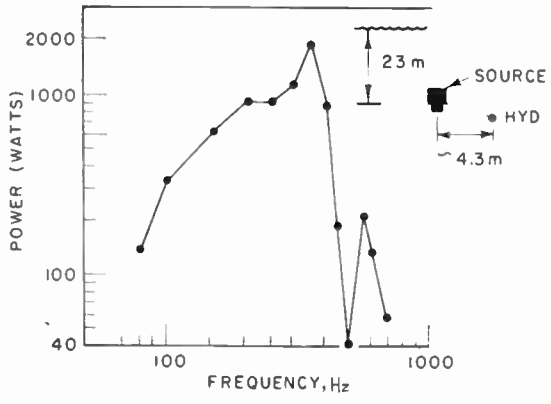


Fig. 6. Frequency response of the hydroacoustic source (after Ewing and Zaunere[22]).

Discharge sparks have been used successfully[18] as sound sources in the band 40 Hz–1 kHz for sub-bottom profiling, sometimes with multiple electrode arrays, and by Anderson[19] for scattering and target strength measurements in deep layers down to 430 m over the band 2–20 kHz. The pressure waveform from an underwater spark (Fig. 7), is similar to that

from an explosive, having a shock pulse followed by one or more sharp bubble pulses and then decaying oscillations, though the initial and first bubble pulses have more comparable amplitudes. The spectrum of the waveform in Fig. 7 is shown in Fig. 8. The peak source level and interval between pulses, and hence the frequencies of the peaks, all depend upon the energy stored in the capacitors, when all other parameters are fixed. The proportion of energy converted to acoustic radiation varies from 0·07%, measured by Anderson, to 32% by Roy and Frolov.[20] The parameters on which this efficiency depends include voltage, storage capacity, circuit inductance, switching loss, electrode geometry and surface condition, water salinity and depth; they are rather too numerous and the data too few to generate reliable scaling laws at present as for explosives. Roy and Frolov showed that efficiency improved with voltage when the spark gap was increased to the maximum for that voltage. The efficiency for the example shown in Fig. 7 was

Fig. 7. Pressure-time waveform from a 3·25 joule spark. (Abscissa 1 div. = 0·25 ms, Ordinate 1 div. = 1100 μbar.)

1·1%, which is equal to that obtained using the results of Caulfield[21] at a similar voltage. By the end of the first bubble pulse 96% of the acoustic energy has been radiated; the remaining energy is radiated more slowly according to the release of a compressed bubble, for which the linear mechanical equivalent circuit of Fig. 3(g) and eqn. (7) apply. Energy levels as high as 100 000 joules have been used and with 60 000 joules the overall pulse train length, which determines the resolution of sub-bottoms, was 25 ms.

Another more recent form of impulsive sound source is the pneumatic device described by Ewing and Zaunere.[22] Up to 300 cm^3 of compressed air at a pressure of 2500 lb/in^2 is suddenly released on demand, producing peak pressures of up to +124 dB (1 μbar) at 1 m. The pressure waveform or spectrum was not described. The volume of the gas at the ambient pressure would be large and so after the initial release pulse, a low-frequency resonance would be expected, e.g. around 30 Hz at 30 m depth. For any gas volume, the resonant frequency will be increased[8] by the near presence of a pressure release boundary or decreased by a rigid boundary, as well as lengthening the pulse by reflection. From the profile record published, the effective pulse length would seem to be approximately 40 ms.

The 'Boomer' sound source[2] developed by Edgerton[3] depends upon the repulsion of a flat aluminium plate from an adjacent flat spiral coil into which a capacitor bank is discharged. Unlike the other electromagnetic sources, such as the loudspeaker and Fessenden oscillator, the 'Boomer' does not have a separate permanent or electromagnetic field system. The force may be considered to arise from the interaction of either the eddy current and the primary current, or the eddy current and the resultant radial magnetic field, and in amplitude is proportional to the square of the coil current. For low-frequency coil currents,

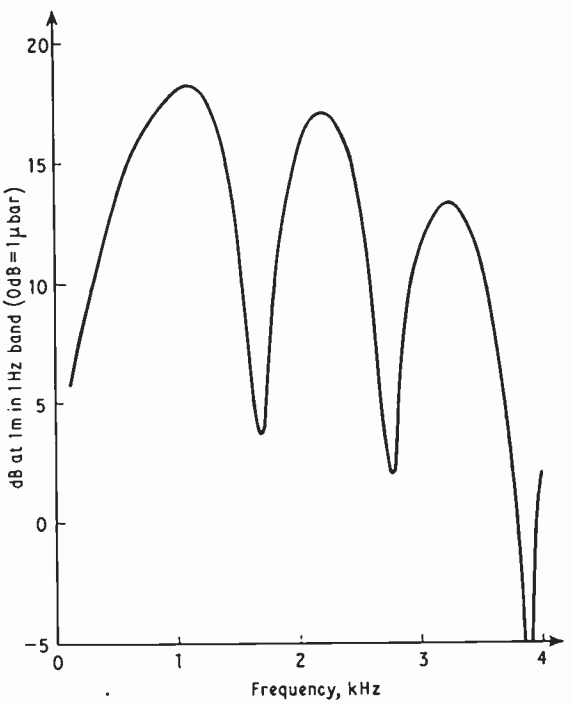


Fig. 8. Sound source spectrum level of the sparker waveform shown in Fig. 7.

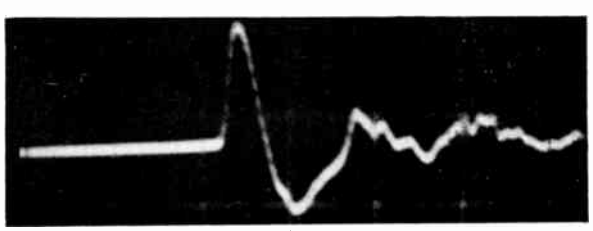


Fig. 9. Pressure-time waveform received at 60 ft below 1000 joule Boomer.
(Abscissa 1 div. = 0·5 ms, Ordinate 1 div. = 3300 μbar.)

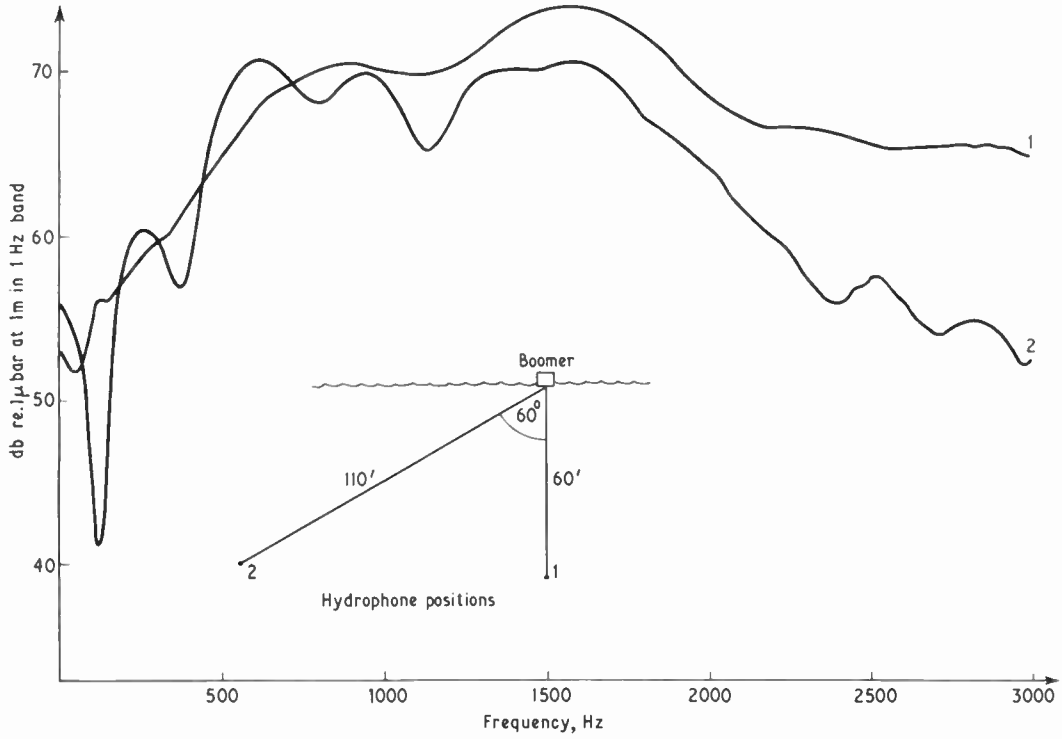


Fig. 10. Sound source spectrum level of Boomer in two positions.

the magnetic field penetrates the plate and the eddy currents, radial field and mechanical force are all small. At very high frequencies, penetration of the plate is small and resultant field is almost completely radial, but the eddy currents are now limited by the increased skin resistance so that the mechanical force is again small. Owing to the 'turn-off' behaviour with reverse current of high-voltage switches, such as thyratrons, ignitrons and spark gaps, the exciting current in the coil is normally the first half cycle of an oscillatory discharge. A typical pressure waveform from a 1000 joule 'Boomer' having an aluminium plate 0·5 m in diameter is shown in Fig. 9 and the spectra at two positions in Fig. 10. The efficiency of converting stored electrical energy to acoustic radiation is around 0·5%. The results of Fig. 4 were obtained with this source. As the forces on the plate are greatest at the centre there is a tendency to excite 'oil-can' modes in the plate. The stiffness of the plate mounting must be low to keep the spring and radiation mass resonance below the working band. Energy levels up to 13 000 joules have been used with single and multiple parallel boomers.

An interesting application of the inherent non-linearity of finite amplitude acoustic waves has been proposed by Berktay[23] to produce low-frequency sound. The interaction of two coincident sound beams at slightly different frequencies (f_1, f_2) produces a secondary wave from an end-fire array of virtual sources at the low difference frequency $(f_1 - f_2)$. As this virtual array may have a great length, determined primarily by the absorption attentuation at the high-frequency, the far-field low-frequency beam may have a narrow beamwidth which varies only slowly with the difference frequency. The far-field axial pressure at the difference frequency, however, is frequency dependent and is given by:

$$P(r) = \frac{\pi P (f_1 - f_2)^2 \exp(-\alpha r)}{c^3 r A} \qquad \ldots\ldots(5)$$

where $P/2$ = power radiated at both f_1 and f_2

r = range

α = absorption attenuation at $(f_1 - f_2)$

and A is an absorption function of f_1 and f_2.

Ignoring the attenuation at $(f_1 - f_2)$, the effective source level at 1 m when $(f_1 + f_2)/2 = 10$ kHz is $3{\cdot}3 \times 10^{-5}$ W $(f_1 - f_2)^2$ μbar. For a level of +90 dB (1 μbar) at 1 m for $(f_1 - f_2) = 300$ Hz, $P = 11$ kW, and for $(f_1 - f_2) = 3$ kHz, $P = 110$ watts. Thus this is an inefficient method of generating the lower frequencies and marginally practical, although valu-

able narrow beamwidths can be obtained from a small active transducer area. For the measurement of small target strengths at short ranges the effective far-field source level calculated above does not strictly apply, since the target is within the virtual array of sources, unless the primary frequencies are very high, when A will be increased anyway.

4. Conclusions

The rather special requirements of a sound source for the measurement of small target strengths in a restricted environment at audio frequencies has been stated. Of the wide variety of sources available none is completely satisfactory and some compromise will need to be made. Of the linear sources a big loudspeaker with a large but rigid diaphragm may be possible, though expensive. Arrays of resonant transducers would have lower Q values than single elements, but again the cost is prohibitive and only the top end of the frequency range could be reasonably handled this way. By using explosives at very shallow depths so that they blow out of the surface the bubble pulse will be lost, but surface reverberation might be troublesome. The pneumatic source may be used, if the bubble resonance is sufficiently low in frequency that it can be filtered out electrically and provided the initial release is fast enough to generate a sharp pulse.

It has not been possible to cover any source in detail; the major limitations of a wide variety of sources have been considered and the acoustic output given where possible, hopefully to facilitate comparison of the wide variety of types.

5. Acknowledgments

It is a pleasure to acknowledge the helpful discussions with, and the assistance and data of, many colleagues at the National Institute of Oceanography.

6. References

1. R. Bowers, 'Modern methods of sub-bottom profiling using sonic reflection techniques', *Science Progress*, **51**, No. 201, pp. 80–4, January 1963.
2. R. Bowers, 'High-power, low-frequency sonar for sub-bottom profiling', *J. Brit. Instn Radio Engrs*, **25**, No. 5, pp. 457–60, May 1963.
3. J. B. Hersey, H. E. Edgerton, S. O. Raymond and G. Hayward, 'Pingers and thumpers advance deep-sea exploration', *J. Instrum. Soc. Amer.*, **8**, No. 1, pp. 72–7, 1961.
4. J. C. Swallow, 'A neutral buoyancy float for measuring deep currents', *Deep-Sea Research*, **3**, pp. 74–81, 1955.
5. D. Webb and M. J. Tucker, Paper in preparation.
6. R. W. G. Haslett, 'Acoustic backscattering cross-sections of fish at three frequencies and their representation on a universal graph', *Brit. J. Appl. Phys.*, **16**, pp. 1143–50, August 1965.
7. C. Devin, 'Survey of thermal, radiation and viscous damping of pulsating bubbles in water', *J. Acoust. Soc. Amer.*, **31**, No. 12, pp. 1654–67, December 1959.
8. M. Strasberg, 'The pulsation frequency of nonspherical gas bubbles in liquids', *J. Acoust. Soc. Amer.*, **25**, No. 3, pp. 536–7, May 1953.
9. I. B. Andreeva, 'Scattering of sound by air bladders of fish in deep sound scattering ocean layers', *Soviet Physics—Acoustics*, **10**, No. 1, July–September 1964.
10. B. S. McCartney, A. R. Stubbs and M. J. Tucker, 'Low-frequency target strengths of pilchard shoals and the hypothesis of swim-bladder resonance', *Nature*, **207**, No. 4992, pp. 39–40, 3rd July 1965.
11. D. E. Weston, 'Acoustic interaction effects in arrays of small spheres', *J. Acoust. Soc. Amer.*, 39, No. 2, pp. 316–22, February 1966.
12. J. B. Hersey and R. H. Backus, 'Sound scattering by marine organisms', pp. 498–539, in 'The Sea', Vol. 1 (ed. M. N. Hill). (Interscience, New York.)
13. C. C. Sims, 'High-fidelity underwater sound transducers', *Proc. Inst. Radio Engrs*, 47, pp. 866–71, May 1959.
14. A. B. Wood, 'A Textbook of Sound', p. 428, 3rd edition. (Bell, London, 1955.)
15. C. C. Sims, 'Bubble transducer for radiating high-power low-frequency sound in water', *J. Acoust. Soc. Amer.*, **32**, No. 10, pp. 1305–8, October 1960.
16. C. S. Clay, W. L. Liang and S. Wisotsky, 'Seismic profiling with a hydroacoustic transducer and correlation receiver', *J. Geophys. Res.*, 69, No. 16, pp. 3419–28, 1964.
17. D. E. Weston, 'Underwater explosions as acoustic sources', *Proc. Phys. Soc.*, 76, Pt. 2, pp. 233–49, August 1960.
18. J. B. Hersey, S. T. Knott, D. D. Caulfield, H. E. Edgerton and E. E. Hayes, 'Adaption of sonar techniques for exploring the sediments and crust of the earth beneath the ocean', *The Radio and Electronic Engineer*, **26**, No. 3, pp. 245–50, March 1963.
19. V. C. Anderson, 'Wide-band Sound Scattering in the Deep Scattering Layer', Scripps Institution of Oceanography, Publication No. 53–36, 1953.
20. N. A. Roy and D. P. Frolov, 'Generation of sound by spark discharges in water', *Proceedings of the 3rd International Congress on Acoustics, Stuttgart*, 1959, pp. 321–3, 1961.
21. D. D. Caulfield, 'Predicting sonic pulse shapes of underwater spark discharges', *Deep Sea Research*, **9**, pp. 339–48, 1962.
22. J. Ewing and R. Zaunere, 'Seismic profiling with a pneumatic sound source', *J. Geophys. Res.* 69, No. 22, pp. 4913–5, 1964.
23. H. O. Berktay, 'Possible exploitation of non-linear acoustics in underwater transmitting applications', *J. Sound and Vibration*, **2**, No. 4, pp. 435–61, 1965.
24. 'Sonics: Techniques for the Use of Sound and Ultrasound in Engineering and Science' (ed. T. F. Hueter and R. H. Bolt). (Wiley, London, 1955.)

7. Appendix

From the equivalent circuit of Fig. 3(d) the resonant frequency of the swim-bladder of a fish can be determined and is

$$f_r = \frac{1}{2\pi a}\sqrt{\frac{3\gamma p_0}{\rho}} \qquad \ldots\ldots(6)$$

where γ = ratio of specific heats of the gas

p_0 = absolute pressure at the depth of the fish

a = radius of a sphere of the same volume as the swim-bladder

ρ = density of water.

Q at resonance with this simple model in which radiation is the only damping, i.e. ignoring the bladder tissue and thermal losses, is given by

$$\begin{aligned} Q &= R_m/X_m \\ &= 4\pi a^2 \rho c / 4\pi a^3 \cdot 2\pi f_r \rho \\ &= \frac{c}{2\pi f_r a} \quad = \left(c \cdot \sqrt{\frac{\rho}{3\gamma p_0}}\right) \\ &= \frac{1}{ka} \end{aligned} \qquad \text{.....(7)}$$

Q is thus dependent on the depth and at 30 m, would be about 40. Taking other losses into account this may fall[9] to about 10. Figure 1 assumes $Q = 9$.

The power intercepted from a plane wave of pressure p_i by a target of acoustic area σ is $p_i^2 \sigma/\rho c$ which must be equal to F_r^2/R_m, the power dissipated by re-radiation. Thus by straightforward circuit analysis:

$$\sigma(f) = \frac{4\pi a^2}{(4\pi^2 f^2 C_m^2 R_m^2)^{-1} + \left(1 - \frac{f_r^2}{f^2}\right)^2} \qquad \text{.....(8)}$$

The first term in the denominator is only significant at resonance, when it is equal to Q^{-2}.

Manuscript received by the Institution on 8th July 1966. (Paper No. 1113/RNA67/EA33.)

STANDARD FREQUENCY TRANSMISSIONS

(Communication from the National Physical Laboratory)

Deviations, in parts in 10^{10}, from nominal frequency for **March 1967**

March 1967	24-hour mean centred on 0300 U.T.			March 1967	24-hour mean centred on 0300 U.T.		
	GBZ 19·6 kHz† GBR 16 kHz‡	MSF 60 kHz	Droitwich 200 kHz		GBZ 19·6 kHz† GBR 16 kHz‡	MSF 60 kHz	Droitwich 200 kHz
1	— 299·6	— 299·1	— 0·5	17	— 300·2	— 300·2	0
2	— 299·6	— 299.9	— 0·1	18	— 300·2	— 300·2	+ 0·1
3	— 299·2	— 300·1	+ 0·5	19	— 300·3	— 300·3	— 0·1
4	— 300·9	— 299·7	+ 0·3	20	— 300·6	— 300·6	— 0·3
5	— 299·8	— 299·4	+ 0·2	21	— 300·8	— 300·8	— 0·7
6	— 300·4	— 299·1	— 0·1	22	— 301·0	— 301·0	— 0·8
7	—	— 300·1	— 0·2	23	— 299·7	— 299·7	— 0·7
8	— 299·9	— 299·9	— 0·2	24	— 299·3	— 299·3	— 0·7
9	— 300·0	— 300·0	— 0·2	25	— 299·6	— 299·6	— 0·7
10	— 300·2	— 300·2	— 0·3	26	— 299·6	— 299·6	— 0·3
11	—	—	—	27	— 299·2	— 299·2	— 0·2
12	— 298·2	— 298·2	— 0·2	28	— 299·1	— 299·1	— 0·1
13	— 297·5	— 297·5	+ 0·1	29	— 299·0	— 299·0	+ 0·5
14	— 299·6	— 299·6	— 0·1	30	— 299·2	— 299·2	+ 0·4
15	— 299·3	— 299·3	— 0·6	31	— 299·5	— 299·5	+ 0·3
16	— 300·0	— 300·0	— 0·3				

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

Mean monthly values: GBZ — 300·6, MSF — 299·6

† 1st to 6th March inclusive only. ‡ 8th to 31st March inclusive only.

Correction: The value of MSF for 19th February should have been $-297{\cdot}8 \times 10^{-10}$

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.

METHODS FOR TRANSMISSION OVER SATELLITE LINKS

The transmission quality required on satellite links is specified by C.C.I.R. and the overall transmission loss varies from case to case (telecommunications, television service, space probes), but is always at least 50 dB greater than in terrestrial radio links. In order to avoid excessive transmitter power levels in the satellites it is usual to effect noise reduction by increasing the bandwidth up to the limit given by the noise threshold of the particular method.

A German paper summarizes different transmission methods for satellite links and shows that in point-to-point links frequency modulation has slight advantages over pulse code modulation and is considerably better than single-side band transmission. In the case of multiple access by many ground stations, however, a single-sideband system on the ground-to-satellite path is a promising alternative to frequency modulation.

'Transmission methods for satellite links', W. Arens, *Nachrichtentechnische Zeitschrift*, **20**, No. 1, pp. 7–11, January 1967.

DOUBLE-BASE DIODE

The range of application of the double-base diode is steadily increasing. By using this semiconductor element essential simplifications of circuitry are often obtained. A German paper treats the fundamentals of this element, especially its application in relaxation oscillators, multivibrators, etc. One part of the paper deals with the problem of temperature stabilization. It is shown that it is possible to design relatively stable circuits at comparatively low cost.

'The double-base diode and its application', E. Strauch, *Nachrichtentechnische Zeitschrift*, **19**, No. 11, p. 680–92, November 1966.

USE OF SUPERCONDUCTORS IN MASERS

The operation of a maser-type amplifier requires the presence of a continuous magnetic field meeting certain conditions of stability, size and orientation, and also of uniformity and controllability.

The author of a French paper examines the advantages and disadvantages of various ways of obtaining this magnetic field and as a result shows the marked superiority of the use of superconductor windings.

An example of a cavity maser using these types of magnetic field generator is given.

'Advantages in the use of superconductor windings in a maser', J. Grangeon, *L'Onde Electrique*, **47**, No. 478, pp. 129–33, January 1967.

MASER FOR TRAVELLING-WAVE AMPLIFIER

A French paper discusses the design of a ruby maser for travelling wave amplifiers having a relatively large pass-band and designed for use as low-noise amplifiers in satellite communications Earth stations.

A large-scale installation using a maser with superconductor magnetic equipment is at present being designed for the station at Pleumeur-Bodou, in Brittany.

'The design of a maser for travelling waves over a large band at centimetric wavelengths', A. Nizery and B. Loriou, *L'Onde Electrique*, 47, No. 478, pp. 121–28, January 1967.

CIRCUIT COMPONENTS FOR USE AT LOW TEMPERATURES

A German paper describes the most important circuit components operated and studied at temperatures below 20°K as far as they come under the following three spheres of application: low-noise amplifiers, data processing, and microwave techniques. Cooled molecular and reactance amplifiers are used in sensitive receivers because of their low thermal noise. Supra-conductive storage cells are used in data processing, large stores and associative stores. In microwave techniques it is possible to design high-Q resonators, rectifiers, amplifiers, and generators by using supra-conductive films. The operation of such circuit components is briefly explained.

'Cryogenics in telecommunications', K. Goser, *Nachrichtentechnische Zeitschrift*, **20**, No. 1, pp. 44–52, January 1967.

RADIATION SPECTRUM OF AN EXTRA-TERRESTRIAL SOURCE

Reception of noise signals through the side-lobes of a phased array is examined in a Russian paper, and it is shown that for a suitable choice of receiver bandwidth the power received through the main beam will be comparable to the power received through the side-lobes. A method for tracking a radiation source is proposed, and is based on a continuous variation of receiver frequency and a simultaneous change of the direction of a side-lobe maximum with respect to the array axis to follow the location of the source. A method for suppression of signal reception through the main beam from undesirable directions is also explained.

The proposed tracking method enables an analysis of radiation spectrum over an octave of frequencies and can also be used to investigate the brightness distribution of complicated but weak sources.

'A method for the determination of the spectrum of an extra-terrestrial source of radiation', A. A. Pistol'kors, *Radio Engineering and Electronic Physics* (English Edition of *Radiotechnika i Elektronika*), **11**, No. 5, pp. 673–77, May 1966.