THE RADIO AND ELECTRONIC ENGINEER

The Journal of the Institution of Electronic and Radio Engineers

Digital Processing of Signals in Communications

THE invention of the transistor in 1948 may be regarded, without exaggeration, as one of the most significant milestones in the history of radio and electronics. A further impressive step was that taken in the middle 'sixties, when the ability to replicate literally thousands of circuits within a hitherto unimaginably small space was realized—and medium and large-scale integration (m.s.i. and l.s.i.) entered the engineering vocabulary.

Through their contribution to the 'digital revolution' m.s.i. and l.s.i. have 'liberated' the systems engineer and, while computer technology generally springs first to the mind in this context, communications systems have reaped comparable gains. Considerable effort is going into developing these systems at the present time and the value of bringing together those working in the broad field of communications to hear about and discuss the latest advances in digital processing of signals is the object of the three-day conference to be held at the Loughborough University of Technology from 11th–13th April next. The conference is organized by the I.E.R.E. with the association of the Institution of Electrical Engineers and the Institute of Electrical and Electronics Engineers.

That the subject is at a crucial stage may be realized from the 'keynote address' to be given by Dr. R. W. Lucky, of Bell Telephone Laboratories, which is sub-titled 'The Promise and the Problems'. Solutions to some of these problems are essayed in the forty-one papers which will be presented by authors from universities, government establishments and industry; authors of eleven of the papers come from outside Great Britain—from the U.S.A., Canada, Switzerland, Germany, Japan and France.

The conference is arranged as four half-day sessions and one whole-day session, which will between them deal both with techniques and with applications. The sessions on techniques cover frequency and time compression and expansion, adaptive methods of detection and equalization, digital filters, signal design, correlation, coding and error correction. Specific application areas within the extensive scope of communications which the conference will hear about include computer simulation and adaptive systems, and digital signal processing applied to data transmission, facsimile and television transmission, satellite systems and speech scrambling. A full list of papers in the sessions is given on page S.40.

It is perhaps worth pointing out that this conference is probably the most important to have been organized by the I.E.R.E. for the *radio* engineer. It relates emphatically to the employment to the full of modern techniques in communications, something which has tended to be overlooked with the opening up of so many new applications of electronics for other purposes in recent years. But with the ever-expanding interdependence between disciplines it is equally certain that many engineers active in these other fields will wish to attend and make their contribution to the discussions.

J. W. R. GRIFFITHS

Contributors
to this issue

Mr. Pabitra K. Roy received the B.Sc.(Hons.) degree in physics and the B.Tech. and M. Tech. in radio physics and electronics in 1966, 1968 and 1969 respectively, from the University of Calcutta. Since 1970 he has been a Research Fellow at the Centre of Advanced Study in Radio Physics in Electronics at the University. His research interests are in microwave interaction with artificial dielectrics and semiconductors.

Dr. J. R. James (Member 1960, Graduate 1956) is at present researching and lecturing at the Royal Military College of Science, where he is a member of the academic staff. He held industrial positions in the development of airborne radar equipment and as a semiconductor applications engineer prior to joining the College in 1961. He obtained the B.Sc. special mathematics degree of the University of London by private study and gained his Ph.D. from the same University in 1968. Dr. James who is the author and co-author of publications on circuitry, microwave components, antennas and speech processing and holds jointly some provisional patents, is currently researching in electromagnetics and on aspects of communication systems design. He serves on the Institution's Papers Committee.

K. H. C. Phillips (Member 1971) received his technical education at Exeter Technical College, Borough Polytechnic and Wimbledon Technical College. Since 1965 his work has been concerned with the design and development of control systems associated with mechanized letter and parcel handling equipment for the British Post Office. His current position is Senior Executive Engineer responsible for the design and development of data translation and information systems concerned with the control of Mechanized Letter Sorting Offices in the United Kingdom.

Dr. L. T. Bruton holds an associate professorship in electrical engineering at the University of Calgary. A full note on his career appeared in the November 1971 issue of the *Journal*.

Mr. I. N. L. Gallett obtained the University of London degree of B.Sc. in physics at the Royal Military College of Science in 1967. He then held a research appointment concerned with dielectric radiators at the College until 1971, and at present is employed on work of a consultative nature.

Dr. Amarnath Datta received the B.Sc. degree in physics in 1961, and the B.Tech., M.Tech., and D. Phil. degree in radio physics and electronics in 1963, 1965 and 1970, respectively, from the University of Calcutta. In 1965 he became a Research Fellow at the Solid State and Microwave Electronics Laboratory of the Institute of Radio Physics and Electronics at Calcutta. He joined the Centre of Advanced Study in Radio Physics and Electronics at the University of Calcutta as a Research Associate in 1968 and he is now also lecturing in the same Centre. Dr. Datta has been engaged in the study of the transport properties of artificial dielectrics and semiconductors at microwave frequencies, and his other research interests include the study of the Faraday effect in inhomogeneous medium.

Mr. David Cattanach studied at Robert Gordon's Institute, Aberdeen, and then Hatfield College of Technology, gaining an H.N.C. in electrical engineering. In 1962 he joined Hawker Siddeley Aviation as an instrumentation engineer and worked for four years on aircraft electrical systems. He then came to the Marine Laboratory of the Department of Agriculture and Fisheries, Scotland. As a Higher Scientific Officer in the Laboratory's Fish Capture Research Team, he has been responsible for the development of instruments which are used to measure the performance of experimental fishing gears.

Point-matched Solutions for Propagating Modes on Arbitrarily-shaped Dielectric Rods

J. R. JAMES,
B.Sc., Ph.D., A.F.I.M.A., C.Eng., M.I.E.R.E.*

and

I. N. L. GALLETT, B.Sc.*

SUMMARY

Every arbitrarily shaped rod is shown to possess a non-radiating guided wave and a new variational solution is derived. A simpler formulation based on the point-matching principle is outlined and likely sources of computational error are scrutinized. In distinguishing between static and wave-like field solutions it is shown that a convergence criterion based on eigenvalues alone can be misleading, point-matching methods being no worse than some other numerical processes in this respect. The validity of the chosen expansion is a critical factor and two new necessary conditions imposed by the boundary geometry are deduced which are of considerable help in selecting an expansion; these being angular periodicity and radial single-valuedness. No sufficient condition for the expansion validity has been given as yet. Relevant mathematical ideas and theories about the point-matching technique are examined and recommendations made on how best to apply the technique. Computations supported by experimental or other evidence are presented including computations of an arbitrarily shaped non-convex guide and a new equilateral triangular guide. The calculations throughout are restricted mainly to the dominant HE_{11} mode and the present computer program is capable of giving acceptable engineering results for a variety of rod shapes which are of interest.

* Department of Electrical and Electronic Engineering, Royal Military College of Science, Shrivenham, Swindon, Wiltshire SN6 8LA.

1. Introduction

Cylindrical and elliptical dielectric rods and tubes possess discrete modal properties as do bounded metal waveguides and an early exploitation of their surface wave characteristics is seen in the dielectric antenna;[1-3] rectangular dielectric rods were also used as antennas[1] but the modal properties were not calculated until very recently.[4] It is now known that surface waves can be launched on to other types of open dielectric waveguides and there is a growing interest in a variety of cross-sectional shapes that lend themselves to railway signalling,[5] easy-to-construct waveguides[6,7] and image lines,[8] etc. It is apparent that only the method of point-matching seems suitable for their calculation but the method itself is currently in doubt in the literature. Thus in view of the increasing engineering importance of open guides it is essential to clarify the engineering usage of point-matching and this is the main aim of the present paper.

The difficulty with open dielectric guides is that the surface wave is unbounded and the usual finite difference methods[9] etc., would required the insertion of some distant metallic boundary; such an approach is suitable for boxed strip lines[10,36] but not ideal for open guides since the box enforces closed waveguide modes. Another technique[11] for open wire guides uses conformal mapping and an electrolytic tank, is very involved and only gives an approximation. Point matching on the other hand is fairly easy to use, involves only the boundary between the air and dielectric rather than operations on the unbounded fields and appears to give reasonable results. The drawback to point-matching is that it may or may not converge to give the correct eigenvalue and even if it does the resulting modal field patterns at least for bounded waveguides, are considered to be invalid.[22] Goell[4] has recently solved rectangular dielectric rods by point-matching giving no justification for the method and in the circumstance it is natural to question the validity of Goell's[4] field patterns and whether they are sufficiently accurate for a useful engineering approximation.

Point-matching has widespread applications and much of the criticism has arisen from electromagnetic scattering problems. The literature is extensive[12-27] and many differing mathematical theories as to why point-matching is or is not valid or how it is constrained, have been disproved or contradicted and a confused situation exists; it is sufficient to summarize the three theories that have so far survived. Bates[22,24] has deduced that point-matched solutions for closed waveguides give incomplete field representations but the eigenvalues so obtained are identical to those calculated by the so-called exact extended boundary condition (e.b.c.) method; Burrows[17,19,21] has proved that the e.b.c. and point-matching methods are mathematically identical for scattering problems but certain reservations are admitted; Millar[18,23] has given necessary and sufficient conditions for the validity of a particular point-matching method.

In contrast there are many electromagnetic problems for which the general point-matching principle has been successfully adopted without question; the principle is the matching of modal expansions on geometrical boundaries which are not compatible with the co-

ordinate system and we mention the solution of tapered waveguides,[28] parallel arrangements of open cylindrical waveguides[29,30] and also cylindrical scatterers.[31,32] It is significant that excellent results have been obtained for a static temperature[33] problem involving Laplace's equation. Application to mechanical problems are numerous and further references are given in the literature.[33,34]

The sequence adopted in this paper is as follow. First we examine the existence of surface waves on an arbitrarily-shaped rod since very sharp corners may cause a slow leak of radiation; in Section 3 we outline our computational point-matched procedure, paying particular attention to numerical error effects which are so often glossed over. A distinction between field problems involving the Laplace as opposed to the Helmholtz equation is emphasized and is recognized as a crucial factor in explaining the effectiveness of point-matching for static field problems. These considerations and some necessary constraints imposed by the rod cross-sectional shape enable a recommended approach to point-matching to be evolved. Throughout the paper particular emphasis is placed on the computed field patterns as well as the eigen-value convergence and in Section 5 additional illustrative examples are calculated which should be helpful to engineers who wish to calculate field data on new types of open dielectric guiding structures with unknown modal properties. To strengthen our results on theoretical grounds we include a critical examination of the three above-mentioned theories in an appendix; as a consequence our recommendations for open dielectric rod guides may be interpreted quite generally for other electromagnetic computations based on point-matching and we briefly discuss this in our conclusions (Section 5).
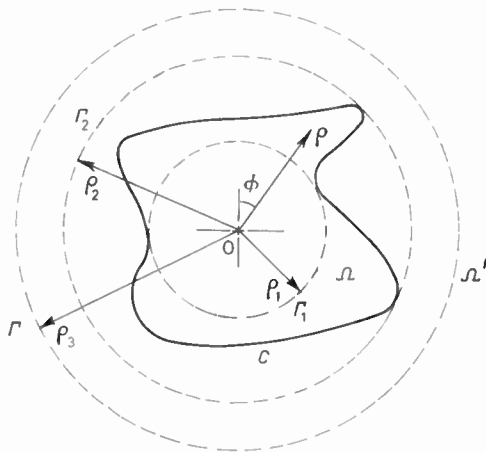

Fig. 1. Arbitrary contour C with origin O; Γ_1 and Γ_2 are the inscribed and excribed circles respectively relative to centre O and $\rho_3 > \rho_2$.

2. Existence of Discrete Modes

It has been shown[35] that any arbitrarily-shaped perfectly conducting metallic waveguide has a discrete propagating mode but we are not aware of a similar generalization for arbitrarily-shaped open dielectric guides. It will be helpful to establish this theoretically

and we adopt a variational approach similar to the bounded guide treatment.[35]

A dielectric rod has a cross-section Ω bounded by a simply connected contour C (Fig. 1). Since the rod is uniform in the z direction we may specify the fields within C by the E and H-wave Hertzian potentials $\Pi_{z_1}^E$ and $\Pi_{z_1}^H$ respectively satisfying

$$\nabla_1^2 \left\{ \begin{matrix} \Pi_{z_1}^E \\ \Pi_{z_1}^H \end{matrix} \right\} + k_1^2 \left\{ \begin{matrix} \Pi_{z_1}^E \\ \Pi_{z_1}^H \end{matrix} \right\} = 0 \qquad \ldots(1)$$

where $k_1^2 = \omega^2 \mu_1 \varepsilon_1 - \beta^2$, β is the axial wave number, μ_1 and ε_1 are the permeability and permittivity of the rod. The fields obey the rod boundary conditions thus Π_z^E and Π_z^H are dependent; we will choose Π_z^E then as some trial solution and on substituting it into the 2-dimensional Green's theorem with $\delta/\delta n$ as differentiation with respect to the outward normal to the boundary contour

$$\int_\Omega \left(\frac{\partial \Pi_{z_1}^E}{\partial x} \right)^2 + \left(\frac{\partial \Pi_{z_1}^E}{\partial y} \right)^2 \, da = k_1^2 \int_\Omega \left(\Pi_{z_1}^E \right)^2 \, da +$$
$$+ \int_C \Pi_{z_1}^E \frac{\partial \Pi_{z_1}^E}{\partial n} \, dC \qquad \ldots(2)$$

External to C $\Pi_{z_2}^E$ and $\Pi_{z_2}^H$ satisfy equation (1) but with k_1 replaced by k_2 where $k_2^2 = k^2 - \beta^2 = k_1^2 - k^2(\varepsilon_r - 1)$, ε_r = relative permittivity of rod, k is free space wave number. Applying the Green's theorem to $\Pi_{z_2}^E$ for the region Ω' bounded internally by C and externally by a circle Γ radius ρ_3 (Fig. 1):

$$\int_{\Omega'} \left(\frac{\partial \Pi_{z_2}^E}{\partial x} \right)^2 + \left(\frac{\partial \Pi_{z_2}^E}{\partial y} \right)^2 \, da = \{k_1^2 - k^2(\varepsilon_r - 1)\} \times$$
$$\times \int_{\Omega'} \left(\Pi_{z_2}^E \right)^2 \, da - \int_C \Pi_{z_2}^E \frac{\partial \Pi_{z_2}^E}{\partial n} \, dC + \int_\Gamma \Pi_{z_2}^E \frac{\partial \Pi_{z_2}^E}{\partial n} \, dC \ldots(3)$$

On adding equations (2) and (3) with subscripts 1 and 2 omitted, the line integrals over C are eliminated and the line integral over Γ vanishes as $\rho_3 \to \infty$ if the trial field does not radiate; sufficient conditions are for the fields to satisfy Sommerfeld's radiation condition and in addition the radial decay of Π_z^E must exceed that of $\rho^{-\frac{1}{2}}$ as $\rho \to \infty$.

On rearranging we have

$$k_1^2 = \frac{\left\{ k^2(\varepsilon_r - 1) \int_{\Omega'} \left(\Pi_z^E \right)^2 da + \int_{\Omega + \Omega'} \left(\frac{\partial \Pi_z^E}{\partial x} \right)^2 + \left(\frac{\partial \Pi_z^E}{\partial y} \right)^2 da \right\}}{\int_{\Omega + \Omega'} \left(\Pi_z^E \right)^2 da} \qquad \ldots(4)$$

All integrals in equation (4) are positive for any trial field which obeys the above radial evanescent constraint and the equation is found to be variational[48] such that for any non-radiating trial solution, k_1^2 is always a real positive number and stationary about the correct value; thus a discrete guided mode always exists for any shape of C. Equation (4) could be used to solve for k_1 and if the contour integrals over C are left in then the trial field need not satisfy the rod boundary conditions; the usual Rayleigh–Ritz formulations etc. also being applicable. However for loosely bound surface waves the integrations over Ω' are not an attractive proposition.

Lossless circular rods are known to possess leaky wave solutions[38] which violate the above radial evanescent constraint but by imposing the latter we exclude these improper modes having k_1 complex. Presumably improper modes also exist for arbitrarily shaped rods and are necessary for the treatment of discontinuities as in planar and cylindrical open guide cases.[39]

3. Point-matching Procedure for Open Dielectric Guide

3.1 Computer Program

Details for bounded guides[14] and rectangular dielectric guides[4] have been described elsewhere; here we outline our formulation and discuss those aspects which are of particular interest. We take the E-wave Hertzian vector potentials with n integer and unknown coefficients a, b, c, d

$$\Pi_{z1}^E = \sum_{n=0}^{\infty} \{c_n^E \cos n\phi + d_n^E \sin n\phi\} \, J_n(k_1\rho) \exp j\beta z$$

$$\Pi_{z2}^E = \sum_{n=0}^{\infty} \{a_n^E \cos n\phi + b_n^E \sin n\phi\} \, H_n^{(1)}(k_2\rho) \exp j\beta z \quad (5)$$

for the interior and exterior fields respectively, with respect to the contour C (Fig. 1). The H-wave vector potentials are similarly defined and obtained from equation (5) by replacing the E superscripts by H; thus for $n = 0$ there are 4 unknown coefficients and 8 for any other value of n. A time variation of $\exp(-j\omega t)$ is adopted and suppressed, $k_1^2 = \omega^2 \mu_1 \varepsilon_1 - \beta^2$, $k_2^2 = k^2 - \beta^2$, k_1 is defined to be a positive real number and k_2 a positive imaginary number. We take a cylindrical coordinate system with origin O (Fig. 1), and the 6 field components on C are readily obtained from the Hertzian vectors (Stratton,[40] chap. 6); the point-matching procedure follows the usual practice n being truncated at a value N. Implicit in this formulation is the assumption that the solution can be obtained by these expansions and this is discussed later on.

It is clear that open dielectric guides require more coefficients than closed guides on account of the separate representations for the internal and external fields and in addition the need for both TE and TM wave solutions. Our formulation required still more coefficients because of its generality regarding the rod geometry and we specify both odd and even wave solutions. The determination of an eigenvalue required an iterative procedure; some of the zeros were very close together and the determinant's value changed rapidly either side of the zeros; in these circumstances we found a simple first-order iterative scheme[41] most suitable. Some pre-knowledge of the root location saves computing time. To identify an acquired mode we gradually deformed a contour with established properties into the required shape, calculating and tracing the desired eigenvalue for each successive deformation.

3.2 Sources of Error

The convergence of eigenvalues and their corresponding fields with increasing N may be inhibited by computer capacity, by the distribution of the points on C, by an ill-conditioned[43] matrix equation or perhaps by a combination of these effects; above all, the assumed field expansion may be only an approximate representation. The present authors (with Slade)[42] have given a qualitative assessment of this situation with recommendation as to how these effects may be isolated.

A field property that has been given insufficient attention is that solutions to Laplace's equation, have an intrinsic error bound. Suppose that ψ satisfies $\nabla^2 \psi = 0$ within a domain D of a plane and on a contour C that bounds D. Let ψ' be a solution obtained by point-matching ψ at discrete points on C. Then $\psi - \psi'$ also satisfies Laplace's equation and achieves its maximum and minimum values on C by the maximum-minimum principle.[44] But $\psi - \psi'$ expresses the error within D and on C and therefore this error has its greatest value on the boundary. Thus for Laplace's equation we may expect point-matched field solutions to be at least as accurate as their fit on the boundary; this property also holds for 3-dimensional fields. Ojalvo and Linzer[33] have used this property to get very accurate point-matched solutions to static temperature problems.
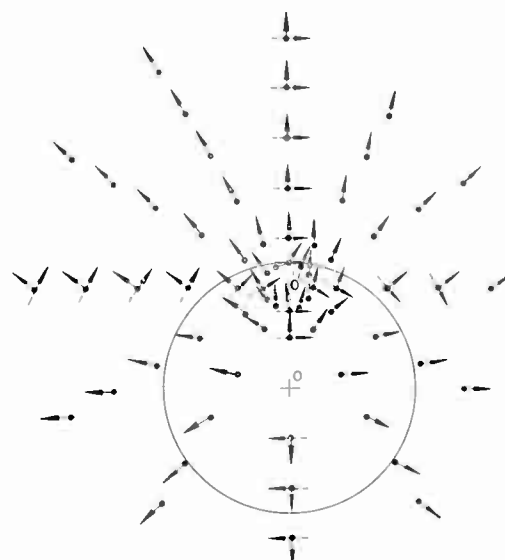

Fig. 2. Direction of transverse fields of TM_{01} mode computed at origin O'. Rod radius = $0.5\lambda_0$, $\varepsilon_r = 2.5$, OO' = $0.4\lambda_0$, λ_0 = free space wavelength, $N = 10$; ——→ electric field; – – –→ magnetic field.

For the Helmholtz equation no such theorem exists as far as we are aware and it is easy to find functions which do not have their maximum and minimum values on the bounding contour; for instance the axial E field of a lossless metal waveguide is zero on the boundary and non-zero within. For the Helmholtz equation then, the good convergence of the fields on the boundary does not guarantee field convergence elsewhere to the same order; this behaviour is likely to be significant for other numerical processes where the field satisfies the boundary conditions at discrete points. For guided waves the eigenvalue convergence is a measure of the field convergence on C.

For the examples shown in this paper we achieved a high degree of eigenvalue convergence using up to
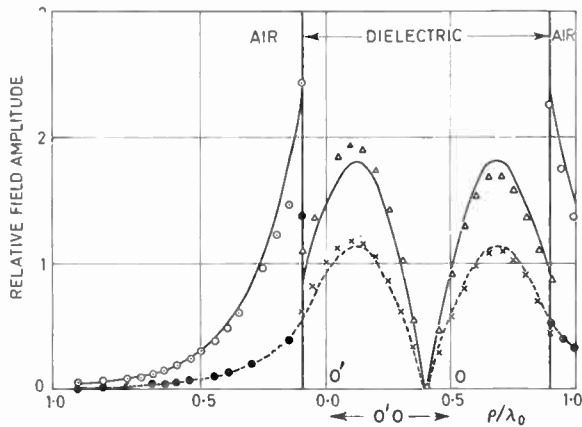
Fig. 3. Transverse field amplitudes corresponding to Fig. 2.
—— E_ρ exact; ———— H_ϕ exact; ○● computed from external expansion; △ × computed from internal expansion. The scales have been adjusted for ease of presentation and H_ϕ has been multiplied by 10^2 relative to E_ρ.

$N = 10$ and our matrix equations were always arranged to be well conditioned;[43] we reason that subsequent field errors are due either to the use of an invalid field expansion or the above-mentioned difficult nature of the Helmholtz equation or perhaps a combination of both effects.

Our first example (Fig. 2) shows the calculated fields for the symmetric TM_{01} mode on a cylindrical dielectric rod. The origin was offset to O' and with N at our computer limit of 10, the fields were matched at $2N+1 = 21$ points on the air/dielectric interface; the points subtended equal angles at O' and were thus more densely spaced on the boundary near to O'. Using this origin the eigenvalue β converged to $1 \cdot 20014k$, compared with a value $1 \cdot 20050k$ which was obtained by repeating the process with the origin O coincident with the rod axis. The transverse field directions agree well with the exact patterns which are radial lines and concentric
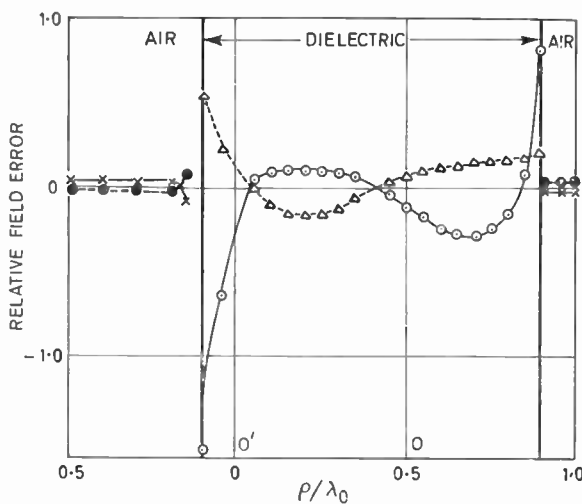

Fig. 4. Relative errors of H_ρ and E_ϕ fields corresponding to Fig. 2.
—— E_ϕ; ———— H_ρ; × ● computed from external expansion; ○ △ computed from internal expansion. The scales correspond to those in Fig. 3.

circles centre O for the transverse E and H fields respectively; some scattered errors occur near O' and to examine these we plot the exact and calculated transverse fields on a line passing through O' and O (Fig. 3). E_ρ and H_ϕ only show appreciable discrepancies near the rod boundary and in this vicinity H_ρ and E_ϕ (Fig. 4) differ considerably from their exact value which is zero. We are confident† that these errors would decrease by an order if N were increased to 12.

For elliptical contours with eccentricity $\leqslant \sqrt{8}/3$ we obtained good field convergence with $N = 10$ and the origin at the intersection of the major and minor axis. Figure 5 shows the HE_{11} odd and even eigenvalues for an elliptical rod with eccentricity $\sqrt{8}/3$ which is in excess of the limit 0·9 experienced by Mullin et al.[12] Eigenvalue convergence in the fourth place was obtained from $N = 8$ and the results are in good agreement with Yeh's[37] exact values which we have transposed as accurately as possible. The corresponding field data for smaller eccentricities, shown in Fig. 6 was computed for $N = 10$ and shows no scattered errors or observable disagreement with backward diode measurements. The experimental results will be referred to later.
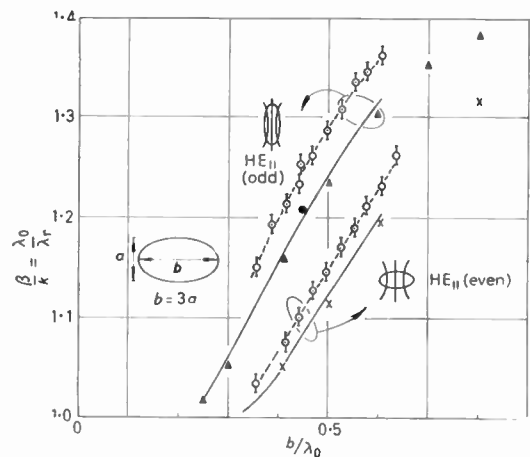

Fig. 5. β/k values for odd and even HE_{11} mode on an elliptical dielectric rod; λ_r = wavelength on rod. —— exact result Yeh[37]; × △ computed with $\varepsilon_r = 2 \cdot 5$, $N = 8$ and origin at centre; ● computed with $\varepsilon_r = 2 \cdot 56$; ––○–– measured on perspex rod with $\varepsilon_r \simeq 2 \cdot 65$.

Computer capacity has prevented us from exploring the mathematical validity of equation (5) in these examples; we estimate that the accuracy of field patterns will readily satisfy engineering requirements if $N = 14$ for the circular rod and $N = 12$ for the elliptical rod computations.

3.3 Necessary Constraints on Rod Geometry

Since k_2 is defined to be a positive imaginary number, Π_{zz}^E (equation (5)) represents an external rod field which is decaying in magnitude in an outward-going radial direction. For a given choice of origin, contour C may not be single-valued; an outward-going radial line may cut C more than once and along this line, in the vicinity

---

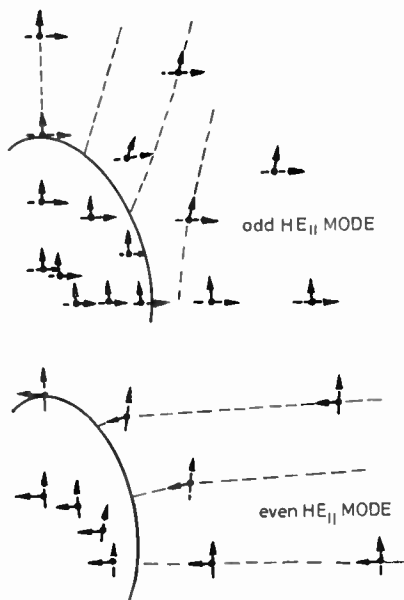† Recent computations on a larger computer substantiate these estimates.

Fig. 6. Transverse fields of elliptical dielectric rod supporting HE_{11} mode; $\longrightarrow$ E $-\longrightarrow$ H field directions computed with $\varepsilon_r = 2\cdot5$; for odd mode $a = 0\cdot16\lambda_0$, $b = 0\cdot25\lambda_0$, $\beta = 1\cdot0980k$ and for even mode $a = 0\cdot14\lambda_0$, $b = 0\cdot25\lambda_0$, $\beta = 1\cdot0295k$. Origin at centre in each case. $----$ electric field measurements on perspex rod $\varepsilon_r \simeq 2\cdot65$, at $\lambda_0 = 4$cm.

of the rod, a monotonic decay will not be obtained. To demonstrate this, two cylindrical dielectric rods were joined by a thin dielectric strip to form the elongated figure-of-eight contour shown in Fig. 7. With origin O, both a theoretical estimate and probe measurement of $|E_z|$ shows that along OP the field does not settle down to a monotonic decay until $\rho/a > 8$.
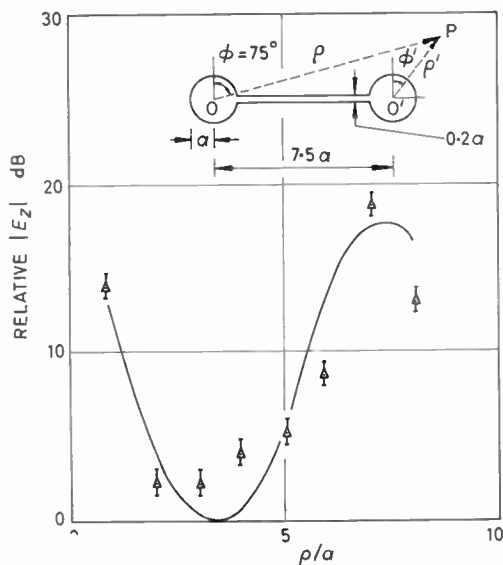
Fig. 7. Variation of E_z field of non-convex dielectric rod along the direction OP. Radius $a = 0\cdot75$cm, $\lambda_0 = 4$cm, polarized with E plane perpendicular to OO'; $\triangle$ measured with probe for perspex rod $\varepsilon_r \simeq 2\cdot65$; —— theoretical estimation of $E_z = H_1^{(1)}(j0\cdot415\rho/a)$ $\cos \phi + H_1^{(1)}(j0\cdot415\rho'/a) \cos \phi'$ with $\varepsilon_r = 2\cdot5$, based on HE_{x1} mode in each circular section.

A convex contour is always single-valued and a non-convex contour may be single-valued for a given choice of origin; in the above example C is never single-valued. A necessary condition for equation (5) to be valid is that C be single-valued. If single-valuedness cannot be obtained then both $H_n^{(1)}$ and $H_n^{(2)}$ Hankel functions must be embraced in an intermediate expansion. We mentioned several successful point-matching calculations in the introduction[29-32] and observe that the choice of field expansions in the latter is compatible with the above constraint. Intermediate expansions have also been used in scattering problems.[53]

Distinct angular periodicity of a contour can constrain n in equation (5) to be non-integer. For instance the contour shown in Fig. 8(a) is composed of a circle and sectors of a circle arranged symmetrically about AOB. In the annular region 1 both the internal and external fields have identical values at $\phi = \phi_1$ and $\phi = 2\pi - \phi_1$.
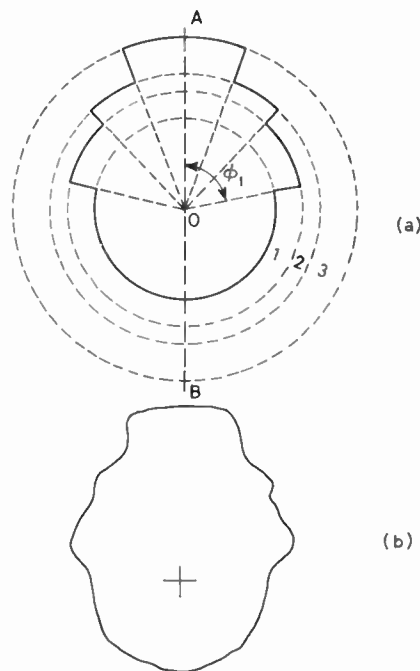

Fig. 8. Contours possessing (a) distinct and (b) approximate angular periodicities which demand fractional values of n in equation (5).

The external field is therefore periodic in the interval $(2\pi - 2\phi_1)$ and the internal field periodic in the interval $2\phi_1$. $\Pi_{z_1}^E$ and $\Pi_{z_2}^E$ (equation (5)) have different periodicities and thus different non-integer values of n; $\Pi_{z_1}^E$ also requires an additional expansion in Y_n functions for this annular region. Annular regions 2 and 3 are similarly treated but the central circular portion and the field external to region 3 can be represented by equation (5) with n integer. This particular contour may therefore be represented exactly by 8 expansions, 6 of which have fractional orders of cylinder functions. Figure 8(b) shows a contour which approximates to the latter but shows no distinct angular properties. For this case the 8 expansions will give better point-matching results than just one internal and one external expansion but the 6

107

additional expansions escalate the computational effort to a level which is no longer attractive and the apparent simplicity of point-matching is lost. We state that distinct angular periodicities, other than 2π, necessitate fractional values of n in equation (5) and thus additional intermediate expansions. For bounded waveguides the situation is less involved and some point-matching calculations based on fractional-order Bessel functions have been described.[22, 54]

3.4 Recommended Procedure

We emphasize that our field solutions appear to be little or no less accurate than some results for bounded guides calculated by other techniques.[45, 46, 47] Recent mathematical theories on point-matching carry the implication that it is notably inferior to other numerical expansion processes and we find it necessary to critically examine the claims in the Appendix. It is the application and not the method itself which leads to difficulties and we recommend that the following points be borne in mind for open guides; the interpretation for other electromagnetic point-matching problems can be readily extracted.
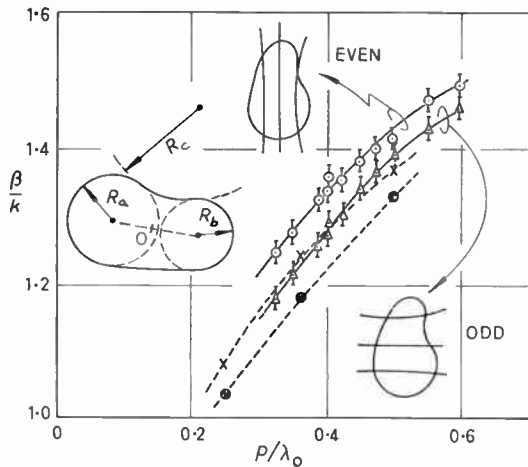

Fig. 9. β/k values for odd and even dominant modes on a dielectric rod, specified by dimensions R_a, R_b, R_c where $p = R_a + R_b$ and $R_b : R_a : R_c = 1 : 2 : 4$. $-\times--\bullet-$ computed with $\varepsilon_r = 2.5$, $N = 8$ and origin at O mid-way between centres. $-\bigcirc--\triangle-$ measured on perspex rod $\varepsilon_r \simeq 2.65$.

(1) Choose a field expansion appropriate to the rod geometry which satisfies the radial evanescent constraint; equation (5) is most suited to nearly circular contours and the radial evanescent constraint is readily satisfied.

(2) Try to choose an origin that makes C single-valued otherwise intermediate expansions will be necessary to accommodate both radial growth and decay.

(3) Identify any distinct or approximate angular periodicities in C which necessarily demand definite periodic behaviour of the field expansions; additional intermediate expansions may be required.

(4) A fundamental difficulty is that no sufficiency condition for the validity of an expansion is available, other than in separable geometries, and hence we may
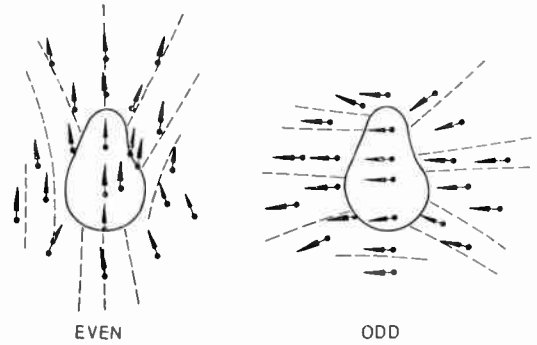

Fig. 10. Direction of transverse electric fields for rod of Fig. 9. $\longrightarrow$ E field direction computed with $\varepsilon_r = 2.5$, $p = 0.5\lambda_0$, $N = 8$. $----$ probe measurement of E field direction on perspex rod $\varepsilon_r \simeq 2.65$, $\lambda_0 = 4$cm. For even mode computation $\beta = 1.3687k$ and for odd mode $\beta = 1.3333k$.

fail to obtain eigenvalue convergence or perhaps the process may appear to converge up to a given N beyond which it diverges. There is no alternative but to adopt a different expansion bearing in mind points (1)–(3) again.

(5) When eigenvalue convergence is obtained to several decimal places, satisfactory field convergence must also be achieved before a solution can be claimed.

4 Illustrative Examples

It is of particular interest to assess the capability of equation (5) which is one of the simplest formulations for open dielectric guides and thus represents a minimum demand on computer capacity. We therefore constrain the examples to have single-valued contours with no prominent angular periodicities. We limit N at 10, thus allowing 21 matching points and necessitating the iterative computation of the zeros of an 84×84 determinant.

We have not carried out a detailed mode study for any particular shape of contour but from a few checks anticipate that a dominant dipole mode with no cut-off frequency is likely always to exist. Since these dipole modes are of considerable practical interest we concentrate our attention in this respect. All the dominant dipole modes shown here have been identified by
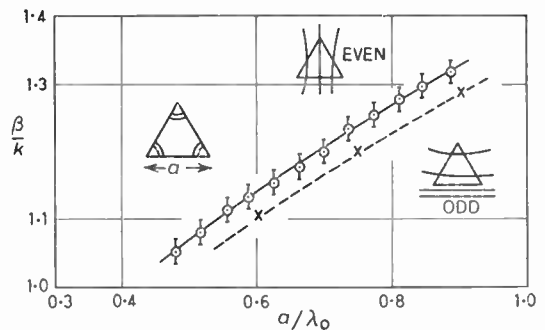

Fig. 11. β/k values for odd and even dominant modes on equilateral triangular dielectric rod. These two different modes had almost coincident eigenvalues; $--\times--$ computed modes with $\varepsilon_r = 2.5$, $N = 6$ and origin at centre; $-\bigcirc-$ measured on perspex rod $\varepsilon_r \simeq 2.65$.

deforming a circular contour into the desired shape, recalculating at each deformation.

The field measurements were carried out on perspex rods machined to within 0·025 mm (0·001 in); the rods were 1 m in length and the dipole modes were launched from a circular horn, paying due regard to the plane of polarization. Lossy screening was used to reduce residual launcher radiation. We were uncertain of the uniformity of the perspex material since the wavelength measurements showed more scatter than usual for this type of experiment. From the elliptical rod results of Fig. 5 we estimate that the perspex has $\varepsilon_r = 2·65$ and our policy throughout has been to compute results with $\varepsilon_r = 2·5$ which is consistent with Yeh's[37] exact calculations. The validity of our computed results may then be judged by comparison with the experiment values together with the comparative information provided by Fig. 5.

Figure 9 shows the odd and even dipole modes on a non-convex dielectric rod, the 17 points subtending equal angles of the origin O; 4th decimal place eigenvalue convergence was readily obtained but the computed fields in Fig. 10 show small scattered errors. We consider that the eigenvalues are valid by comparison with the experimental curve and the results of Fig. 5. The calculation of this particular shape did not appear to present any problems and we estimate that if $N = 12$ the scattered field errors would be difficult to detect in comparison with measurement error.
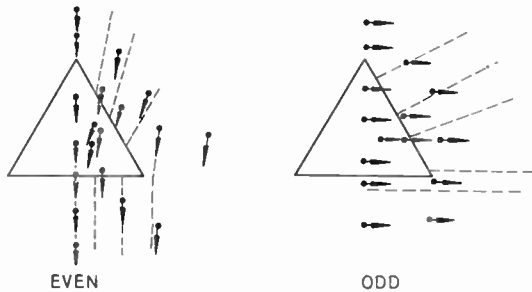

Fig. 12. Direction of transverse electric fields for triangular rod of Fig. 11. —→ E field direction computed with $\varepsilon_r = 2·5$, $a = 0·6\lambda_0$, $N = 6$. ---- probe measurements of E field direction on perspex rod $\varepsilon_r \simeq 2·65$, $\lambda_0 = 4$cm. For computed even mode $\beta = 1·1999k$ and for computed odd mode $\beta = 1·1902k$.

Next we consider a contour with sharp corners and Fig. 11 shows the dipole mode eigenvalues for an equilateral triangular contour which converged to 4th place of decimals for $N = 6$. An interesting feature is that the eigenvalues of the odd and even dipoles modes were almost coincident over a wide range of dimensions and the experimental values in Fig. 11 are a mixture of both. One explanation is that the odd mode is the superposition of the even modes with respect to the other two apexes of the triangle; the odd and even eigenvalues are therefore identical and the difference experience here is a measure of the computational error. The theoretical aspects of this situation are being studied. The fields for the even mode were surprisingly accurate for $N = 6$ but the odd mode fields (Fig. 12) show a disagreement that might be anticipated for so few matching points. We
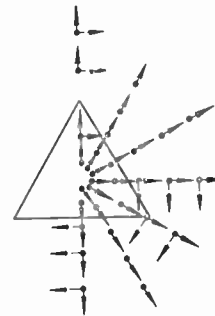

Fig. 13. Direction of transverse electric and magnetic fields for pseudo-symmetric mode on triangular rod of Fig. 11. —→ E field; --→ H field computed with $\varepsilon_r = 2·5$, $a = 0·9\lambda_0$, $N = 10$, $\beta = 1·02347k$.

estimate† that $N = 12$ would give satisfactory field values although a residue of error will always be present in the corners. In Fig. 13 we show a pseudo-symmetric mode, so called because it was obtained by the deformation process from the symmetric mode on a circular rod in Fig. 2; we did not check this experimentally because of launching difficulties.

Returning to the question of the corners we find open dielectric guides more difficult in this respect than closed guides. For the latter, the fields may be inspected to trace out a contour C'' upon which the Dirichlet or Neumann boundary conditions are satisfied. The given number of matching points may then be distributed to make C'' approximate well to the waveguide cross-sectional shape. For open guides we find that the boundary conditions are only satisfied at the matching points and elsewhere the continuity of tangential E fields follows a different contour to the continuity of the H fields. The points can be distributed for a given N to make both these continuity contours approximate well to C but the process is tedious. We have found the following quick method very useful although we offer no theoretical justification for its sufficiency. For a single-valued contour the points may be regarded as lying on some contour C' which has a unique representation as a Fourier series. If the number of terms is $2N + 1$ then they can be evaluated from the
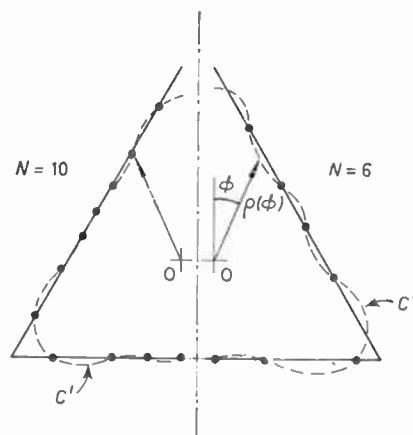

Fig. 14. Distributions of 21 and 13 points on an equilateral triangular rod corresponding to $N = 10$ and 6 respectively. C' is the contour defined by a Fourier representation.

† See footnote on page 106.

$2N+1$ points by a matrix procedure. We calculate C' several times and distribute the $2N+1$ points until C' is visually a good approximation to C. The distributions of points used for the results of Figs. 11, 12 and 13 are given in Fig. 14 and were arrived at by calculating and adjusting C' which is also shown. When C' fitted C more closely for a given N then in general better field fits were obtained, particularly near the corners.

Finally we have calculated rectangular dielectric rods with similar success; Goell[4] has effected computer capacity economy by utilizing the symmetry of rectangular rods and has been able to use more matching points; we consider that his results are very accurate.

Our experience indicates that the calculation of contours similar to the above based on equation (5) presents little problem provided sufficient points are taken to obtain field convergence, the criterion for the latter being comparison with measured or other data. The distribution of points remains very much a cut-and-try process.

5 Discussion and Conclusions

As a first step in this investigation we have proved that a non-radiating surface wave exists on every arbitrarily-shaped open dielectric waveguide and have revealed a new method of solution for open dielectric guides based on variational techniques. The variational integrals must be taken over the infinite transverse plane $\Omega+\Omega'$ and our earlier comments that the point-matching method is easier to apply still hold.

Our good point-matching results have been obtained at a time when the literature reflects serious doubts about the employment of the technique and we attribute our success to the identification of those issues which are significant: Millar's necessary condition is seen to be weak for open guides. The outcome of our critical review of recent mathematical ideas is that we can find no evidence that point-matching is itself a questionable technique; it is purely a question of choosing field expansions which can be fitted to the boundary conditions to achieve a desired accuracy. The user is committed to making some guess at these expansion and there appears to be no mathematical constraint which is sufficient to test their validity; it is emphasized however that other numerical expansion techniques share this difficulty. We have been able to relieve the difficulty somewhat by revealing two necessary constraints of single-valuedness and angular periodicity, imposed by the rod geometry. Finally we have shown that even with appropriate expansions the eigenvalues often converge faster than their corresponding fields; this can be misleading but we emphasize that other numerical processes can experience this difficulty.

The impression is often conveyed that point-matching is an inherently simple technique and this is untrue. From Section 3.3 it is seen that several intermediate expansions are required to fit contours with complex shapes in which case point-matching offers no simplicity over any other process; to try to force a few simple expansions to fit a complex contour is a misuse of the technique and accounts for much of the adverse criticism.

Our particular formulation is simple in that only two expansions are used and our choice of cylinder functions

means that it is ideally suited to nearly circular contours. Our results show however that the contours can digress appreciably from circular and the computer program is capable of calculating a variety of interesting shapes with an accuracy that will satisfy most engineering requirements. We conclude that point-matching is eminently suitable for the evaluation of these open dielectric waveguides and we believe that our recommendations can be usefully interpreted for other electromagnetic calculations where the technique is being employed.

6 Acknowledgments

7 References

1. Kiely, D. G., 'Dielectric Aerials' (Methuen Monograph, London, 1953).

2. James, J. R., 'Theoretical investigation of cylindrical dielectric-rod antennas', *Proc. Instn Elect. Engrs*, 114, pp. 309–19, 1967.

3. Gallett, I. N. L., 'Radiation patterns of tubular dielectric structures', European Microwave Conference September 1969 (IEE Conference Proc. No. 58).

4. Goell, J. E., 'A circular harmonic computer analysis of rectangular dielectric waveguides', *Bell Syst. Tech. J.*, 48, pp. 2133–60, 1969.

5. Young, L., 'Advances in Microwaves', pp. 191–300 (Academic Press, New York, 1969).

6. Tischer, F. J., 'Analysis of the multiple-strip H guide', *Proc. Instn Elect. Engrs*, 116, pp. 1509–13, 1969.

7. Diament, P., Schlesinger, S. P., and Vigants, A., 'A dielectric surface wave structure: the V-line', *I.R.E. Trans. on Microwave Theory and Techniques*, MTT-9, pp. 332–7, 1961.

8. Schlesinger, S. P. and King, D. D., 'Dielectric image lines', *I.R.E. Trans.*, MTT-6, pp. 291–9, 1958.

9. Wexler, A., 'Computation of electromagnetic fields'. *I.E.E.E. Trans.*, MTT-17, pp. 416–39, 1969.

10. Davis, J. B. and Corr, D. G., 'Computer analysis of the fundamental and higher order modes in single and coupled microstrip', *Electronics Letters*, 6, pp. 806–8, 1970.

11. Irikov, V. A., 'Single wire lines for surface waves whose cross section has a complex shape', *Radio Engng Electron Phys.* 12, pp. 424–31, 1967.

12. Mullin, C. R., Sandburg, R. and Velline, C. O., 'A numerical technique for the determination of scattering cross sections of infinite cylinders of arbitrary geometrical cross section', *I.E.E.E. Trans. on Antennas and Propagation*, AP-13, pp 141–9, 1965.

13. Harrington, R. F., 'On the calculation of scattering by conductors', *I.E.E.E. Trans.*, AP-13, pp. 812–3, 1965.

14. Yee, H. Y. and Audeh, N. F., 'Uniform waveguides with arbitrary cross section considered by the point-matching method', *I.E.E.E. Trans.*, MTT-13, pp. 847–51, 1965.

15. Bates, R. H. T., 'The point-matching method for interior and exterior two dimensional boundary value problems', *I.E.E.E. Trans.*, MTT-15, pp. 185–7, 1967.

16. Bates, R. H. T., 'Modal expansions for electromagnetic scattering from perfectly conducting cylinders of arbitrary cross section', *Proc. Instn Elect. Engrs*, 115, pp. 1443–5, 1968.

17. Burrows, M. L., 'Equivalence of the Rayleigh solution and the extended boundary-condition solution for scattering problems', *Electronics Letters*, 5, pp. 277–8, 1969.

18. Millar, R. F., 'Rayleigh hypothesis in scattering problems', *Electronics Letters*, 5, pp. 416–7, 1969.

19. Burrows, M. L., 'Rayleigh hypothesis in scattering problems', *Electronics Letters*, 5, pp. 417–8, 1969.

20. Bates, R. H. T., 'Rayleigh hypothesis; the extended boundary condition and point-matching', *Electronics Letters*, 5, pp. 654–5, 1969.

21. Burrows, M. L., 'Example of the generalized-function validity of the Rayleigh hypothesis', *Electronics Letters*, 5, pp. 694–5, 1969.

22. Bates, R. H. T., 'The theory of the point-matching method for perfectly conducting waveguides and transmission lines', *I.E.E.E. Trans.*, MTT-17, pp. 294–301, 1969.

23. Millar, R. F., 'On the legitimacy of an assumption underlying the point-matching method', *I.E.E.E. Trans.*, MTT-18, pp. 325–6 and 327, 1970.

24. Bates, R. H. T., 'On the legitimacy of an assumption underlying the point-matching method', *ibid.*, pp. 326–7.

25. Mittra, R. and Wilton, D. R., 'A numerical approach to the determination of electromagnetic scattering characteristics of perfect conductors', *Proc. Instn Elect. Engrs*, 117, pp. 2064–5, 1969

26. Lewin, L., 'On the inadequacy of discrete mode-matching techniques in some waveguide discontinuity problems', *I.E.E.E. Trans.*, MTT-18, pp. 364–9 and pp. 371–2, 1970.

27. Nielsen, E. D., 'On the inadequacy of discrete mode-matching techniques in some waveguide discontinuity problems', *ibid.*, pp. 396–71.

28. Shockley, T. D., Haden, C. R. and Lewis, C. E., 'Application of the point-matching method in determining the reflection and transmission coefficients in linearly tapered waveguides', *I.E.E.E. Trans.*, MTT-16, pp. 562–4, 1968.

29. Cook, K. R. and Teh-Ming Chu, 'Mode coupling between surface wave transmission lines', *I.E.E.E. Trans.*, MTT-17, pp. 265–70, 1969.

30. Veselov, G. I. and Krekhtunov, V. M., 'Natural oscillations in a system of N open dielectric waveguides', *Radio Engng Electron Phys.*, 14, pp. 1211–7, 1969.

31. Olaofe, G. O., 'Scattering by an arbitrary configuration of parallel circular cylinders', *J. Opt. Soc. Amer.*, 60, pp. 1233–6, 1970.

32. Olaofe, G. O., 'Scattering by two cylinders', *Radio Science*, 5, pp. 1351–60, 1970.

33. Ojalvo, I. U. and Linzer, F. D., 'Improved point-matching techniques', *Quart. J. Mech. Appl. Maths.*, 18, pp. 41–56, 1965.

34. Laura, P. A., 'Application of the point-matching method to waveguide problems', *I.E.E.E. Trans.*, MTT-14, p. 251, 1966.

35 Montgomery, C. G., Dicke, R. H. and Purcell, E. M., 'Principles of Microwave Circuits', p. 44, Radiation Series No. 8 (McGraw-Hill, New York, 1948).

36. Hornsby, J. S. and Gopinath, A., 'Fourier analysis of a dielectric-loaded waveguide with a microstrip line', *Electronics Letters*, 5, pp. 265–7, 1969.

37. Yeh, C., 'Elliptical dielectric waveguide', *J. Appl. Phys.*, 33, pp. 3235–43, 1962.

38. James, J. R., 'Leaky waves on a dielectric rod', *Electronics Letters*, 5, pp. 252–4, 1969.

39. Tamir, T. and Oliner, A. A., 'Guided complex waves, Pt. 2: Relation to radiation patterns', *Proc. Instn Elect. Engrs*, 110, pp. 325–34, 1963.

40. Stratton, J. A., 'Electromagnetic Theory' (McGraw-Hill, New York, 1941).

41. Noble, B., 'Numerical Methods', pp. 20–26 (Oliver and Boyd, Edinburgh, 1964).

42. Gallett, I. N. L., Slade, M. G. and James, J. R., 'Some necessary conditions for point-matching techniques', Royal Military College of Science, Tech. Note RT50, 1970.

43. 'Modern Computing Methods', Chap. 1, 2nd Ed. (H.M. Stationary Office, London, 1961).

44. Petrofskii, I. G., 'Partial Differential Equations', p. 169 (Interscience, New York, 1954).

45. Beaubien, M. J. and Wexler, A., 'An accurate finite-difference method for higher order waveguide modes', *I.E.E.E. Trans.*, MTT-16, p. 1016, 1968.

46. Bulley, R. M. and Davies, J. B., 'Computation of approximate polynomial solutions to TE Modes in an arbitrarily shaped waveguide', *I.E.E.E. Trans.*, MTT-17, p. 444, 1969.

47. Thomas, D. T., 'Functional approximations for solving boundary value problems by computer', *I.E.E.E. Trans.*, MTT-17, pp. 451–3, 1969.

48. Morse, P. M. and Feshback, H., 'Methods of Theoretical Physics', Pt. II, Chap. 9 (McGraw-Hill, New York, 1953).

49. Waterman, P. C., 'Matrix formulation of electromagnetic scattering', *Proc. Inst. Elect. Electronics Engrs*, 53, pp. 805–12, 1965.

50. Erma, V. A., 'An exact solution for the scattering of electromagnetic waves from conductors of arbitrary shape: I. Case of cylindrical symmetry', *Phys. Rev.*, 173, pp. 1243–57, 1968.

51. Erma, V. A., 'Exact solution for scattering of electromagnetic waves from conductors of arbitrary shape: II. General case', *Phys. Rev.*, 176, pp. 1544–53, 1968.

52. Erma, V. A., 'Exact solution for scattering of electromagnetic waves from conductors of arbitrary shape: III. Obstacles with arbitrary electromagnetic properties', *Phys. Rev.*, 179, pp. 1238–46, 1969.

53. Hizal, A. and Marincic, A., 'New rigorous formulation of electromagnetic scattering from perfectly conducting bodies of arbitrary shape', *Proc. Instn Elect. Engrs*, 117, pp. 1639–47, 1970.

54. Nagenthiram, P., 'Computer study of the point matching method for hollow waveguides', University College London, Report, November 1970.

8. Appendix

Many attempts have been made to put point-matching on a firmer theoretical basis but a clearer picture has not yet emerged   We have examined the more recent theories very carefully and find various inconsistencies which will now be pointed out.

8.1 *Millar's Theory*

Millar[18, 23] has given a necessary and sufficient condition for the validity of certain point-matched solutions; for open guides the condition is that the singularities of the continued internal and external fields must not lie in the annular region between circles Γ_1 and Γ_2 (Fig 1)   We accept the necessity of the condition but point out that the sufficiency has not been established since convergence of the expansions does not ensure that the latter are valid field representations satisfying the boundary conditions.   Field singularities can only be easily recognized for elliptical and circular contours; we have carried out several tests on these shapes and find that Millar's necessary condition is extremely weak; that is any error that occurs when the condition is violated is too small to be perceptible.   The circular rod of Fig. 2 was computed several times with $0 \leqslant OO' \leqslant 0·4\lambda_0$ and when the condition was violated at $OO' = 0·25\lambda_0$ no increased field error could be detected; 4th place eigen-

value convergence was always achieved. A series of ellipses with eccentricity $\leqslant \sqrt{8}/3$ were computed and when an eccentricity of $1/\sqrt{2}$ was exceeded no perceptible error occurred; in fact very good field convergence was always obtained as the example in Figs. 5 to 6 shows Clearly for open guides Millar's necessary condition is very weak The fact that the condition is not sufficient invalidates deductions by Nielsen[17] concerning closed waveguides.

8.2 Burrows's Theory

The e.b.c. principle has been defined elsewhere[49] and several computational procedures have been based on it which we will refer to here as e.b.c. methods. Burrows[17,19,21] has proved for a 2-dimensional scattering problem that the point-matching and e.b.c. methods are mathematically equivalent in a generalized function sense but in numerical practice this equality cannot always be realized, the point-matching failing for certain shapes of boundary contours C. Implicit in this proof is the assumption that the e.b.c. method is exact and free from geometric constraints. The surface current density F is expanded as a series of functions orthogonal over C with unknown coefficients (equation (4); ref. 17); F is defined to be a function of angular position only, thus an outward going radial line can only cut C once. C is therefore constrained to be a convex contour or alternatively a non-convex contour with a suitable choice of origin; C is thus single valued with respect to O and the constraint arises in the formulation of the e.b.c. method contrary to the implicit assumptions.

8.3 Bates's Deduction

Bates[22] has applied both e.b.c. and point-matching methods to bounded waveguides and proves that they yield identical eigenvalues; he deduces that the resulting point-matched field solutions are incomplete compared to those obtained by the e.b.c. method which by implication is considered to be exact. To examine this deduction we adopt Bates's[22] notation. Equations (12) and (5) of ref. 22 can be expressed as $\mathbf{Jf} = 0$ and $\mathbf{J}^T\mathbf{a} = 0$ respectively; τ is the transpose operation, $\mathbf{a}$ and $\mathbf{f}$ are the column vectors $(A_M, A_{-(M-1)}, \cdots A_{(M-1)}, A_M)$ and $F_1, F_2, \cdots F_v)$ respectively, $v = 2M+1$ and $\mathbf{J}$ is a non-singular matrix of dimensions $v \times v$ with elements $\mathbf{J}_m(kr_n) \exp(-jm\theta_n)$; $n = 1,2, \cdots v$; $m = -M$, $-(M-1), \cdots (M-1), M$. Using an elementary transformation we can relate $\mathbf{a}$ and $\mathbf{f}$ by $\mathbf{a} = \lambda\mathbf{Rf}$ where λ is any scalar multiplier and matrix $\mathbf{R} = \mathbf{P}^T\mathbf{Q}^{-1}$, the superscript -1 being the inverse operation; $\mathbf{P}$ and $\mathbf{Q}$ are non-singular matrices of dimension $v \times v$ which reduce $\mathbf{J}$ to its normal form. All terms are generally complex throughout.

Next expand $H_0^{(2)}(kR)$ in equation (1) of ref. 22 as†:

$$H_0^{(2)}(kR) = \sum_{m=-\infty}^{\infty} H_m^{(2)}(kr) \, J_m(k\rho) \exp(\pm jm(\phi-\theta))$$

and rearrange as a numerical integration as before to give a harmonic expansion of $E_z(\rho, \phi)$ valid in the largest

† The function $H_0^{(2)}(kR)$ in eqn. (1) of ref. 22 is mathematically invalid for internal fields and should be $Y_0(kR)$; the matrix $\mathbf{R}$ above is then real.

inscribed circle Γ_1 centre at the origin, within the wave-guide contour C, thus:

$$E_z(\rho, \phi) = \sum_{m=-M}^{M} J_m(k\rho) \exp(-jm\phi) \times$$
$$\times \left[\frac{-\omega\mu\Delta C}{4} \sum_{n=1}^{v} F_n H_m^{(2)}(kr_n) \exp(jm\theta_n) \right] \quad \ldots(6)$$

where ΔC is the incremental arc length along C. For the point-matched solution $E_z(\rho, \phi)$ is defined (equation (5), ref. 22), as

$$E_z(\rho, \phi) = \sum_{m=-M}^{M} A_m \, J_m(k\rho) \exp(-jm\phi) \quad \ldots(7)$$

everywhere within and on C. These expansions for $E_z(\rho, \phi)$ are equivalent within Γ_1 if their coefficients can be equated. To test this equate the coefficients of equations (6) and (7) to give $\mathbf{Hf} = \mathbf{a}$ where $\mathbf{H}$ has elements $(-\omega\mu\Delta C/4) H_m^{(2)}(kr_n) \exp(-jm\theta_n)$. We now have two matrix equations relating $\mathbf{a}$ and $\mathbf{f}$ and eliminating $\mathbf{a}$ from them we obtain the matrix eigenvalue equation

$$(\mathbf{R}^{-1}\mathbf{H} - \lambda\mathbf{I})\mathbf{f} = 0 \quad \ldots(8)$$

$\mathbf{I}$ being the identity matrix. Field solutions based on $\mathbf{a}$ are considered incomplete by Bates in comparison with those obtained from $\mathbf{f}$, implying that the latter fields converge as close to the unique field solution as we please for M sufficiently large; $\mathbf{f}$ is therefore uniquely determined, in which case equation (8) is satisfied by a unique value of λ which in turn uniquely relates $\mathbf{a}$ and $\mathbf{f}$. This unique value of $\mathbf{a}$ is compatible with the above equations and establishes the equality of coefficients in equation (6) and (7). The e.b.c. and point-matching methods therefore give identical fields within Γ_1; for the remainder of the field interior to C it is not evident as to which method is more accurate.

8.4 Expansion of Surface Current Density

Other numerical processes using field expansions are not without constraints as we now show. Waterman[49] expands the surface current density in vector eigenfunctions of the vector Helmholtz equation, (equation (6), ref. 49), these being outgoing wave functions. Clearly the current does not satisfy this equation; the expansion is only justified if the scattered fields near the surface are expressible as outward going waves because the jump in the scattered tangential magnetic fields across the surface, defines the current, (equation (2c), ref. 49). A necessary condition for such a field representation is that the surface be single-valued with respect to the chosen origin, otherwise incident waves can be scattered towards the origin. The sufficiency of this condition has yet to be established. We see then that Waterman's formulation is restricted to scatterers with a single-valued surface. Similar constraints have been found necessary by Erma[50-52] in an exhaustive treatment on scattering. For a 2-dimensional scattering problem based on the e.b.c. principle Bates[16] expands the surface current density in a series of functions which are mutually orthogonal over an arbitrary contour C (equation (7), ref. 16). Mathematically little is known about such an expansion but its existence is established if the series can be uniquely and independently defined. In order to obtain a numerical solution Bates constrains the terms of the expansion to

be products of cylindrical and trigonometric function by virtue of his equation (19), ref. 16. This in turn constrains the scattered fields since the jump in scattered tangential magnetic fields defines the surface current. The solutions obtained are therefore dependent on this field constraint and this is not made clear.

8.5 Point-matching in relation to other Expansion Methods

We can find no evidence that the technique of point-matching is itself unreliable, incomplete, inexact, etc., in relation to e.b.c. or any other expansion method. The outstanding difficulty is choosing field expansions which will give solutions to the desired accuracy; there appears to be no sufficiency condition which will guide the user in this matter. However it must be appreciated that other expansion techniques suffer from this difficulty as the above review shows, and point-matching is no worse in this respect.

CORRECTIONS

Paper: 'Predicting servomechanism dynamic errors from frequency response measurements'

(Vol. 42, No. 1, January 1972).

Page 7 (List of Symbols)

x_f should be $(m+n+1) \times (1)$.

Page 14, Fig. 9

Correct graphs are as shown. (Positions of sample mean arrows have been amended.)

Page 15, Equations (8)

Matrix C, element of 1st row, 2nd column should be

$$\left(\frac{1}{\omega_0^2} \frac{\partial x_{18}}{\partial a_2} \right)^2$$

Matrix D, element of 1st row, 1st column should be

$$\left(\frac{1}{\omega_0^2} \frac{\partial I_p}{\partial a_1} \right)^2$$

Page 16, Fig. 10

Leftmost 'With position loop' legend should refer to the lower curve.

Page 19, Equation 26

Should be $EM \left[\sigma_{x_f}^2 \right] = \ldots$

Page 19

Sentence following equation (30) should read ... and $\sigma_{\omega 0}^2 / \omega_0^2$ between ...

Fig. 9.

Paper: 'Relative merits of quadratic and linear detectors in the direct measurement of noise spectra'

(Vol. 42, No. 2, February 1972).

In some copies of the issue two type characters became displaced during printing.

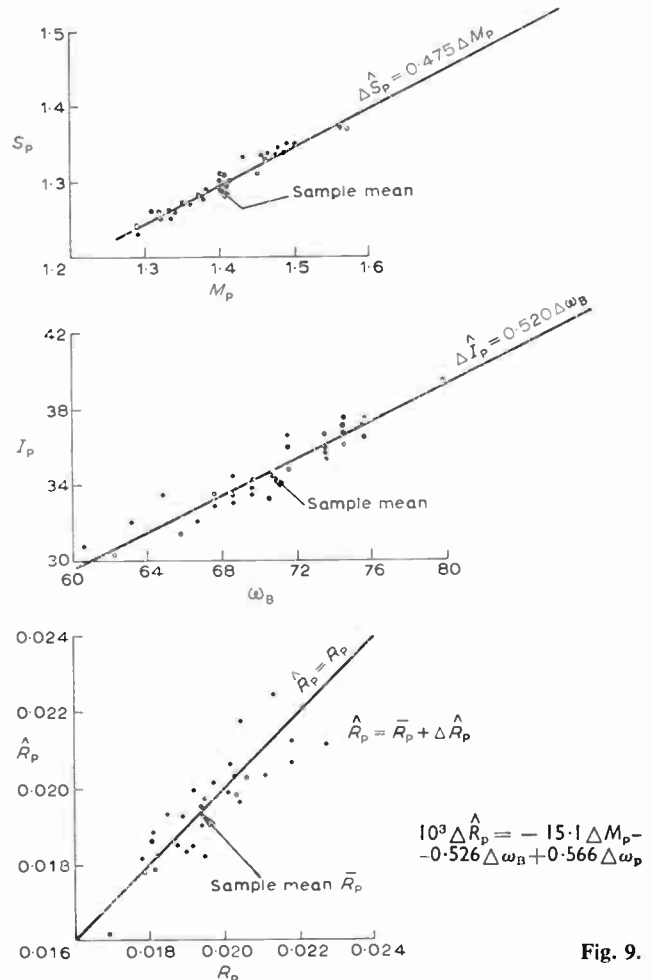Page 67 Equation (6L) should read $G(0) = L = \frac{b^2}{4\pi v^2} \int_0^\infty S^2(x) \, dx$

'Zero' Decay Analogue Store using Guarded Gate MOSFETs

B. L. HART, B.Sc., C.Eng., M.I.E.R.E.*

and

R. W. J. BARKER, M.Sc., C.Eng., M.I.E.E.†

SUMMARY

An analogue store technique, which uses a guarded gate m.o.s.f.e.t. amplifier assembly and a novel bipolar transistor switching scheme, permits the simultaneous attainment of an acquisition time limited by amplifier slewing rate and a mean 'hold' decay rate of only 170μ V/s for a 10 pF storage capacitor.

* Department of Electrical Engineering, North East London Polytechnic, Dagenham, Essex RM8 2AS.

† Department of Electrical and Electronic Engineering, Portsmouth Polytechnic, Anglesea Road, Portsmouth PO1 3DJ.

1. Introduction

Analogue stores find a diversity of applications in the fields of instrumentation and measurement. They play an important role in analogue-to-digital interface circuitry, feature in p.c.m. communication systems equipment and have been used in drift correction circuits for d.c. coupled amplifiers. In the normally encountered type of analogue store, a charge proportional to an input signal voltage is rapidly established on a storage capacitor during the 'sampling' period and this charge—which ideally remains constant until a further sampling period—subsequently decays, comparatively slowly, as a result of parallel leakage paths during the 'hold' interval. The acquisition time, t_a, sets a limit to the minimum sampling pulse width, τ, and the output voltage decay rate sets maximum limit to the hold time, t_h, for a specified fidelity of output voltage level. For a given t_a, t_h is maximized by reducing to a minimum the effects of any leakage paths in parallel with the storage capacitor; this presents a number of practical problems.

This paper will show how modern circuit techniques permit t_h to be maximized. The approach adopted relies on (i) the use of a special guarded gate m.o.s.f.e.t. assembly, (ii) methods for device selection and amplifier design to achieve a low offset voltage temperature co-efficient in a unity gain buffer stage incorporating the m.o.s.f.e.t., (iii) a novel interconnexion scheme for the analogue switch. Hold time decay rates as low as 170 μV/s have been measured with a 10 pF storage capacitor.

2. Circuit Operation

A simple analogue store is shown in the circuit of Fig. 1(a): this comprises a bilateral analogue switch, SW, controlled by a sampling pulse of duration, τ, a capacitor, C, and an operational amplifier, A, with small signal, l.f. gain of A_v, connected as a voltage follower. A practical scheme would differ in detail from the schematic of Fig. 1(a). Thus, SW might be preceded by a V-I converter and τ made dependent on V: the circuit as shown might be included in a d.c. feedback loop to reduce possible errors in SW. Whatever variations employed, an equivalent circuit for capacitor discharge during the 'hold' period is shown in Fig. 1(b) in which v_s = stored voltage, I_{SW} = switch leakage current, R_x = capacitor leakage resistance and R_i, C_i, I_i are respectively the incremental input resistance, input capacitance and input current of A when operating in the follower mode.

Thus,

$$C_T(dv_s/dt) + (v_s/R_T) + I_{SW} + I_i = 0 \qquad \ldots.(1)$$

where $R_T = R_x$ in parallel with R_i, $C_T = C + C_i$.

To minimize (dv_s/dt) for a given C_T, and hence maximize t_h, it is necessary to maximize R_x, R_i and to minimize I_{SW} and I_i. R_x is inversely related to C and is a function of capacitor dielectric quality; R_i and I_i are optimized by using a guarded gate m.o.s.f.e.t. structure, type MBH 1,[1] connected up with an operational amplifier as a voltage follower, as shown in Fig. 2, for which, typically, $I_i \simeq 10^{-16}$A; $C_i \simeq 0.4$pF; R_i is not measurable because of its extremely high value.
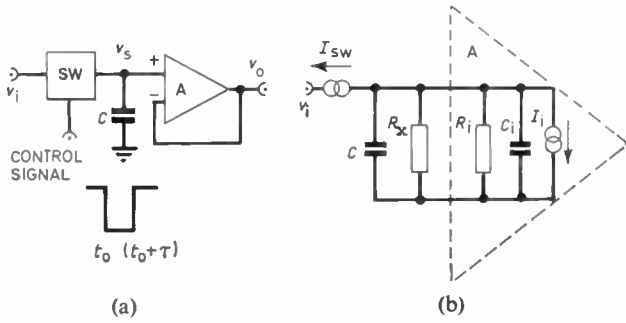
Fig. 1. (a) A simple analogue store.
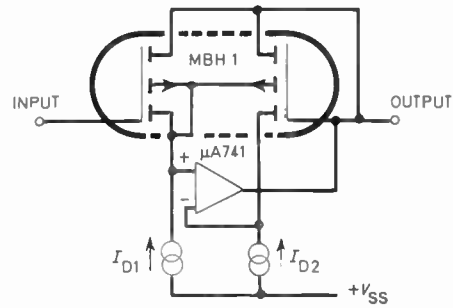(b) Equivalent circuit of Fig. 1(a) during hold time.

Fig. 2. Voltage follower circuit.

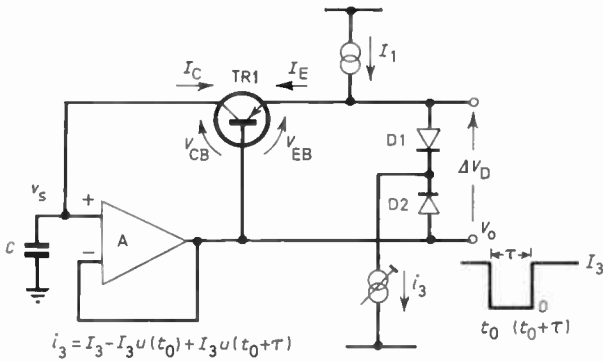$$i_3 = I_3 - I_3 u(t_0) + I_3 u(t_0+\tau)$$
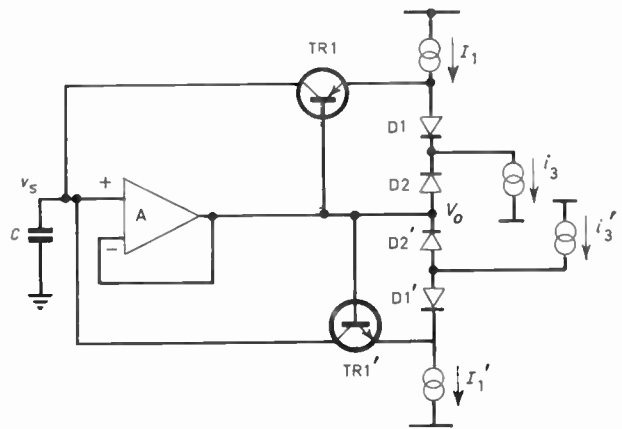
Fig. 3. Transistor switching scheme.

Fig. 4. Bilateral drive for C.

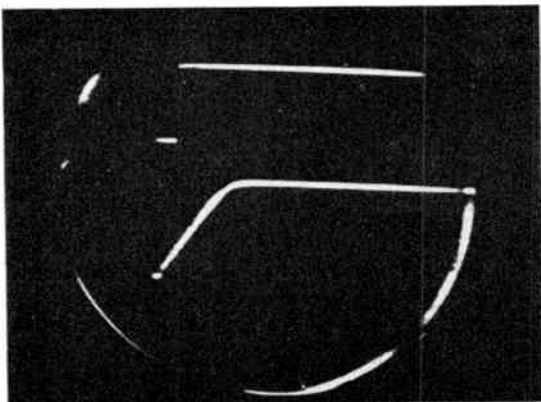Fig. 5. Transient response of Fig. 3 showing storage effects in TR1.
Upper trace: Sampling pulse: amplitude 0 to -4V, width 1μs.
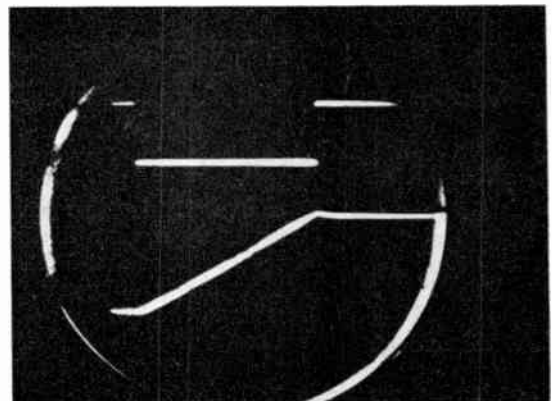Lower trace: Output pulse: -2.5V to 0V.

Fig. 6. Transient response of Fig. 3 with Schottky barrier diode added between base and collector of TR1.
Upper trace: Sampling pulse: amplitude 0 to -1.5V, width 10μs.
Lower trace: Output pulse: amplitude -7V to -2V.

A technique for minimizing I_{sw}, which is likely to be the dominant cause of leakage is illustrated in Fig. 3*, in which TR1 and matched diodes D1, D2 constitute the switch. If A has $V_{os} = 0$ and $A_v \gg 1$, then $V_{CB} = 0$, irrespective of v_s. For $t_0 > t > (t_0 + \tau)$, $I_C = I_{sw} = 0$, if $V_{EB} = \Delta V_D = 0$. This means that all terminals of TR1 are at the same potential and requires the setting, $I_3 = 2I_1$.

When $V_{os} \neq 0$, I_c can still be made zero if $\Delta V_D \neq 0$: this means setting I_3 for

$$V_{os}/\alpha_R V_T \simeq (2I_1 - I_3)/(I_3 - I_1) \qquad \ldots (2)$$

Equation (2) follows from the use of the logarithmic V-I law for D1, D2, and the application of the Ebers-Moll transistor model for TR1. The adjustment $\Delta V_D \neq 0$ also allows the condition, $I_{sw} = -I_i$, to be realized, thus giving the possibility of zero decay. By design, I_1, $I_3 \neq f(T)$; however (2) still remains valid for devices selected[2] for an appropriate (dV_{os}/dT).

For $t_0 < t < (t_0 + \tau)$, $i_3 = 0$, D1 and D2 are cut off and $I_E = I_1$. Hence C charges in a positive sense. Precise details of the charging process depend on the relationship between I_1/C_T and the slewing rate, S, of A. If $I_1/C_T \ll S$, a condition easily met in practice by suitable choice of C, then during charging, $dv/dt = \alpha_F I_1/C_T$; at the end of the sampling interval, TR1 cuts off promptly. For the case $I_1/C_T > S$, TR1 saturates to such an extent that I_C is just sufficient for the capacitance, C_T, to be charged at the maximum rate the amplifier will allow; at the end of the sampling interval, TR1 must desaturate. This means that C_T will continue to charge for a time, t_x, where $t_x < (I_1 \tau_{BS})/C_T$.

The effects of saturation can be minimized by using a device type with a low τ_{BS} for TR1 and if necessary the use

---

* The circuits shown in Fig. 3 and Fig. 4 are the subject of patent applications.

of a Schottky barrier saturation clamping diode between the base and collector of TR1.

The switch in Fig. 3 is not bilateral. To charge, as well as discharge, C, an n-p-n transistor TR1' and an additional diode pair D1', D2' are required as shown in Fig. 4. It is apparent that such a scheme automatically eliminates possible damage to the m.o.s.f.e.t. gate due to induced charges. In Fig. 4, TR1, TR1' may be replaced by other suitable active devices and the diode pairs replaced by appropriate switches (e.g. symmetrical silicon alloy chopper transistors).

3. Performance

Measurements were made on the circuit of Fig. 3 using $C = 10 \text{pF(air)}$, TR1 was a 2N2605A, D1 and D2 were 1N5261; I_1 was $95 \mu A$ and I_3 was set to give minimum (dv_0/dt) at room temperature. In a 10 minute test interval, $(d\bar{v}_0/dt) = 170 \mu V/s$.

Figure 5, taken with the aid of a storage oscilloscope, shows the transient response of the analogue store when a sampling pulse width of $1 \mu s$ is used. The effect of storage in TR1 is evident. Figure 6, in which a $10 \mu s$ sampling pulse width was used, shows how the addition of a Schottky barrier diode between the base and collector of TR1 largely eliminates this effect. The gradient of the transient ramp in Fig. 5 and Fig. 6 corresponds approximately with the maximum slewing rate of A.

4. References

1. Barker, R. W. J., 'Techniques for minimizing m.o.s.f.e.t. gate leakage current', *Proc. Inst. Elect. Electronics Engrs*, 59, pp. 283-4, February 1971 (Letter).
2. Barker, R. W. J. and Hart, B. L., 'Offset-voltage temperature-coefficient in dual m.o.s.f.e.t. source followers', *Electronics Letters*, 7, pp. 123-5, 25th March 1971.

The Authors

Mr. R. W. J. Barker took a B.Sc. in electrical engineering at Nottingham University in 1957, which was followed by industrial experience with Creed & Co., Mullard and Edwards High Vacuum. During this period he was mainly concerned with the design of specialized electronic equipment, including control and measuring circuitry for prototype vacuum systems. He took up a lecturing appointment with North East London Polytechnic in 1967, and read part-time for a M.Sc. degree of London University. In 1970 Mr. Barker moved to his present post as a Senior Lecturer in the Department of Electrical and Electronic Engineering, Portsmouth Polytechnic. His current research interests include field effect transistor electronics and electrometry.

Mr. B. L. Hart (Member 1961, Graduate 1955) received the B.Sc. degree in physics from Queen Mary College, London, in 1951. Following two years National Service in the Royal Corps of Signals he studied at Edinburgh University, where he obtained the Postgraduate Diploma in electronics and radio in 1954. After a period as a graduate trainee with the Marconi Co. he worked as a research engineer in the radio display group of the Marconi Research Laboratories, Great Baddow, on the design and development of interscan marker systems, alpha-numeric displays and logic elements. From 1961 to 1964 he was a sales engineer for Transitron Electronic Ltd. He joined the staff of the electrical engineering department of West Ham College of Technology (now incorporated in North East London Polytechnic) in 1964 and is currently a Senior Lecturer in charge of the Postgraduate Electronics Laboratory. His research interests include device modelling, waveform circuit techniques, and ultrahigh-impedance wideband amplifiers.

Some Turning-Points in Infra-Red History

Professor R. V. JONES,
C.B., C.B.E., M.A., D.Phil., F.Inst.P., F.R.S. *

Opening Address at the I.E.R.E. Conference on Infra-red Techniques held at the University of Reading from 21st to 23rd September, 1971

SUMMARY

The address covers significant advances from the earliest work, by William Herschel, to the development of detectors in the past decade which bridge the gap between light waves and radio waves. The theoretical discussions of the nineteenth and early twentieth centuries, which led to an understanding of radiation, are followed by descriptions of applications in World Wars I and II and brief mention is made of thermography, the invention of laser generators and infra-red meteorology.

* Department of Natural Philosophy, University of Aberdeen, Aberdeen AB9 2UE, Scotland.

Newton's Speculations

One of the first clues to the existence of infra-red radiation was gained between 1704 and 1718, the dates of the first two editions of Newton's 'Opticks'. Thanks largely to Robert Boyle it was known that a vacuum would not transmit sound but that it would transmit light and magnetic force, and this may have prompted Newton to enquire, at some time after his first edition, whether heat would be transmitted as well. The experiment was evidently performed, for Newton added some extra Queries in his second edition,[1] including the Eighteenth. This starts:

> 'If in two large tall cylindrical Vessels of Glass inverted, two little Thermometers be suspended so as not to touch the Vessels, and the Air be drawn out of one of these Vessels, and these Vessels thus prepared be carried out of a cold place into a warm one; the Thermometer *in vacuo* will grow warm as much, and almost as soon as the Thermometer which is not *in vacuo*. And when the Vessels are carried back into the cold place, the Thermometer *in vacuo* will grow cold almost as soon as the other Thermometer.'

This experiment does not occur in the works of Boyle, who died in 1691, and it was therefore most probably undertaken by someone else, quite possibly by Francis Hauksbee, who made Boyle's last vacuum pump and who is known to have performed experiments under Newton's direction. The result of the experiment led Newton to speculate:

> 'Is not the Heat of the warm Room conveyed through the vacuum by the vibrations of a much subtiler Medium than Air, which after the Air was drawn out remains in the vacuum? And is not this Medium the same with that Medium by which Light is refracted and reflected, and by whose Vibrations Light communicates Heat to Bodies, and is put into Fits of easy Reflection and easy Transmission?'

William Herschel's Experiments

Newton's inspired query remained unanswered for nearly a century, until in 1800 William Herschel (1738–1822) performed his famous experiments on the solar spectrum. Herschel, incidentally, had started his career as an oboe and violin player in the Hanoverian Foot Guards and had emigrated to England in 1757. After a period as music instructor to the Durham Militia, he attained security as an organist at Bath where he educated himself in other subjects; and he progressed from a study of harmony to mathematics and thence to optics and astronomy. By 1781 he had made his own reflecting telescopes and had discovered Uranus.

In the course of observing the Sun he studied the efficacy of various coloured filters in cutting down the dangerous heating effects of the Sun's rays on the human eye. His subjective impressions led him to a vital doubt, as he related in his celebrated paper[2] in the *Philosophical Transactions* for 1800:

> 'It is sometimes of great use in natural philosophy, to doubt of things that are commonly taken for granted; especially as the means of resolving any doubt, when

once it is entertained, are often within our reach. We may therefore say, that any experiment which leads us to investigate the truth of what was before admitted upon trust, may become of great utility to natural knowledge.... It will therefore not be amiss to mention what gave rise to a surmise, that the power of heating and illuminating objects, might not be equally distributed among the variously coloured rays.'

Sir William Herschel, Bt., F.R.S.
1738–1822

By Herschel's time the idea seems to have developed that, since the light of the Sun was accompanied by heat, the amount of heat carried by a given quantity of light was directly proportional to the luminosity of the light. Herschel's experience had led him to doubt whether this was so, and he began to speculate whether light of different colours might carry different amounts of heat for the same luminosity. He therefore devised his famous experiment (probably during the winter of 1799–1800) to show the heating effect of various parts of the solar spectra produced by a prism. His detectors were simple thermometers, with their bulbs blackened; his two best specimens were made by Alexander Wilson, M.D. (1714–1786), the Professor of Practical Astronomy at Glasgow.

Wilson was both a brilliant technician and a good observer; besides cutting new type founts for the university printers (including the Greek fount for the Foulis edition of Homer), he went into partnership with the university technician to set up a shop for scientific instruments. The technician, incidentally, was James Watt. Wilson, who was an expert glassblower, improved the sensitivity of thermometers by mastering the difficulty of making elliptical capillaries; and he invented a marine barometer, based on a thermometer that was especially sensitive in the region of the boiling point of water, so that the depression of the boiling point could be used as a measure of the reduction in atmospheric pressure.

As far as I know, the thermometers used by Herschel have been lost; but at least one of Wilson's thermometers survives in the Royal Scottish Museum in Edinburgh. It has a Fahrenheit scale and such a scale seems to have been used by Herschel. Typically the temperature rises that he measured in his spectrum were between 1° and 10°F, and he read the scale to the nearest $\frac{1}{4}$°F; the ambient temperature was usually between 50 and 60°F, and the response time of the Wilson thermometers that he used was five minutes. Because of the difficulty caused by variations in the ambient temperature during the course of an experiment, Herschel devised a method of differential measurement in which two closely similar thermometers were placed in juxtaposition, with one of them exposed to the light reflected by a mirror placed in the appropriate portion of the spectrum, while the other was shielded. He measured the intensity of the heating effect due to the radiation by taking the difference in temperature between the two thermometers. This was perhaps the first systematic use of differential measurement in experimental physics; it has ever since been an almost essential feature of infra-red technique.

Herschel quickly found that the maximum heating effect did not coincide with the obvious maximum of luminosity in the middle of the visible spectrum, but that the heating effect increased as he moved the sampling mirror to the red end of the spectrum, and that it increased still further when the mirror was substantially beyond the deepest red that he could see. By a rapid sequence of experiments he established that the unseen rays were subject to the same laws of reflexion and refraction as visible light and his first paper, which is some 70 pages long, concluded that:

'if we call light, those rays which illuminate objects, and radiant heat, those which heat bodies, it may be enquired, whether light be essentially different from radiant heat? In answer to which I would suggest, that we are not allowed, by the rules of philosophizing, to admit two different causes to explain certain effects, if they may be accounted for by one.... It remains only for us to admit, that such of the rays of the Sun as have the refrangibility of those which are contained in the prismatic spectrum, by the construction of the organs of sight, are admitted, under the appearance of light and colours; and that the rest, being stopped in the coats and humors of the eye, act upon them, as they are known to do upon all the other parts of our body, by occasioning a sensation of heat.'

So at the end of his first paper Herschel appears to have believed that radiant heat and visible light were part of the same spectrum, and he even postulated that the main difference might be in the momenta of the corpuscles appropriate to the various parts of the spectrum. But during the summer of 1800 he sought to establish

whether the identity was real or not: and he was now to become a victim of his own thoroughness. He conscientiously searched for any difference that might be shown up experimentally between the two kinds of radiation, and he in effect asked the question whether it was simply accidental that radiant heat and visible light were refracted to much the same extent by a prism. Ultimately, as he reported in his third paper of 1800,[3] he thought that he had devised a crucial experiment. Unfortunately, he did not describe it in detail, but he said that:

'We are now to appeal to our prismatic experiment on the subject, which is to decide the question. First, with regard to light, I must anticipate a series of highly interesting observations I have made, which, though they certainly claim, cannot find room in this Paper. These have given me the means of acting separately upon either of the extremes, or on the middle of the prismatic spectrum; and by them I am assured that red glass does not stop red rays. Indeed the appearance of objects seen through such coloured glasses, till I can give those observations will be a sufficient proof to everyone that they transmit red light in abundance. Next, with regard to the rays of heat, the case is just the reverse; for, by our preceding table, the red glass stops no less than 692 out of a thousand of such rays as are of the refrangibility of red light.'

It would therefore seem that, in the red region of the spectrum in which he made this observation, the transmission of his red glass filter was, by bad luck, only 30·8%. This was the answer given him by his thermometry, but he could not reconcile it with the 100% optical transmission that he assumed from the visual measurements which he did not describe (and which unfortunately he never described later). He therefore concluded that the radiant heat and the visible radiation coming out of the prism at the same angle of refraction were absorbed differently by the red glass and were therefore of different natures.

The conclusion is distinctly surprising, in that he had also made a photometer to check his optical transmission, but the error can be feasibly explained if his thermometric observations were made in the region of the absorption edge of the red filter while his visual observations would have integrated over the whole of its visual transmission range. As a further difference, he noted that the transmission of radiant heat seemed to increase with time, whereas the visual transmission remained constant; the transmission of the red glass as measured by the thermometer had appeared to be 30·8% after ten minutes or so exposure, but it had only been 25% after two minutes. Assuming that he had allowed for the response time of his thermometer, we can see in retrospect that the rise in apparent transmission may well have been due to re-radiation from the filter itself as it heated up, for this was a frequent bugbear in early measurements. Unhappily it reinforced Herschel's change of mind in believing that he had found differences between radiant heat and visible light, and he therefore left in some perplexity the subject that, in the course of a single year, he had so brilliantly discovered.

The Wave Properties of Radiant Heat

A year later Thomas Young[4] in the celebrated Bakerian Lecture of 1802 in which he established the wave properties of light, went on to speculate:

'At present, it seems highly probable that light differs from heat only in the frequency of its undulations or vibrations; those undulations which are within certain limits, with respect to frequency, being capable of affecting the optic nerve and constituting light; and those which are slower, and probably stronger, constituting heat only.'

Despite the accuracy of this speculation, the nature of infra-red was still uncertain when Baden-Powell reviewed the state of the subject for the British Association in 1832. Baden-Powell's survey[5] is notable for its record of the demonstration by W. Ritchie, the Rector of Tain Academy and the inventor of the Ritchie Photometer, that 'The absorptive power of surfaces is precisely proportional to their radiating power'. This demonstration, made with the aid of Leslie's cube and differential air thermometer, anticipates by at least 20 years the statement of the law that was given a theoretical basis by Kirchhoff,[6] and which is generally ascribed to him.

Also contained in Baden-Powell's survey was a summary of Prévost's theory of exchanges[7] in which, on the assumption that 'heat is a discrete fluid every particle of which moves in a straight line' Prévost correctly supposed that the phenomena of radiation might be explained by the idea that all bodies radiated to one another, the radiation from hotter bodies being the greater, so that the exchange would ultimately bring them all to the same temperature.

The B.A. Report also speculated about the possibility that thermal radiation might be polarized by reflexion, but a convincing demonstration of this effect had to wait for a few years until J. D. Forbes, the Professor of Natural Philosophy at Edinburgh, showed that linear polarization could be obtained by reflexion from mica and tourmaline, and circularly polarized radiation by a Fresnel rhomb of rocksalt.[8] The historian of science may note in passing that Forbes, Leslie, Ritchie, and Wilson were all Scots as were also Balfour, Stewart, and Brewster; the Clarendon Code, which required universities to be staffed by active members of the Church of England, preferably in holy orders, did not apply after 1685 in Scotland—to the benefit of Scottish science. It was not abolished in England until the University Tests Act was passed in 1871.

Forbes, who had Clerk Maxwell as one of his pupils, had been able to perform his experiments because of the thermopile which had been invented by Nobili in 1829, and improved for radiation measurement by his friend Melloni.[9] They developed the thermopile to the stage where with a simple collecting cone it could detect a human body at about 10 metres. Melloni systematically explored the properties of thermal radiation and among other things discovered the transparency of rocksalt. After publishing many papers, he surveyed his work in an attractive monograph 'La Thermochrôse', published by Joseph Baron of Naples in 1850.

Although the thermopile now became the outstanding detector, the most remarkable instrumental achievement up to 1850 was that of the two great experimenters, Fizeau and Foucault,[10] who succeeded in making an alcohol-in-glass thermometer so small that it could be placed in the separate fringes produced in an interference experiment, and so sensitive that it could detect the temperature rise due to the radiation in a single fringe. The diameter of its bulb was less than 1·1 mm, and the capillary was such that 1 degC caused a movement of the alcohol meniscus of 8mm; the capillary must have been no more than ten microns in diameter. Reading with a microscope Fizeau and Foucault were able to detect a change of about 0·00025 degC, and they thereby established that thermal radiation could give rise to interference phenomena from which its wavelength could be shown to be greater than that of visible light. The temperature rise in the central fringe produced in a double mirror experiment was 39·5 divisions, with the first fringe on either side giving about 20 divisions, each division corresponding to 0·0025 degC.

'Infra-red'

In 1879 the famous T^4 Law was formulated empirically and rather luckily by Stefan[11] from some measurements of Tyndall which showed that the radiation from a heated platinum wire at 1473°C was 11·7 times that at 798°C. It was, incidentally, about this time that the term 'Infra-red' began to appear as an alternative term to 'radiant heat'. Its French and German equivalents *Infra-rouge* and *Ultra-rot* had been used earlier; English and American workers had first tended to follow the German practice and talk of 'Ultra-red', J. W. Draper had thus used 'Ultra-red' in 1877[12] but had changed[13] to 'Infra-red' by 1880. W. Abney had used[14] 'ultra-red' in 1878 and again in the abstract of his Bakerian Lecture of 1880; but in the main text[15] of the Lecture, dated 17th December 1879, he used 'Infra-red'. This is the earliest mention of infra-red that I have found: neither Abney nor Draper remarked the change of usage, which thus seems to have occurred naturally by common consent just before 1880.

New detectors appeared about this time, notably the radiometer[16] of Crookes (1876) and the bolometer[17] of Langley (1881). Langley was President of the American Association for the Advancement of Science in 1888, and his Presidential Address[18] surveyed the history of radiant heat. It concluded:

'Finally, if, turning to the future, I were asked what I thought were the next great steps to be taken in the study of radiant heat, I should feel unwilling to attempt to look more than a very little way in advance. Immediately before us, however, there is one great problem waiting solution. I mean the relation between temperature and radiation; for we know almost nothing of this, where knowledge would give new insight into almost every operation of nature (nearly every one of which is accompanied by the radiation or reception of heat), and would enable us to answer enquiries now put to physicists in vain by every department of science, from that of the naturalist as to the enigma of the brief radiation of the glow-worm, to that of the geologist who asks as to the number of million years required for the cooling of a world.'

The Radiation Formulae

Langley's conclusion proved remarkably prescient. Within five years Wien discovered[19] his astonishingly simple relation between the absolute temperature T of a black body and the wavelength λ_m at which it emits maximum energy:

$$\lambda_m T \simeq 2900$$

where λ_m is measured in microns, and T in degrees absolute. Just how relatively recent these events are may be realized from the fact that one of Wien's pupils, E. O. Lowenstein, is currently Professor of Zoology and Comparative Physiology at Birmingham: Wien was his Professor at Munich in the 1920s.

Wien[20] subsequently produced a formula for the wavelength distribution of radiation from a black body, but Lummer and Pringsheim[21] in 1899 showed that this was inconsistent with their spectral measurements in the range 12–18 microns. Incidentally wavelength measurements had now become reliable thanks to the efforts of various workers, notably in France where Pierre Curie[22] was one of those who followed up the work of Fizeau and Foucault. Rayleigh[23] and Jeans[24] produced an alternative formula which was consistent with observations at these longer wavelengths but which was completely invalid for short wavelengths. The long wavelength agreement had been established by Rubens and Kurlbaum,[25] who reported their confirmation to Planck in 1900. This was decisive in convincing Planck that some new assumptions would have to be made to produce a formula which was consistent with Wien's at short wavelengths and Rayleigh's at longer wavelengths; and it was, of course, one of the turning points in the history of science when Planck[26] discovered that the necessary assumption was that the states of oscillators had to be quantized. Although he was as dissatisfied with his method of deriving the black body formula as he was satisfied with its remarkable agreement with observation, Planck himself believed, so his son recorded,[27] that he had made 'a discovery comparable perhaps only to the discoveries of Newton'; it was without doubt the greatest achievement to come out of infra-red studies.

Photons

At the time, and for long afterwards, Planck did not believe that the quantization extended from the elementary oscillators to the radiation that they emitted. This was first postulated by Einstein in his famous paper[28] on the photo-electric effect 'On a heuristic viewpoint concerning the production and transformation of light'. Although Einstein's concept of a photon was to prove exceedingly fruitful, Planck as late as 1913 did not accept it. Indeed, in recommending the election of Einstein in 1913 to the Prussian Academy of Science, Planck, along with Warburg, Nernst and Rubens, concluded[29] their declaration of support:

'Summing up, we may say that there is hardly one among the great problems, in which modern physics

is so rich, to which Einstein has not made an important contribution. That he may sometimes have missed the target in his speculations as, for example, in his hypothesis of light quanta, cannot really be held too much against him, for it is not possible to introduce fundamentally new ideas, even in the most exact sciences, without occasionally taking a risk.'

Although this reservation now appears distinctly droll we ought to remember that the resistance to the idea of light quanta persisted for at least another ten years. Niels Bohr, for example, with all his great insight and imagination, refused to accept the idea as late as 1924, when he is reported to have said that even if Einstein had sent him a telegram informing him that he had found an irrevocable proof of the existence of light quanta, even then, Bohr[30] held, 'The telegram could only reach me by radio on account of the waves which are there'. What seems ultimately to have dispelled Bohr's opposition was the discovery of the Compton effect, combined with Heisenberg's formulation of indeterminacy and Bohr's own theory of complementarity.

The difficulty that had so beset physicists was that interference experiments seemed to demand that a single photon must pass through two slits and that it could therefore not be localized as a particle or quantum. When it was shown that hitherto-accepted particles such as electrons could be diffracted, the difficulty was accepted as a fact of nature rather than resolved. It may, incidentally, be significant that Einstein's Nobel Prize was awarded to him for his explanation of the photo-electric effect, and not for relativity.

We should be getting too far away from the subject of our conference to pursue these matters further, but they serve to remind us of the part that infra-red studies have played in the development of physics and of science generally, and they show something of the difficulties that beset the greatest of physicists in establishing the concepts that we now take for granted.

Beginnings of Application

So far our survey of the infra-red has been concerned with its discovery and its influence on fundamental thinking in physics. There were also pointers during the nineteenth century to its prospective influence on other branches of pure science. In their first paper Nobili and Melloni[31] recorded their observation that a caterpillar radiated more than the chrysalis and butterfly which proceeded from it, presaging the interaction between infra-red and biology and medicine that has recently developed. When Ångström[32] showed in 1889 that the absorption spectrum of carbon monoxide differed from that of carbon dioxide, he established that infra-red absorption spectra were mainly molecular in origin; this fact is the basis of all the contribution that infra-red studies have made to our knowledge of molecular structure and to process control in chemical industry. The widespread discovery that solids varied in their ability to transmit infra-red, and Rubens's[33] work on reststrahlen, indicated that infra-red could give valuable clues to the structure of crystals.

World War I

The onset of World War I encouraged the application of science for military purposes, and this fact—as is so often the case—ultimately led to remarkable developments in technology. On the German side an infra-red detector was developed which, according to Dr. H. Gaertner, picked up British torpedo boats off Ostend at about 10 kilometres' range. It also stimulated the work of Hoffmann[34] in America on the detection of men and of aircraft, and the invention by Case of the 'thalofide' photo-conductive cell. As regards ideas in Britain, Lord Cherwell told me in 1935 that he had suggested the use of infra-red for aircraft detection in 1916, but that no action had resulted. After World War I the British Admiralty took up the Case type of photo-conductive cell, and for 18 years they made a fruitless search for materials that would surpass it in sensitivity or spectral range. In the meantime they developed a system of ship-to-ship signalling using photo-conductive cells; they also developed a phosphor which glowed when near infra-red radiation fell on it after it had previously been activated with ultra-violet. This device was in a sense the forerunner of the modern 'multiple photon counter'.

Noise Limits

The decade following World War I saw various improvements in methods of detecting radiation. Among the most important of these was the work of Moll and Burger[35] on the development of vacuum thermoelements and the diagnosis by Ising[36] that in their use of a thermoelement to amplify the deflexions of their galvanometer, Moll and Burger had for the first time reached the important limit of instrumental sensitivity set by Brownian Motion. Another significant advance of this period, although it hardly received due recognition, was the development by Dobson and Perfect[37] of the technique of chopping the radiation entering a spectrometer and thus imprinting a label on the radiation such that its effects in the detector could be uniquely recognized by a phase-sensitive commutator synchronized with the chopper. Although Dobson and Perfect applied their technique to ultraviolet spectrometry, it has since proved even more valuable in infra-red work; the phase-sensitive detection on which it depended was actually an old idea, going back to Ayrton and Perry[38] who applied it to Wheatstone Bridge measurements in the 1880s.

My own contact with infra-red started in 1932, when I was inveigled into working with an infra-red spectrometer largely because the preceding research student had become impatient with it and wanted an excuse for persuading our Professor to let him go on to a different line. Within a few weeks I found that the thermopile was broken, and since our laboratory funds did not extend to purchasing a new one, I started to make one. On and off I was to go on making infra-red detectors for the next 25 years.

World War II: Britain

It was not difficult to make detectors that were better than the best that were then commercially available, and by 1935 I was drawn to the problem of providing some

means of air defence by detecting aircraft by the heat radiated from their engines. By that time I could make thermopiles with time-constants down to 10 milliseconds, and within a year or so I had made an equipment which could be flown experimentally in an aircraft. On 27th April 1937 I made what appears to have been the first flight in which one aircraft was detected by another in flight by infra-red.[39] The range was modest—only 500 yards—but the limitation was not due to the detector but to the effect of bumps on the input transformer that I used to couple the thermopile into the electronic amplifier. Again, this may have been the earliest application of conventional electronic techniques to the amplification of signals from thermal infra-red detectors, the application being made possible by the relatively high speed of the detector (Fig. 1).

Several other ideas came out of this work, including optical radar, where I planned to pulse a searchlight and use a gated image converter so as to measure the range of the target and to ignore the scatter reflected from the nearer regions of the beam. I also did some work on a thermal image converter, dependent on the increase in thermionic emission from the hotter parts of a surface of low work function, and achieved crude pictures of my finger with a tube made specially by EMI. I also suggested using a semiconducting mosaic in an iconoscope, but these rather far-fetched ideas were not pursued because we were beginning to see that radar offered better promise.

In the meantime we had to deal with several criticisms. One was that it would be easy to screen aircraft engines so that they would radiate very little infra-red. I succeeded in showing that even if this was so, one could still detect aircraft flying at 200 miles/h by the unavoidable aerodynamic heating of their frontal surfaces. What we could not do, however, was to offer any hope of detecting aircraft through cloud, or of giving a measurement of range. In these respects radar was clearly going to be superior, and infra-red was almost completely dropped.

World War II: Germany

In Germany, things went differently. That they did so was largely due to the discovery by B. Lange[40] in 1930 that galena (lead sulphide) was sensitive down to about 3 microns wavelength. In retrospect I can see that I myself had the necessary clue for, as a boy at school in 1925 or 1926, I had focused an image of the Sun on the contact of a cat's whisker in a crystal set, and had obtained a crackle in the earphones. I had, however, not seriously considered the possibilities of photo-electric conduction of infra-red, largely because I was thinking particularly of the problem at 10 microns, and I thought that if a photo-emissive surface was to be sensitive at this wavelength it would present a problem analogous to trying to use a photocell with its cathode incandescent, unless one worked at very low temperatures—and it seemed beyond the bounds of possibility that we should ever be able to fly liquid air generators in aircraft. My belief that it would be difficult to find photo-emissive surfaces with sensitivities beyond 2 microns, or even photoconductive ones, seemed entirely borne out by the Admiralty experience in the search for better materials than thalofide. Lange's work was, of course, published, but the Admiralty workers seemed positive that it must have some alternative explanation. My last work before I left infra-red for Intelligence in September 1939 was to make a vacuum system for one of them who proposed to reinvestigate infra-red photo-conductors.

My main contact with infra-red was now on the German side, and from time to time new developments began to appear. First came the WPG15, an equipment made by Zeiss which included a 'chopped' bolometer at the focus of a 60 cm mirror. This was mainly used for detecting ships, the most spectacular project in which it was involved being the attempt by the German Secret Service to mount it, along with an infra-red burglar alarm system, near Gibraltar to monitor our ships passing through the Straits. The story of that episode, and how we frustrated it, has been told by Kim Philby in his 'My Silent War'.

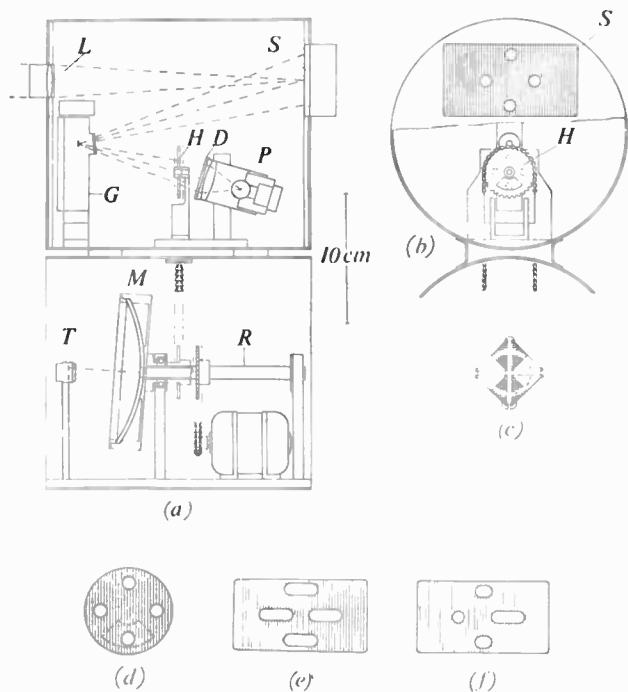Towards the end of the war the 'Kiel' apparatus began

Fig. 1. Scheme of first air-to-air infra-red detector, 1937. (After ref. 39).

(a) Vertical section through detector. The mirror M is offset on its axis of rotation R so as to cause the image of the target to trace a circular path over the four-element thermopile T 20 times per second. Signals from the latter are amplified and fed to the short period mirror galvanometer G which is illuminated by a lamp P through a diaphragm D containing four holes and a sector H which rotates in synchronism with the mirror M. The effect is to illuminate cyclicly four spots on the screen S. If the thermopile is receiving a signal at the time when a particular spot is being illuminated, this spot is drawn out into a band by the motion of the galvanometer mirror. For trials in daylight an image of the target can be thrown onto the screen S by means of the lens L.

(b) Observer's view of detector if no signal is being received.
(c) Arrangement of thermopile elements. Manganin foil (unshaded) butt-welded to constantan foil (shaded).
(d) Arrangement of diaphragm and sector.
(e) Indication to observer if target is dead ahead.
(f) Indication to observer if target is to right.

to appear in German nightfighters, and it was clear that they must have a detector which was fast enough to give some form of display on a cathode-ray screen. This detector turned out to be the lead sulphide cell which had been developed following Lange's discovery. The surprise precipitated a frenzy of work both in Britain and America; for, although the Kiel apparatus was not nearly as good as our radar equivalent for nightfighting purposes, the development of aircraft with larger engines, particularly jets, made them much easier targets; moreover, infra-red detection was peculiarly promising as a homing system for missiles because the signal increased as the missile closed, and there was no minimum range problem as with radar. The Germans themselves had, of course, realized these advantages; and with the priority that they had been forced to give to defence in the last years of the war they were already well on with missile design.

New Detectors

It had been found by German workers that lead selenide was even more promising than the sulphide, and it seemed likely that the telluride would be even better. For some years after the war our own developments were therefore in this direction until Welker[41] suggested that, since the lead sulphide group were II-VI compounds, and germanium and silicon (which also showed near infra-red conductivity) were Group IV elements, the III-V compounds might be worth looking at, with the result that indium antimonide captured the major interest.

Two factors contributed to the advance of detectors at this time. The first was the availability of techniques for growing crystals, developed notably by Pohl's group in Germany before World War II and by Stockbarger at MIT; I myself started to grow crystals in 1937 to make infra-red transmitting materials in connexion with the air defence work. The techniques could be applied to growing a whole range of materials, including the semiconductors, so that these were available for single-crystal studies of the solid state, and thus for the understanding of solid state theory. In turn this understanding, which was the second factor, could be applied to the development, at the Royal Radar Establishment and elsewhere, of detectors whose characteristics could be largely predicted, and which could be reproduced without the uncertainties that had been inherent in the earlier empirical methods. It was thus possible to progress from the intrinsic photoconductors to those which depended on impurity levels, such as gold-doped germanium.

A further advance was that by Rollin[42] and, independently, by Goodwin.[43] This was to use the free carriers in n-type indium antimonide to absorb long-wave photons, giving an increase in carrier mobility that would show up as an increase in conductivity. The absorption could be enhanced by placing the crystal in a magnetic field, and so producing a cyclotron resonance that would enable the detector to be tuned; such fields could now be achieved by the development of superconducting magnets. Instruments embodying these principles were made in 1961 and 1962, and they could be tuned up to 2000 microns. They provided a further experimental bridge between optics and radio waves. The gap had first been

bridged by Glagoleva Arkadieva[44] in 1920 when she showed that radiation of about 90 microns wavelength could be generated by using iron filings as Hertzian oscillators sparked in oil and Rubens[45] had already shown that a wavelength substantially greater than this— 330 microns—could be generated in mercury vapour. A second bridge was provided by Dicke[46] in 1946 when he showed that it was possible to use standard microwave techniques to detect the thermally-excited long-wave radiation from a black body.

Besides all these developments in photo-conductive detectors, and thermal devices such as the pyro-electric detector of Putley, we must remark the very ingenious detector of Luft[47] which he used in his gas analyser and which depends on the selective absorption of radiation by gas in the detector. This was a modern adaptation of an idea due to Tyndall[48], who used to demonstrate selective absorption of the radiation from hot carbon dioxide in a Bunsen flame by allowing it to pass through a flask itself containing carbon dioxide. Tyndall showed that when a sector was rotated at a suitable frequency between the flame and the flask, the alternate expansion and contraction of the gas in the flask as it heated and cooled due to the chopping of the radiation could be audibly manifested by acoustic resonance in the flask. The elegance of the device lies in the specific nature of its response to radiation that is characteristic of the gas in the flask. Luft's analysers have thus proved very effective in detecting small traces of gaseous impurities.

New Applications

The development of detectors, although it forms a subject in itself, is really a means towards the end of studying the spatial, temporal, and spectral aspects of the radiation emitted by various kinds of source, and thence to learning something about the source—be it the position of an aircraft, the spectrum of a molecule, the radiation from a breast cancer or a hot point on a transmission line. Much development has taken place in all these respects since 1945. Scanning techniques for missile homing heads, which appeared in rudimentary form during World War II, are now highly sophisticated. Spectroscopy for molecular and astrophysical studies has been much improved by interferometric devices pioneered by Fellgett.[49]

Thermography

The conversion of infra-red pictures into visible ones has been an objective ever since the earliest days of infra-red studies. William Herschel's son, John, devised a thermographic system[50] which depended on the differential drying of dampened paper by the local temperature rise caused by incident thermal radiation. Many years later Czerny[51] devised a more promising system depending on the differential evaporation of an oil film, and this was explored by the Germans as a possible device during World War II. In view of the success of more recent devices, notably that produced by the Barnes Engineering Corporation, it may be worth mentioning that the Barnes system was anticipated in surprising detail by another German development during World War II; and since this is not generally known, it
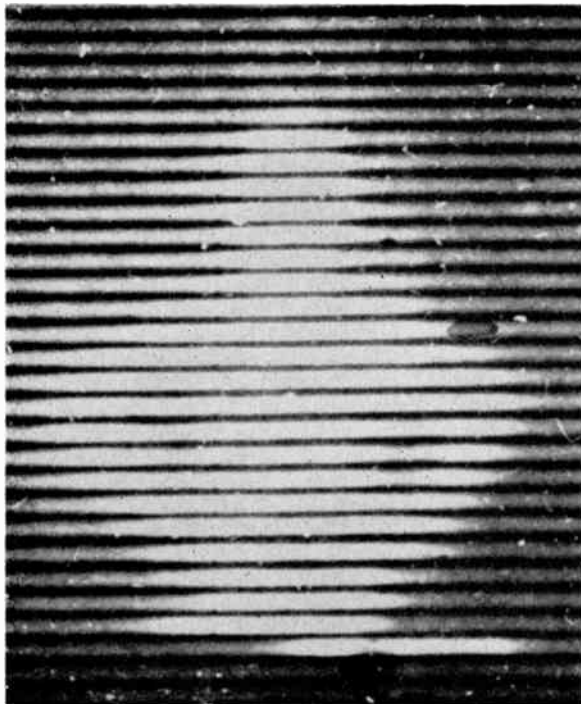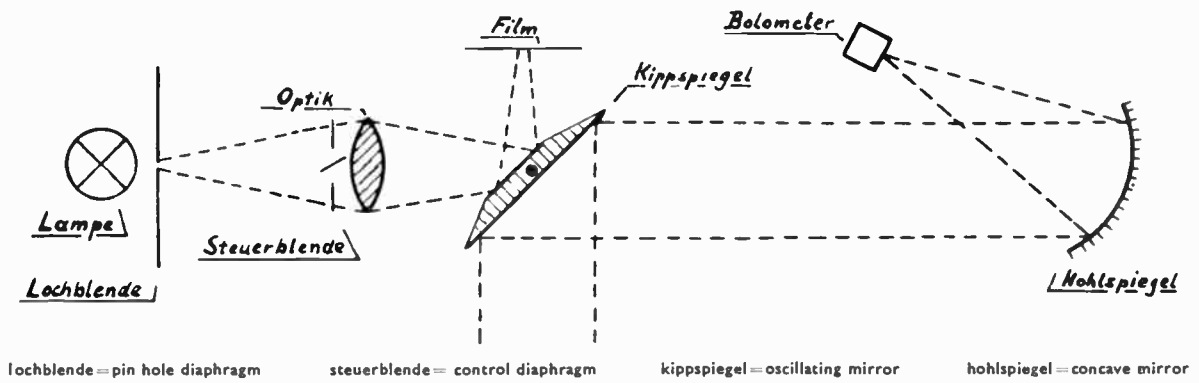
lochblende = pin hole diaphragm   steuerblende = control diaphragm   kippspiegel = oscillating mirror   hohlspiegel = concave mirror

Fig. 2. (a) Scheme of thermographic apparatus 'Potsdam L' made by E. Leybold's Nachfolger, from their memorandum of December 1944. (b) and (c) Thermographs produced by 'Potsdam L'.

may help if I give some details. The development was undertaken by E. Lebold's Nachfolger of Clausthal, and was known as the 'Potsdam L' apparatus. The sensitive element was a quick response bolometer designed by Czerny, and the scanning and recording system was very similar to that later used in the Barnes infra-red camera. The scheme, along with sample thermographs, is shown in Fig. 2. The military objective of the development was aerial thermal reconnaisance; and although the apparatus was far from attaining the necessary performance before the War ended, it represented an impressive step towards the modern development of thermography. Radiation from the human body and other surfaces can now be pictorially presented by a range of devices which approach photographic resolution.

New Generators

In contrast with what has happened in detection, there have been relatively few developments in the generation

of infra-red radiation; but there is, of course, one which is at least comparable in scientific and technical interest with anything that has occurred in detectors. This is the laser, where remarkable powers have been achieved with carbon dioxide at 10·6 microns and with hydrogen cyanide at 337 microns; and the spin-flip Raman laser of C. K. N. Patel's group at Bell Telephone Laboratories and S. D. Smith's at Heriot-Watt University[52] offers the prospect of tuning with high resolution.

Infra-red Meteorology

Among the many applications of infra-red, one of the most remarkable has been that by J. T. Houghton's group at Oxford and S. D. Smith's group,[53] who have adopted a suggestion by L. D. Kaplan[54] of MIT, and have demonstrated that it is possible to survey the temperature profile of the atmosphere by a gas-analyser

type of detector in a satellite. Its mode of operation depends on the small changes in the emission spectrum of carbon dioxide at 15 microns due to the change of fine structure with temperature. At first sight, I think, many physicists would have said that this would be impossible; but more than two million profiles have already been obtained.

Natural Sensors

Finally we may note that, as has so often happened with other forms of detection, infra-red has been anticipated in the animal world. The *Crotalidae*, or pit vipers, have pit-organs on either side of their heads which are sensitive to thermal radiation and which are used in the location of prey; Bullock and Barrett[55] report a threshold sensitivity of about 10^{-3} watt cm^{-2} for a detectable neural response from these organs. The *Boidae*, or boa constrictors, have labial pits which are about five times less sensitive, but which appear to be used for the same purpose.

Epilogue

After nearly 40 years' association with infra-red, I might well be asked what developments have impressed me the most. Solid-state detectors, lasers, interference spectrometry, thermography, synthetic crystals, homing missiles, and atmospheric sounders have all appeared in that time. If I had to pick the most surprising, these would be the laser, the tuned long wave detector, and the atmospheric temperature sounder. The cumulative effect of all these developments is astonishing: when I recall the sheer practical difficulty of getting a simple thermal detector to work in an aircraft in 1937, I find the present state of the art almost incredible—even though we still have a little way to go in producing ideal detectors, and we still lack good techniques beyond 50 microns, and especially a good system of rapid pictorial presentation.

We used to bewail the tantalizing state of infra-red technique. If it had been ten times better we could have been reasonably sure that it would be useful. And if it had been ten times worse, we could have given it up. In the meantime, thanks to great efforts by many workers it has become much more than ten times better, and we are beginning to see some remarkable results.

References

1. Newton, I., 'Opticks', p. 323, 2nd Edition (W. & J. Innys, London, 1718).
2. Herschel, W., 'Investigation of the powers of the prismatic colours to heat and illuminate objects; with remarks, that prove the different refrangibility of radiant heat. To which is added, an inquiry into the method of viewing the sun advantageously, with telescopes of large apertures and high magnifying powers, *Phil. Trans.*, 90, pp. 255–92, 1800.
3. Herschel, W., 'Experiments on the solar, and on the terrestrial rays that occasion heat; with a comparative view of the laws to which light and heat, or rather the rays which occasion them, are subject, in order to determine whether they are the same, or different', *Phil. Trans.*, 90, pp. 437–538, 1800. (See p. 521).
4. Young, T., 'On the theory of light and colours', *Phil. Trans.*, 92, pp. 12–48, 1802.
5. Baden-Powell, 'Report on the present State of our Knowledge of the Science of Radiant Heat', Reports of the First and Second Meetings of the British Association, pp. 259–301 (Murray, London, 1833).
6. Kirchhoff, G. R., 'Über den Zusammenhang zwischen Emission und Absorption von Licht und Wärme', *Monatsberichte der Akademie der Wissenschaften zu Berlin*, pp. 783–7, 1859. See also Balfour Stewart, 'Account of some experiments on radiant heat', *Proc. Roy. Soc. Edin.*, 22, pp. 1–3, 95–7, 426–9, 1858.
7. Prévost, P., 'Du Calorique Rayonnant' (Paris, Geneva, 1809).
8. Forbes, J. D., 'On the refraction and polarization of heat', *Phil. Mag.*, 6, pp. 134–42, 205–14, 284–91, 866–71, 1835.
9. Nobili, L. and Melloni, M., 'Recherches sur plusiers phénomènes calorifique enterpris au moyen du thermo-multiplicateur', *Ann. Chim. Phys.*, 48, pp. 198–218, 1831.
10. Fizeau, H. and Foucault, L., 'Recherches sur les intérferences des rayons calorifiques', *C.R. Acad. Sci. Paris*, 25, pp. 447–50, 1847.
11. Stefan, J., 'Über die Beziehung zurischan der Wärmestrahlung und der Temperatur', *Wien Berichte*, 79, pp. 391–428, 1879.
12. Draper, J. W., 'On the fixed lines in the ultra-red invisible region of the spectrum', *Phil. Mag.*, 3, pp. 86–9, 1877.
13. Draper, J. W., 'On the phosphorograph of a solar spectrum and on the lines in its infra-red region', *Phil. Mag.*, 11, pp. 157–69, 1881.
14. Abney, W. de W., 'On the photographic method of mapping the least refrangible end of the solar spectrum', *Proc. Roy. Soc.*, 30, p. 67, 1880.
15. Abney, W. de W., *op. cit.*, *Phil. Trans.*, 171, pp. 653–67, 1880 ('Infra-red' occurs on p. 655).
16. Crookes, W., 'On repulsion arising from radiation', *Phil. Trans.* 166, pp. 325–76, 1876.
17. Langley, S. P., 'The bolometer and radiant energy', *Proc. Amer. Acad.*, 16, pp. 342–59, 1881.
18. Langley, S. P., 'The history of a doctrine', *Amer. J. Sci.*, 37, pp. 1–23, 1889.
19. Wien, W., 'Temperatur und Entropie der Strahlung', *Wied. Ann. der Phys.*, 52, pp. 132–65, 1894.
20. Wien, W. 'Über die Energievertheilung im Emissionsspectrum eines schwarzen Körpers', *Wied. Ann. Phys.*, 58, pp. 662–9, 1896.
21. Lummer, O. and Pringsheim, E., 'Die Strahlung eines'schwarzen Körpers' Zwischen 100°C and 1300°C', *Wied. Ann. der Phys.*, 63, pp. 395–410, 1897.
22. Desains, P. and Curie, P., 'Recherches sur la détermination des longuers d'onde des rayons calorifiques à basse température', *C.R. Acad. Sci., Paris*, 90, pp. 1506–9, 1880.
23. Rayleigh, Lord, 'Remarks on the law of complete radiation', *Phil. Mag.*, 49, pp. 539–40, 1900.
24. Jeans, J. H., 'On the partition of energy between matter and aether', *Phil. Mag.*, 10, pp. 91–8, 1905.
25. Rubens, H. and Kurlbaum, F., 'Anwendung der Methode der Reststrahlen zur Prüfung des Strahlungsgesetzes', *Ann. der Phys.*, 4, pp. 649–61, 1901.
26. Planck, M., 'Über eine Verbesserung der Wienschen Spektralgleichung', *Verh. Deutsch. Phys. Gesellschaft*, 2, pp. 202–4, 1900.
27. Planck, M., Reported in Jammer, Max., 'The Conceptual Development of Quantum Mechanics', pp. 22–3 (McGraw-Hill, New York, 1966).
28. Einstein, A. 'Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt', *Ann. der Phys.*, 17, pp. 132–48, 1905.
29. Jammer, Max., *loc. cit.*, p. 44.
30. Jammer, Max., *loc. cit.*, p. 187.
31. See reference 9.
32. Ångstrom, K. J., 'Étude des Spectres Infra-rouge de l'Acide Carbonique et de l'Oxyde de Carbone', pp. 549–57 (Öfvers, Stockholm, 1889).
33. Rubens, H. and Nichols, E. F., 'Versuche mit Wärmestrahlen von grosser Wellenlänge', *Ann. der Phys.*, 60, pp. 418–62, 1897.
34. Hoffmann, S. O., 'The detection of invisible objects by heat radiation', *J. Amer. Phys. Soc.*, 24, pp. 163–6, 1919.
35. Moll, W. J. H. and Burger, H. C., 'The thermorelay for recording small rotations', *Phil. Mag.*, 50, pp. 624–6, 1925.
36. Ising, G., 'A natural limit for the sensibility of galvanometers', *Phil. Mag.*, 1, pp. 827–34, 1926.
37. Dobson, G. M. B. and Perfect, D. S., 'A method of comparing very small amounts of light by means of a photoelectric cell and a valve amplifier'. 'Physical and Optical Societies Joint Dis-

cussion on Photoelectric Cells and their Applications', pp. 79–84 (The Physical and Optical Societies, London, 1930).

38. Ayrton, W. E. and Perry, J., 'Modes of measuring the coefficients of self and mutual induction', *J. Soc. Telegr. Engrs*, **16**, pp. 292–343, 1888.

39. Jones, R. V., 'Infrared detection in British Air Defence 1935–38', *Infrared Physics*, **1**, pp. 153–62, 1961.

40. Lange, B., 'New photo-electric cell', *Phys. Zeits.*, **31**, p. 964, 1930.

41. Welker, H., 'Über neue halbleitende Verbindungen', *Z. Naturforsch.*, **7a**, pp. 744–9, 1952 and **8a**, pp. 248–51, 1953.

42. Kinch, M. A. and Rollin, B. V., 'Detection of millimetre and submillimetre wave radiation by free carrier absorption in a semi-conductor', *Brit. J. Appl Phys*, **14**, pp. 672–6, 1963.

43. Goodwin, D. W. and Jones, R. H., 'Far infrared and microwave detector', *J. Appl. Phys.*, **32**, pp. 2056–7, 1961.

44. Arkiedeva, A. A. Glagoleva, 'A new source of very short electromagnetic waves', *Z. fur Phys.*, **24**, pp. 153–165, 1924.

45. Rubens, H. and von Baeyer, O., 'On extremely long waves emitted by the quartz mercury lamp', *Phil. Mag.*, **21**, pp. 689–95, 1911.

46. Dicke, R. S., 'The measurement of thermal radiation at microwave frequencies', *Rev. Sci. Instrum.*, **17**, pp. 268–75, 1946.

47. Luft, K. F., 'Recording method of gas analysis by means of infrared absorption without spectral splitting', *Zeits. f. tech. Phys.*, **24**, pp. 97–104, 1943. See also Fastie, W. G. and Pfund, A. H., 'Selective infra-red gas analyzers', *J. Opt. Soc. Amer.*, **37**, pp. 762–8, pp. 1651–1653, 1947.

48. Tyndall, J., See for example 'Heat as a Mode of Motion', 4th Edn., p. 403 (Longmans, London, 1870).

49. Fellgett, P. B., Ph.D. Thesis, Cambridge University, 1951.

50. Herschel, J. F. W., 'On the chemical action of the rays of the solar spectrum on preparations of silver and other substances, both metallic and non-metallic, and on some photographic processes', *Phil. Trans.*, **130**, pp. 1–60, 1840.

51. Czerny, M. and Mollet, P., 'New method in infra-red spectrography', *Phys. Zeits.*, **38**, pp. 1008–10, 1937.

52. Allwood, R. L., Smith, S. D., *et al.*, 'A tunable spin-flip magneto-Raman laser', Proceedings of Conference on Infra-red Techniques, Reading 1971, pp. 107–114. (I.E.R.E. Conference Proceedings No. 22); *The Radio and Electronic Engineer*, **42**. (To be published.)

53. Houghton, J. T., Ellis, P. J., *et al.*, 'Infra-red atmospheric temperature sounding from satellites', *loc. cit.*, pp. 257–270; *The Radio and Electronic Engineer*, **42**. (To be published.)

54. Kaplan, L. D., 'Inference of atmospheric structure from remote radiation measurements', *J. Opt. Soc. Amer.*, **49**, pp. 1004–7, 1959.

55. Bullock, T. H. and Barrett, R., *Commun. Behav. Biol.*, Part A1, p. 19, 1968. See also Harris, J. F. and Gamow, R. I., 'Snake infrared receptors: thermal or photochemical mechanism?', *Science*, **172**, pp. 1252–3, 1971.

Additional Bibliography

Ballard, S. S., (Ed.) Special issue on Infra-Red Physics and Technology, *Proc. Inst. Radio Engrs*, **47**, pp. 1415–1649, 1959.

Barr, E. Scott, 'Historical survey of the early development of the infra-red spectral region', *Amer. J. Phys.*, **28**, pp. 42–54, 1960.

Barr, E. Scott, 'The infra-red pioneers':
  I. 'Sir William Herschel', *Infrared Physics*, **1**, pp. 1–4, 1961;
  II. 'Macedonio Melloni', *Infrared Physics*, **2**, pp. 67–74, 1962;
  III. 'Samuel Pierpoint Langley', *Infrared Physics*, **3**, pp. 195–206, 1963.

The Author

Professor R. V. Jones was educated at Wadham College and the Clarendon Laboratory, Oxford, where he took his D.Phil. degree in Infrared Spectroscopy in 1934. He was subsequently Skynner Research Student in Astronomy at Balliol College, and worked on the detection of aircraft by infrared in the years before World War II. During the war he was Head of Scientific Intelligence, and was Director of Intelligence on the Air Staff and later Director of Scientific Intelligence in the Ministry of Defence. He was appointed to the Chair of Natural Philosophy at Aberdeen University in 1946.

Many governmental and other bodies have called upon Professor Jones's knowledge and experience during the past twenty-five years. His appointments have included the Chairmanship of the Safety in Mines Research (Advisory) Board, of the Research Advisory Council of the Transport Commission, of the Paul Instrument Fund Committee, of the Infrared Committee in the Ministry of Aviation, of the Electronics Research Council in the Ministry of Technology, of the Air Defence Working Party of the Ministry of Defence, and of the British National Committee for the History of Science, Medicine and Technology. He is a Vice-President of the Royal Society and is Editor of the Society's Notes and Records. He is an Honorary Fellow of Wadham College and an Honorary D.Sc. of the University of Strathclyde.

His research interests include particularly the measurement of small displacements, and fundamental phenomena in optics such as radiation pressure and the aberration of light.

Electronics in Letter Handling Systems

K. H. C. PHILLIPS, C.Eng., M.I.E.R.E.*

Based on a paper presented at the Conference on Electronic Control of Mechanical Handling held in Nottingham from 6th to 8th July 1971.

SUMMARY

Within the past decade, electronic devices have superseded electromechanical devices in the control of complex mechanical machinery. The benefits gained include improved performance, increased reliability, reduced costs and wider application of automatic control. This paper describes some aspects of electronic control incorporated in the Letter Handling Systems currently being installed by the British Post Office throughout the U.K.

* *Post Office Headquarters, Postal Mechanization and Buildings Department, London EC2V 7JL.*

1 Introduction

The sorting of 35 million letters and packages posted throughout the United Kingdom during each weekday presents a formidable mail handling task to the British Post Office. This task when undertaken using traditional hand sorting methods incurs a significant man-power commitment, since more than 250 million individual letter handling operations are involved during each 24-hour period of postal operations. The undertaking is aggravated by the receipt of more than 75% of the total daily postings in rather less than 25% of the 24-hour work period available.

In order to maintain the standard of its services and meet the increasing demands on its resources, the Post Office has embarked on a nation-wide programme of mechanizing a number of large letter sorting offices.[1] A range of specially designed letter handling machines has been developed which together form an integrated mail handling system. The tasks of some of these machines are complex, and the consistent achievement of the high degree of operational reliability essential to the service is conditional on the simplicity, reliability and performance of the machine control systems involved.

Much of the electronics associated with letter handling machinery is concerned with position sensing, synchronization and controlling the paths of letters through each type of machine. Other electronic systems are concerned with the conversion of address or routing information as read from the envelope by machine operators and electronic means, into signals used to control the operation of associated machines. Computer systems are currently finding application in the monitoring of machine performance and also in machine operator training facilities.

2 Letter Handling Processes

Before describing applications of electronics in postal automation it is pertinent to mention briefly the letter handling processes involved in mechanized sorting.

2.1 Segregation

Mail collected from street pillar boxes arrives at the sorting office in bags which may contain letters and packets having a wide diversity of thickness, weight, shapes and sizes. The speedy separation of this mail into groups suitable for machine handling is an essential pre-requisite for any letter mail mechanization system. This process is known as segregation and is achieved by mechanical gauging and measuring processes.[2] Letters can be separated into 2 groups:

(1) items having a thickness greater than $\frac{1}{4}$ inch

and (2) items $\frac{1}{4}$ inch thick or less having

    (a) width greater than 5 inches

    (b) length greater than 7 inches

    (c) length less than 7 inches

    (2(b) and 2(c) embrace the Post Office Preferred letter sizes.)

Ninety per cent of the mail traffic is of type 2(c) which is suited to further mechanized processing.
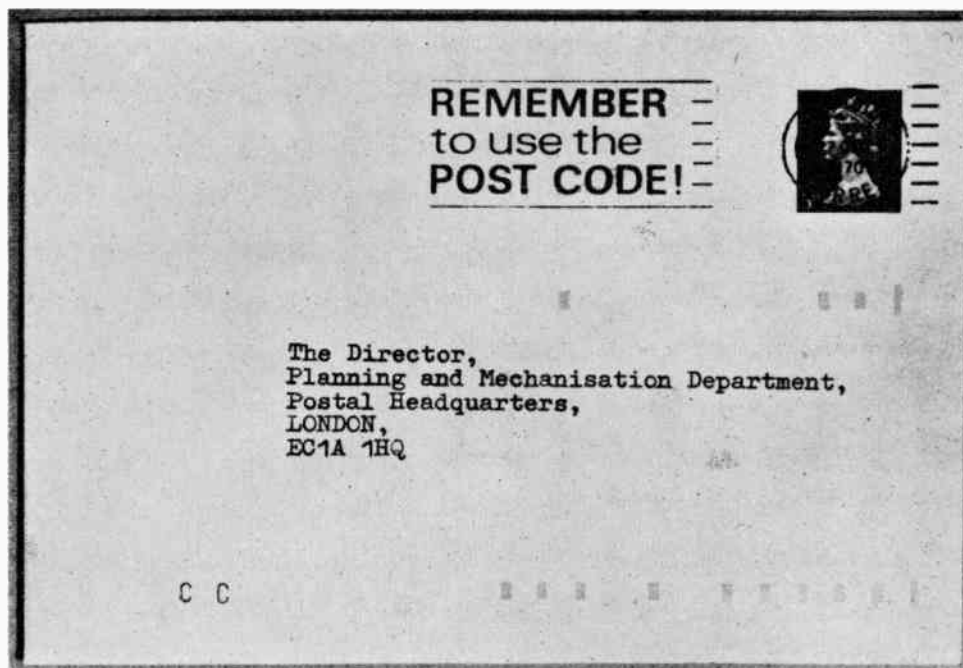
Fig. 1. Envelope codes.

In segregation, photocell detection and logic circuits are used to detect mail flow stoppages in the letter transport conveyors thereby limiting interruptions to the flow of mail.

2.2 Letter Facing

After segregation suitable mail is 'faced'. The automatic letter facing machine, ALF[3, 4] has three functions:

1. To orientate all letters in such a manner that the address can be easily read.
2. To cancel or obliterate the stamp and date mark the items.
3. To separate 1st from 2nd class letters.

Unlike segregation, facing involves high speed (20 000 items per hour) inspection of each letter.

The reliable facing of letters in the ALF is governed by the position of the stamp on the envelope, which is expected to be in the top right-hand corner of the envelope relative to the address. Stamps are over-printed with a non-toxic phosphorescent resin which, after irradiation by ultraviolet light of 2537 Å wavelength, glows for up to 500 ms. Stamps are coded by over-printing the stamp appropriate to the first weight step (4 oz)* in the second class mail (at present $2\frac{1}{2}$p) with a single central vertical bar. All other stamps have two phosphorescent bars, one on each side of the stamp face. The presence of the stamp on the envelope and the coding information it contains is detected by photomultipliers positioned in the machine letter path immediately following the u.v. irradiation section.

The ALF has two stamp scanning sections, in the first each letter is examined for the position of the stamp. The information obtained is used to control a diverter in the letter path, which directs the letter either through a twist path or a straight through path, such that when the letter stream is re-combined, all stamps will be located near a common long edge of the letter. The second scanning position is used to detect the code information on the stamp, and its position, i.e. either close to the leading edge of the letter or close to the trailing edge. The information obtained is subsequently used to control the routing of the letters into one of four stacks, i.e. 1st class stamp leading; 1st class stamp trailing; 2nd class stamp leading; and 2nd class stamp trailing. A fifth stack is used for collection of unstamped or mis-read items.

Some items of mail which bear no stamps, e.g. Official Paid and pre-printed Business Reply items, are detected by optical scanning equipment located adjacent to each phosphor scanner. The optical scanners detect the contrast in the intensity of the light reflected by a fine beam of light directed at a pre-defined area on the face of each mail item. The contrast in the reflected light obtained from the printed pattern and the background colour of the paper of such items is used to determine the orientation of the item in the facing process described earlier. In the case of Official Paid mail the stylized Official Paid medallion located in the normal position of the stamp is used for facing purposes. Business Reply items are faced by detecting the position of the specially printed vertical lines on the card or envelope.

The main function of the electronics in the ALF involves letter position sensing, scanning, registration of detected data with the letter transport systems, data analysis, and letter path diverter control. Sequencing and synchronization of the machine control signals are strictly related to machine speed by control pulses generated within the machine.[4]

2.3 Letter Coding

In any mechanized letter coding and sorting system the address information on every letter must be encoded and

*2oz with effect from 6th March 1972.

Fig. 2. Schematic diagram of the coding desk translator.
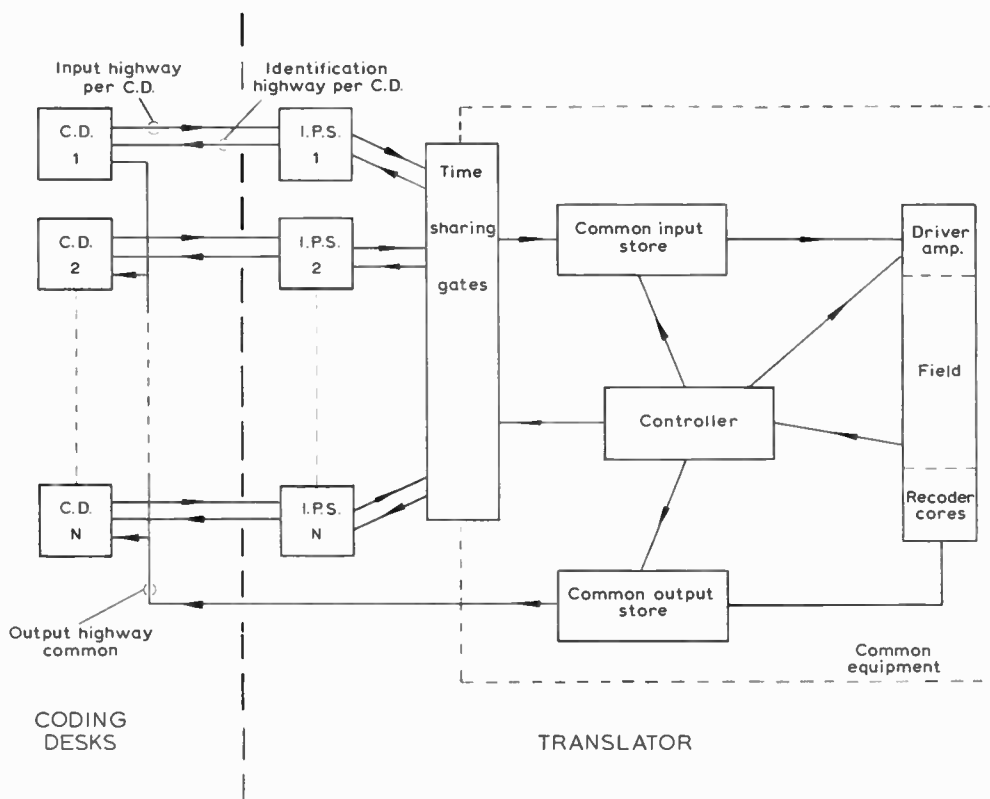
printed in some form on the envelope which facilitates high-speed reading and sorting by machine. Since optical character recognition techniques cannot yet be reliably applied to letter mail circulating in the U.K., the process of reading address information prior to code-marking the envelope is a manual task, and code-marks are printed on to individual letters at an operator-controlled letter coding desk. These letters are presented one at a time to an operator who types on an almost conventional typewriter keyboard the postcode or specified parts of the address appearing on the envelope. The characters keyed have to be converted or translated into a machine readable code.[5] The British Post Office uses two 14-bit binary codes printed on the envelope as shown in Fig. 1. The code-mark nearer the lower edge of the envelope identifies the city/town or postal delivery area to which the letter is to be delivered. The other code-mark identifies a street, part of a street or thoroughfare, or building within the postal delivery area identified by the lower code-mark. Code-marks are printed[6] on the envelope using a material having phosphorescent properties similar to, but not the same as that used for stamp detection. The equipment which translates the keyboard information into data signals used to control the pattern of binary code-marks printed at the coding desk is known as the coding desk translator.

2.4 Sorting

After being coded letters are loaded into the automatic sorting machines. In each machine letters are individually fed at speed (approx. 8000 letters/hour) into a code-mark reader where envelopes are irradiated with u.v. light of 3650 Å wavelength which stimulates the phosphor to emit a blue light peaking in the 4500 Å region. After irradiation letters are passed in front of a photomultiplier which detects the code-marks during the after-glow period. Since the letter is in motion while passing the photomultiplier, the readout is sequential and the reader circuitry assembles and staticizes the code-mark pattern. The first of the code-marks read (start mark) is used to synchronize the operation of the reader circuitry. The staticized code-mark data are subsequently used to determine into which of the sorting machine letter sortation boxes the letter is to be deposited.

In mechanized sorting it is clearly impracticable and uneconomic to provide an individual sortation box to collect together mail for delivery to each of the 1600 postal delivery districts throughout the U.K., or even for each delivery point within the local postal service area. Further, it would be uneconomic to install sorting machines which have fixed sorting methods. Normally the allocation of sortation boxes is determined by mail volume, distribution, transportation and delivery considerations. In meeting these requirements it must be possible to route letters bearing any code-mark to any sorting machine sortation box.

The function of converting code-mark data into any sorting machine instruction code whilst providing flexibility of sorting arrangements is provided by the sorting machine translator.

3 Translation Systems

In the British code/sorting system both types of translation equipment currently in use employ a threaded core field as a memory device in the translation process.[7] A

generalized description of the threaded core coding desk translator operation will serve to illustrate some of the principles employed. The coding desk translator is used in conjunction with a number of coding desks to convert the letter address information as input via the coding desk keyboards, into one or two 12-bit binary codes for printing on the envelope.

Input codes keyed by the coding desk operator can comprise up to 7 alpha-numeric characters. Codes can be copy-typed postcodes (e.g. BS10 2AB), or alphabetic codes extracted from address information (ABEON for ABERavON, or QUENS for QUEEN Sт). Other variations comprising fewer characters are also permissible.

Figure 2 shows a schematic diagram of the coding desk translator. The keyboard on each coding desk is connected directly to an individual input store at the translator. Alpha-numeric key depressions are binary encoded at the keyboard and fed individually to the input store where input codes are compiled, staticized and recognized. Each input store is connected to the translator common equipment via time sharing gates on a sequential basis. All translator operations within the common equipment are executed sequentially by the controller.

When the code in the input store is recognized as valid an 800 µs time slot is allocated during which the complete translation operation is completed. Initially up to 5 binary encoded characters are transferred and staticized in key depression order in the common input store register. If the input code contains more than 5 characters, all except the last three characters are transferred during the first half of the time slot (i.e. the first 400 µs), the remainder during the second half. The two parts of the code are treated as separate translation requests, and two separate translation output codes are generated within the same time slot. If the input code has fewer than 5 characters, the unused character positions in the common input store are occupied by a filler code.

The threaded core translation field consists of 5 columns of mu-metal uncut 'C' cores having linear B/H characteristics. Each of the 5 columns is associated with a character position in the common input store register. One core is assigned to each alphabetic and numeric character in each column. All cores in the field have an associated core drive amplifier. Logic gates on the input to each core drive amplifier in each column are arranged to decode the binary input code staticized in each character position of the common input store register. Thus the 5-character code staticized in the common input store register will prime the input to only one core drive amplifier in each column of cores. Filler codes in the main register are associated with filler cores in each column.

Translation is effected by applying a simultaneous drive pulse to all core drive amplifiers which results in core drive current being switched into only those cores with primed input decoding gates. Each driven core acts as a pulse transformer which results in a voltage being induced into each wire threading that core. A unique wire which threads all five driven cores would thus have

five units of voltage induced into it. Other wires threading the field would have 4 or less units of voltage induced depending upon how many of the driven cores are threaded. A constant current generator connected to all wires threading the field allows 100 mA to flow in the field. From the simple explanatory circuit given (Fig. 3) it can be seen that when the core drive amplifiers are energized, current will flow only in that wire having 5 voltage units induced in it, since the diodes in all the other threading wire paths will be reverse biased.
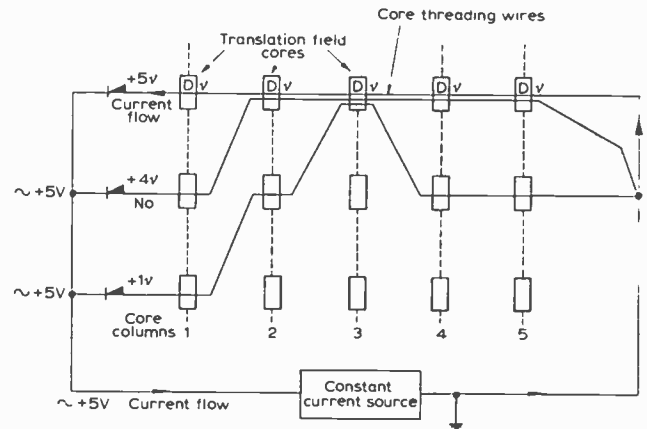

Fig. 3. Simplified explanation of current flow in translation field.

The 12-bit binary translation code to be associated with the alphabetic or alpha-numeric code held in the common input store is generated by routing the 100 mA field current through 12 selected secondary or recoder cores corresponding to the required 12-bit code-mark. The recoder field comprises 24 cores, one core corresponding to the '1' and '0' logic state of the 12-bit binary pattern. The signals generated in the secondary windings of the recoder cores are amplified and staticized in the output store register. The outputs of this register are connected to a common data highway feeding all coding desks associated with the translator.

In order to prevent erroneous data being selected at the recoder field outputs and subsequently routed to the coding desk printer store, a validity check is made on the core drive conditions in the translator field which caused the output data to be staticized. This is achieved by comparing the voltage established across the translation field against a pre-determined threshold voltage. The voltage established across the field by 5 units of voltage induced into a single wire threading each driven core is higher than that attainable in one or many wires having 4 units of voltage induced. Since valid translations are only possible when a single wire threads 5 driven cores, the threshold voltage is set at a level in excess of that attainable in wires having 4 units of voltage induced, but lower than the minimum voltage established in a wire threading 5 driven cores.

When the voltage across the field exceeds the threshold level the state of the threshold sensing amplifier output changes causing a coding desk identification, or routing
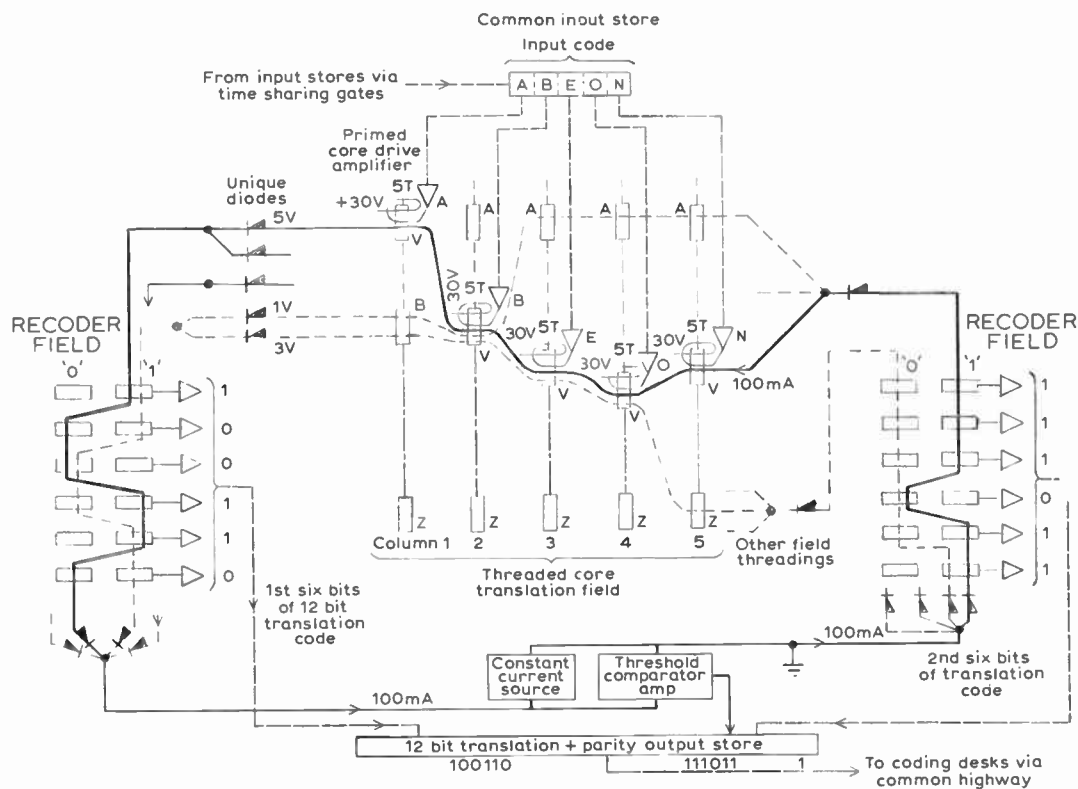
Fig. 4. General arrangement of threaded core field.

signal to be generated. The identification signal causes the code-mark printer control data available on the translator output highway to be routed into the calling coding desk printer store. The absence of a threshold comparison amplifier output prevents data in the output store being routed to any coding desk.

Translator output data comprises a 13-bit binary code having odd parity, the parity bit being generated in the output store by the parallel addition of the recoder output data. The 14-bit (start bit) printed on the envelope is added by the coding desk printer.

A more complete circuit arrangement of the coding desk translator field is given in Fig. 4.

In practice the 12-bit recoder field is divided into two 12-core groups each capable of generating a 6-bit binary code. This simplifies the core threading arrangements in the translation field.

In each recoder group a core is used to represent both the '1' and '0' state of each bit in the 6-bit binary code. Each of the 64 possible binary outputs from the 6-bit field is generated by supplying field current to one of the 64 field threading wires. The threading wires associated with both fields are terminated on each side of the translation field. Thus by connecting a translation field threading wire to one of 64 terminating points on each side of the translation field, any five-character alpha or alpha-numeric coding desk input code can be associated with any one of (64 × 64) 4096 different 12-bit binary output codes.

The coding desk translators currently in the field have a translation field threading capacity of approximately 16 000 wires in two separate fields, and can service up to 63 coding desks each working independently within any 50 ms period.

Self-checking and system monitoring facilities are incorporated which enable rapid fault detection, location and cure. The detection of a system malfunction inhibits access from all connected coding desks since the potential presentation of erroneous data to the code-mark printer units would result in subsequent mis-sorting and incur delays in the delivery of mail.

The data translation technique using threaded core fields was adopted for the British letter mechanization systems for the following reasons:

(1) The equipment could be manufactured relatively cheaply, involving no complex manufacturing processes.

(2) The data storage media is 'non-volatile' (threaded wires), and security of translation data is assured.

(3) Simplicity of the technique gives reliability.

(4) The data conversion process in the fields is affected quickly (less than $10 \mu s$) which facilitates time sharing of the common translation equipment amongst a large number of associated user machines.

4 Electronic Control Elements

The performance and efficiency of letter handling machinery is largely dependent upon mechanical design and methods used to transport, orientate and control the movement of letters. The design of related electronic control systems is, in comparison, less difficult since

logic circuits can be easily configured to accommodate most control requirements in complex machinery.

The range of electronic logic circuit functions used in letter processing machinery are the same as those found in many industrial control applications. Experience with early generations of British letter handling machinery showed that the adoption of a variety of electronic techniques to realize essentially the same logic circuit functions promoted maintenance difficulties in equipment familiarization, fault location, and component replacement. Rationalization of logic circuits for use in all types of machinery and equipment was thus necessary to achieve an overall higher degree of performance and reliability consistent with a lower electronic equipment capital, spares and maintenance costs.

A range of diode transistor logic elements[8] having a defined performance specification and designed specifically for postal engineering applications was consequently introduced, and are currently used in the majority of letter sorting equipment control systems.

The features of the logic system developed are:

(1) All logic circuits are designed to a standard of performance consistent with the requirements of mechanized letter processing equipment control systems.

(2) All circuits are mounted on one of two sizes of printed circuit plug-in boards and incorporate a limited range of high-quality discrete circuit components, facilitating the adoption of mass production manufacturing techniques and giving attendant benefits of reduced costs.

(3) The number and complexity of logic circuit functions mounted on single plug-in boards is restricted to facilitate extensive circuit testing using an inexpensive 'go/no-go' tester, thereby simplifying fault locating procedures.

(4) Each circuit has designed high noise immunity to give reliable performance within the adverse electrical interference environment encountered in mechanized sorting installations.

(5) Both NAND and NOR logic functions are included in the range in order to minimize the number of circuit elements required to realize complex control circuits.

(6) Uniformity of circuit interface signal levels, circuit symbols and documentation standards aids circuit comprehension and simplifies equipment testing disciplines.

There are, however, situations where a particular logic circuit function cannot be configured efficiently or economically using the standard range of logic circuits. In such circumstances special logic circuit boards are used which have essentially the same circuit design standards, signal interface levels, and are mounted on the same size of printed circuit boards as the standard range. The variety of functions and small number of each type in use precludes the design of a 'go/no-go' tester, and circuit testing of these boards is accomplished by establishing test conditions and measuring perform-

ance using a diagnostic test facility available on the standard logic range test equipment.

Experience in operational translation equipments has shown that the 'go/no-go' circuit testing philosophy has simplified fault finding procedures and contributed to a lower equipment down-time. The measured reliability of the logic circuits has been higher than the calculated single failure per 200 circuits per year.

5 Future Trends

This paper has described some elements of system control electronics as applied to letter sorting equipment currently being installed in Sorting Offices in the U.K.

The decision to use a purpose-designed range of plug-in logic circuits in the majority of machine control applications has been justified by the performance and reliability obtained in the field. In an operating environment where short periods of machine 'downtime' could quickly contribute to the degradation of the overall sorting performance, efficient maintenance procedures are essential. The philosophy of incorporating a common range of inexpensive, easily tested and exchanged control elements into all types of machine has increased machine serviceability and reduced maintenance and spares costs.

It is evident that the low cost, greater reliability and better performance obtainable with microelectronic devices can be used to advantage in letter sorting equipment, and second generation systems will incorporate this technology where there is a useful advantage to be gained.

Equipment currently under development makes greater use of the flexibility, adaptability and development potential of small digital computers for system control and monitoring.

6 Acknowledgments

7 References

1. de Jong, N. C. C., 'Progress in postal engineering, Part 3, Letter mail', *Post Office Elect. Engrs J.*, 63, No. 4, pp. 203–11, January 1971.

2. Hills, E. G., 'Segregating operational requirements and development', British Postal Engineering Conference, May 1970, Paper No. 7. (Institution of Mechanical Engineers, London.)

3. Copping, G. P., 'Automatic letter facing', *loc. cit.*, Paper No. 17.

4. Wicken, C. S., 'Control systems for automatic letter facing machines', *loc. cit.*, Paper No. 18.

5. Andrews, J. D., 'Coding principles and practice', *loc. cit.*, Paper No. 12.

6. Hewett, J. W. and Muckett, R. G., 'Code-marking of letter mail', *loc. cit.*, Paper No. 38.

7. Goodison, H., 'Threaded core translators', *loc. cit.*, Paper No. 9.

8. Goodison, H., 'A review of postal engineering logic units', *loc. cit.*, Paper No. 26.

A Transistor Realization of the Generalized Impedance Converter

Professor L. T. BRUTON,
B.Sc., M.Eng., Ph.D.[*]

SUMMARY

A transistor realization of a generalized impedance converter (g.i.c.) is developed from an ideal nullor representation. The use of g.i.c.s for realizing low-sensitivity RC-active filters is well established and this new circuit overcomes the requirement for using two operational amplifiers in order to realize a g.i.c. The proposed transistor realization is a single-loop structure with the practical advantage of low output impedance; the use of this circuit in the realization of typical elliptic transfer functions is considered.

* Department of Electrical Engineering, The University of Calgary, Alberta, Canada.

1. Introduction

The generalized impedance converter (g.i.c.) has been proposed and used as a basic building-block in the synthesis of low-sensitivity RC-active filters.[1-5] It has been demonstrated that g.i.c.s using operational amplifiers result in high Q-factors and extended wideband operation.[6,7] The g.i.c. is defined, for the purpose of this contribution, as a linear two-port with a transmission matrix T given by

$$T = \begin{bmatrix} 1 & 0 \\ 0 & \dfrac{Z_2 Z_4}{Z_1 Z_3} \end{bmatrix}$$

and the well-known two-nullor[8,9] representation is shown in Fig. 1(a). This contribution is concerned with obtaining a simple transistor realization of the g.i.c. and
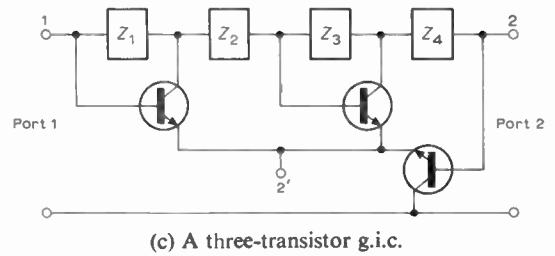

(a) A two-nullor g.i.c.

(b) An equivalent three-nullor g.i.c.

(c) A three-transistor g.i.c.
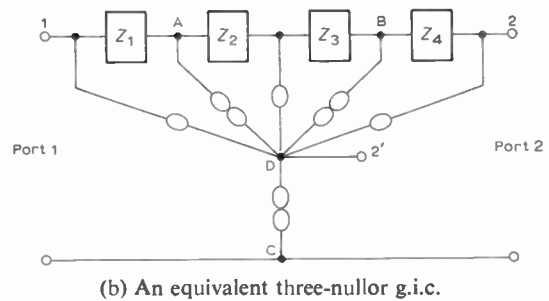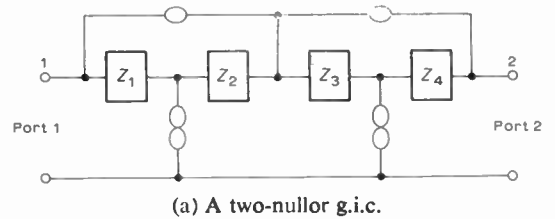

(d) An alternative realization.
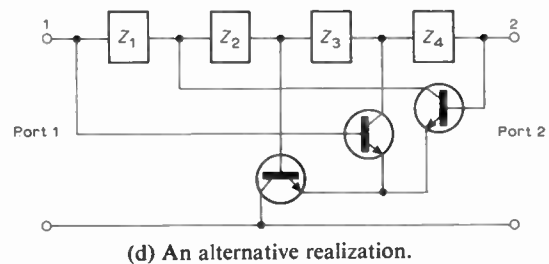
Fig. 1.

with the design and application of the resultant circuit to the synthesis of RC-active filters. Particular attention is given to the g.i.c. realization in which Z_1 and Z_3 are capacitive and Z_2 and Z_4 are resistive because it is this particular realization that has been proposed in several active filter synthesis techniques.

2. A Basic Three-transistor G.I.C.

A direct transistor realization of Fig. 1(a) is not possible because all possible nullator-norator pairs have four terminals whereas the nullor representation of the ideal transistor must obviously correspond to a three-terminal nullor. However, using the concepts of nullor equivalence[10] it is straightforward to prove that the three nullor network proposed in Fig. 1(b) is identical to the g.i.c. network in Fig. 1(a). Then, direct replacement of the nullors of Fig. 1(b) by ideal bipolar transistors leads to a three-transistor g.i.c. realization, as shown in Fig. 1(c). Although this transistor-g.i.c. realization is indeed physically realizable, it is necessary to consider the practical advantages that may be gained from this transistor-g.i.c.
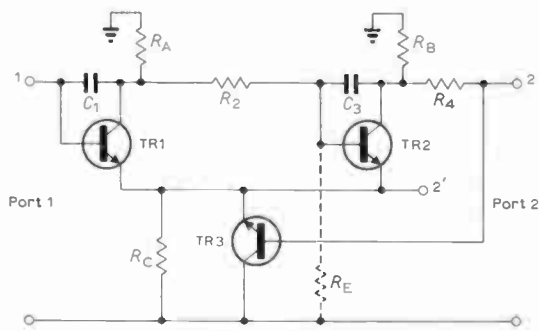
(a) Introduction of arbitrary resistors R_A, R_B, R_C.
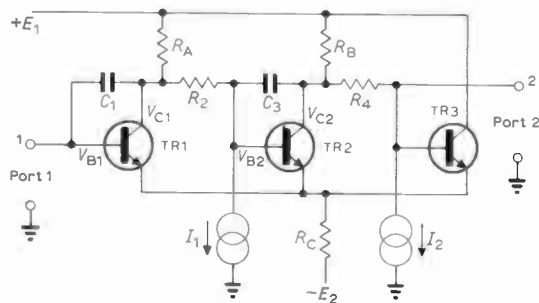
(b) A practical bias arrangement.

Fig. 2.

There is presently much published work concerning proposals for transistor-impedance converters[11–13] and it is the opinion of the author that it is not difficult to propose original transistor realizations of impedance converters by simply manipulating nullor realizations and then substituting transistors. The importance of any particular new transistor realization must be verified by suitable reference to the ease with which a corresponding practical circuit may be obtained. For example, an original three-transistor g.i.c. is given in Fig. 1(d) but the

author would not wish to pursue the relevance of the network because it is essentially a complicated multiple-loop structure for which satisfactory stable operation has not so far been achieved.

The three-transistor g.i.c. of Fig. 1(c) is basically three cascaded common-emitter stages that may be suitably biased by introducing bias resistors such that the small-signal operation of the network remains unaltered. Thus, it is noticed that nodes A, B, C and D in the nullor version (Fig. 1(b)) are interconnected by norators. Thus, an arbitrary connexion of additional impedances between nodes A, B, C and D will leave the network unaltered because the norators themselves allow completely arbitrary voltages and currents between these nodes. Three resistors R_A, R_B and R_C, shown in Fig. 2 (a), are so chosen that a resistive current path exists between every collector and every emitter and ground, thereby facilitating a direct current biasing arrangement via power supply voltages $+E_1$ and $-E_2$, as shown in Fig. 2(b). In the final design of this transistor-g.i.c. it has been necessary to provide direct current-sources I_1 and I_2 in order to establish constant bias voltages at the bases of transistors TR2 and TR3. It is a simple matter to determine I_1 and I_2. For example, if the circuit is designed for a collector voltage V_{C1} at transistor TR1 and a base voltage V_{B2} at transistor TR2, then we choose

$$V_{C1} - V_{B2} = I_1 R_2$$

where

$$Z_2 = R_2.$$

A complete circuit design is given in Fig. 3(a) where, for the purpose of measurement, two terminating resistors R_s and R_5 are included. The resonant frequency ω_0 is given by[13]

$$\omega_0^2 = \frac{R_5}{C_1 R_2 C_3 R_4 R_s}$$

and direct substitution of the element values in Fig. 3(a) gives $\omega_0 = 10^4$ radians/second. This result was confirmed experimentally. The Q-factor of the circuit has been improved by employing two-transistor n-p-n–p-n-p pairs in place of TR1, TR2, and TR3. The measured Q-factor of Fig. 3(a) was 70. No attempt has been made here to optimize the design of this transistor realization and the bandwidth and harmonic distortion performance has not been investigated.

2.1. Q Enhancement

For many applications, the circuit in Fig. 3(a) is a satisfactory g.i.c. building-block for the synthesis of RC-active filters. However, in stringent high-Q applications it may be desirable to enhance the Q-factor[14, 15] of the circuit. This may be achieved by means of a large-valued resistor R_E (0·5–2 MΩ) connected from the base of TR2 to ground as shown dotted in Fig. 2(a). The use of this Q enhancement technique is easily derived. Consider the driving-point admittance at port 1, $Y_1(j\omega)$, in terms of the branch elements. Ideally, the nullor network gives

$$Y_1(j\omega) = -[\omega^2 C_1 C_3 R_2 R_4]/R_5 \qquad \ldots (1)$$

which, when resistively terminated at port 1, corresponds to the ideal infinite Q realization. In practice, the finite
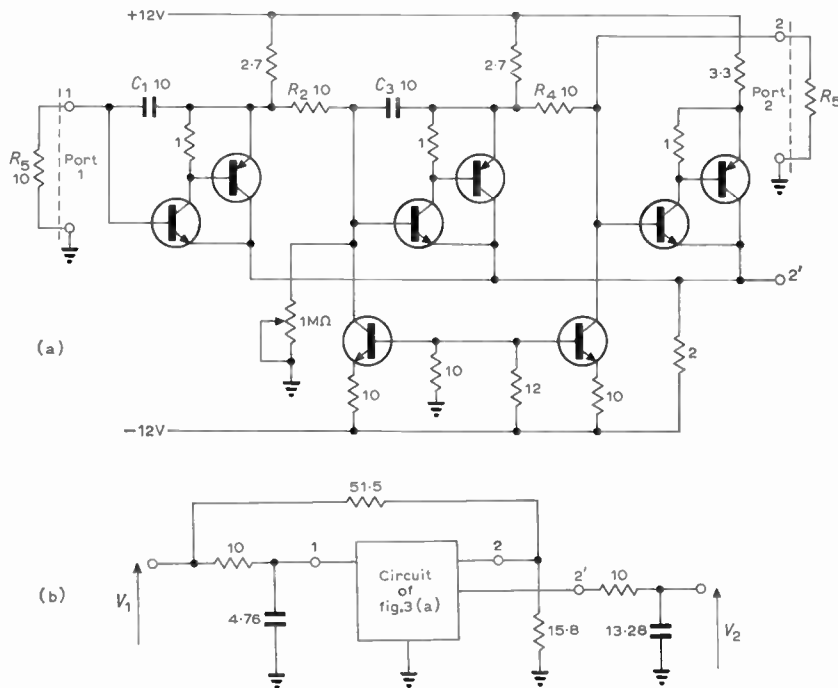
Fig. 3.

(a) An experimental g.i.c. realization.

(b) The basic elliptic transfer function realization (after Mitra[12]). (The resistor values are in kΩ and capacitor values are in nF.)

positive measured Q-factor, Q_0, corresponds to an additional effective shunt *positive* capacitance[13] at port 1 such that

$$Y_1(j\omega_0) = -\omega_0^2 C_1 C_3 R_2 R_4 \left[1 - \frac{j}{Q_0}\right]/R_5 \quad \ldots (2)$$

where ω_0 is the resonant frequency. Consider now the driving-point admittance obtained by adding the Q enhancement resistor R_E. Direct analysis of Fig. 2(a) gives

$$Y_1(j\omega_0) = \frac{-\omega_0 C_1 C_3 R_2 R_4}{R_5} \left[1 + j \frac{\omega_0 C_3 R_4 R_5}{R_E}\right] \quad \ldots (3)$$

Thus, by selecting

$$\frac{\omega_0 C_3 R_4 R_5}{R_E} = \frac{1}{Q_0}$$

we can arrange for R_E to introduce an effective shunt *negative* capacitance in equation (3) across port 1 that will exactly cancel the effective shunt positive capacitance given by equation (2). Thus, R_E is given by

$$R_E = Q\omega_0 C_3 R_4 R_5 \quad \ldots (4)$$

To enhance the measured Q_0 of 70, it is now a straightforward calculation to obtain R_E. Direct substitution into equation (4) gives an R_E of 700 kΩ. It was confirmed experimentally that a resistance R_E of this value enhanced the Q-factor to a value of several thousand.

3. Applications to RC Active Filter Synthesis

The proposed g.i.c. circuit may be used in any of the synthesis techniques in which a g.i.c. is the basic building block.[1-5] Apart from its simplicity, the proposed ransistor g.i.c. has an advantageous feature that is not shared by the class of operational amplifier realizations that have so far been used; that is, there exists a node terminal 2′, Fig. 1(b), which is connected to the conventional terminal 2 by a nullor and to ground via a

norator. Thus, terminal 2′ exhibits the same voltage as terminal 2 but, due to the norator, independently of the impedance between terminal 2′ and ground. In other words, terminal 2′ is a voltage source output from the g.i.c. and may be directly cascaded to another network without the requirement for buffer amplifiers.

A recently proposed g.i.c. cascade synthesis technique,[16] capable of realizing elliptic transfer functions, suffers from a practical disadvantage that interstage buffer amplifiers are required. Using the proposed transistor g.i.c., direct cascading of g.i.c.s is achieved. For example, suppose we wish to realize a low-pass elliptic transfer function by the cascaded-g.i.c. technique with the following characteristics:

pass-band ripple	0·4 dB
stop-band attenuation	30 dB
cut-off frequency	1590 Hz

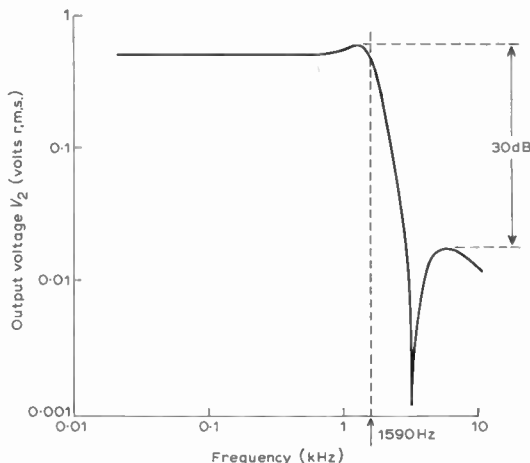**Fig. 4.** Measure transfer function.

A suitable normalized transfer function is given by

$$T(s) = \left[\frac{1+0 \cdot 194s^2}{0 \cdot 827s^2+0 \cdot 476s+1}\right] \left[\frac{1}{1+1 \cdot 328s}\right]$$

This may be achieved by cascading the basic g.i.c. structure with a first-order RC lag network. A practical circuit realization is given in Fig. 3(b) where terminal 2' is used to cascade directly with the RC lag network. The practical circuit has been denormalized to the appropriate cut-off frequency and is realized exactly as described in a previous paper.[16] The measured transfer function is given in Fig. 4 from which it is observed that the curve is in excellent agreement with the specified transfer function. For this particular low-Q design, the composite transistor pairs and Q enhancement are unnecessary refinements of the basic g.i.c. circuit.

4. Conclusions

A new transistor-g.i.c. circuit realization has been suggested that for many practical applications is preferable to the class of two operational amplifier g.i.c. realizations that have been previously employed. For applications in which high Q (greater than a few hundred) is not required and/or where a voltage source output terminal is desirable, the proposed g.i.c. offers a relatively simple circuit realization. In integrated circuit form, it is probably less complex than *one* operational amplifier. It has the additional practical advantage that it is stable at low frequencies and should be free from latch-up difficulties. The Q-factor may be enhanced to an arbitrarily high value by means of a simple resistive adjustment.

Finally, it will be observed that this g.i.c. realization corresponds to the ideal nullor network of Fig. 1(b) and that this latter network is also a three-nullor version of the three-operational amplifier 'biquad' or 'ring-of-three' circuits. Thus, the transistor-g.i.c. is also equivalent, in nullor form, to the three-operational amplifier realizations.

5. References

1. Antoniou, A., 'Gyrators using operational amplifiers', *Electronics Letters*, 3, pp. 350–2, August 1967.

2. Riordan, R. H. S., 'Simulated inductance using differential amplifiers', *Electronics Letters*, 3, pp. 50–51, February 1967.

3. Bruton, L. T., 'Network transfer functions using the concept of frequency-dependent negative resistance', *I.E.E.E. Trans. on Circuit Theory*, CT-16, pp. 406–8, August 1969.

4. Antoniou, A., 'Realization of gyrators using operational amplifiers and their use in RC-active network synthesis', *Proc. Instn Elect. Engrs*, 116, pp. 1838–50, November 1969.

5. Gorski-Popiel, J., 'RC-active synthesis using positive-immittance converters', *Electronics Letters*, 3, pp. 381–2, August 1967.

6. Bruton, L. T., 'Nonideal performance of two-amplifier positive-impedance converters', *I.E.E.E. Trans. on Circuit Theory*, CT-17, pp. 541–9, November 1970.

7. Bruton, L. T., 'Non-ideal performance of a class of positive immittance converters', *I.E.E.E. Trans. on Circuit Theory*, CT-16, pp. 572–4, November 1969.

8. Mitra, S. K., 'Analysis and Synthesis of Linear Active Networks', Chap. 11, pp. 492–3 (Wiley, New York, 1969).

9. Antoniou, A., 'New gyrator circuits obtained by using nullors', *Electronics Letters*, 4, pp. 87–9, 8th March 1968.

10. Braun, J., 'Equivalent NIC networks with nullators and norators', *I.E.E.E. Trans. on Circuit Theory*, CT-12, pp. 441–2, September 1966. (Letters).

11. Bendik, J., 'Equivalent gyrator networks with nullators and norators', *I.E.E.E. Trans. on Circuit Theory*, CT-14, p. 98, March 1967. (Letters).

12. Mitra, S. K., 'Equivalent circuits of gyrators', *Electronics Letters*, 3, pp. 333–4, July 1967.

13. Bruton, L. T., 'Frequency selectivity using positive impedance converter-type networks', *Proc. I.E.E.E.*, 56, pp. 1378–9, August 1968.

14. Orchard, H. J., 'Inductorless filters', *Electronics Letters*, 2 pp. 224–5, June 1966.

15. Holmes, W. H., Gruetzmann, S., and Heinlein, W. E., 'Direct-coupled gyrators with floating ports', *Electronics Letters* 3, pp. 46–7, February 1967.

16. Bruton, L. T., 'Biquadratic sections using generalized impedance converters', *The Radio and Electronic Engineer*, 41, pp. 510–12, November 1971.

Property of Nickel-powder Artificial Dielectric in Transverse Magnetic Field

A. N. DATTA, M.Tech., D.Phil.*

and

P. K. ROY, B.Sc., M.Tech.*

SUMMARY

This paper describes the experimental results on the non-reciprocal transmission property of an X-band waveguide filled with an E-plane slab of a new type of gyrotropic medium, namely nickel-powder artificial dielectric. Both attenuation and phase shift are found to indicate a non-reciprocal nature. The interesting transmission property of the device suggests its potentiality in the construction of some reciprocal and non-reciprocal devices, e.g., wideband variable attenuators and isolators. The material itself is found to have a loss tangent nearly an order of magnitude lower than that of the conventional artificial dielectrics. The results have further been interpreted theoretically from an exact solution of this loaded waveguide problem assuming a complex tensor susceptibility for the dielectric.

* *Institute of Radio Physics and Electronics, 92 Acharya Prafulla Chandra Road, Calcutta 9, India.*

1. Introduction

In the presence of a d.c. magnetic field, ferrites, semiconductors and gas plasma become anisotropic. In ferrites the permeability assumes a tensorial nature, while in the other two permittivity is a tensor quantity. Waveguides containing these materials exhibit non-reciprocal transmission properties and permit the realization of several non-reciprocal microwave components, e.g. isolators, circulators, attenuators and phase-shifters.[1,2] An artificial dielectric composed of ferromagnetic metal particles is another medium which becomes anisotropic in a d.c. magnetic field. It has been experimentally demonstrated that rotation of a plane polarized signal in Faraday configuration takes place in such a medium.[3] To assess the suitability of this cheaper form of anisotropic material in the construction of the above mentioned microwave components, some experiments have been performed on the transmission characteristics of an X-band waveguide filled with an E-plane slab of the material using a d.c. magnetic field parallel to the E-plane. The present paper describes the results of this investigation. An exact theoretical analysis of the problem has also been presented in support of the observed transmission property of the loaded waveguide.

2. Experiment

The artificial dielectric was prepared by mixing nickel powder (grain size 5μm) and paraffin wax in the volumetric ratio of 1 : 6.4. Care was taken to make the mixture a homogeneous one. Samples of different thicknesses were then prepared in rectangular plate form having dimensions of 76mm × 10.2mm (3in × 0.4in) with tapered ends. The transmission property of the sample-filled waveguide section was determined using a transmission bridge. The d.c. magnetic field was applied by
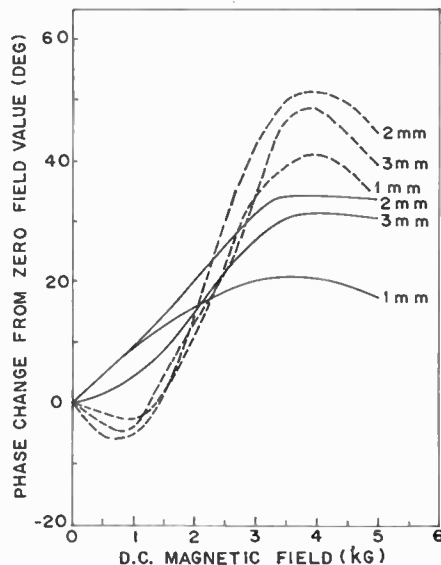

Fig. 1. Phase change from zero magnetic field value *vs.* magnetic field for different positions of the sample.
Frequency 8375MHz. Sample thickness 1·6 mm.
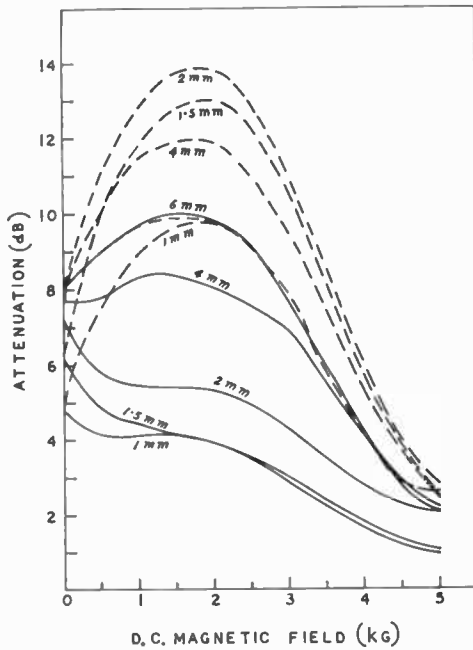—— forward direction ---- reverse direction.

Fig. 2. Attenuation as a function of the d.c. magnetic field for different positions of the sample.
Frequency 8375 MHz. Sample thickness 1·6 mm.
—— forward direction ---- reverse direction.

means of an electromagnet and the field was measured using a Hall probe calibrated against a n.m.r. probe. All the measurements were taken avoiding the hysteresis effect in the sample.

3. Results and Discussion

Figure 1 shows the nature of variation of phase with magnetic field from the zero magnetic field value for different positions of the sample 1·6 mm thick, while the attenuation as a function of the magnetic field has been plotted in Fig. 2. The experimental results may be summarized as follows.

The phase shifting property of the sample-filled section is non-reciprocal in nature and is dependent on the position of the sample. There is an optimum sample position for maximum non-reciprocity in phase. However, the differential phase shift for the two directions of the magnetic field is only a small fraction of the actual phase shift at zero field value which is found to be more than 200°.

The attenuation characteristics exhibit a more pronounced non-reciprocal behaviour. In the reverse direction it resembles the characteristics of a resonant circuit with the peak occurring at a magnetic field of nearly 1·75 kilogauss. It may be mentioned here that the ferromagnetic resonance in nickel particles comprising the artificial dielectric occurs for X-band signal in this region.[4] One can, therefore, identify the loss in the reverse direction with the ferromagnetic resonance loss in the nickel particles, and employ this loss characteristic in the study of the ferromagnetic resonance phenomenon in metal in powder form. It should be recalled that the standard method for observing this phenomenon is to study the property of a resonant cavity loaded with flat specimens.[5]

The peak value of attenuation at resonance depends strongly on the position of the sample and there exists an optimum position for which the highest attenuation is observed. In the forward direction, however, the sample behaves simply as a magnetic field dependent lossy material where the attenuation increases with its depth of insertion.

To study the amount of isolation of power-flow attainable in this configuration the ratio R of reverse attenuation to forward attenuation has been plotted in Fig. 3 as a function of the magnetic field using the sample position as a parameter. It can be seen that there is an optimum sample position for the highest non-reciprocal effect. To optimize the isolation the experiment was repeated using three sample thickness (1 mm, 1·27 mm and 1·9 mm). Experimental data on the reverse-to-forward attenuation ratio R are also shown in Fig. 3, for optimum position of these samples. The highest figure (nearly 6) for the above ratio was observed with the thinnest sample. The optimum sample position was found to move away from the nearest sidewall with decreasing thickness.

The constancy of the reverse-to-forward-attenuation ratio over a large magnetic field variation suggests a large bandwidth of operation for the device. Experiments were performed at three frequencies, 8375 GHz, 8930 GHz and 9700 GHz, with the thinnest sample for three positions and the 3 dB bandwidth for optimum condition was found to be more than 1·5 GHz.

Besides the application as a broad-band isolator described above, the loaded waveguide structure may work as an electrically variable microwave attenuator. To examine the performance of these components a plot of phase change v. attenuation has been made for different positions of the sample (Fig. 4). It is observed that under optimum condition the attenuation changes from 2·5 dB to 10·5 dB with a phase change of only 8 deg.
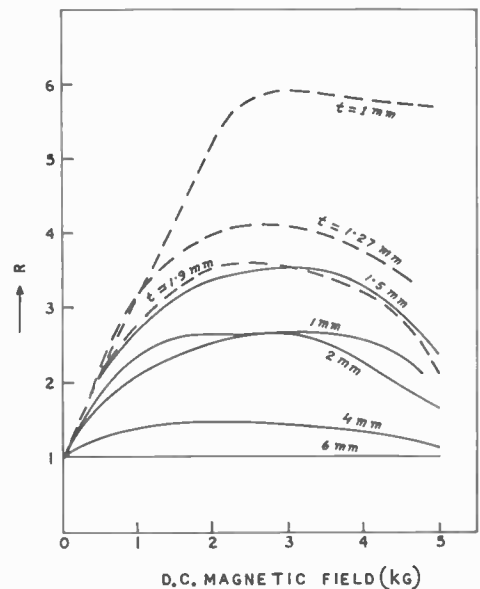

Fig. 3. Reverse-to-forward-attenuation ratio as a function of the d.c. magnetic field. Solid curves are the results for different positions of the sample 1·6 mm thick, the dotted curves are for different sample thickness. Frequency 8375 MHz.
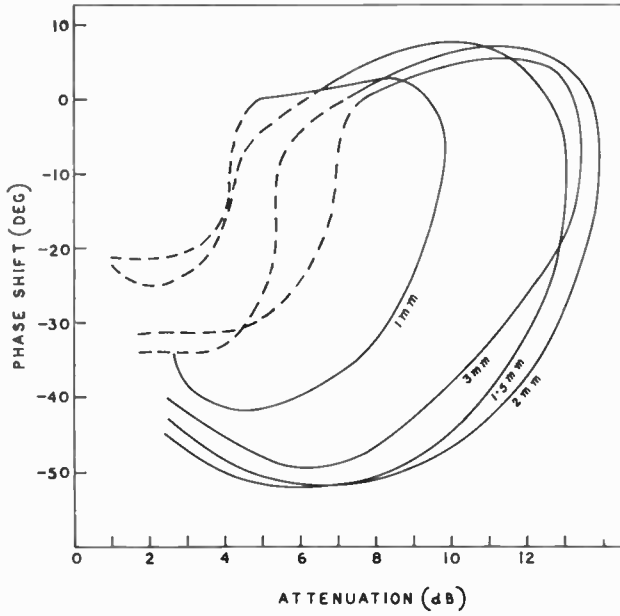
Fig. 4. Phase shift v. attenuation for different positions of the sample 1·6 mm thick. Frequency 8375 MHz.

Larger ranges of attenuation variations have been realized using higher concentration of metal particles in the artificial dielectric. Although the insertion loss of these components appears higher compared to those using ferrites, these components have marked advantages in respect of bandwidth.[6]

The property of the nickel-powder artificial dielectric in a transverse magnetic field is attractive so far as its losses are concerned. The experimental data indicated a reduction in losses by an order of magnitude with the application of a high d.c. magnetic field. The total loss tangent at a field of 7 kilogauss was found to be only 0·009 as compared to a figure of nearly 0·09 for an artificial dielectric using the same concentration of non-magnetic materials like copper, aluminium and zinc.[7] The nickel powder artificial dielectric with a polarizing d.c. magnetic field will, therefore, be suitable for applications where low-loss high dielectric constant materials are required.

4. Theoretical Interpretation

The non-reciprocal attenuation and phase-shifting property of the waveguide loaded with artificial dielectric can be explained assuming the susceptibility of the medium to be complex and tensorial of the form given by

$$\chi = \begin{bmatrix} \chi_{xx} & \chi_{xy} & 0 \\ -\chi_{xy} & \chi_{yy} & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \dots (1)$$

where $\chi_{xx} = \chi_{yy}$.

The different elements of this tensor can be determined through a knowledge of the effective scalar permeability $\mu_{m\pm}$ of the medium to circular polarizations using the relations

$$\mu_{m\pm} = 1 + \chi_{\pm} = \mu'_{m\pm} - j\mu''_{m\pm} \qquad \dots (2)$$

and

$$\chi_{\pm} = \chi_{xx} \pm j\chi_{xy} \qquad \dots (3)$$

where $\mu_{m\pm}$ and $\mu^{*}_{m\pm}$ are respectively the real and imaginary part of the effective scalar permeability $\mu_{m\pm}$ and $\chi_{\pm}$ is the effective scalar susceptibility.

The effective permeability is related to various parameters of the ferromagnetic metal particles and the host medium. Based on the concept of effective permeability and permittivity of a mixture as given by Lewin[8] the expression for the effective permeability of nickel-powder artificial dielectric biased with a d.c. magnetic field has been obtained by Datta, Engineer and Nag[9] in the following form

$$\mu'_{m\pm} = \mu_1 \frac{\begin{aligned}&\{c\mu_{i\pm}(1+2f)+\mu_1 2(1-f)\} \\ &\times \{c\mu_{i\pm}(1-f)+\mu_1(2+f)\} \\ &\quad + c^2\mu^2_{r\pm}(1+2f)(1-f)\end{aligned}}{\{c\mu_{i\pm}(1-f)+\mu_1(2+f)\}^2+c^2\mu^2_{r\pm}(1-f)^2} \qquad (4)$$

$$\mu''_{m\pm} = \frac{9fc\mu^2_1\mu_{r\pm}}{\{c\mu_{i\pm}(1-f)+\mu_1(2+f)\}^2+c^2\mu^2_{r\pm}(1-f)^2} \qquad (5)$$

where f is the fractional volume of the particles in the mixture

μ_1 is the specific magnetic permeability of the host medium

$$c = \frac{\lambda\sqrt{\omega\varepsilon_0}}{\pi r\sqrt{\sigma}}$$

λ, ω are the free space wavelength and angular frequency

r is the radius of the particle

σ is the conductivity of the metal particles

$$\mu_{r\pm} = \frac{1}{\sqrt{2}}[\sqrt{\mu''^2_\pm+\mu'^2_\pm}+\mu''_\pm]^{\frac{1}{2}}$$

$$\mu_{i\pm} = \frac{1}{\sqrt{2}}[\sqrt{\mu''^2_\pm+\mu'^2_\pm}-\mu''_\pm]^{\frac{1}{2}}$$

$\mu_\pm$, $\mu''_\pm$ are respectively the real and imaginary part of scalar permeability of nickel.

The propagation constant Γ through the waveguide containing an E-plane slab of the artificial dielectric with a tensorial susceptibility described by equation (1) can be determined from the solution of the following transcendental equation:[10]

$$\tfrac{1}{2}\left(\frac{k_a^2}{\rho^2}-\frac{\Gamma^2}{\theta^2}-k_m^2\right)\cos k_a(L-\delta-2a)+\frac{\Gamma k_a}{\rho 0}\sin k_a(L-\delta-2a)+$$

$$+\tfrac{1}{2}\left(\frac{k_a^2}{\rho^2}+\frac{\Gamma^2}{\theta^2}+k_m^2\right)\cos k_a(L-\delta)+$$

$$+\frac{k_a k_m}{\rho}\cot(k_m\delta)\sin k_a(L-\delta)=0 \qquad \dots (6)$$

where

$$\rho = \frac{1+\chi_{xx}}{(1+\chi_{xx})^2+\chi^2_{xy}}$$

$$0 = \frac{1+\chi_{xx}}{\chi_{xy}}$$

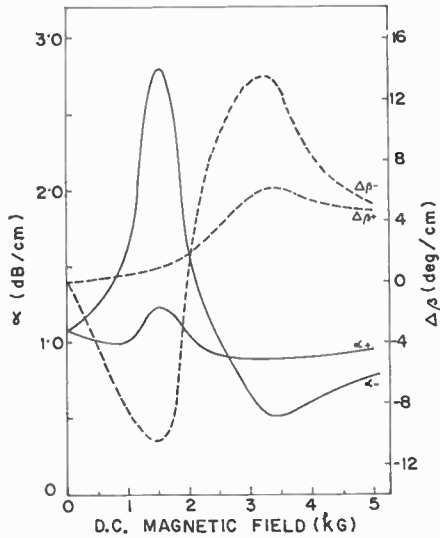L is the internal width of the waveguide

Fig. 5. Theoretical curves for attenuation constant and change in phase constant from zero field value *vs.* d.c. magnetic field. The ± sign has been used to specify the forward and reverse direction of propagation.

a is the distance of the slab from the nearest sidewall

δ is the thickness of the slab

$$k_a^2 = \omega^2 \varepsilon_0 \mu_0 + \Gamma^2$$

$$k_m^2 = \omega^2 \varepsilon \mu_0 \left[\frac{(1+\chi_{xx})^2 + \chi_{xy}^2}{1 + \chi_{xy}} \right] + \Gamma^2$$

ε_0, μ_0 are the free space permittivity and permeability

ε is the permittivity of the slab.

Because of the presence of a linear term in Γ in equation (6) it will have two distinct solutions for the two directions of propagation.

The equation has been solved graphically using the computer data of real and imaginary part of l.h.s. of the equation (6) for different chosen values of α and β.

The equation was solved under the following conditions: $L = 2 \cdot 286$ cm, $a = 0 \cdot 2$ cm and $\delta = 0 \cdot 16$ cm. The susceptibility tensor was obtained using the following values: $c = 10$, $f = 0 \cdot 15$ and $\mu_1 = 1$. The r.f. permeability of nickel as a function of d.c. magnetic field was

assumed to follow the characteristics of ferrite. The results of the theoretical analysis have been presented in Fig. 5. The general nature of these curves are found to be identical to those obtained experimentally.

5. Conclusion

The present communication describes some interesting reciprocal and non-reciprocal transmission properties of waveguides loaded with artificial dielectric. As a non-reciprocal device neither the reverse attenuation nor the reverse-to-forward attenuation ratio are satisfactory for the present moment. It is, however, expected that with proper choice of particle size, concentration, sample thickness and with dielectric loading these figures may be improved.

6. Acknowledgment

7. References

1. Lax, B. and Button, K. J., 'Microwave Ferrites and Ferrimagnetics'. (McGraw-Hill, New York, 1962.)

2. Kuno, H. J. and Hershberger, W. D., 'Solid-state plasma controlled non-reciprocal microwave devices', *I.E.E.E. Trans. Microwave Theory and Techniques*, MTT-15, No. 1, p. 57, 1967.

3. Engineer, M. H., Datta, A. N. and Nag, B. R., 'Microwave Faraday rotation in nickel-powder artificial dielectrics', *J. Appl. Phys.*, 38, No. 2, p. 884, February 1967.

4. Bozorth, R. M., 'Ferromagnetism'. (Van Nostrand, New York, 1951.)

5. Bloembergen, N., 'On the ferromagnetic resonance in nickel and supermalloy', *Phys. Rev.*, 78, p. 572, 1950.

6. Reggia, F., 'Amplitude and phase modulators in rectangular waveguides for 5 to 7 Gc/s', *I.E.E.E. Trans. Microwave Theory and Techniques*, MTT-14, No. 3, p. 154, March 1966.

7. Kelly, J. M., Stenoien, J. O. and Isbell, D. F., 'Waveguide measurements in the microwave region on metal powders suspended in paraffin wax', *J. Appl. Phys.*, 24, p. 258, 1953.

8. Lewin, L., 'Electrical constants of a material loaded with spherical particles', *J. Instn Elect. Engrs*, 94, pt. III, p. 65, 1947.

9. Datta, A. N., Engineer, M. H. and Nag, B. R., 'Microwave Faraday rotation in nickel-powder artificial dielectrics—a suggested explanation', *J. Appl. Phys.*, 38, No. 6, p. 2422, May 1967.

10. Ref. 1, Chap. 9, p. 358.

A 16-channel Digital Acoustic Telemetry System

D. CATTANACH*

Based on a paper presented at the Conference on Electronic Engineering in Ocean Technology, held in Swansea from 21st to 24th September 1970

SUMMARY

The system has been specifically developed for the real-time measurement of the physical and geometric characteristics of full-scale fishing gears and consists of:

(a) Instruments attached to various positions on the gear for the measurement of tensile loads, distances, vertical heights, speed and depth.

(b) Encoding and telemetry equipment which interrogate the instruments in sequence and transmit data by acoustic link to the listening ship.

(c) Shipboard signal receiving equipment which processes received data for display visually on line printer, or through a computer interface.

Information is transmitted over the acoustic link in an unambiguous way as each digital data word contains a specific reference to the source instrument. A degree of accuracy within 0·5% overall is achieved.

* Marine Laboratory, Department of Agriculture and Fisheries for Scotland, Aberdeen AB9 8DB.

1. Introduction

The fishing gear research engineer is faced with the problem of measuring various features of fishing gears and retrieving data from underwater instruments under arduous working conditions. Techniques available for this include self-recording instruments[1,2] which do not convey real-time information, and an umbilical-type system (separate cable) which is difficult to use in practice, particularly in bottom trawls, and further is expensive and has limited versatility. As an alternative to these, the acoustic link provides real-time data without the need to have special cables and associated machinery.

The advent of integrated circuitry has now opened up a new concept in underwater instrumentation technology, and has made possible the design of complex miniature systems which would not have been practical only a few years ago.[3] This has led to the development of an underwater telemetry system by the Marine Laboratory's Gear Research and Engineering Unit. The system has been designed specifically for work on trawl fishing gear, and is a development from an earlier single-channel instrument[4] to accommodate a maximum of 16 measurement channels.

By using a coded pulse duration modulated carrier, information can be transmitted from the fishing gear to the towing vessel without any degradation whatsoever. Further, a higher measuring resolution can be achieved and a more reliable link established, since readability is good with a poor signal/noise ratio.

Figures 1(a) and (b) show a typical arrangement of instruments on a trawl net with the control and transmitter unit, Fig. 1(c), located on the centre of the headline. Each instrument is connected by cable to the control unit, which transmits coded instrument data via the acoustic link to the towing vessel (Fig. 2). The shipboard receiving equipment is capable of operating on a signal which fades over a 40 dB range and of discriminating against spurious noise caused by multi-path effects etc. The receiving equipment decodes the signal and outputs to a digital read-out unit, giving a visual display of channel number and data value. Permanent records are made on an on-line printer, but the signal may also be connected to an on-line computer for further real-time processing. The specification for the system is shown in Table 1.

Table 1. Specification for underwater telemetry system

(a) 16 data channels scanned sequentially or by pre-programmed selections.

(b) Maximum information rate 3 channels per second.

(c) Digital accuracy 0·1%.

(d) Maximum working range 1500 metres.

(e) Parameters to be measured:
   (i) Tensile loads of 0–5 tons.
   (ii) Vertical heights 0–10 metres.
   (iii) Speed 0–5 knots.
   (iv) Depth 0–400 metres.
   (v) Distance 0–30 metres.

The overall accuracy is dependent on the individual accuracy of each instrument, this being better than 1% of the f.s.d. in the worst cases (speed and depth).
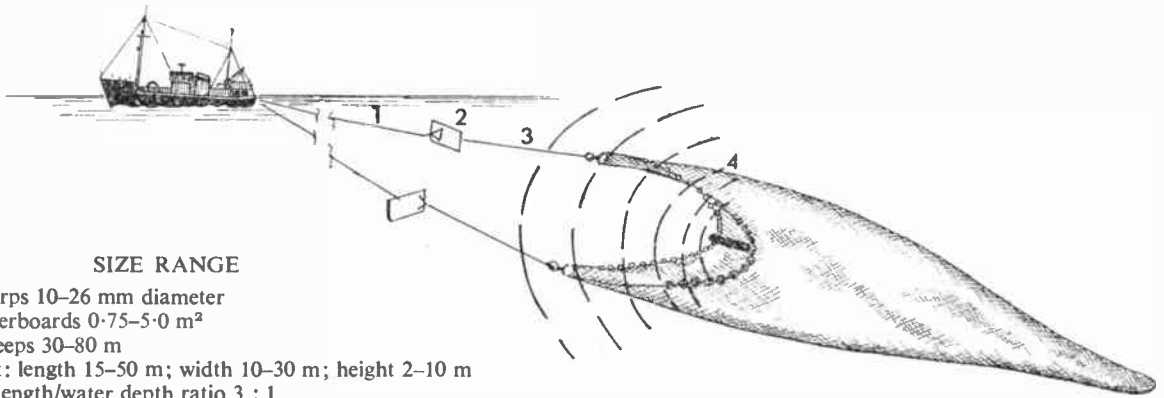
KEY   SIZE RANGE
1 Warps 10–26 mm diameter
2 Otterboards 0·75–5·0 m²
3 Sweeps 30–80 m
4 Net: length 15–50 m; width 10–30 m; height 2–10 m
Warp length/water depth ratio 3 : 1
Typical fishing depth 30–500 m

(a) Simplified diagram of fishing gear.
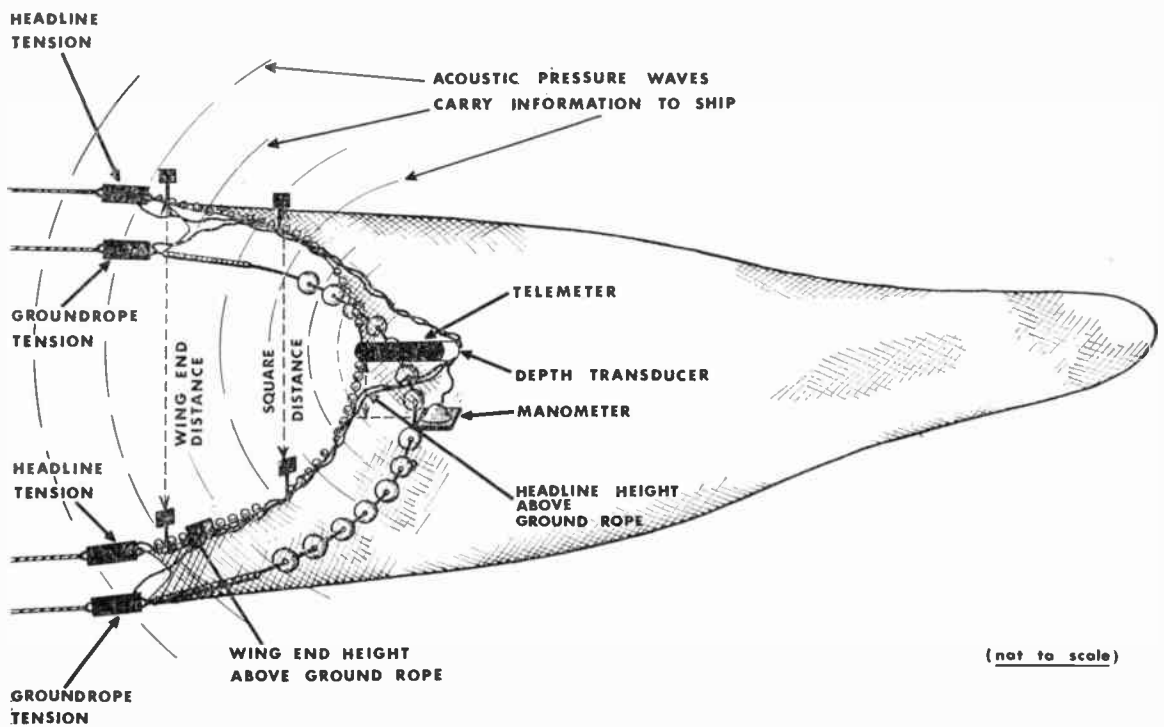
(b) Arrangement of instruments on fishing net.

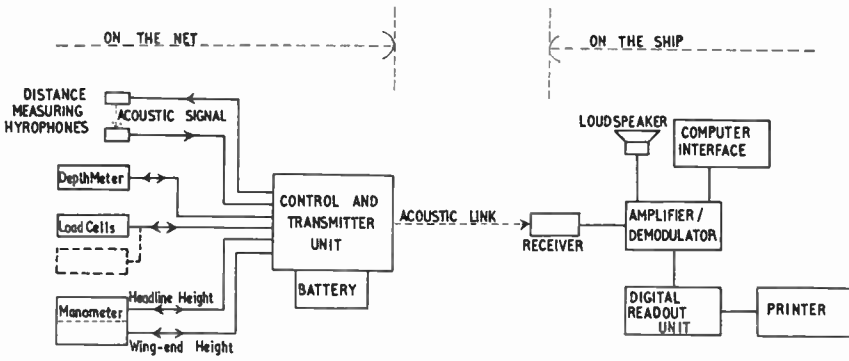**Fig. 1.**   (c) Underwater control and transmission unit with housing.

Fig. 2. Block diagram of net telemeter system.

2. The Underwater Control and Transmitter Unit

Figure 3 shows a simplified block diagram of the underwater transmitter unit. This communicates with the instruments on the net by connecting cables going to either the power and signal separator or to the distance channel multiplexer. The unit controls timing and sequencing of all measurements made, codes the data and transmits the results over the acoustic link to the ship.

2.1. Encoder Unit

The unit has two modes of operation control—the measurement cycle and the transmission cycle. At the beginning of each measurement cycle the channel counter

is advanced one count, except that at the end of the full sequence (16 or less) of measurement the counter is reset to zero. The decode of this count is used to select the appropriate multiplexer driver, thus selecting the channel to be measured. At the same time, a coded number representing that channel is set up on the channel address lines which preset the channel identifier digit (4 bits) within the encoder master counter. The system is now ready to perform a measurement. Halfway through the cycle $\bar{P}$ goes true, initiating the interrogator and firing the distance transmitter. The interrogator controls only those instruments having an analogue output (load cells, etc.), while the distance transmitter relates to the measurement of distance using hydrophones (see Sect. 2.3).
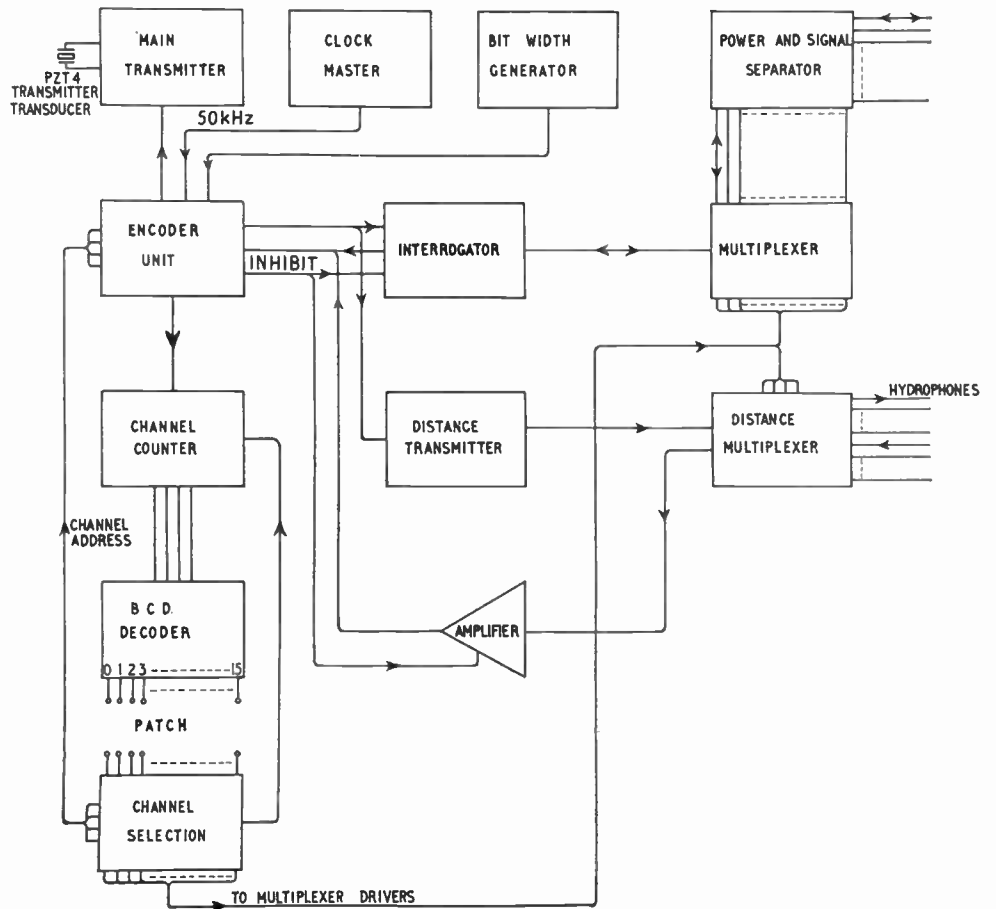

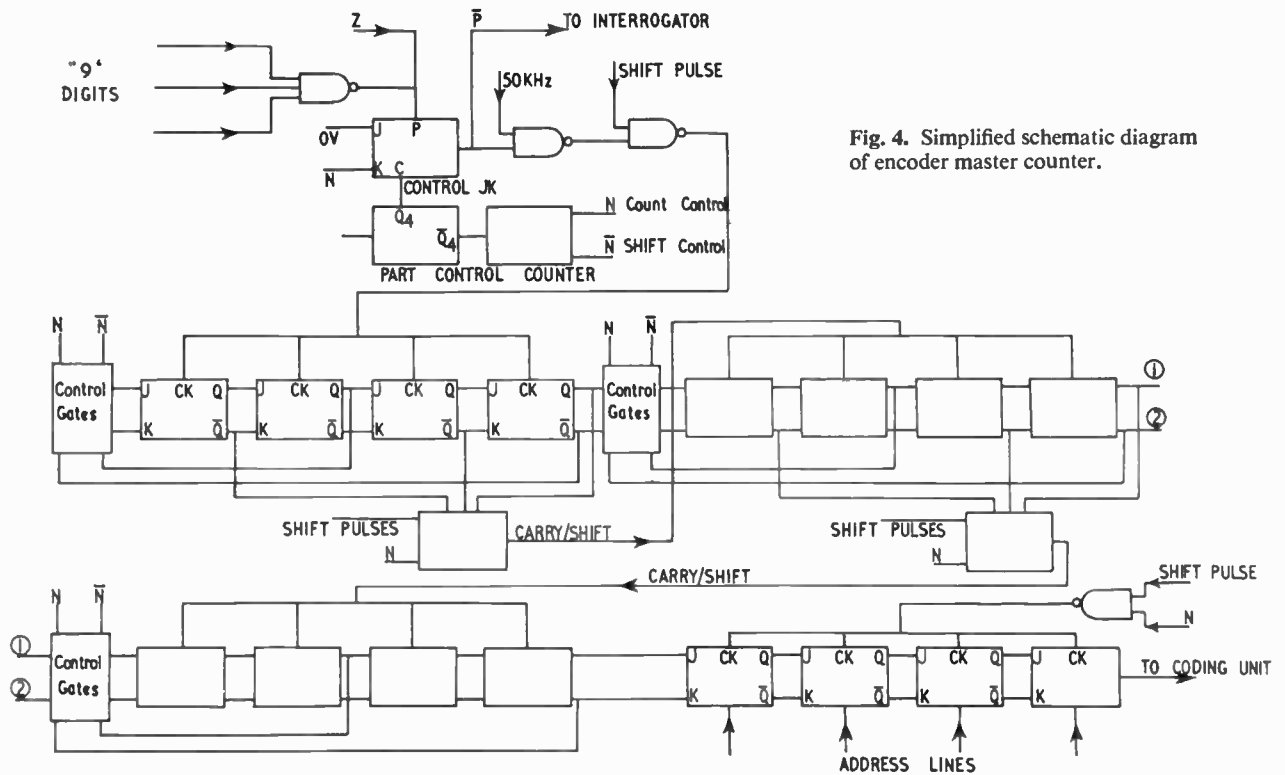Fig. 3. Simplified block diagram of underwater control and transmitter unit.

Fig. 4. Simplified schematic diagram of encoder master counter.

2.1.1. Measurement accuracy

As in any digital frequency meter or counter the encoder measures the period between pulses by counting the number of pulses produced from a highly stable and accurate frequency source, normally termed the 'clock'. The timing accuracy is defined by the maximum count obtainable since the resolution will always be ± 1 count assuming the inaccuracy of the clock to be several orders better than the staticizing error.

A three-decade count will meet the required specification (Sect. 1) giving an accuracy of $\pm 0.1\%$, i.e. ± 1 count in 1000. Clocking with a period of 20 µs fills the counter in 20 ms, this being the time for an acoustic pulse to travel a distance of about 30 m in water. This is a convenient figure since the clock frequency can be used to provide the carrier for transmission. In this application 50 kHz is quite suitable, being outside the worst part of the sea-noise spectrum and yet low enough not be be attenuated to any appreciable extent (approximately 3 dB/100 m).

2.1.2. Circuit description

Each decade consists of four JK flip-flops wired to give a divide-by-10 function (see Fig. 4). The output of each decade is connected to the input of the next through a gating network. By controlling these gates, the decades work as a synchronous counter (measurement cycle) or as a serial shift register (transmission cycle). An additional four JK flip-flops hold the channel identifier digit preset from the channel address lines. The logic separates these last flip-flops from the counting decades during the measurement cycle, but connects them into the long shift register during the transmission cycle. Over a complete

cycle (measurement followed by transmission), the sequence of operation is as follows.

The control counter line N sets the control gates, which in turn sets the decades to the counting model. The control counter signal Q_4 now set the control JK output to one, opening the clock gate and initiating the interrogator at the same time. The counter proceeds to clock 20 µs intervals. On receipt of a return signal, either from the interrogator or the distance amplifier, line Z presets the control JK closing the clock gate. At some time later N switches changing the decades into the shift register mode, gating in the shift pulses and suppressing the 'carry'. The number held in the register is now shifted out serially into the PDM (pulse duration modulator) unit, starting with the channel identifier digit whose 4 bits were previously loaded with the code for the appropriate channel number during the measurement cycle. After 16 shift pulses the register is empty and N reverts to the counting mode. In the event of a malfunction of any channel whereby no return signal is received within the allowed measurement of 20 ms, a full house is detected and the control JK reset. The signal transmitted in these circumstances would be the channel number followed by 999. This ensures the counter always starts empty for the next measurement, and the failure of an instrument to respond to interrogation will not result in an arbitrary data value.

2.2. Transmitter Data Format

2.2.1. Channel selection

The 16 channels are allocated specifically, in that four must be used for hydrophones (distance measurement, two hydrophones per channel), and twelve for instru-
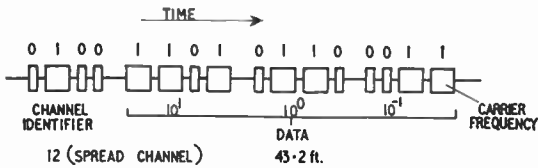
Fig. 5. Typical data transmitter output.

ments having transducers with analogue output (load cells, etc).

The patchboard used for channel selection permits the choice of any combination to be measured, for example, four load measurements—two distance measurements—and one speed measurement in any order. This occupies seven channels. At the end of the seventh channel the channel counter can be reset back to 1 by patchboard connexion thus preventing the system scanning the unused nine channels.

2.2.2. Channel speed

The operational speed of the system is determined by the bit-width (Bw) generator. This is a square-wave generator from which the basic digital format (pulse length modulation) is derived. The data from one measurement is presented as a serial word containing four characters. The first character designates the channel number, and the remaining three the data value. Each character is made up of four pulses or bits. The bit length determines the binary state, i.e. a binary '1' is a pulse three times as long as a '0'. Each bit is separated by a space the same length as a '0' pulse, and each character separated by a space three times as long as this. A typical format is given in Fig. 5, and the code used in Table 2.

Although the bit-width can be set accurately, the data rate is dependent on the decoded value being presented, but approximately a bit length of 20 ms for a binary '0' gives a data rate of about one channel per second.

The band-width allowed by the acoustic properties of the sea is the limiting factor regarding maximum information rate, but three channels per second have been achieved in some experimental trials with success, although recording may present problems at this speed unless special equipment is used.

Table 2. B.c.d. as used in telemetry system

Decimal equivalent				
0	0	0	0	0
1	0	0	0	1
2	0	0	1	1
3	0	1	1	0
4	1	1	0	1
5	1	0	1	1
6	0	1	1	1
7	1	1	1	0
8	1	1	0	0
9	1	0	0	0
10	1	0	1	0
11	0	0	1	0
12	0	1	0	0
13	0	1	0	1
14	1	0	0	1
15	1	1	1	1

2.2.3. Data format logic

A square wave which is not synchronized with the 50 kHz clock determines the bit width. From this waveform the format of the transmitted digital word is produced in the following manner, referring to Figs. 6 and 7.

The basic digit of four bits is generated by a two-input NAND gate G1, a binary counter (the control counter), an Exclusive OR, and a bistable. The control counter determines the number of bits per digit and digits per word, and controls the logic accordingly. Assume for the moment that G2 acts as an invertor only (which is the case when the data are all zeros). Initially $\bar{Q}_2$ and Q_3 are true, the latter allowing the bit-width square wave (Bw) to gate through G1 clocking the control counter. On the second negative edge $\bar{Q}_2$ changes state, setting the bistable clock line false (via the Exclusive OR) which toggles Q_3. Q_3 now being true allows the clock line to be set true by Bw, but since Q_3 is false, this pulse is inhibited by G1. The next negative excursion of Bw sets CL false, resetting Q_3. Q_3 now being true opens G1 allowing Bw to

Fig. 6. Simplified diagram of format generator.
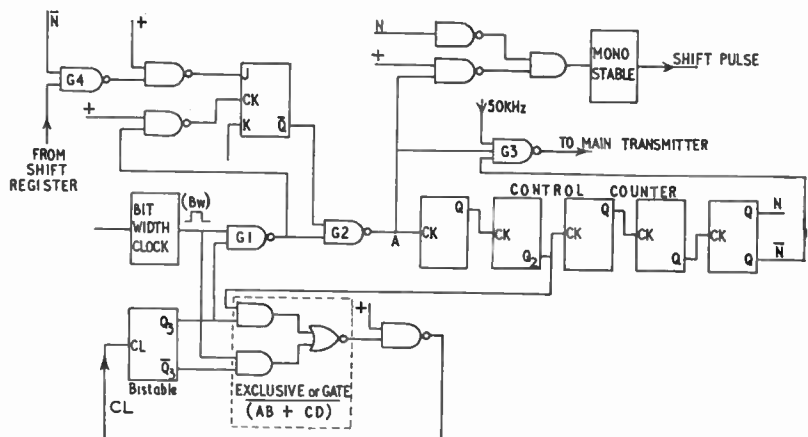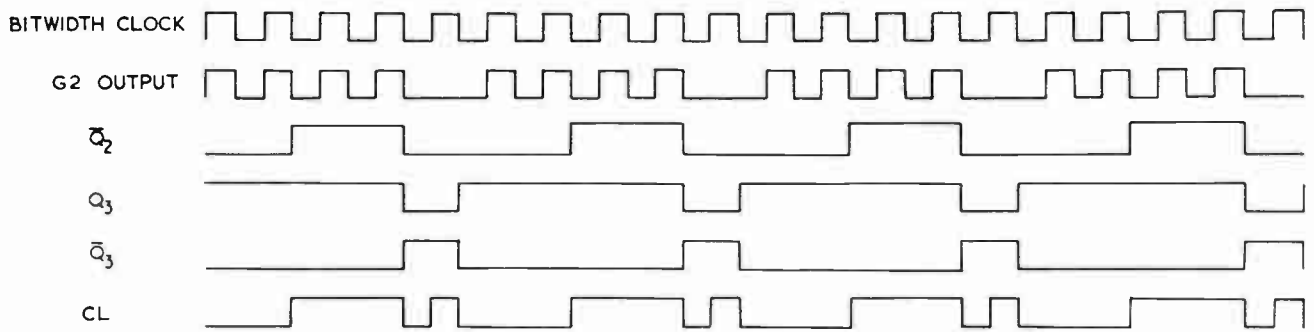
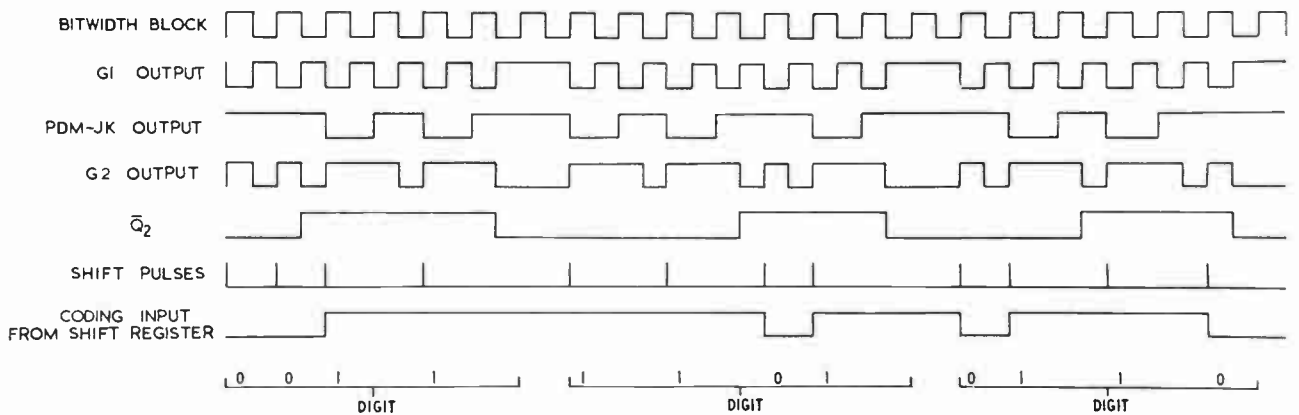(a) Waveform diagram when PDM–JK input is false.

Fig. 7. (b) Typical waveform diagram for normal operation.

clock the control counter once more. After two counts $\bar{Q}2$ goes true, resetting CL true and bringing the cycle back to the start again. Hence a group of four pulses, each separated by one pulse length and the group terminated by a space of 3 pulse lengths, have been generated. This signal is further gated by G3 from a waveform produced by bit 5 of the control counter, the resultant being a group of four digits spaced from the next group by a time interval of 16 basic pulse lengths. The 50 kHz being fed to G3 is modulated by the final PDM signal before being applied to the main transmitter.

2.2.4. PDM circuitry

The number held in the 3-decade counter and channel identifier register is serially shifted into the coding unit when the logic switches to transmission mode. If the input to G4 (see Fig. 6) is false this sets the PDM JK output to 1 and has no influence on the bit-width being produced at point A. Consider the case when the input of G4 is true. This sets the PDM JK to a bistable which toggles when clocked by the Bw pulse, setting the output of G2 true. This is held in this state until the end of the next Bw pulse when it reverts to a low state. So the shifting of a '1' into the PDM unit causes the Bw pulse to be stretched a factor of 3.

Since the control counter is triggered from the PDM unit gate, G2, it will count the pulses irrespective of the length, so the basic format of 4 bits to a digit and 4 digits to a word will be maintained. The timing of this sequence may be more easily appreciated by noting that the shift pulses are synchronous with the clocking pulse to the control counter.

Figure 7(b) shows an idealized set of waveforms when the example data shown are shifted into the coding unit from the shift register.

2.3. Distance Measurement

The measurement of distance is achieved by measuring the time for an acoustic pulse to travel between two omnidirectional hydrophones. Since the basic clock frequency is 50 kHz the resolution obtained by the counter is ± 3 cm assuming the velocity of sound in water to be approximately 1500 metres per second.

2.3.1. Transmitter

The distance transmitter provides 750 V from a stabilized static convertor used to shock-excite the transmitting hydrophone. Two thyristors in tandem switch a charged capacitor on to the hydrophone. The thyristors switch off when the capacitor current falls below the 'hold-on current' of these devices. This technique ensures a low impedance drive and can supply appreciable power to the hydrophone, even in the event of (limited) cable leakage.

Measurements made of the hydrophone exciting current have shown peak values in the region of 15 A with a rise time of a few microseconds. It can therefore be appreciated that careful screening and earthing is of paramount importance to prevent radiation and earth loop noise arising from the distance transmitter circuit.

2.3.2. Distance receiver

The signal picked up on the receiving hydrophone is amplified and applied to the gate of a high-speed thyristor which provides a fast rise-time pulse for the closure of the 'end of count' gate.

The receiving amplifier and interrogator return line is inhibited for 200 µs after P̄ goes true. This prevents possible premature closure of the 'end of count' gate due to system noise associated with firing of the distance transmitter.

2.4. *Main Transmitter*

The main transmitter is a push-pull Class B stage of conventional design developing 10 W r.m.s. into a pair of piezo-electric transducers in parallel. These produce a fantail beam 60° × 30° to the half-power points, the larger angle being in the horizontal plane. This beam shape allows for the skewing effects that may occur on the fishing gear, and reduces the problem of multi-path effects,[5] since a narrower vertical beam reduces the intensity of surface and bottom reflexions.

The transducers are made from PZT5 disks which are mass-loaded to lower the mechanical resonance to 50 kHz. Half-wave resonance mode is utilized by loading with $\lambda/4$ steel backing plates in a sandwich construction. In order to maintain a backing of low acoustic impedance necessary for the successful operation of this mode, the transducer assembly is potted in a mix of epoxy resin and micro-miniature hollow glass spheres. The glass spheres are a by-product of furnace fly-ash collected by electro-precipitators. Only spheres with specific gravity of less than 0·6 are used in the potting. This mix has proved most successful, withstanding pressures of 1000 lb/in² (70 kgf/cm²) with no water ingression.

2.5. *Interrogator*

The interrogator acts as a line driver providing a low output impedance pulse of sufficient amplitude to trigger the selected measuring instrument. After being routed by the multiplexer the pulse enters a filter network which provides isolation from the main power lines, since the instrument power and signals are shared by the same conductors in the connecting cable.

At the instant P̄ goes true a 50 kHz signal from the crystal clock is gated into the encoder's three digit (4 bits per digit) decimal counter.

Circuitry within each instrument compares the voltage output of the measuring transducer with a ramping voltage, which is started on the receipt of the interrogation pulse. When the two are detected to be equal, a pulse is generated which returns along the connecting cable to the interrogator and activates the 'end-of-count' gate via line Z. The count held at this point is therefore proportional to the instrument transducer signal, and hence is the measurement required.

3. Instruments

The instrument array on some of the larger trawls sometimes requires cable runs of 50 metres in length. Under normal circumstances it would be quite un-

desirable to convey low-level analogue signals from transducers along such lengths of runs. In underwater applications especially, where all cables tend to suffer in time from insulation degradation, it would be most unwise indeed to do this. For this reason all instruments with primary analogue output used in the system include circuitry to convert this to a time analogue (proportioned pulse delay) so that communication over the connecting cable is by high-level pulses.

Interconnecting cables have a high failure rate in any system, as they are exposed to rough treatment, and require terminations such as plugs and sockets. In order to improve reliability, the number of conductors are kept to a minimum by superimposing the pulse signals on the instrument power lines, separating these at either end with high- and low-pass filters respectively.

3.1. *Voltage-to-time Convertor*

The block diagram of Fig. 8 shows the standard circuit used in each instrument.

The analogue voltage produced by the measuring transducer, normally a strain gauge bridge, is amplified by the differential comparator. When a negative-going edge is transmitted down the connecting cable, the flip-flop toggles, releasing the clamp on capacitor C1. The voltage on C1 then ramps as it is fed from a constant-current generator. When the ramping voltage equals the voltage being produced from the amplifier, the comparator changes state, firing the pulse generator which sends a pulse back up the connecting cable (power line). The pulse also resets the bistable and clamp, bringing the circuit back to its initial state. By controlling the constant current the slope of the voltage ramp can be adjusted to alter the calibration of the instrument.

Standardizing on such a circuit simplifies instrument design since it is only necessary to convert the parameter to be measured into an analogue voltage, using standard pressure or other transducers.

3.2. *Outline of Instrument Techniques*

Tensile loads are measured by developing a proportional hydraulic pressure in a closed chamber, through the action of a piston on a hydraulic fluid. This technique has been successfully used in a number of commercially-available load cells. The special merit of such a system is its inherently low hysteresis and freedom from temperature effects, provided that the displacement of the fluid in operation is small. The effect of temperature on the properties of hydraulic fluid is, in general, quite negligible. The pressure is monitored using a CEC Type 4-326 high-quality strain gauge pressure transducer, which requires a much smaller fluid displacement than, for example, a Bourdon gauge. During laboratory tests changes of only 9 kg (20 lb) were detectable on a static load of 3½ tonnes, demonstrating the sensitivity of the instrument. Quite possibly a better figure than this is obtainable, but very sensitive calibration equipment would be necessary to show this.

Vertical heights are measured using a hydrostatic manometer. The pressure developed by a head of low viscosity silicone oil is monitored using a CEC strain
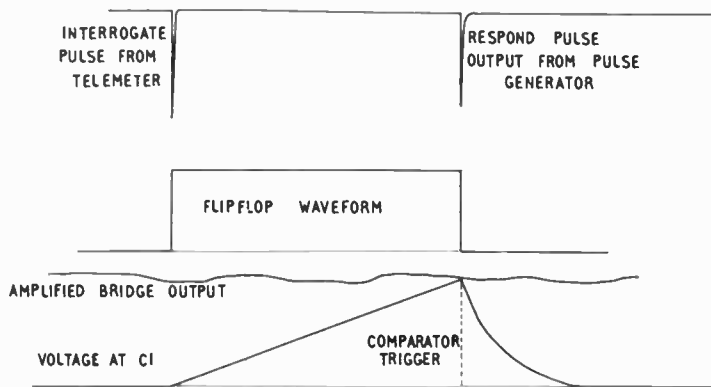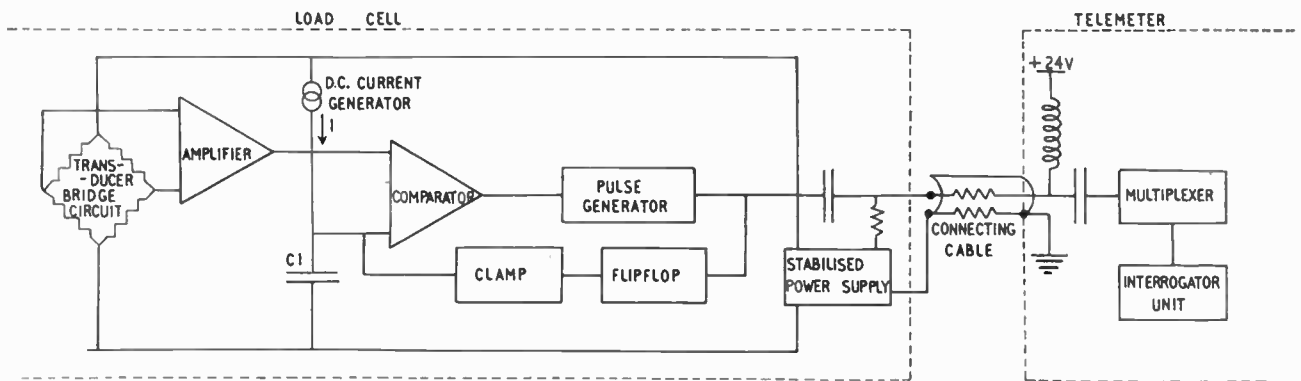
Fig. 8. Schematic diagram of voltage-to-time convertor.

gauge pressure transducer. To ensure successful operation, great care is required in bleeding the system since any air bubbles suspended in the oil prove detrimental to the accuracy. The instrument is bulky, and on small nets affects the readings to be taken. The measuring system requires protection from sea water pressure, but if the connecting hose between the main body and sensing head is damaged, then the whole instrument can be written off. For these reasons a new instrument is under development which will measure the differential static pressure between two fluids, one of which is the surrounding sea water. This system should be capable of discerning 1 in of water pressure $(25 \cdot 4 \text{ kgf/m}^2)$ on an absolute static head of 500 lb/in^2 $(3 \cdot 5 \text{ kgf/cm}^2)$, and the effects of hose damage should be far less severe.

Net speed through the water is measured on a high quality propeller type log capable of measuring speeds down to 1/10th of a knot. The propeller revolutions are measured by a low torque tacho-generator, the output of which is smoothed before being applied to the voltage-to-time convertor.

4. Shipboard Receiving Equipment

4.1. Towed Receiver

The retrieval of the acoustic signal on board ship presents the designer with a choice of either using a towed body which houses the receiving hydrophone, or a hull-mounted array. The latter is certainly preferable for ease of handling, but the additional cost of modifying an existing vessel can outweigh the advantages to be gained.

For this reason a towed body has so far always been employed in experimental work using this system. If a new ship were to be fitted at the building stage with a hull-mounted transducer, the situation would of course be different.

For most efficient signal reception the receiving transducer must be aligned approximately in the direction of the transmitter. With a towed receiver it is not possible to exercise fine control over transducer orientation. However, the present design uses borehole-type towing cable which is fairly heavy, and when taken together with the weight of the receiver itself the transducer is in fact depressed to a satisfactory angle at normal trawling speeds. Fin depressors were tried at one stage, but gave no significant improvement.

The transducer is tuned to 50 kHz, using the same techniques as described in Sect. 2.4, and has a beam width of 60°. The hydrophone is towed behind the ship at a distance of approximately 60–80 m in order to keep well clear of propeller noise, etc.

A band-pass amplifier, having a voltage gain of 10, is incorporated in the hydrophone to improve the signal/noise ratio. The amplifier output is transmitted to the ship via the towing cable (see Fig. 9).

4.2. Signal Amplifier and Demodulator

The first stage of the shipboard receiver consists of a monolithic a.g.c. controlled amplifier, having an a.g.c. range of approximately 30 dB. The first stage output feeds a tuned stage before being sampled by the a.g.c.
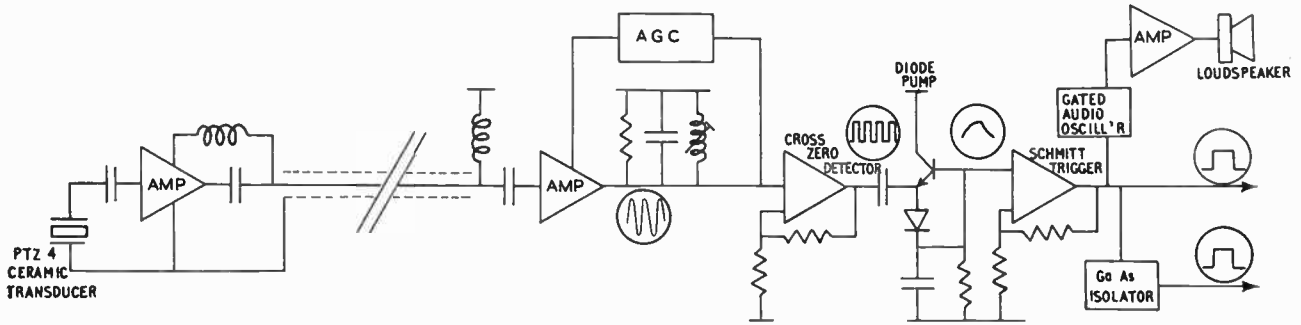
Fig. 9. Schematic diagram of shipboard receiver.

detector. The a.g.c. time-constant is selectable so that a choice can be made to suit the operating conditions at the time. The signal is then squared by the cross zero detector and fed to the diode pump. In order to achieve a degree of noise discrimination, a charging time-constant of 1 ms has been chosen to provide delay before the Schmitt trigger switches. Pulse length information is preserved provided there is no significant difference between the charging and discharging time constants of the diode pump. This demodulated output is now suitable for driving direct into the digital read-out unit. If an on-line computer is being used to monitor the signals[6] an additional output is provided using a gallium arsenide isolator to ensure complete protection from earth loops. Finally, a gated oscillator is incorporated to give the operator an audible indication of signal quality.

4.3. *Digital Read-out Unit*

The function of this unit (Fig. 10) which is connected to the pulse output of the main receiver (Sect. 4.2 and Fig. 1) is to convert the PDM signal to normal binary levels and then decode this to give a decimal display. In the following text a 'bit' refers to any pulse applied to the input of the digital read-out unit. Binary '0' is a bit of unit length and '1' is a bit 3 units long. A character is 4 bits terminated by a space 3 units long, and a word is 4 characters terminated by a space 16 units long.

4.3.1. Detectors

The input signal to the digital read-out unit is buffered and then applied to four detector circuits in parallel. These detect the time information (pulse and space
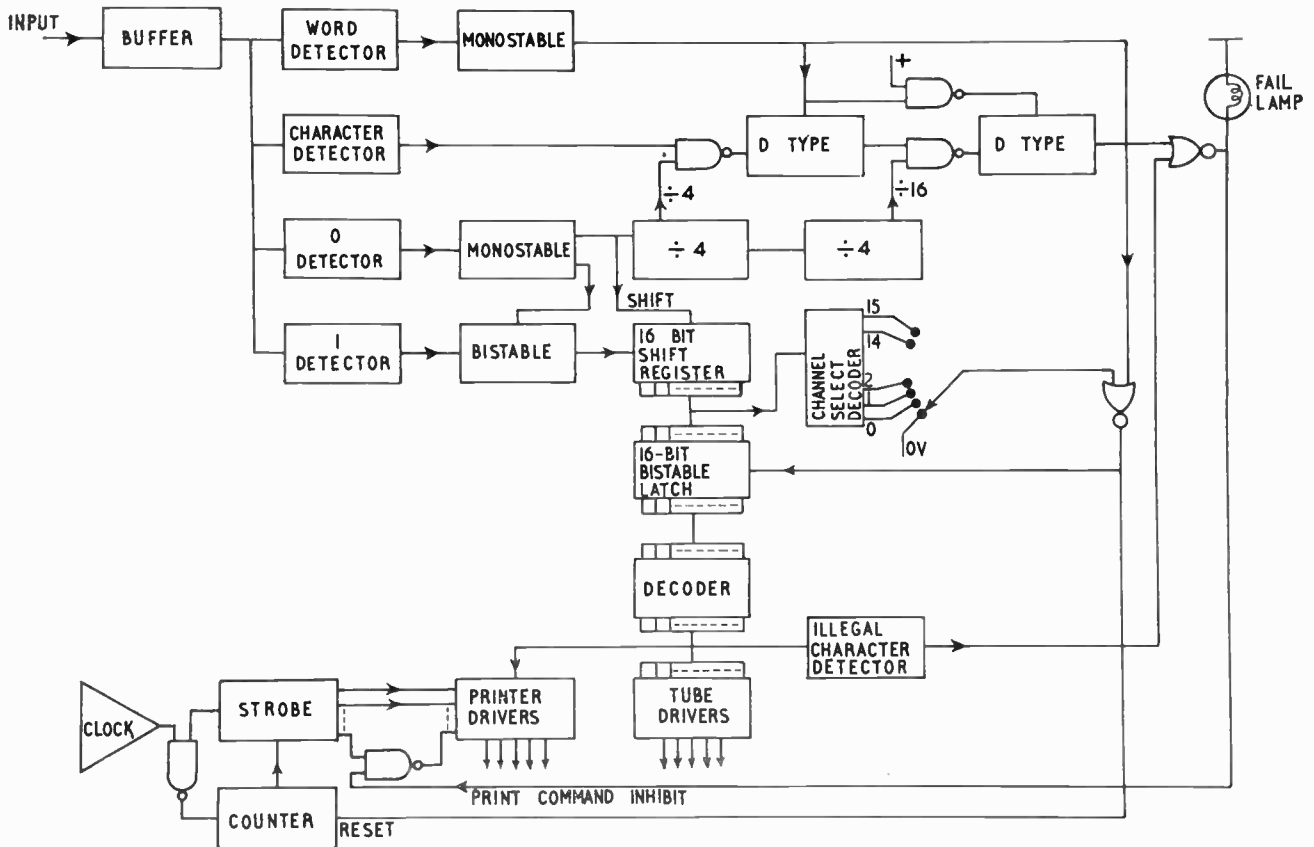

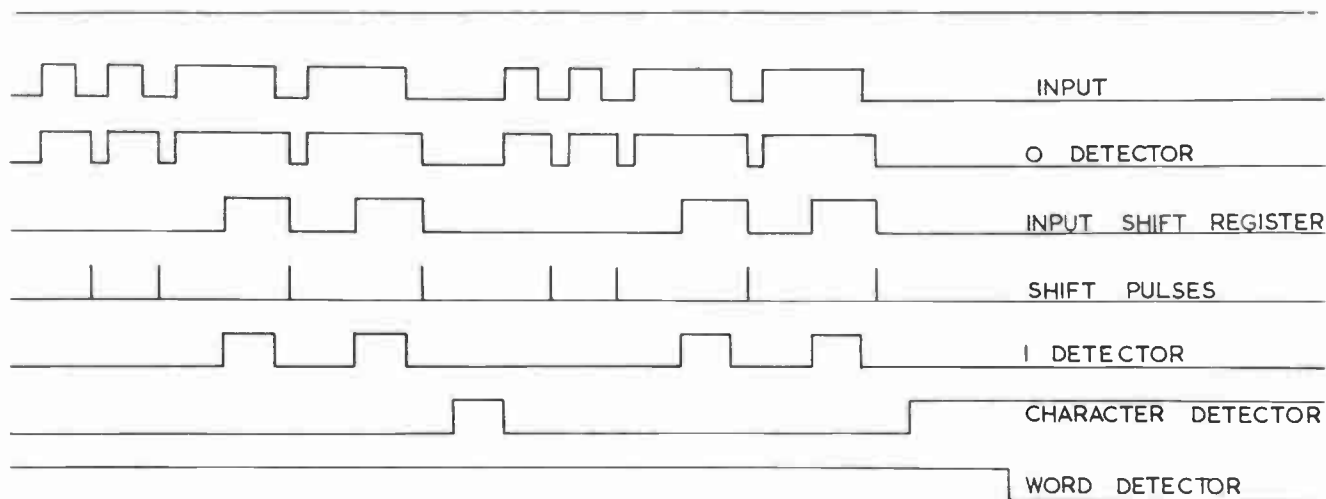**Fig. 10.** Simplified block diagram of digital read-out unit.

Fig. 11. Waveform diagram for digital read-out unit.

lengths) in the signal. There is no essential difference between the pulse length ('0' and '1') and the space length (character and word) detectors. The former are the same signal as the latter with an inverter added to the front end. The basic detector is a resettable integrator with an output from a Schmitt trigger connected to the integrating capacitor. The detector output switches only when the input pulse (or space, as the case may be) length exceeds a certain value related to the integrating time constant. The outputs of the four detector circuits are used in the following circuitry to convert the PDM signal to normal binary levels.

4.3.2. Circuit description

Figure 11 shows idealized waveforms relating to the last two characters of a word. If the bit presented is a '0' (short pulse) the '1' detector remains low, leaving the input to the shift register a logical 0. If the bit present is a '1', this detector goes high before the end of the pulse. The state of the detector is stored in a bistable. On the negative excursion of each bit waveform, the '0' detector changes state after the integrating period, generating a shift pulse to the shift register. This clocks in the state of the '1' detector which previously sampled the bit-width. The back edge of the shift pulse clears the bistable ready for the next bit. After 16 shifts the register is filled with a series of binary levels representing the transmitted word. The data are now in a suitable form ready for decoding. As half of the cycle period is taken up with transferring the data into the shift register, this would severely restrict the display period and time available for printing out before the next series of data arrives. This is overcome by holding the shift register data in a 16-bit bistable latch which acts as temporary data storage. Transfer of data through the latch to the decoder is effected by a signal generated every time the work detector triggers thus giving a display time of one complete cycle. Finally, the telemeter code is converted to decimal by a monolithic 1-in-16 decoder, whose outputs control the numerical indicator drivers.

A facility for allowing the choice of a particular channel to be displayed is incorporated in the circuitry. The shift register lines that hold the channel number are decoded by a 1-in-16 decoder, whose outputs are selectable by a 16-way rotary switch. This is used to gate through the appropriate data transfer pulse to the bistable latch only when the correct channel number appears in the shift register. For normal operation the gate is held permanently open by a switch.

4.3.3. Fail detection

In order to guard against misreadings due to bits being added to or dropped from a word (due to severe multipath conditions, noise, etc.), the design includes circuitry for checking each character for bit content and also illegal data numbers. The logic required is evolved by considering the waveform conditions. Firstly, after every 4 shift pulses (S_4) there should be a character detection pulse (Ch Det), and every 16 shift pulses (S_{16}) a word detection pulse (W Det). Further, if any character, bar the first one, has a decoded decimal value outside the range 0–9, the word is illegal. The Boolean expression to generate a fail condition is therefore:

$$\text{Fail} = S_4.\text{Ch Det} + \bar{S}_{16}.\text{W Det} + F_1 + F_2 + F_3$$

The character detector waveform is therefore NAND-gated with a waveform produced by every 4th shift pulse. If the channel detector pulse arrives other than after the 4th shift pulse, the D type bistable is clocked, setting the D input to the next bistable a 1. This is clocked through when the word detector pulse arrives. Gates coupled to the appropriate outputs from the last three 1 in 16 decoders generate F_1, F_2 and F_3 signals showing that the word is illegal. These are NOR-connected with the timing check signal. The resultant controls the visual 'fail indicator', and inhibits the print command pulse. Similarly the condition is set up again if the 16th shift pulse and word detector are not synchronized.

4.3.4. Printer interface

The printer used is a Clary model 35 serial entry printer which requires its data lines to be strobed sequentially in order to set up the correct type number for each column.

A 1-in-16 decoder controlled by a 4-bit counter produces the correct waveform for the data strobe lines

and print command pulse. Strobing is terminated by using the last strobe line to close the clocking gate to the counter. Strobing is initiated by using the word detector pulse to reset the strobe counter, allowing the clock gate to open.

5. Conclusions

The system has been in use now for over two years, operating for approximately 1000 hours in the sea. Failures have been confined mainly to mechanical damage to the instruments which must only be expected in this type of work due to the arduous working conditions. For this reason the mechanical construction must be extremely robust to withstand the knocks and bangs received when the equipment is shot and hauled onboard. Plenty of spare cable and underwater plugs and sockets must be carried as these are vulnerable to catching in the netting. The only failure to date found in the logic circuitry was due to accumulation of salt deposits between the printed circuit conductors. Spraying the boards with a printed circuit lacquer has subsequently proved most successful and no further problems have been encountered. The acoustic link has proved to be very reliable and ideally suited to this work. Better than 1 misreading per 1000 has been readily obtainable working in Sea State 6 and 7 which is far better than our expectations. The only disadvantage found is the loss of signal when the towing vessel is turning sharply but this seldom occurs in practice.

Now that the system has undergone its initial proving stage a few modifications are planned. The number of discrete components and battery power could be reduced by using m.s.i. low-power logic to replace the s.s.i. logic which was only available at the time of design. A smaller, more compact instrument would also be desirable.

The use of the telemeter has not only allowed the gear technologist to study the performance of the gear with greater accuracy, but to be given the data in real time with far better correlation than was possible with self-recording instruments. By using the telemeter in conjunction with an on-line computer it is possible to plot the instantaneous parameters of the fishing gear, and indeed display the shape of the net graphically if the computer facilities are available.

Although this instrument was specifically designed for data acquisition the acoustic digital technique lends itself for many other underwater applications where random noise discrimination and remote interrogation are primary features, e.g. underwater data buoys and stations or programmed acoustic release mechanisms.

A modified version of the telemeter is presently being built at Stirling University's Engineering Department for the remote control of an underwater tracking vehicle. It is anticipated that the first field trials will take place in the summer of 1972.

6. Acknowledgments

The author wishes to acknowledge the contribution made by all colleagues who participated in this project, and also to N.B.A. Controls, Farnborough, who manufactured the first two instruments, for their helpful suggestions.

7. References

1. Nicholls, J., 'Trawl gear instrumentation and full-scale testing', 'Modern Fishing Gear of the World', Vol. 2 (Fishing News (Books) Ltd, London, 1964).

2. Foster, J. J., 'Data collection in fishing gear', Proceedings I.E.R.E. Conference on Electronic Engineering in Oceanography, Southampton, 1966. Paper No. 19, pp. 19/1–19/5. (I.E.R.E. Conference Proceedings No. 8.)

3. MacLennan, D. N., 'Recent developments in the use of multi-channel telemetry for fisheries research', Symposium on Micro-electronics in Underwater Acoustic Systems. University of Birmingham, February 1970. (In press.)

4. Mowat, M. J. D., 'A digital acoustic telemeter for fishing gear research', Proceedings I.E.R.E. Conference on Electronic Engineering in Oceanography, Southampton 1966. Paper No. 20, pp. 20/1–20/3. (I.E.R.E. Conference Proceedings No. 8.)

5. Tucker, D. G. and Gazey, B. K., 'Applied Underwater Acoustics' (Pergamon Press, London 1964).

6. MacLennan, D. N., 'Instrumentation for the on-line measurement of the environment and performance of fishing gears', Oceanology International 69 Conference, Brighton, February 1969. Technical sessions, day 3.

The MacRobert Award
Engineering's 'Nobel Prize'

The MacRobert Award consists of £25,000 and a Gold Medal, and is given annually for an outstanding contribution by way of innovation in engineering or the physical technologies or in the application of the physical sciences, which has enhanced, or will enhance, U.K. prestige and prosperity.

The Award was instituted by the MacRobert Trusts in 1968 with the aim of honouring individuals or small teams of individuals. The MacRobert Trusts were founded by Lady MacRobert of Douneside and Cromar, wife of Sir Alexander MacRobert, Bt., head of the British India Corporation.

The Rules and Conditions for the MacRobert Award lay down that the prize winners are to be selected by the Council of Engineering Institutions in conjunction with the MacRobert Trusts and that in making their selection they will be advised by a committee consisting of a chairman and nine members. The chairman and six of the members will be nominated by the Council of Engineering Institutions and the remaining three members by The Royal Society.

When the Award is given to a team working for a firm, organization or laboratory, the medal is awarded to the firm, organization or laboratory and the money prize is divided between the smallest possible number of employees, and in no case more than five, nominated by the organization as having played a pre-eminent part in the team responsible for the innovation. Each of these individuals receives a replica of the medal, the division of the prize money being determined by the Council in consultation with the organization. Where the

The Medal, which measures 75 mm, was designed by Mr. Arnold Machin, O.B.E., R.A. (who was responsible for the Queen's head on the British decimal coinage). The well-known Archimedean quotation 'Give me a fulcrum and a lever long enough and I will move the world' which encircles the obverse, implies the realization of man's vision, limited only by current technology. The figure bestriding the globe symbolizes the overcoming of gravitational force and the beginning of man's exploration of the Universe, while the upraised callipers signifies that measurement is essential for realization of his vision. The reverse of the Medal, shown above, bears the MacRobert Arms.

Award is given to an independent team, CEI determines the membership of the team and the sharing of prize money.

Submissions for the Award may be made to CEI by firms, organizations, laboratories or individuals, but the Council reserves the right to consider innovations in connexion with which no submission has been made. Submissions must give the following information:

(a) Particulars of individual or independent team giving full name, age, address and, if any, academic qualifications and membership of professional institutions.

(b) In the case of a submission by a team working for a firm, organization or laboratory, a list of the individuals to be considered as a party to the submission giving full name, age, status in the firm or laboratory, and address.

(c) An appropriate description of between 200 and 500 words of the achievement in relation to the purpose of the Award, supported by any published or other relevant evidence.

(d) The names of at least two independent referees from whom further supporting statements could be sought.

Submissions for the MacRobert Award should be addressed to: MacRobert Award Office, CEI, 2 Little Smith Street, London SW1P 3DL, to arrive not later than 1st May 1972.

Previous Winners

1969. The first Award was made to two teams (five men in each) for innovations in engineering: the novel superstructure of the Severn Bridge by Freeman, Fox and Partners; and the *Pegasus* engine by Rolls-Royce, Bristol, which is used in the v.t.o.l. Hawker *Harrier* and which includes among its many new features the achievement of vectored thrust by four rotating nozzles spaced so that the resultant thrust line passes approximately through the centre of gravity of the engine.

1970. The physical sciences were honoured when the three geophysicists responsible for directing British Petroleum's exploration of the Alaskan North Slope Oilfield won the Award for the techniques they pioneered for surveying accurately through the permafrost layer, i.e. the permanently frozen layer of subsoil which exists in much of the sub-arctic area.

1971. The Award was made to the Gas Council, five of whose research engineers were honoured for their work on new processes of gas production, notably the Catalytic Rich Gas process.

Each year the Awards have been presented to the winners by the President of CEI, H.R.H. Prince Philip, at Buckingham Palace.