Founded 1925

Incorporated by Royal Charter 1961

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

VOLUME 42 No. 6 JUNE 1972

THE RADIO AND ELECTRONIC ENGINEER

The Journal of the Institution of Electronic and Radio Engineers

The Schools Links with Industry

BRIDGING the schools–industry communications gap was the theme of a recent report from the Electronics Economic Development Committee (E.E.D.C.). Through a link scheme, by work experience, the support of local science and technology centres, and well produced careers literature, the aim of 'The Electronics Industry and The Schools'* is to promote the long term availability of high calibre recruits into the electronics industry.

The advocated link scheme between the industry and schools would, initially, promote a greater understanding by both sides of problems and needs and, secondly, co-ordinate the preparation and execution of necessary measures. Schools would be 'linked' with a varied selection of industries within their area. Through industry-based contacts, or 'link men', staff and pupils would be assisted in making science and technology teaching closely related to everyday life in science-based industries. Schools would be actively involved in actual industrial processes.

The implementation of a link scheme with schools has been initiated during recent years by the National Electronics Council. In March of this year, at the request of N.E.C., the Institution of Electrical Engineers in co-operation with the I.E.R.E. and the Institute of Physics held a one-day symposium at the City University in London. This brought out some of the practical implications, in particular those involved in organizing school projects.

By encouraging young people's natural interest in electronics, the scheme could result in an increase in the technological relevance of the school curriculum and lead to greater understanding by careers and other teachers of industry's requirements and the way of life of others. Industry, too, could benefit—by becoming more aware of the consequence of changes in education and school curricula.

The E.E.D.C Report contains, as an appendix, an interesting syllabus for an 'A' level course in electronics, which, again, was originally prepared for discussion by N.E.C., by Professor G. B. B. Chaplin (Fellow) of the University of Essex. This aims at relating the main fields of communication, computer and control systems to human aspects and everyday experience while maintaining an appropriate intellectual level for 'A' level students.

The report points out that work experience has a place in a comprehensive programme of careers education and recommends legislation to be introduced providing experience in a pupil's final year at school but after the raising of the school leaving age in order to prevent the interruption of existing schemes.

Careers literature is a major medium of communication and the Electronics EDC has, in collaboration with the Central Youth Employment Executive, produced a companion booklet to the main report. 'Careers Literature: its preparation and distribution' outlines a wide range of fundamental, but often overlooked, points to be remembered by firms preparing publications, whether glossy booklet or less ambitious leaflets for local schools, as well as practical hints to aid distribution. It also provides guidance on the quality and quantity of information needed on occupations to appeal to the average pupil.

Initiated by the Electronics EDC Working Group on Scientific and Technological Manpower, the report is prefaced by the Group's Chairman, Professor G. D. Sims (Fellow), with the observation that one of the greatest problems is making clear what industry has to offer and requires. This is especially so in setting out the whole range of technical posts from engineer to technician and he believes that it is of the highest importance to show that there are good careers in electronics for those who are intelligent but not academically in the highest flight.

*The report and supplement are both available free, from NEDO, Millbank Tower, London SW1P 4QX.

World Radio History

Contributors to this issue

Mr. J. D. Martin received his B.Sc. (Eng.) degree in electrical engineering from Northampton Engineering College, London (now the City University) in 1957. He subsequently was awarded an M.Sc. (Eng.) degree for post-graduate research on transistor blocking oscillators. His early training was with Kelvin-Hughes Ltd., and from 1960 to 1965, he was concerned with remote control systems for railways and public utilities with Westinghouse Brake & Signal Co. Ltd., Chippenham. In 1965 Mr. Martin was appointed a lecturer in the School of Electrical Engineering at Bath University of Technology. His principal current research interests are digital systems and signal processing.

Mr. Geoffrey Baker studied applied physics at Brighton College of Technology, graduating with a 2nd Class Honours B.Sc. degree in 1961. From 1961 to 1964 he was employed at S.R.D.E., Christchurch, on pulse circuitry for radio communications. In 1964 he joined Mullard Southampton and for the past 5 years has been engaged on the research and development of pyroelectric detectors. Prior to that he was concerned with detailed measurements on photoconductive infra-red detectors.

Dr. S. L. Hurst (Member 1963) has been a lecturer in the School of Electrical Engineering at Bath University of Technology since 1961. He obtained his initial engineering training and education with British Thomson-Houston, Rugby, and graduated in electrical engineering in 1949. From 1950 to 1961 he was with the Westinghouse Brake & Signal Company at Chippenham, working on varied design and development projects, principally control and supervisory systems for the G.P.O., and home and overseas railways. In 1967 Mr. Hurst was awarded the M.Sc.(Eng.) degree by London University for research into ternary logic systems. He is currently engaged in investigations into the synthesis of single-gate and multi-gate threshold realizations of Boolean functions and in 1971 gained his doctorate for research in these subjects.

Mr. David Charlton entered Downing College, Cambridge, in 1957 after completing military service and he graduated in natural science in 1959. He has been employed at Mullard Southampton since 1960. His initial activities included studies of noise in transistors and photoconductors and he later lead a section developing doped germanium photodetectors. At various times he has held project responsibility for research on indium antimonide and cadmium mercury telluride photodetectors. He has worked on pyroelectric detectors since 1968.

Dr. J. Stephen graduated from the University of the Witwatersrand in 1947. Hr then carried out research on electronic techniques for seismological measurements, obtaining his Ph.D. in 1951. Since 1952 he has worked on transistor digital computers and on various aspects of magnetic materials. For the past five years he has been involved in the ion-implantation programme at Harwell, in which his main responsibility is the electrical measurements on ion implanted semi-conductors and the application of implantation to the fabrication of solid-state devices.

Dr. P. J. Lock obtained the B.Sc. degree in chemistry of the University of Leicester in 1965. He carried out research in infra-red and laser-Raman spectroscopy at Leicester and presented his doctoral thesis on 'The vibrational spectra of some halogeno- and aquohalogeno-complexes' in 1968. From 1968 to 1971 he was employed in the Electro-Optical Department at Mullard Southampton, on research and developmental work on the pyroelectric infra-red detector. In 1971 he joined the Materials Division of the Electrical Research Association, Leatherhead, where he is carrying out research on electrostatics, in particular on charge-transport mechanisms in polymers.

Dr. J. R. James (Member 1960, Graduate 1956) is on the teaching and research staff of the Royal Military College of Science. A note on his career was published in the March 1972 issue of the *Journal.*

World Radio History

U.D.C. 621.396.677.75

Engineering Approach to the Design of Tapered Dielectric-rod and Horn Antennas

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

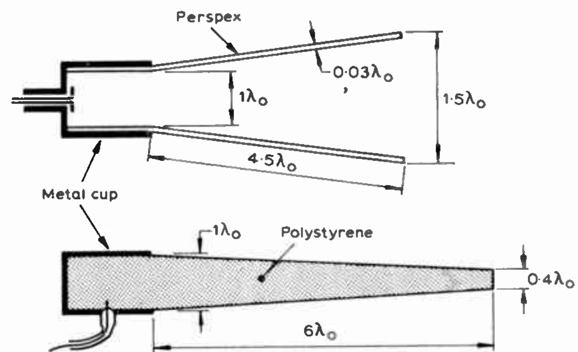
SUMMARY

The taper profile of optimized dielectric-rod and horn antennas is synthesized as a series of non-interacting planar radiating apertures. The method is semi-empirical, straightforward to apply, enables the dielectric-rod antenna to be satisfactorily optimized and provides a means of evaluating and optimizing a dielectric-horn antenna with variable wall thickness. The optimum profiles are taken as those which smoothly transform the surface-wave power from the launcher to the radiating aperture. The optimization of the dielectric-rod antenna considerably improves the radiation pattern while computations supported by measurements confirm earlier reports that a dielectric-horn can have a higher gain than a metal horn of similar dimension but side-lobe level is seen to be an important issue. Wide flare-angle horns give ideal E-plane patterns at the expense of a high side-lobe level in the H-plane; for small flare angles the dielectric horn gives similar patterns to the tapered rod antenna and thus preserves rotational symmetry. Calculations throughout are restricted to cylindrical geometry but other geometries and variations on the dielectric-horn principles are described. Useful engineering design data have been compiled for both the dielectric-horn and rod antennas and curves are given which determine near-optimum parameters for gains up to about 20 dB which is seen to be a practical operating limit for these surface wave devices. A unified impression of dielectric antennas emerges with the important conclusion that, when optimized, dielectric-rod and horn antennas are in fact competitive wth small metal horns for some applications; furthermore the dielectric-horn antenna, used singly or in arrays, is an ideal device for producing a low side-lobe level in the E-plane.

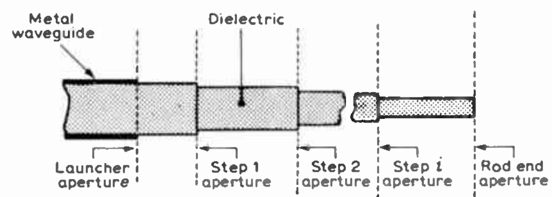
* *Department of Electrical and Electronic Engineering, Royal Military College of Science, Shrivenham, Swindon, Wiltshire.*

1. Introduction

The dielectric-rod antenna (Fig. 1(a)) has been extensively studied in recent years[1–5] and has proved to be a useful device in certain specialized applications; the dielectric-tube antenna has also received some attention[1,6] and is found to be similar in performance to the uniform dielectric-rod antenna. The dielectric-horn antenna also shown in Fig. 1(a) is perhaps the least known of all dielectric antennas and little has been established about it apart from the early reports[1,7] that its gain appears to exceed that of a metal horn of similar dimensions and this could be a significant property. Dielectric antennas bring about some saving in weight, have good sealing and corrosion properties and are now attracting renewed interest in view of the ease with which dielectrics can be manufactured with present-day techniques. The theoretical solution of these antennas is difficult because the radiation does not emanate from a simple plane aperture as in the metallic horn case; many isolated aspects of the problem have been successfully analysed[26] but until the overall radiating system can be treated as a whole the optimization of these antennas rests largely with the experimenter. In this paper an engineering approach to this problem is adopted whereby theoretical techniques are supplemented by experimental information. Optimized versions of these antennas are then compared with one another and also with the conventional metal horn.



(a) Dielectric-horn antenna with uniform wall thickness (Kiely[1]) and dielectric-rod antenna with linear taper.



(b) Approximating the dielectric-rod profile by a series of steps, each one assumed to produce a planar aperture of infinite extent.

Fig. 1.

The investigation concentrates on tapered dielectric-rod and horn antennas and as little is known about the latter the paper is concerned with their evaluation. The dielectric-horn antenna is certainly no less involved a device than the tapered dielectric-rod antenna for which no accurate design data exists; the difficulty lies in

World Radio History

calculating the radiation from the tapered structure to a sufficient accuracy and much the same problem occurs for dielectric-horn antennas. We describe a semi-empirical approach which assists in the selection of an optimum taper profile for the dielectric-rod antenna and useful design curves are obtained as a consequence. The dielectric-horn antenna may be calculated in a similar manner by synthesizing its profile by a system of short concentric dielectric tubes; the predominant radiation characteristics are, however, exhibited by the radiation from the end aperture alone and we evaluate the dielectric-horn on this simpler basis. Suitable applications for the antenna are pointed out and the essential design parameters identified. Calculations are confined to cylindrical geometry for simplicity and the treatment of other cross-sectional shapes follows in a similar manner but is computationally involved.

2. Semi-empirical Approach to Tapered Dielectric-rod Antennas

The diameter of the dielectric-rod is tapered along the antenna length to obtain a large efficient launching aperture in conjunction with a large radiating aperture at the rod end.[4] Unfortunately radiation leaks out along the taper and it is shown below that this radiation does not contribute usefully to the radiation pattern. The main design problem is therefore to select a taper profile which radiates least power and no exact means of calculation appears to exist. The writer has previously outlined[9] the following step-synthesis procedure whereby a step discontinuity on the rod is regarded as a radiating planar aperture of infinite extent in the plane of the discontinuity; with this approximation a tapered dielectric-rod antenna whose profile is assumed to be a series of m small non-interacting steps, may be treated as a system of m radiating planar apertures (Fig. 1(b)). The radiated E-field is given by the sum of the radiation from each aperture, thus

$$\mathbf{E} = -\frac{jk\,e^{-jkR}}{4\pi R}\mathbf{R}_1 \times \sum_{i=0}^{m} K_i \int_A [(\mathbf{n}\times\mathbf{E}_i) - (\mu_0/\varepsilon_0)^{\frac{1}{2}}\mathbf{R}_1 \times (\mathbf{n}\times\mathbf{H}_i)]\, e^{jk\rho_i\cdot\mathbf{R}_1}\, da \qquad (1)$$

where the far-field vector Kirchhoff integral is employed with the notation of Silver[12] and ($\mathbf{E}_i$, $\mathbf{H}_i$) is the aperture field of the ith aperture. The aperture field is taken as the unperturbed surface wave on each rod segment and for the diameters and dielectric constant of interest only the dominant HE_{11} wave exists (see, for instance, Fig. 8.6 on page 72 of ref. 13). K_i for $i = 1, 2, 3, \ldots, m$, is a factor which defines the proportion of incident power radiated at each step. The launcher radiation corresponding to $i = 0$ is obtained approximately from a chopped-surface-wave-distribution as described in ref. 4. The radiation lost at each step can be estimated by the Lorentz reciprocity theorem[13] but in the present author's experience the resulting radiation pattern predictions are still too inaccurate to be useful; this is attributed to the fact that reflected waves at each step are neglected as is the interstep radiation coupling.[9] It was thought that a better approximation would result by taking the aperture field as the difference between the unperturbed fields on either side of the step but poor results were again obtained. Eventually experiments indicated that the leakage of radiation along a smooth monotonically decreasing taper was to a large extent governed by the rate of change of power flow ratios along the taper and the following empirical formula

$$P_i^{\mathrm{RAD}} = (P_i^{\mathrm{I}} + P_i^{\mathrm{O}})\left[1 - \frac{\gamma_{i+1}}{\gamma_i}\right]^W \qquad (2)$$

where

$$\gamma_i = P_i^{\mathrm{I}}/(P_i^{\mathrm{I}} + P_i^{\mathrm{O}}),$$

was found to give some approximate functional description of the radiating process. P_i^{RAD} is the power lost at the ith step and P_i^{I} and P_i^{O} are respectively the surface wave power flowing in the interior and exterior regions of the ith segment, $i = 1, 2, 3, \ldots, m$; W is an empirical constant which depends on rod geometry and dielectric constant and is to be determined by curve fitting the computed to the measured radiation patterns. From a physical standpoint W may be interpreted as a parameter embodying the complex effects which arise when the small steps are spaced closer together.

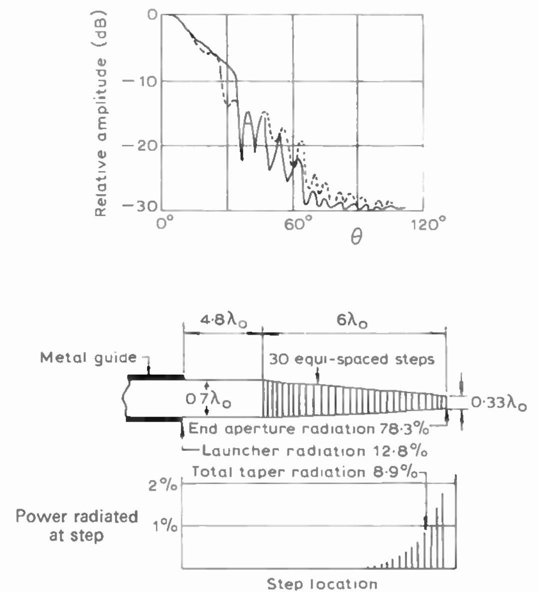


Fig. 2. H-plane radiation patterns of dielectric-rod antenna with linear taper, approximated by thirty equally-spaced steps. $\varepsilon_r = 2{\cdot}56$ (perspex) —— measured, ---- computed. $\phi = 0°$.

2.1. *Results for a Typical Dielectric-rod Antenna*

The usefulness of this empirical approach for determining the K_i factors in eqn. (1) is demonstrated in Fig. 2 where a linear taper has been synthesized into 30 equally spaced steps; a value of $W = 2{\cdot}2$ gave a best fit for the calculated and experimental H-plane radiation patterns. The angle θ lies in the H-plane for which $\phi = 0°$; θ and ϕ are spherical angular co-ordinates. Many other cases have been examined and a notable feature is that for smooth tapers both the experimental and computed patterns begin to converge for as little as three equally-spaced steps. A similar situation occurs for E-plane computations. The radiation from the taper increases the side-lobe level, is undesirable from an operational standpoint and is characteristic of a typical interference

World Radio History

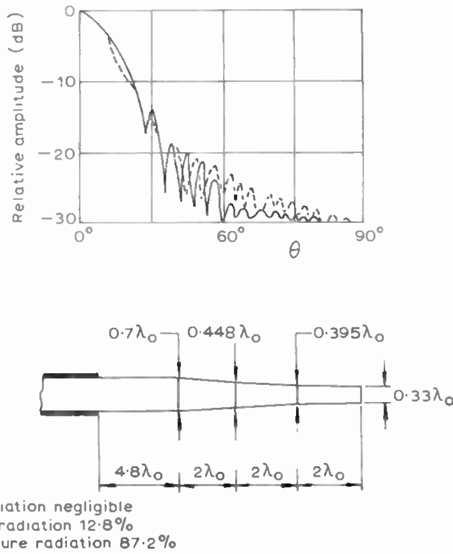


Fig. 3. H-plane radiation patterns of dielectric-rod antenna with optimized taper profile. $\varepsilon_r = 2{\cdot}56$ (perspex) —— measured, ---- computed. $\phi = 0°$.

pattern; that is to say there is little evidence that leaky wave radiation[2,24] is occurring at distinct angles. In the case shown the curve-fitted-computations show that the taper radiates near the end with a loss of 8·9% of its available surface wave power. It is evident from Fig. 2 that a concave profile would provide a more gradual taper near the rod end and hence reduce this premature loss of radiation at the expense of encouraging the taper to radiate nearer to the launcher region. Several profiles were computed and fitted to the measured radiation patterns to see if a compromise could be reached whereby the taper radiated uniformly along its length but at a negligible level; the concave profile in Fig. 3 gave best results and consists of three linearly tapered portions rather than a gradual curve, in order to ease machining difficulties. The curve fitting procedure indicated that negligible surface wave power was radiated along this taper and as such this is a well optimized dielectric-rod antenna; the computed gain was 19·6 dB but the measured gains were 2 dB less mainly due to the use of perspex. In these experiments the tapered section was separated from the launcher by a section of uniform diameter rod to isolate the various sources of radiation; in practice the antenna would consist of the tapered section only.

Marcuse[10] has recently calculated the radiated power from a tapered cylindrical dielectric-rod using a similar step synthesis approach. A particular case that he has treated (Fig. 8 in reference 10) is not too dissimilar to that in Fig. 2 and some comparison is permitted. Taking the same taper length as in Fig. 2 Marcuse calculates that a linear and an exponential taper lose 8% and 6% respectively of the incident power which is of the same order as the results given in Figs. 2 and 3. The percentage of radiation from the taper must be accurately known to within about 1% if the side-lobe level of the antenna is to be calculated within acceptable engineering limits; since Marcuse's mathematical model neglects radiation coupling effects[9,11] between the step apertures, doubts must be raised as to whether it is sufficiently accurate for the purposes of designing the dielectric-rod antenna without some recourse to measured data.

To increase the gain of the dielectric-rod antenna it is necessary to taper a longer rod to a smaller terminal diameter. Marcuse's results do not indicate that this will present difficulties and neither do any assessments based on the Lorentz reciprocity theorem,[13] yet in practice it has not been possible so far to increase readily the gain beyond about 20 dB; the reason seems to be that the taper dimensions are critical for small diameter rods and the rod lengths are excessive for a small increase in gain and are greater than $10\lambda_0$. An estimation by Snyder[14] for a very gentle linear taper shows that as the cut-off frequency is approached the taper radiation increases steeply thus supporting the above observation. This view is also compatible with other analytical results for a surface-wave structure.[26] The dielectric materials such as perspex are very convenient for experimental work but materials of lower loss such as polystyrene or polythene are necessary for operational devices. Whatever material is used the loss will increase with taper length but even for perspex this does not significantly contribute to the threshold effect.

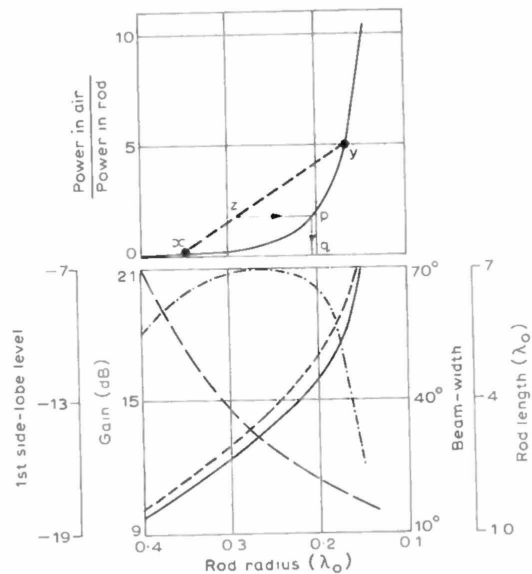


Fig. 4. Computed curves for tapered dielectric-rod antenna. $\varepsilon_r = 2{\cdot}5$ and launcher aperture radius $= 0{\cdot}4\lambda_0$; the following details refer to lower graph: —— antenna gain, ---- rod length, — — — 3 dB beamwidth, -.-.-.- 1st side-lobe level.

There is therefore substantial evidence that the tapered dielectric-rod antenna and also other surface-wave types, have a threshold performance beyond which they cease to be viable practical antennas. For the simple monotonically tapered short rod under consideration here we find that this threshold is in the vicinity of 20 dB gain. For longer rods it may be necessary to modulate the diameter periodically as in the 'cigar' antenna[15] in order to obtain higher gains but the antenna would then be several orders longer.

World Radio History

2.2. *Design Curves for Tapered Dielectric-rod Antennas*

The curves in Fig. 4 are based on the radiating properties established above and it is assumed that the optimum taper profile is one for which the ratio γ_{i+1}/γ_i in equation (2) is constant from step to step. With this criterion the taper profile is immediately defined once the terminal rod diameters are stated. The curves embrace a useful range of antenna performances up to the above mentioned gain threshold of 20 dB below which we have been able to optimize the tapers. The launcher radius is taken as $0{\cdot}4\lambda_0$ and we assume that the rod will fill the latter in which case the launcher radiation is known to be about 10% of the incident metal guide power;[4] the curves are based on this figure and the remaining power is assumed to emanate from the end aperture. The side-lobe can be improved by launching from a larger aperture or alternatively utilizing a ring source launcher.[13] There are many other techniques for increasing the launching efficiency such as placing a flange or choke[17] around the metal waveguide aperture but all improvements necessitate additional metallic components, increase the space occupied by the launcher and in some cases narrow the operating bandwidth. Practical system constraints will determine whether or not the launcher efficiency is to be optimized beyond that of the simple metal cup arrangement considered here (Fig. 1(a)); antenna gain and beamwidth will not vary appreciably with more efficient launching but the performance figures for side-lobe level in Fig. 4 are to be regarded as a conservative assessment of what can be achieved.

To demonstrate the use of Fig. 4 suppose that an antenna gain of 15 dB is required. A rod terminal radius of $0{\cdot}223\lambda_0$, 1st side-lobe level of $-7{\cdot}3$ dB, beamwidth 24° and rod length of $4{\cdot}4\lambda_0$ are given by the lower set of curves. To obtain the taper profile mark the power ratio curve at points x and y that correspond to the launcher and terminal radii of $0{\cdot}4\lambda_0$ and $0{\cdot}223\lambda_0$. Intermediate radii along the taper are then defined by projecting the line xy on to the abscissa via the power ratio curve (solid line in upper graph). To demonstrate this further the line xy already drawn in Fig. 4 corresponds to the terminal radii of the antenna in Fig. 3 and the point z when projected to q via p predicts an optimum radius of $0{\cdot}21\lambda_0$ at a distance of λ_0 along the taper as opposed to the value of $0{\cdot}224\lambda_0$ obtained by curve fitting. This is in good agreement considering the dielectric constant and launcher radius do not correspond exactly to the conditions of Fig. 4; similar agreement is obtained for other points along this taper.

Since the dielectric-rod antenna possesses radiation patterns with rotational symmetry,[4] the 1st side-lobe level quoted may be taken with good approximation as the value for either the E or H plane. It is believed that the curves are accurate enough for most design purposes and certainly much more informative than existing data based on a simplified sin x/x calculation.[2] A complete antenna design must embody the bandwidth and matching aspects of the launcher feed but our experiments have shown[27] that the radiation patterns are not sensitive to conditions within the launcher provided the latter is not heavily over-moded; for the launcher radius considered here it is easy to maintain the H_{11} mode and the feed design can be pursued independently.

3. The Dielectric-horn Antenna

Early references to the dielectric-horn antenna are Kiely[1] (1953) and Prochazka[7] (1959). These experiments were confined to horns with uniform wall thickness and calculations based on the assumed simplified radiation mechanism gave some agreement with measurements[7] but were generally insufficiently representative; as far as is known no attempt has previously been made to investigate the effect of horn wall thickness. The present improved knowledge of dielectric-rod and also dielectric tube antennas[6,17] now enables the dielectric-horn antenna with or without varying wall thickness to be evaluated in the following simplified way: the horn is regarded as a system of short concentric non-overlapping dielectric tubes with various external and internal radii b and a respectively, which approximate to the actual horn dimensions in a similar fashion to the synthesis of the stepped dielectric-rod profile (Fig. 1(b)). Experiments show that the horn could be justifiably represented in this way and furthermore the radiating system closely parallels the situation for the tapered rod antenna in that low side-lobe patterns corresponded to negligible radiation from the horn conical surface. From an evaluation standpoint it is therefore sufficient to consider only the radiation from the launcher and horn mouth and the subsequent computations carry this simplifica-

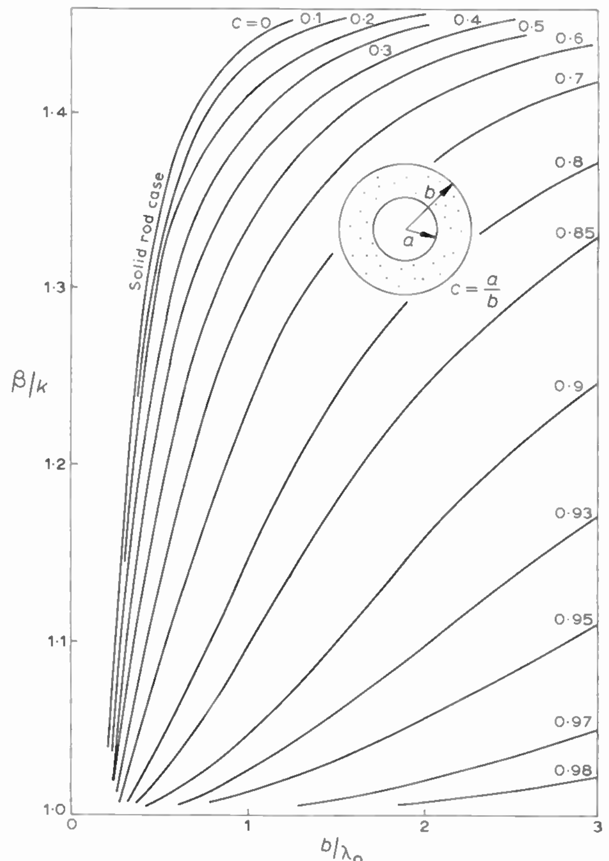


Fig. 5. Values of β for HE_{11} mode on dielectric tube. $\varepsilon_r = 2{\cdot}2$.

tion. The horn mouth is taken as a radiating dielectric tube and since a detailed study of the dielectric-tube antenna has already been described elsewhere[1,6,16] we emphasize only those details that are relevant to the horn situation.

3.1. *Surface-wave Modes on a Dielectric Tube*

One physical interpretation of the tapered dielectric-rod antenna is that the launcher field is progressively transformed along the taper until the surface-wave wavelength approximately matches that of free space. This view may be extended to the dielectric-horn antenna and to illustrate this further we have computed a range of wavelengths for the dominant hybrid mode on dielectric tubes whose dimensions are relevant to the horn situation. In Fig. 5 values of the phase constant β normalized by k, are plotted as a function of tube radius and thickness for $\varepsilon_r = 2{\cdot}2$ since most of our experiments use paraffin wax. $\beta/k = \lambda_0/\lambda$ where λ_0 and λ are the free-space and surface-wave wavelengths respectively. Taking the simplified stepped-tube model for the horn we see that the horn wall thickness must taper to a very small value at its mouth if the terminal radius is to be large whereas at the launcher end the wall needs to be very thick for good launching. In the designs described below we take $c(=a/b) = 0$ at the launcher in which case the launching efficiency from a metal cup as in Fig. 1(a) should be at least as good as for the dielectric rod case. It was conjectured that the efficiency would be higher since some rays emanating from the launcher would be diffracted forward by a wide flare angle dielectric horn and evidence of this effect is apparent in the experimental results below.

A second point of interest is whether it is reasonably valid to assume that only the HE_{11} mode need be considered on each elemental tube in this simplified model. To examine this, the cut-off dimensions for the next four higher modes on a dielectric-tube are plotted as a function of tube radius and thickness in Fig. 6; also superimposed in this figure is a horn profile which is typical of those used in the subsequent experiments. We see that only the symmetric modes and in particular the H_{01}, can exist on any of the elemental tubes in addition to the HE_{11} mode but with careful launching from the H_{11} mode in the dielectric-filled waveguide these symmetric modes should not be excited. For the cases considered below there is no evidence that symmetric modes are being excited and the assumption that only the HE_{11} mode need be considered is justified.

3.2. *Theoretical Evaluation of the Dielectric-horn Antenna*

The radiation calculation is carried out in a similar manner to that for the dielectric-rod antenna but the manipulations are necessarily more extensive due to the additional boundary condition; the salient details are summarized as follows. The launcher radiation is calculated from the unperturbed H_{11} mode in the latter, as was done in equation (1), and the launching efficiency is again based on a chopped-surface-wave distribution.[4] The horn mouth is regarded as a dielectric tube aperture for the purposes of calculating the radiation pattern; the relevant β is given in Fig. 5 and the unperturbed HE_{11} mode is taken as the aperture field. Using equation (1) but with $m = 0$ and 1 only, since we are assuming that an optimum non-radiating taper profile exists, the radiation pattern resolves into functions of Lommel integrals as in the rod case;[4] these equations are given in detail elsewhere.[17] As the antenna has a pencil-type beam, the gain can be readily calculated from E- and H-plane radiation patterns.[18]

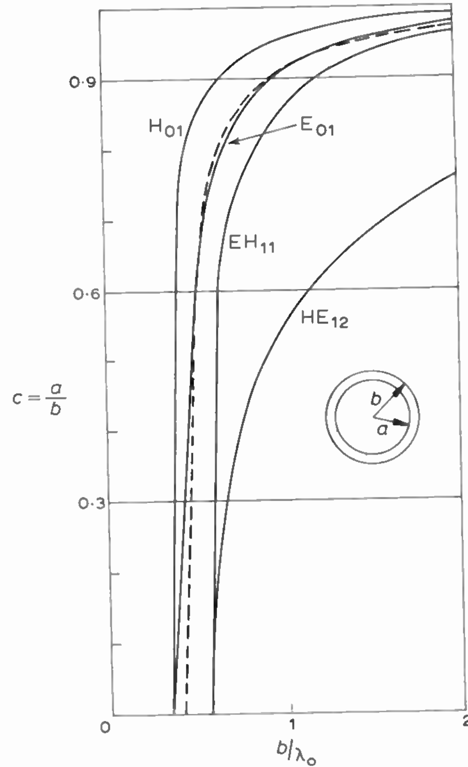


Fig. 6. Cut-off conditions of the next four modes above the HE_{11} mode. $\varepsilon_r = 2{\cdot}26$, ---- profile of typical dielectric horn.

Since the dielectric-horn antenna would appear to be a natural replacement for metal horns we will use the latter for the purposes of comparison. The radiation patterns of the metal horn have been obtained by applying the vector Kirchhoff integral[12] to the unperturbed H_{11} mode in the horn mouth; other modes have been neglected but the results agree well with measurements and geometric optics calculations[19] in the main lobe region which is of interest here.

In Fig. 7 the gain of an optimized dielectric-horn antenna is plotted as a function of the radius and wall thickness of the horn mouth; also given is the gain of a metal horn of identical mouth radius b. If a horn taper profile can be selected which is optimum, i.e. radiation is emitted solely from the launcher and horn mouth, then these curves show that the dielectric horn can always be arranged to have a higher gain than a metal horn of identical mouth dimensions by tapering the wall thickness to a very small value. This is in agreement with the earlier observations.[1]

World Radio History

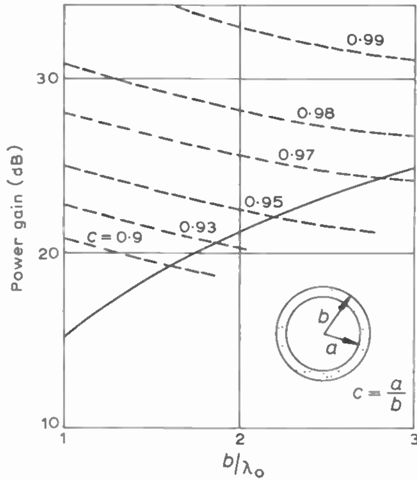


Fig. 7. Antenna power gain computed using the vector Kirchhoff integral. —— cylindrical metal horn aperture radius b;---- dielectric horn, $\varepsilon_r = 2{\cdot}2$, launcher radius $= 0{\cdot}4\lambda_0$, launcher loss = 10%.

Unfortunately antenna gain is not the only system parameter of interest and the computations of the first side-lobe level (Fig. 8) show that the dielectric-horn is superior to the metal horn in the E-plane but considerably inferior in the H-plane; the high side-lobe level in the H-plane is not due to interference with the launcher since radiation from the latter is deliberately omitted in these curves. The explanation is that the E-field is concentrated in the dielectric walls, thus the field distribution tends to resemble two vertical line sources and creates an interference pattern in the H-plane; however this field distribution appears to be ideal as regards the E-plane. The inference from Figs. 7 and 8 is that a superior antenna to the metal horn can be obtained if the dielectric horn terminal radius is large and the wall very thin but this

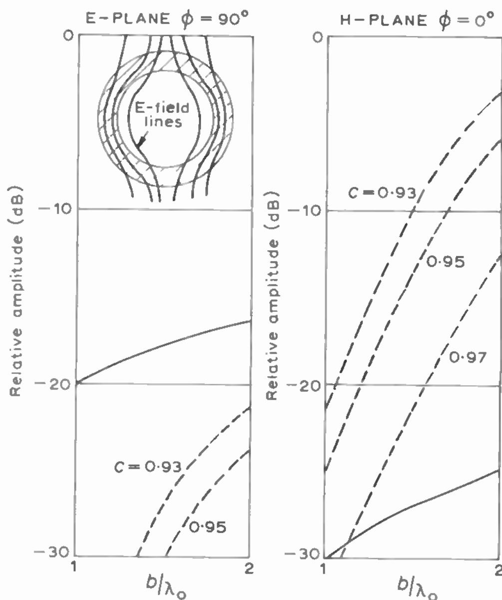


Fig. 8. First side-lobe level relative to main-lobe amplitude computed using the vector Kirchhoff formula. —— metal horn aperture radius b, ---- dielectric horn aperture, $\varepsilon_r = 2{\cdot}2$. Inset sketch shows concentration of E field in the dielectric tube walls.

depends on whether a non-radiating taper profile can be found.

3.3. *Experimental Substantiation of Dielectric-horn Antenna Design*

After much experimenting it was found that the situation showed little departure from the dielectric rod case discussed above: that is, horn profiles that produced negligible radiation along their length could be found for antenna gains below about 20 dB while gains in excess of this figure are difficult to achieve. Similarly, radiation emanating from the taper increased the overall side-lobe level and did not contribute usefully to the radiation patterns. This immediately constrains the wall thickness in Fig. 7 to values of $c < 0{\cdot}93$ that produce very high H-plane side-lobe levels in Fig. 8 although the E-plane pattern continues to have a low side-lobe level. The H-plane side-lobe level can be improved by reducing the dielectric horn terminal diameter until the horn has zero flare angle.

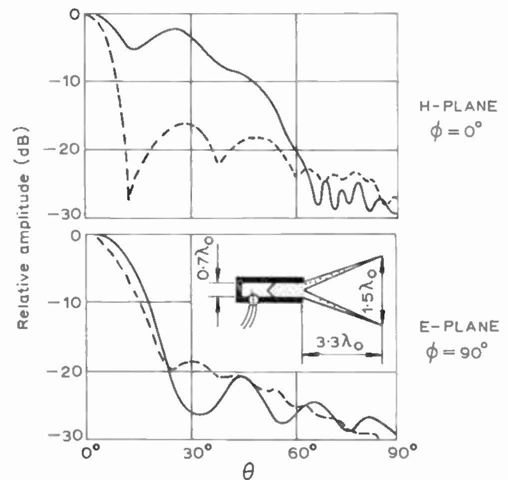


Fig. 9. Radiation patterns of wide flare angle dielectric horn. $\varepsilon_r = 2{\cdot}2$ (paraffin wax), $\lambda_0 = 3{\cdot}2$ cm, $c = 0{\cdot}95$; —— measured, ---- computed assuming zero taper radiation loss and 12·8% launcher loss.

Several dielectric horns have been tested[23] which bear out these theoretical predictions and it is sufficient here to show three cases. In Fig. 9 the horn was too short for the given terminal wall diameter and thickness; it is considered that appreciable radiation occurs along the horn taper and this is responsible for the disagreement between the measured and computed patterns. Of particular interest here is the high side-lobe level in the H-plane as opposed to the low level in the E-plane and the fact that the contribution from the taper spoils the radiation characteristics in both planes. On reducing the flare angle and thickening the horn walls (Fig. 10), less radiation is lost along the horn taper and the measured results tend more to the computed values. Finally, Fig. 11(a) shows the patterns of a zero flare angle dielectric horn; the radiation along the horn taper has been considerably reduced and there is a low H-plane side-lobe level. The E-plane pattern was very similar thus exhibiting rotational symmetry of the beam. In fact this zero flare angle horn gives similar results to the linearly-tapered dielectric-rod antenna

World Radio History

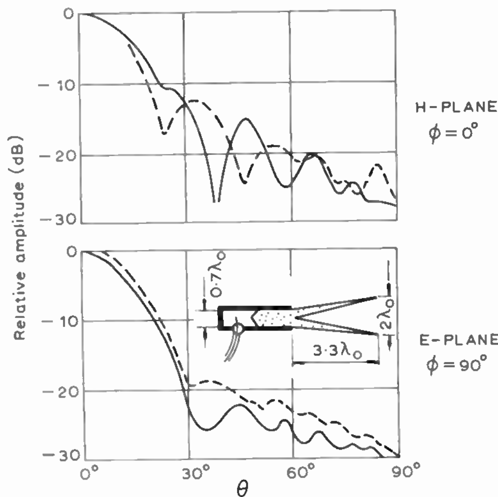


Fig. 10. Radiation patterns of moderate flare angle dielectric horn. $\varepsilon_r = 2{\cdot}2$ (paraffin wax), $\lambda_0 = 3{\cdot}2$ cm, $c = 0{\cdot}9$; —— measured, ---- computed assuming zero taper radiation loss and 12·8% launcher loss.

(Fig. 2). The radiation along the horn taper could be reduced further by optimizing the profile but this presented machining difficulties and was not attempted.

The dielectric-horn antenna may therefore be summed up as follows: if equal beamwidths and low side-lobe levels are required in both the E and H-planes then the dielectric horn must have zero flare angle and is similar in length and performance to the tapered dielectric-rod antenna. Since the latter is slightly easier to manufacture it is a natural choice but the additional physical strength of the dielectric-horn should not be overlooked. There may be many applications where an optimized dielectric-rod or zero flare angle horn antenna may be preferred

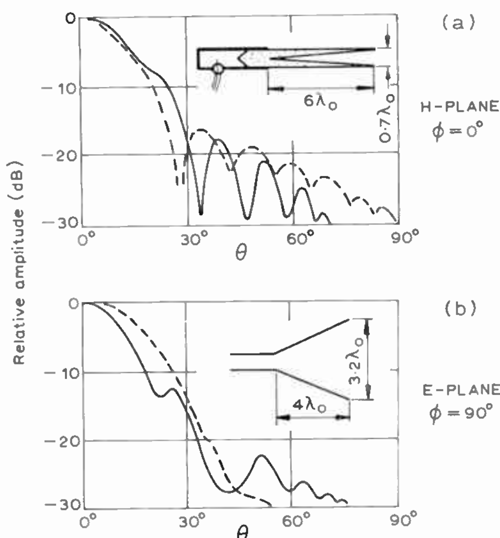


Fig. 11.

(a) H-plane radiation pattern of dielectric horn with zero flare angle. $\varepsilon_r = 2{\cdot}56$ (perspex), $\lambda_0 = 3{\cdot}2$ cm, $c = 0{\cdot}82$, —— measured with a gain of 17 dB, ---- computed with a gain of 19·2 dB assuming no radiation from or dissipative losses in the taper and launcher loss of 12·8%. $\phi = 0°$.

(b) Measured radiation patterns of cylindrical metal horn. Measured gain = 18·5 dB; —— E-plane, ---- H-plane.

to a conventional cylindrical metal horn and comparison of Fig. 11(b) with Fig. 3 emphasizes the competitive performance of these dielectric antennas which can be brought about by optimizing the various taper profiles. If however a narrow beam with low side-lobe level is required in the E-plane only, the dielectric horn is without doubt more suitable than either the dielectric-rod or cylindrical metal horn antenna. From the experiments shown here it appears that a dielectric antenna will generally be 50% longer than a metal horn of similar performance; the physical diameter will of course be much less. Similar results hold for dielectric antennas and metal horns in rectangular geometry.

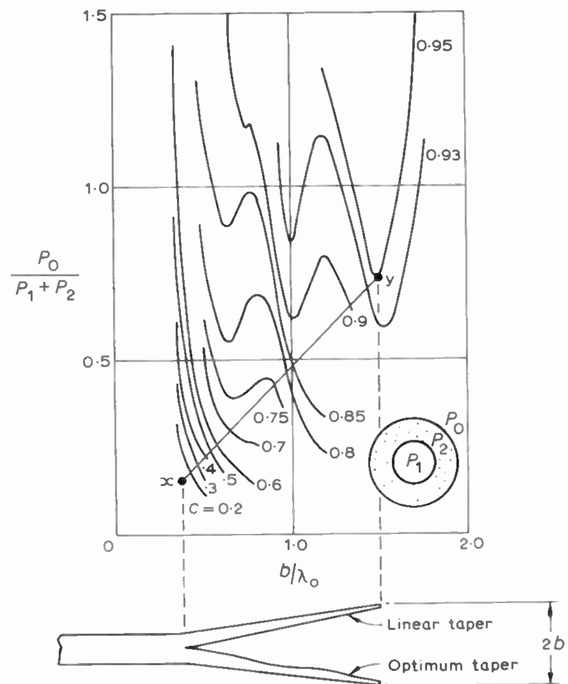


Fig. 12. Fraction of surface wave power within horn boundary based on step model approximation. Launcher and horn mouth radii correspond to point x and y; c values on line xy as a function of radius b define the near optimum profile, $\varepsilon_r = 2{\cdot}2$.

3.4 *Optimization of the Dielectric-horn Taper*

In principle the dielectric horn could be flared out in a variety of ways but here we have confined our study to flares with linear exterior walls in which case only the interior wall profile can be optimized to minimize the taper radiation losses. Tests have shown that as for the dielectric rod case, the fraction of surface wave power within the horn antenna exterior boundary wall as a function of radius b gives some indication as to how to choose the inside wall profile; these curves are given in Fig. 12, and define the inner wall profile as follows: the values of b and c that correspond to the horn dimensions at the launcher and mouth are marked on the curves at x and y; c values on the line xy then define the inner wall radius since b varies linearly along the horn length. We find however that the profile differs little from a straight line as the sketch in Fig. 12 shows and since a linear wall flare is easier to construct it has been adopted in most of our designs; the radiation along

World Radio History

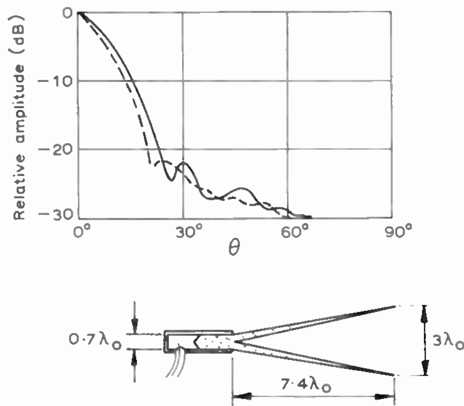
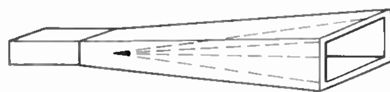


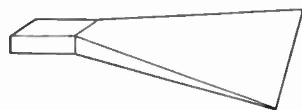
Fig. 13. E-plane radiation patterns of long wide flare angle horn. $\varepsilon_r = 2{\cdot}2$ (paraffin wax), $\lambda_o = 3{\cdot}2$ cm, $c = 0{\cdot}96$; —— measured, ---- computed with gain 24 dB assuming no radiation from or dissipative losses in the taper and 12·8% launcher loss. $\phi = 90°$.

the horn taper is then controlled by the terminal wall thickness and horn length. To design a zero flare angle horn profile the line xy becomes parallel to the ordinate axis and the c values vary in a smoother monotonic fashion giving a concave profile as in the rod case.

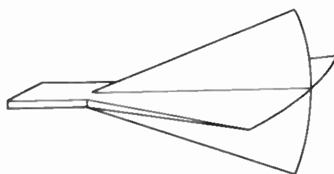
An example of the type of E-plane pattern that can be obtained from a long horn with linear walls is shown in Fig. 13. The radiation along the length is not negligible but is reduced to well below the launcher radiation level and computations agree well with measurements; further optimization of this horn was not attempted because of machining difficulties. The possibility of using an array of such horns suggests itself but grating lobe constraints need to be considered. It was anticipated that a wide flare-angle horn would collect some of the launcher radiation and for the wide flare angles used here some weak evidence exists in this respect since the computed side-lobe levels away from the main beam are generally higher than those actually measured. This is particularly so for the E-plane patterns of Figs. 9 and 10.



(a) rectangular horn antenna



(b) wedge antenna with converging and diverging taper profiles



(c) double wedge-shaped antenna

Fig. 14. Various types of rectangular dielectric antenna suitable for rectangular waveguide launchers.

Another possibility is to maintain a constant value of c along the flare so that the internal and external walls diverge in a radial manner. It is evident from the curves above that it is not possible to choose c to satisfy both good radiation pattern and efficient launching conditions. To develop this principle further it is necessary to increase the size of the launcher and length of the antenna; a practical realization of this idea is the infra-red detection antenna found in moths which is tens of wavelengths long and has a relatively large diameter launcher.

4. Variations on the Dielectric Horn

Rectangular waveguide is more common than circular and it is possible to develop the above antennas for rectangular launchers. In fact a large variety of surface wave antennas may be developed using rectangular, elliptical. triangular and other cross-sectional shapes of rods and tubes. Unfortunately the trapped surface waves cannot be expressed in analytic form and recourse must be made to point-matching calculations.[20] Aperture fields are then obtained as a series and radiation patterns

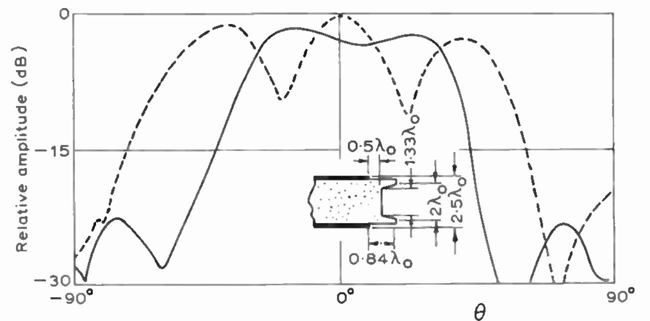


Fig. 15. Measured H-plane radiation patterns of dielectric filled overmoded cylindrical waveguide. $\varepsilon_r = 2{\cdot}2$ (paraffin wax), $\lambda_o = 3$ cm; ---- dielectric-filled waveguide aperture, —— with dielectric filling extended to form a short horn as sketched. $\phi = 0°$.

can only be obtained by numerical integration. Some examples of variations on the cylindrical dielectric horn are shown in Fig. 14 and we have tested them experimentally: these antennas are interesting but show no marked advantages over conventional forms although they may have specialized applications.

Much attention has been focused on the use of dielectric inserts in the mouth of the metal horns in order to reduce side-lobe level. For instance a recently reported dielectric wedge-shaped antenna[21,28] is very similar to the zero flare-angle horn (Fig. 11(a)). The insertion of dielectric strips into metal horn mouths[22] condenses the E-field in a similar way to the dielectric tube (Fig. 8). Another interesting effect is shown in Fig. 15 whereby a short thick dielectric horn protrudes from an overmoded cylindrical waveguide and enables a broad beam pattern to be obtained from what was previously a multi-lobed radiation pattern. One explanation[25] of this mode filtering action is that radiation coupling between the launcher and the horn adjusts itself to minimum for a certain set of modes in the metal launcher and a given geometry of the dielectric insert; another explanation is that the dielectric extension forms a resonant open cavity

World Radio History

for these dimensions. Many other dielectric insert configurations are possible but their mechanism, as in the above case, is generally involved and much reliance is necessarily placed on actual measurements.

5. Discussion and Conclusions

The optimization of the taper profile of dielectric antennas is seen to present considerable theoretical difficulties; the approach developed here is to supplement theoretical considerations based on a simplified model with actual measured data and a high degree of optimization has been found possible. A conclusion of practical importance is that the optimization of the taper profile is well worthwhile and places both the dielectric-rod and zero flare-angle dielectric-horn antennas on a competitive footing with conventional metal horns of moderate gain. Design curves have been presented for commonly occurring values of dielectric constant but details of the general computer programs are available from the author. It is not envisaged that these dielectric devices will generally be alternatives to metal horns because robustness and power handling requirements obviously favour the metal antenna. However there are many applications where the low cost and weight, ease of manufacture and good corrosion and sealing properties associated with dielectric antennas will be an advantage; one such application is in large arrays.

Evaluation of the dielectric-horn antenna has shown that it could be particularly useful when directional properties in one plane only are required as is so often the case. An array of these devices is an interesting possibility.

When the flare angle is reduced to zero the dielectric-horn has similar patterns to the tapered rod antenna and exhibits rotational symmetry. The bandwidth of the dielectric-horn antenna is comparable to that of the tapered dielectric-rod antenna and is typically 1 GHz at X band, but this is more often than not constrained by the bandwidth of the feed system. An alternative launcher design has not been considered since the method used here seems to be widely applicable in practice to both horn and rod antennas.

An interesting fact has emerged that the dielectric-rod and horn antennas have similar actions despite their widely differing geometries; in essence the principle is to transform the launcher aperture into a larger one having an acceptable aperture distribution. The transformation may also be carried out by any one of a variety of open dielectric structures which are suitably tapered along their length. From a practical point of view there seems little to choose between the performance of these various forms of surface-wave antenna and the final choice rests with the overall system constraints; in particular the gain limitation experienced for the dielectric-rod and horn antennas seems to be a general feature.

6. Acknowledgments

Thanks are due to Mr. I. N. L. Gallett and Mr. E. Sigee for assistance with computations and measurements respectively, and to Professor M. H. N. Potok for his continued interest and advice.

7. References

1. Kiely, D. G., 'Dielectric Aerials' (Methuen Monograph, London, 1953).
2. Zucker, F. J., 'Surface and leaky-wave antennas', in Jasik, H. (Ed.), 'Antenna Engineering Handbook', Chap. 16. (McGraw-Hill, New York, 1961.)
3. Brown, J. and Spector, J. O., 'The radiating properties of end-fire aerials', *Proc. Instn Elect. Engrs*, **104B**, pp. 27–34, 1957.
4. James, J. R., 'Theoretical investigation of cylindrical dielectric-rod antennas', *Proc. Instn Elect. Engrs*, **114**, pp. 309–19, 1967.
5. Bach Andersen, J., 'Metallic and Dielectric Antennas' (Polyteknisk Forlag, Denmark, 1970).
6. Gallett, I. N. L., 'Radiation patterns of tubular dielectric structures'. European Microwave Conference, London, September 1969.
7. Prochazka, M., 'Die dielektrische hornantenna', *Hochfrequenztechnik u. Electroak.*, **68**, pp. 93–103, 1959.
8. James, J. R., 'Studies of cylindrical dielectric-rod antennas' Tech. Note RT41, Royal Military College of Science, June 1968.
9. James, J. R., 'Aperture coupling in cylindrical dielectric-rod antennas', *Electronics Letters*, **4**, pp. 39–41, 1968.
10. Marcuse, D., 'Radiation losses of the dominant mode in round dielectric waveguides', *Bell Syst. Tech. J.*, **49**, pp. 1665–93, 1970.
11. James, J. R., 'Some further properties of cylindrical dielectric-rod antennas', Tech. Note RT37, Royal Military College of Science, July 1967.
12. Silver, S., 'Microwave Antenna Theory and Design', p. 161 (McGraw-Hill, New York, 1949).
13. Barlow, H. M. and Brown, J., 'Radio Surface Waves', (Clarendon Press, Oxford, 1962).
14. Snyder, A. W., 'Radiation losses due to variations of radius on dielectric or optical fibres', *I.E.E.E. Trans. on Microwave Theory and Techniques*, **MTT-18**, No. 9, pp. 608–15, September 1970.
15. Barlow, H. M. and Brown, J. *loc. cit.*, p. 164.
16. Kharadly, M. M. Z. and Lewis, J. E., 'Properties of dielectric-tube waveguides', *Proc. I.E.E.*, **116**, No. 2, pp. 214–24, February 1969.
17. Potok, M. H. N., James, J. R. and Gallett, I. N. L., 'Dielectric antennas' 2nd Annual Research Report ELS/PR/No. 18, R.M.C.S. February 1970.
18. Silver, S., *loc. cit.*, p. 581.
19. Hamid, M. A. K., 'Diffraction by a conical horn', *I.E.E.E. Trans. on Antennas and Propagation*, **AP-16**, No. 5, pp. 520–8, September 1968.
20. James, J. R. and Gallett, I. N. L., 'Point-matched solutions for propagating modes on arbitrarily shaped dielectric-rods', *The Radio and Electronic Engineer*, **42**, No. 3, pp. 103–13, March 1972.
21. Boulanger, R. J. and Hamid, M. A. K., 'A new type of dielectric-loaded waveguide antenna', *Microwave J.*, December 1970, pp. 67–8.
22. Hamid, M. A. K. *et al.*, 'Radiation characteristics of dielectric-loaded horn antennas', *Electronics Letters*, **6**, pp. 20–1, 1970.
23. James, J. R., Gallett, I. N. L. and Sigee, E., 'New ideas and results on dielectric horns', Proceedings of the European Microwave Conference, Stockholm, August 1971.
24. James, J. R., 'Leaky waves on a dielectric rod', *Electronics Letters*, **5**, No. 11, pp. 252–4, May 1969.
25. James, J. R., 'Aperture coupling in surface wave antennas', Tech. Note RT33, Royal Military College of Science, February 1967.
26. Collin, R. E. and Zucker, F. J., 'Antenna Theory', Parts I and II. Inter-University Electronics Series (McGraw-Hill, New York, 1969).
27. Potok, M. H. N., James, J. R. and Gallett, I. N. L., 'Dielectric antennas' 1st Annual Research Report ELS/PR/No. 16, R.M.C.S., November 1968.
28. Hamid, M. A. K. *et al.*, 'A dielectric loaded circular waveguide antenna', *I.E.E.E. Trans.*, **AP-20**, 1, pp. 96–97 January 1972.

Manuscript first received by the Institution on 29th February 1972 and in final form on 27th March 1972. (Paper No. 1451/CC130).

© The Institution of Electronic and Radio Engineers, 1972

World Radio History

U.D.C. 535.23.08: 537.227

High Performance Pyroelectric Detectors

G. BAKER, B.Sc. *

D. E. CHARLTON, B.A. *

and

P. J. LOCK, B.Sc., Ph.D.†

Based on a paper presented at the Conference on "Infra-red Techniques" held in Reading from 21st to 23rd September 1971

SUMMARY

A survey of a number of possible materials for use in pyroelectric detectors shows that strontium barium niobate, lithium sulphate and triglycine sulphate (TGS) are the most interesting materials. Over a large range of frequencies and sensitive areas TGS and its derivatives offer the highest sensitivities. The major disadvantage of TGS, a tendency to depole, has been overcome by the addition of organic dopants. This can lead to biased hysteresis loops and improved dielectric performance. An increased range of operating working temperature is obtained by deuteration. A brief survey of some of the applications of pyroelectric detectors is given.

* *Mullard Southampton, Millbrook Industrial Estate, Southampton SO9 7BH.*

† *Formerly with Mullard Southampton, now with the Electrical Research Association.*

1. Introduction

Many years have passed since pyroelectricity was first observed but it is only within the last decade that any great interest has been shown in the use of this effect to detect infra-red radiation. Work up to 1969 has been described by Putley[1] and further developments have been reported by Schwarz and Poole.[2] This paper describes the present state of the art; in particular the influence of the material properties on the detector performance will be discussed.

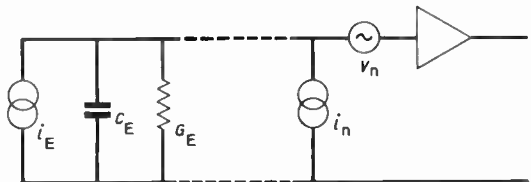


Fig. 1. Equivalent circuit of the detector.

The theory of the pyroelectric detector has been given initially by Cooper[3] and more completely in Putley's review.[1] It will be summarized here in order to establish the basis on which pyroelectric materials are judged for their suitability and the associated amplifier designed. Pyroelectric materials are those with a temperature-dependent spontaneous electrical polarization. Under equilibrium conditions this electrical asymmetry is compensated by the presence of free charges. If, however, the temperature of the material is changed at a rate faster than these compensating charges can redistribute themselves, an electrical signal can be observed. This means that the pyroelectric detector is an a.c. device, unlike other thermal detectors (such as thermistors) which detect temperature levels rather than temperature changes. It can be shown that at frequencies above that corresponding to the reciprocal thermal time-constant the current responsivity is constant; this leads to a good high frequency performance.

2. Noise Equivalent Power of Pyroelectric Detector

Noise in the system arises from the detecting element and the associated amplifier; in order to minimize it the device is used in the voltage mode. The equivalent circuit of the system is shown in Fig. 1. As the load on the signal current is capacitive above the electrical turnover frequency the voltage responsivity falls with increasing frequency. However, as the current noise

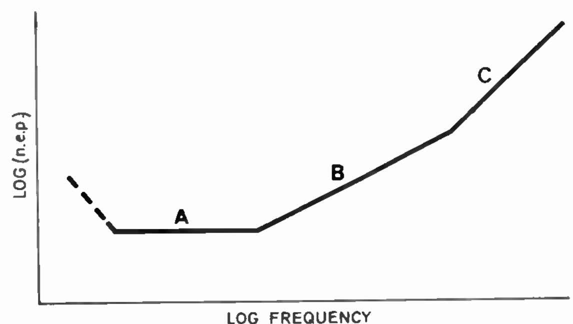


Fig. 2. Variation of noise equivalent power with frequency.

World Radio History

sources are affected in the same way, the signal/noise ratio is expected to be constant over a wide frequency range (up to the frequency where the voltage noise contribution dominates).

The noise equivalent power (n.e.p.) of the system can be shown to be given by

$$\begin{array}{l}\text{n.e.p.}\\ \text{(unit bandwidth)}\end{array} = \frac{1}{\eta}\sqrt{4kT\left\{\underbrace{TG_T}_{\text{thermal noise}} + \frac{C_T^2+(G_T/\omega)^2}{\lambda^2A^2}\left[\underbrace{G_E}_{\text{element noise}}+\underbrace{G_N+R_N(G^2+\omega^2C^2)}_{\text{amplifier noise}}\right]\right\}}$$

where η = emissivity of the element
G_T = thermal conductance between the element and its surroundings
C_T = thermal capacity of the element
λ = pyroelectric coefficient
A = electrode area
G_E = electrical conductance of the element
G_N = noise conductance of the amplifier $\left[\frac{i_n^2}{\Delta f} = 4kT\,G_N\right]$
R_N = noise resistance of the amplifier $\left[\frac{V_n^2}{\Delta_f} = 4kT\,R_N\right]$
G = total conductance across the element
C = total capacitance across the element (including strays)

If the thermal noise contribution is disregarded and only frequencies $\omega > 1/\tau_T$ are considered we get

$$\text{n.e.p.} = \frac{xdC_P}{\eta\lambda}\sqrt{4kT\{\underbrace{G_E}_{\text{element noise}}+\underbrace{G_N}_{\text{amplifier current noise}}+\underbrace{R_N(G^2+\omega^2\ C^2)}_{\text{amplifier voltage noise}}\}}$$

where x = element thickness
d = density
C_p = specific heat.

The three noise contributions vary with frequency in different ways. The amplifier current noise limited case gives an n.e.p. independent of frequency (region A, Fig. 2). It is found that the conductance of the element varies with frequency in a similar way for most pyroelectric materials ($\sigma \propto f$); this leads to an n.e.p. proportional to $\sqrt{}$(frequency) (region B, Fig. 2). It is usual to use an f.e.t. as a high impedance amplifier input, and this device typically has a voltage noise behaviour leading to n.e.p. proportional to $\sqrt{f}$ at low frequencies and f at higher frequencies (regions B and C, Fig. 2).

3. Choice of Detector Material

It is apparent from the above that it is not possible to separate the effects of the pyroelectric material and the amplifier. Additionally, figures of merit which can be derived for different materials for each of the three types of noise limited cases are of limited use as their individual applicability varies from frequency to frequency and detector area to detector area.

We have adopted the approach of computing the n.e.p. for a range of areas and frequencies for those pyroelectric materials for which sufficient data are available. In order to do this, figures for the noise contribution of the amplifier were estimated. The best voltage noise achieved to date in a j-f.e.t. is 20 nV/$\sqrt{}$Hz at 10 Hz dropping to 2 nV/$\sqrt{}$Hz above 1 kHz. F.e.t.s with leakage currents down to 0·5 pA can also be obtained. Although these two specifications are not as yet available in one f.e.t. they have been used as it is unlikely that both will be equally important for the majority of situations and therefore one or other can be relaxed.

When the computation is carried out it is found that three materials are superior to all the others. These are triglycine sulphate, strontium barium niobate, and lithium sulphate (Figs. 3(a) to (d). In choosing which of these is likely to be most suitable for a certain application several factors must be considered. These are:

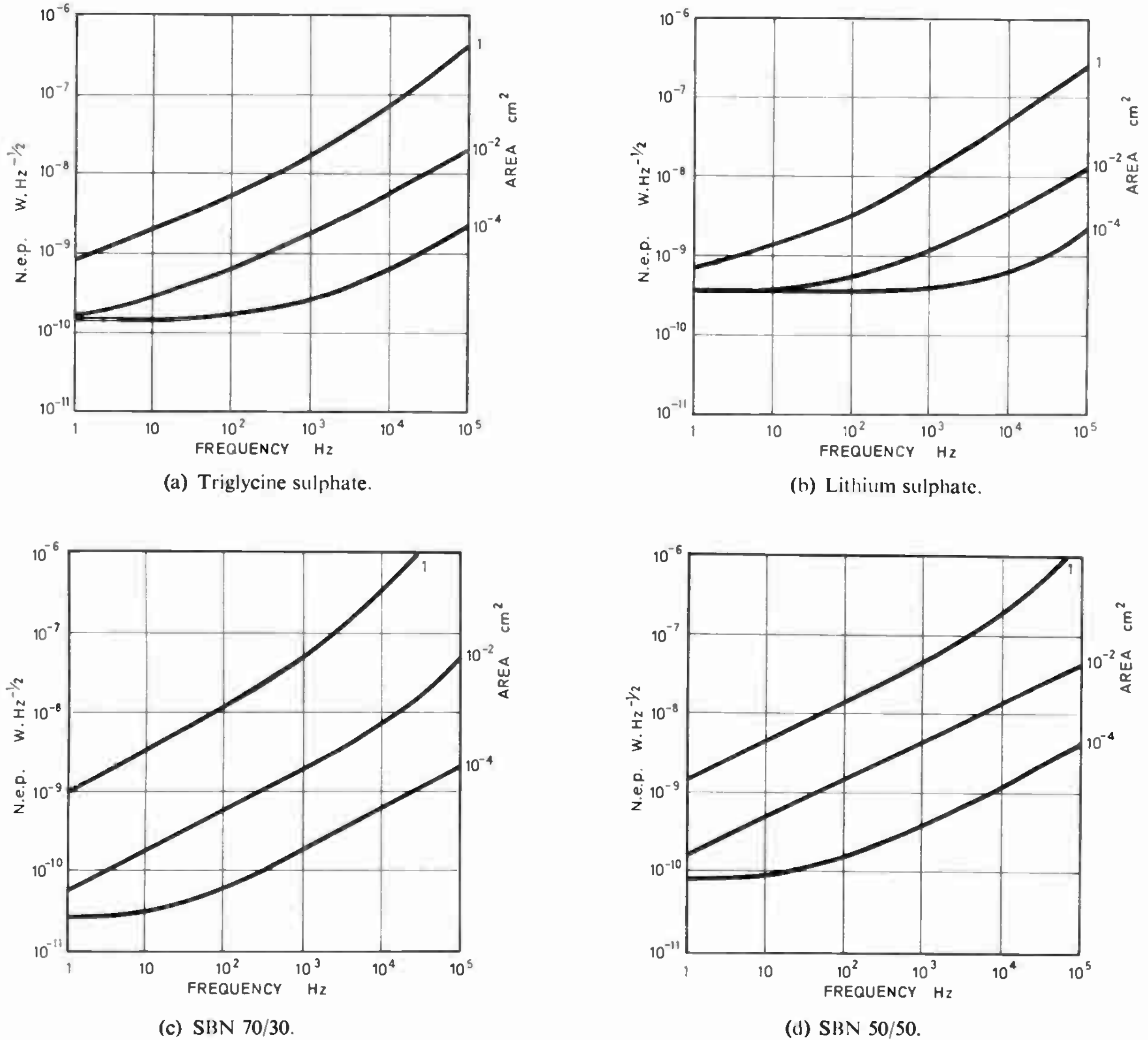
(1) bandwidth of operation,
(2) element area required,
(3) environmental conditions; e.g. temperature range.

Strontium barium niobate (SBN) is most suitable for use at low frequencies with small elements because of its high dielectric constant which leads to a relatively large voltage noise contribution from the f.e.t. Lithium sulphate is however most suitable for the larger area, higher frequency end of the range (because of its low dielectric constant and therefore a lower voltage noise contribution from the f.e.t.).

Over a large part of the frequency-area domain triglycine sulphate (TGS) is seen to give the best sensitivity. However, this material does have several disadvantages. These include a low Curie point (49°C) above which the pyroelectric effect is no longer present and, due to it being ferroelectric, a tendency to undergo partial reversal (depoling). These properties are also present in SBN, although somewhat higher Curie points can be achieved at the expense of some loss of sensitivity. Lithium sulphate has no Curie point and can be used up to higher temperatures; it is not ferroelectric and cannot therefore depole.

However, we have now developed a new material based on TGS which is completely resistant to depoling and has the further advantages of lower dielectric loss and dielectric constant.[4] When 1-alanine is added to the solution from which the TGS crystals are being grown these crystals have their properties and sometimes also their habits modified. The most significant change is seen in the P/E hysteresis loop (Fig. 4) where it can be seen that the doped material has a loop displaced along the field axis. This behaviour has been seen in other

World Radio History



(a) Triglycine sulphate.

(b) Lithium sulphate.

(c) SBN 70/30.

(d) SBN 50/50.

Fig. 3. N.e.p. *vs.* frequency curves.

ferroelectric materials (e.g. colemanite[5]) and is known to arise from the presence of an internal biasing field. As can be seen from Fig. 4(b), the polarization in the absence of an external field can have only one value. Therefore, whenever the material is below its Curie point and no large external depoling fields are present, it is fully poled. This effect is permanent; repeated cycling through the Curie point, both with and without external field applied, has no effect. Detectors made from this material regain their performance rapidly on cooling to below the Curie point even after prolonged periods above it.

The lower dielectric loss of the alanine doped material (LATGS) can be seen from the measurements shown in Fig. 5. Although there are large spreads in the values, the conductivity of the doped material is about an order of magnitude lower than that of pure TGS. The dielectric constant is also lower; values of ε' down to about 15–20 have been observed at audio frequencies. As the Johnson noise due to the conductance of the pyroelectric element is the major factor limiting the performance of TGS, this improvement in the resistivity is of considerable value. The lowering of the dielectric constant also leads to an improvement in performance, in this case at the higher frequency end of the range.

If the performance *vs.* frequency curve is recalculated for doped TGS and normalized for area by expressing as the specific detectivity D^*, the result is as shown in Fig. 6. This compares the predicted performance of the doped TGS pyroelectric detector with the performance found for other uncooled thermal detectors. Detectors have been made which fit this curve very well and in some cases surpass it; a 0·5 mm square element has been made which has a D^* of about $1{\cdot}8 \times 10^9$ at a few hertz. It is apparent from the figure that above about 20 Hz the pyroelectric detector can give a higher sensitivity than any other uncooled thermal detector. Only the Golay cell is capable of a higher D^* and this only in the frequency range 5–20 Hz. With further developments in the fabrication of thinner elements it is probable that some further improvements are possible.

World Radio History

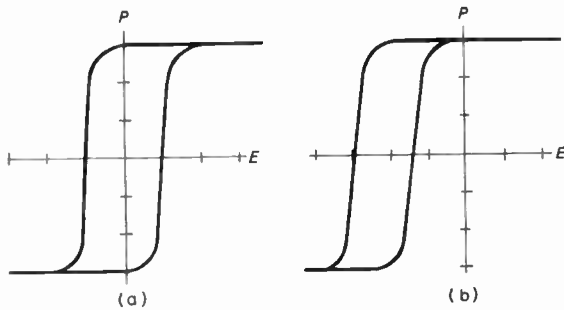


Fig. 4. Polarization *vs.* field hysteresis loops for (a) TGS and (b) doped TGS.
x-axis; 3 kV.cm^{-1} per division (*E*)
y-axis; 1 μC.cm^{-2} per division (*P*)

To extend the possible applications we have also developed detectors with deuterated TGS elements. These have a working range up to about 56°C rather than about 45°C as for pure TGS and can also be made with non-depoling material.

4. Applications

The present state of the art on pyroelectric detectors can therefore be summarized as:

(1) $D^* > 10^9$ cm Hz$^{\frac{1}{2}}$ W^{-1} at 10 Hz,
$> 10^8$ cm Hz$^{\frac{1}{2}}$ W^{-1} at 1 kHz.

(2) Detectors completely resistant to depoling.

(3) Temperature range up to 56°C.

Two fields in particular have now become accessible by virtue of these improvements. These are infra-red spectrophotometry and a variety of satellite applications.

Most infra-red spectrophotometers use either a Golay cell or a thermocouple-type detector, both of which are slow devices (5–20 Hz). The pyroelectric detector can be used as a direct replacement for either of these devices but a further possibility also arises: that of much faster spectral scanning. An improvement of an order of magnitude in the scan rate is quite feasible. This would give a 2–15 μm spectrum in about 10 seconds. Process control uses are therefore quite possible and further increases in speed could be achieved with some loss in resolution.

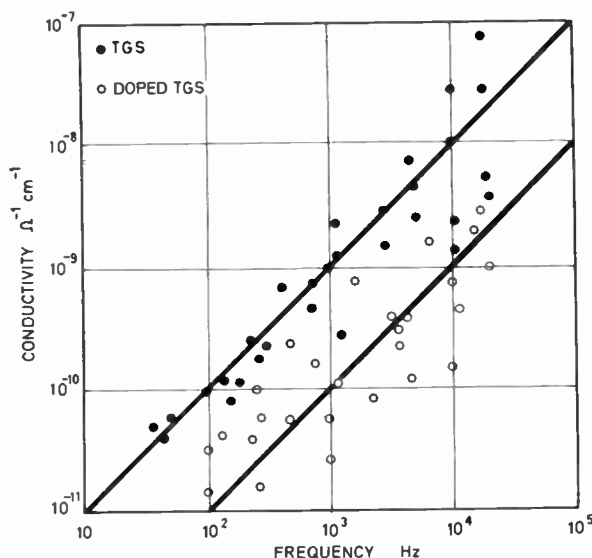


Fig. 5. Conductivity *vs.* frequency for pure and doped TGS.

Satellite instrumentation can require infra-red detection for several purposes, in particular horizon sensing and meteorological measurements such as the vertical temperature profile radiometer being developed for *Nimbus E*.[6] The prevention of depoling is of great value in these applications as alternative methods which have

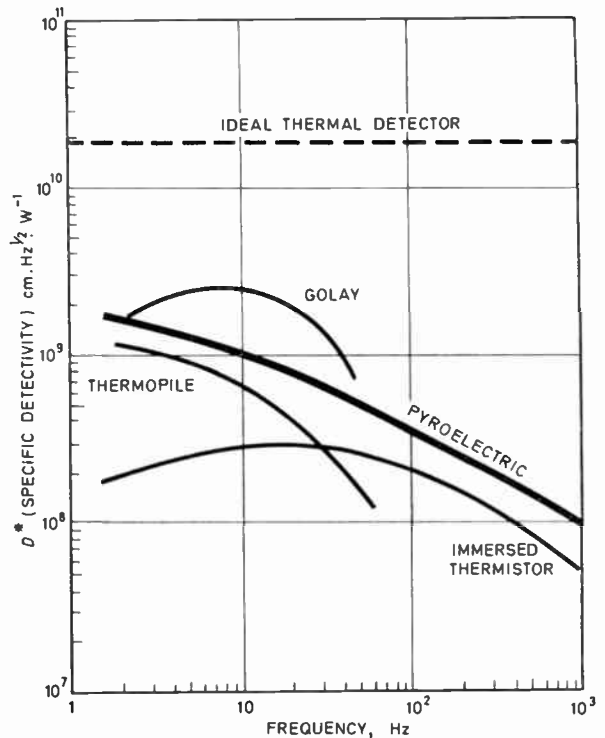


Fig. 6. *D** *vs.* frequency for thermal infra-red detectors.

been tried are less satisfactory and require more amplifier components. Another advantage of pyroelectrics is their very low power requirements (9 V, 5 mA), which is easily available in space systems.

Passive intruder alarms are another use to which the device can be put. Using a simple Cassegrainian optical system a range of detection in excess of 100 yards can be achieved for detection of persons crossing the field of view.

Laboratory applications include work with lasers, especially CO_2 (10·6 μm)[7] and HCN (337 μm). The detectors can be used for general problems such as power measurement and checking optical alignment. Bandwidths of operation out to megahertz and beyond can be simply achieved with some loss of performance or, by adding further electronic stages to the amplifier, retaining the high sensitivity if it is required. The possibility of heterodyne detection is also being investigated[8] but, ideally, lower dielectric loss materials are required for the high-frequency applications.

World Radio History

In this paper we have described the advances made in pyroelectric detectors and pointed out some of their present uses. The state of the art is now such that these devices must be strongly considered on all future occasions when an uncooled thermal detector is capable of meeting a requirement.

5. Acknowledgments

The authors wish to thank the Directors of Mullard Limited and the Ministry of Defence for permission to publish this paper. They also wish to thank their colleagues who have contributed to this work. Particular thanks are due to Dr. B. E. Bartlett, Dr. F. A. Capocci, Mr. L. R. King and Mr. R. A. Lockett of Mullard Southampton, Mr. A. E. Jenkinson and Dr. E. T. Keve of Mullard Research Laboratories, and Dr. E. Putley of R.R.E., Malvern.

6. References

1. Putley, E. H., 'The pyroelectric detector', in 'Semiconductors and Semimetals', Vol. 5, pp. 259–285. Willardson, R. K. and Beer, A. C. (Ed.) (Academic Press, New York, 1970).
2. Schwarz, F. and Poole, R. R., 'Performance characteristics of a small TGS detector operated in the pyroelectric mode', *Appl. Optics*, **9**, pp. 1940–1, 1970.
3. Cooper, J., 'A fast-response pyroelectric thermal detector', *J. Sci. Instrum.*, **33**, pp. 467–72, 1962.
   Cooper, J., 'Minimum detectable power of a pyroelectric thermal receiver', *Rev. Sci. Instrum.*, **33**, p. 92, 1962.
4. Lock, P. J., 'Doped triglycine sulphate for pyroelectric applications', *Appl. Phys. Lett.*, **19**, pp. 390–1, 1971.
5. Wieder, H. H., 'Ferroelectric properties of colemanite', *J. Appl. Phys.*, **30**, pp. 1010–18, 1959.
6. Ellis, R. J., *et al.*, 'Infra-red atmospheric temperature sounding from satellites', *The Radio and Electronic Engineer*, **42**, pp. 155–61, April 1972.
7. Hadni, A., Thomas, R. and Perrin, J., 'Response of a triglycine sulphate pyroelectric detector to high frequencies (300 kHz), *J. Appl. Phys.*, **40**, pp. 2740–5, 1969.
8. Eng, T. and Gudmundsen, R. A., 'Theory of optical heterodyne detection using the pyroelectric effect', *Appl. Optics*, **3**, pp. 161–6, 1970.
   Leiba, E., 'Optical heterodyne system with pyroelectric detectors', *C.R. Acad. Sci. Paris*, B, **268**, pp. 31–3, 1969.

Manuscript first received by the Institution on 17th June 1971 and in final form on 4th April 1972. (Paper No. 1452/CC 131).

© The Institution of Electronic and Radio Engineers, 1972

World Radio History

U.D.C. 537.311.33:537.534.79

Ion Implantation in Semiconductor Device Technology

J. STEPHEN, Ph.D., C.Eng., M.I.E.E.*

SUMMARY

The role of ion implantation in the fabrication of semiconductor devices is reviewed with special reference to silicon planar technology. The applications of ion implantation to various devices is discussed in detail. A summary is given of the present state of the art in the compound semiconductor field.

* *Electronics and Applied Physics Division, Atomic Energy Research Establishment, Harwell, Didcot, Berkshire.*

Contents

1 Introduction
2 The formation and electrical behaviour of ion implanted layers
 2.1 Method of ion implantation
 2.2 Profile of implanted ions
 2.3 Electrical properties of implanted layers
 2.4 Implanted junction properties
 2.5 Masking against the ion beam
 2.6 Implantation through oxide layers
 2.7 Summary
3 Applications of ion implantation
 3.1 Controlled surface doping
  3.1.1 Diffusion source
  3.1.2 High-value resistors
  3.1.3 Channel regions of metal oxide silicon transistors
  3.1.4 Large area applications
 3.2 Buried layers
  3.2.1 Varactor diodes
  3.2.2 Impatt diodes
  3.2.3 Isolating layers
  3.2.4 Narrow base bipolar transistors
 3.3 Abrupt junctions
  3.3.1 Impatt diode structure
  3.3.2 Avalanche photodiodes
 3.4 Precision alignment of metal oxide silicon transistors
 3.5 Doping of imperfect semiconductor materials
 3.6 Doping of high-resistivity silicon
4 Compound semiconductors
5 Production equipment for ion implantation
6 The future
7 Acknowledgments
8 References

1. Introduction

The development of silicon planar technology over ten years ago, revolutionized the fabrication of semiconductor devices. It opened the way to integrated circuits and the *m*etal *o*xide *s*ilicon (m.o.s.) technology which, during the past year, have together given rise to some extremely complex digital circuits on a single piece of silicon. During this period of development, ion implantation, an alternative method to thermal diffusion for introducing impurities into semiconductors, has been under investigation and it was soon realized that it offered some attractive advantages over diffusion when making certain devices. This paper is concerned with the fabrication of semiconductor devices by ion implantation, with a special reference to silicon.

Table 1 shows the chronological development of the semiconductor technology and the ion beam technology. Shockley, in 1954, patented the use of ion beams for producing the buried base layer in a bipolar transistor.[1] In the patent he foresaw many of the advantages and problems of ion implantation. The coalescence in the

World Radio History

Table 1 Chronological development of the semiconductor and ion-beam technologies

Semiconductor technology	Date	Ion-beam technology
Silicon point contact crystal diode	1940	Development of Calutron for separation of fissile materials.
Point contact transistor	1948	
Crystal growing; Ge and Si	1950	
Grown junction Ge transistor	1952	Nuclear research including β and λ spectroscopy, fission and nuclear decay studies.
Alloy transistor	1952	Influence of ion bombardment studied by Ohl.
Shockley patent on ion implantation[1]	1954	
Diffused transistor (silicon)	1956	
Oxide masking for diffusion	1957	
Surface stabilization by oxide layer	1959	
Planar process Epitaxial layers M.o.s. transistors	1960	Fission chemistry and radiation damage
Integrated circuits	1961 1962	Study of penetration of high energy heavy ions into solids
Ion implantation for nuclear particle detectors and solar cells	1965	Development of Freeman source with fine focus ion beam. Channelling experiments
Mo.s.t. arrays	1966	
Ion implanted m.o.s.t. u.h.f. implanted transistor	1966 1967	Ion implantation at MeV energies
Large scale integrated circuits Multi-layer interconnexion techniques	1967	
Ion implanted m.o.s. arrays	1968	Development of commercial machines for implantation
Silicon gate m.o.s. arrays	1969	
Complex ion implanted m.o.s. arrays placed on the market	1970	Commercial ion implantation equipment available
Threshold control used in commercial implanted m.o.s. arrays	1970	
High yield all-implanted m.o.s. array	1971	
Ion implanted silicon vidicon target	1971	

early 1960s was due to the interests of physicists who were studying the interaction of high energy particles with matter. For their experiments they required solid-state detectors with thin entry windows for detecting these particles.[2] The application of ion implantation to microelectronics started in 1964/5 when it was realized that buried layers and p-n junctions could be made. The bipolar transistor was one of the first devices to be considered,[3,4] and this was followed by the *m*etal *o*xide *s*ilicon *t*ransistor (m.o.s.t.) with the self aligned gate.[5,6]

The semiconductor specialist may well ask the question: why introduce another method of doping semiconductors when the existing epitaxial and thermal diffusion techniques are being perfected and extended to meet present and future demands for new devices with better yields?

One of the main purposes of this paper is to attempt to answer this question and to demonstrate that ion implantation is making a contribution to the production of some present-day semiconductor devices as well as having a promising development potential for new devices and more complex circuits.

In the fabrication of semiconductor devices by the planar technology, the first step is to introduce electrically active impurities into an ultra-pure semiconductor to form discrete isolated p and n regions with their exposed surfaces passivated by an insulating layer to reduce leakage currents. Next, the p and n regions are contacted and interconnected with metal layers running over the insulating surface. By this method, devices ranging from simple diodes to complex integrated circuits, incorporating many thousands of transistors, can be made.

There are four methods of introducing impurities into a semiconductor crystal:

(i) during the growth of the crystal or an epitaxial layer,

(ii) alloying with a suitable metal such as indium or aluminium,

(iii) thermal diffusion and by

(iv) ion implantation.

Only the last two methods permit accurate control over the *concentration* and *position* of the impurities.

Before considering the part that ion implantation can play in the fabrication of semiconductor devices, particularly those made by silicon planar technology, some of the basic characteristics of implanted impurities must be considered. This is a very broad subject if compound semiconductors are included. The behaviour of boron and phosphorus in silicon will be discussed in detail as these two impurities have figured prominently in the initial exploitation of ion implantation in device fabrication. Many of the statements made concerning the implantation of silicon apply to the other semiconductors. A short summary of the present state of ion implantation into compound semiconductors is included at the end of the paper. A comprehensive bibliography of ion implantation has been prepared at Harwell by Morgan and Greenhalgh.[7]

2. The Formation and Electrical Behaviour of Ion Implanted Layers

2.1 *Method of Ion Implantation*

An ion implanting machine is comprised of three basic components; a source, an analyser and a target chamber.[8] The dopant atoms are ionized in the source using a supporting plasma of argon ions. Ions from the source are extracted through an aperture, accelerated to a high energy and then mass-analysed, usually by a strong magnetic field which deflects the various ions through different angles. The pure dopant ion beam is focused to enter the target chamber and bombard the semiconductor wafer. The beam is scanned over the surface of the wafer to achieve a uniform implantation dose. The total dose is measured by the charge deposited on the semiconductor.

World Radio History

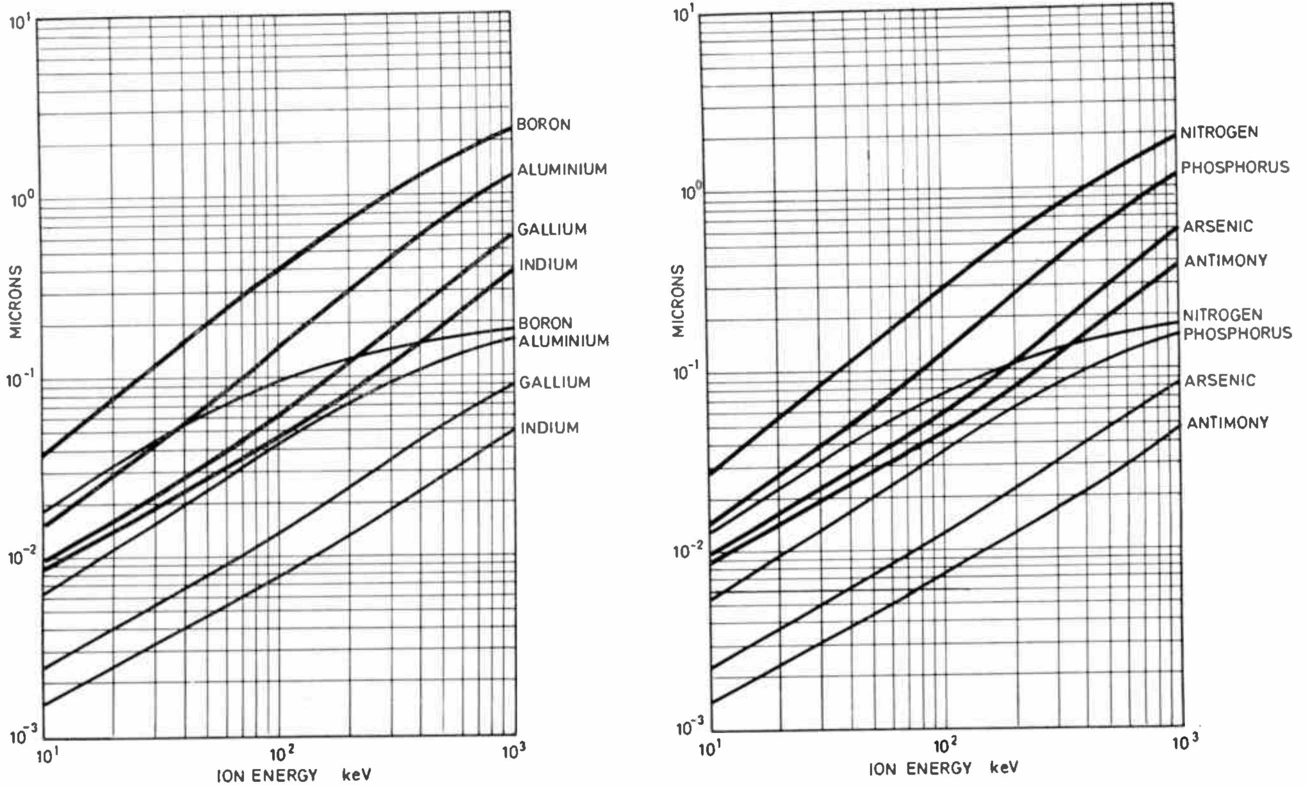


Fig. 1. Projected ranges (upper curves) and standard deviations (lower curves) of acceptors and donors in silicon.

The depth and concentration of the implanted atoms are determined by the ion, its energy, the total dose, the substrate material and its crystal orientation.

The implantation procedure is independent of the substrate temperature and therefore the restrictions imposed by thermal diffusion on certain dopants and masking materials no longer apply. Thus, a very wide range of dopants and masking methods can be employed for implantation. Due to the directional nature of the ion beam, the areas to be implanted on a wafer can be defined laterally very accurately by using an in-contact mask.

2.2 *Profile of Implanted Ions*

The profile in depth of implanted ions in an amorphous solid is Gaussian in shape and Lindhard, Scharff and Schiøtt[9] produced a theory, known as the L.S.S. theory, for the range of implanted ions. By applying their formulae it is possible to calculate the range profiles of a large number of ion-substrate pairs to an accuracy of about 20%. Johnson and Gibbons[10] have produced a series of tables covering most common semiconductors and dopants for ion energies from 10 to 1000 keV, while Smith[11] has produced additional data in the form of curves. Figure 1 shows the ranges and standard deviations of the four acceptor and four donor impurities in silicon. A phenomenon known as channelling occurs when the ion beam is aligned to a principal axis of the semiconductor crystal particularly the ⟨110⟩ and ⟨111⟩ axes. The range can be increased by over an order of magnitude for ions such as phosphorus as shown by Dearnaley.[2]

Most ion implanting machines have a maximum acceleration potential of about 200 kV and implantations

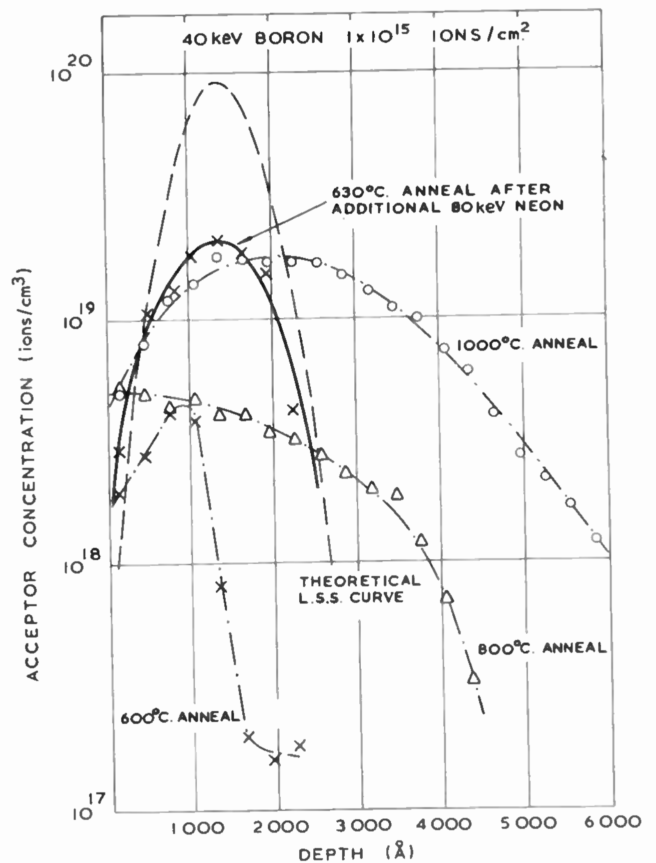


Fig. 2. Profiles of 40 keV room temperature boron implantation in silicon annealed at various temperatures.

World Radio History

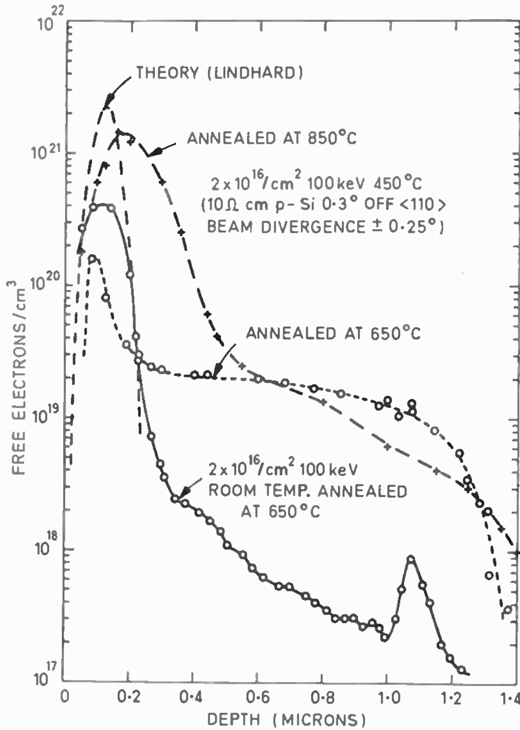


Fig. 3. Profiles of 100 keV room temperature and 450°C phosphorus implantations in silicon annealed at various temperatures (after Shannon *et al.*[13]).

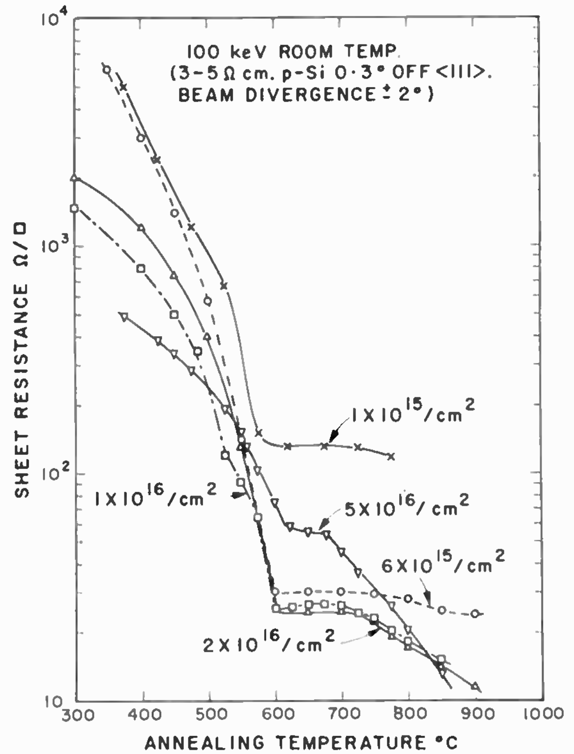


Fig. 5. Annealing curve for 100 keV phosphorus implant in silicon for a range of doses (after Shannon *et al.*[13]).

at these energies result in layers that are 1 μm or less in depth. Deeper implants can be achieved by using higher energies or deliberately channelling ions into the crystal. Typical profiles for boron implanted at room temperature and annealed at various temperatures are shown in Fig. 2 (after Blamires *et al.*[12]). The electrical activity increases

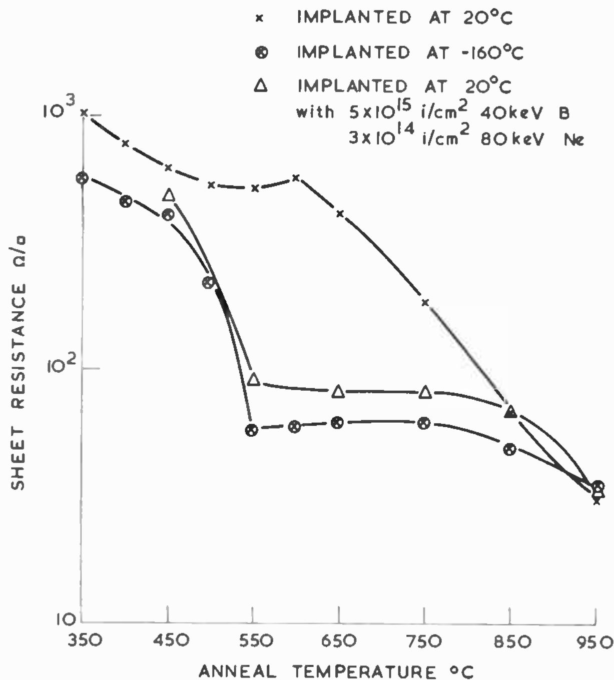


Fig. 4. Annealing curve for 40 keV boron implant in silicon for two implantation temperatures (after Blamires *et al.*[12]).

with temperature and thermal diffusion is evident from the increasing depth particularly for the 1000°C profile. The combined boron and neon profile is an example of the implantation of an electrically inert ion moderating the annealing behaviour of boron. The three phosphorus profiles in Fig. 3 have two distinct regions, a Gaussian distribution near the surface agreeing with the L.S.S. theory and long channelling tails extending nearly ten times deeper into the silicon. The crystal axes and the ion beam were accurately aligned to enhance the channelling for all three profiles.

2.3 *Electrical Properties of Implanted Layers*

When an energetic ion is stopped in a solid some of its energy is dissipated by displacing atoms of the solid producing radiation damage. The degree of damage depends on the solid, the ion, its energy, the total ion dose and the temperature during implantation. The damage may be so severe as to render the implanted volume amorphous. It is necessary to anneal the implant at a suitable temperature for about 30 minutes to allow the crystal structure to recover and a fraction of the implanted impurities to occupy substitutional sites and become electrically active. The damaged crystal structure in silicon recovers at 650°C and germanium at about 400°C. Figures 4 and 5 show the recovery of implanted silicon, on annealing, as a *decrease* in the sheet resistance (i.e. increase in conductivity) for boron[14] and phosphorus[13] implantations. A pronounced recovery is seen for both dopants at temperatures below 650°C and the sharp recovery at 550°C for boron implanted at 77°K (Fig. 4) was first observed by Davies.[15] The sheet

World Radio History

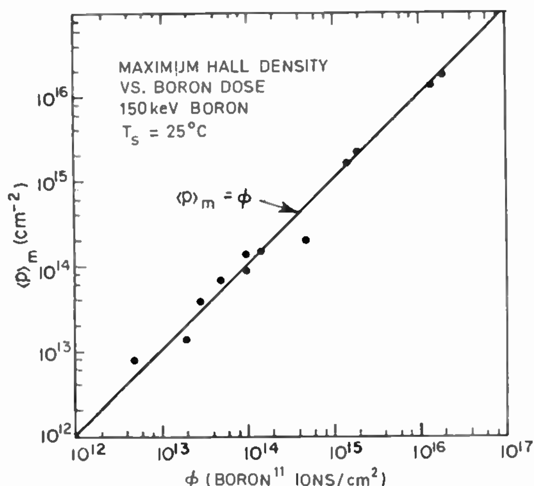


Fig. 6. Relation between the number of electrically active centres and the number of implanted ions (after MacRae[16]).

resistance obtained after annealing at 500°C is of particular importance for the self-aligned gate m.o.s.t. as this is the maximum temperature to which the wafer can be exposed without the aluminium conductors chemically reacting with the silicon dioxide insulator.

MacRae[16] has made measurements of the electrical activity in boron (Fig. 6) and phosphorus implants that indicate a 1:1 correspondence between the number of implanted ions and the number of electrically active impurity centres after annealing at 950°C.

Recently, measurements have been made by Pickar and Dalton[17] and Davies and Roosild[18] of the minority carrier lifetime in implanted silicon and its recovery on annealing. Shallow boron diffused p^+-n diodes were used in both series of experiments, and carbon and other light ions were implanted at very high energies into the lightly doped n region beyond the junction. The carrier lifetimes were measured directly by the charge extraction method[18] (Fig. 7) and indirectly by measuring the leakage current, at a fixed reverse bias, which is approximately proportional to the number of recombination centres[17] (Fig. 8). Figure 7 shows the recovery of the minority carrier lifetime with annealing temperature for increasing doses of carbon ions, while Fig. 8 shows the decrease of the leakage current, on annealing, to nearly its original value for a 300 keV, $1 \times 10^{15} C^+/cm^2$ implant. These results are consistent and show that for light ions with doses up to about 10^{13} ions/cm² complete recovery is possible by annealing above 600°C. Only partial recovery of carrier lifetime is obtained for larger ion doses which exceed 10^{14} ions/cm² and the lifetime in silicon that was made amorphous is low, about 10^{-8}s as shown by MacRae.[16]

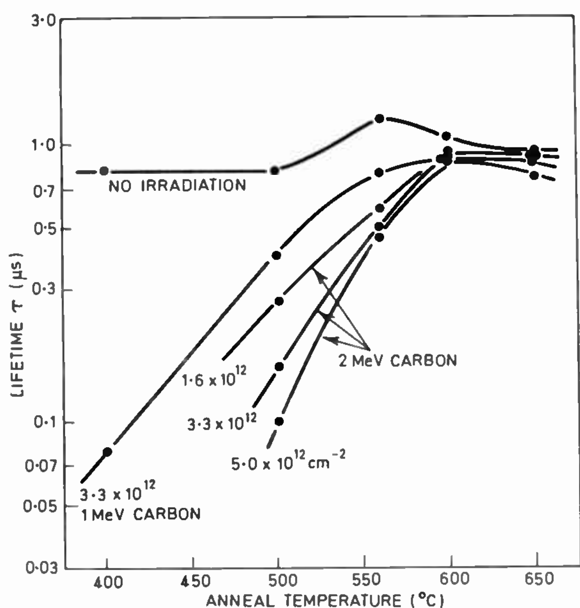


Fig. 7. Recovery of minority carrier lifetime on annealing (after Davies and Roosild[19]).

2.4 *Implanted Junction Properties*

In many devices the characteristics of the implanted layer and the junction formed between it and the substrate are of equal importance. Implanted p-n junctions have been reported for some time but only recently have more detailed studies been made of their characteristics.

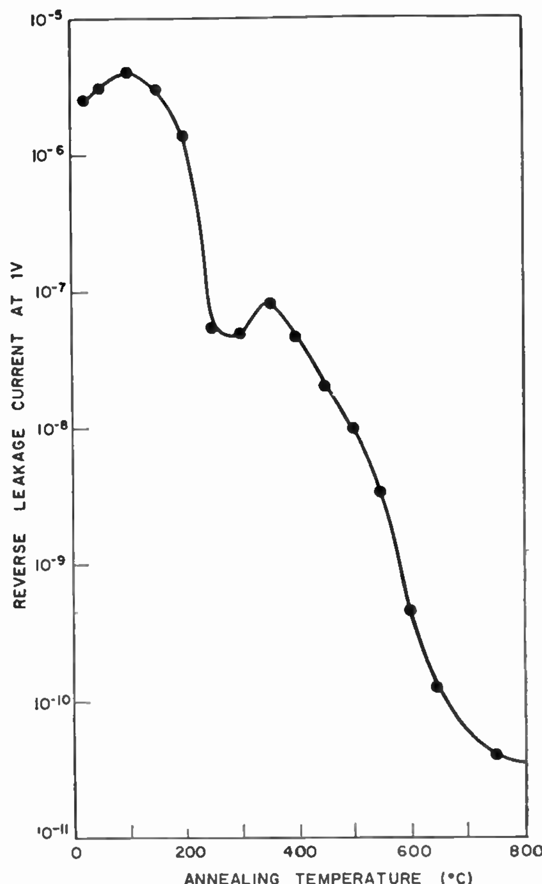


Fig. 8. The recovery of the leakage current on annealing (after Pickar and Dalton[17]).

From the nature of the shape of an implant profile it is clear that the junction is hyperabrupt, i.e. dN/dZ is very steep, (Figs. 2 and 10). The edges of the implant end abruptly immediately below the edge of the mask and there is no sidewards penetration of the dopant under the mask edge as occurs with diffusion. The sharp edge in the implant profile produced by the mask leads to high electric fields and reduced breakdown voltages.

World Radio History

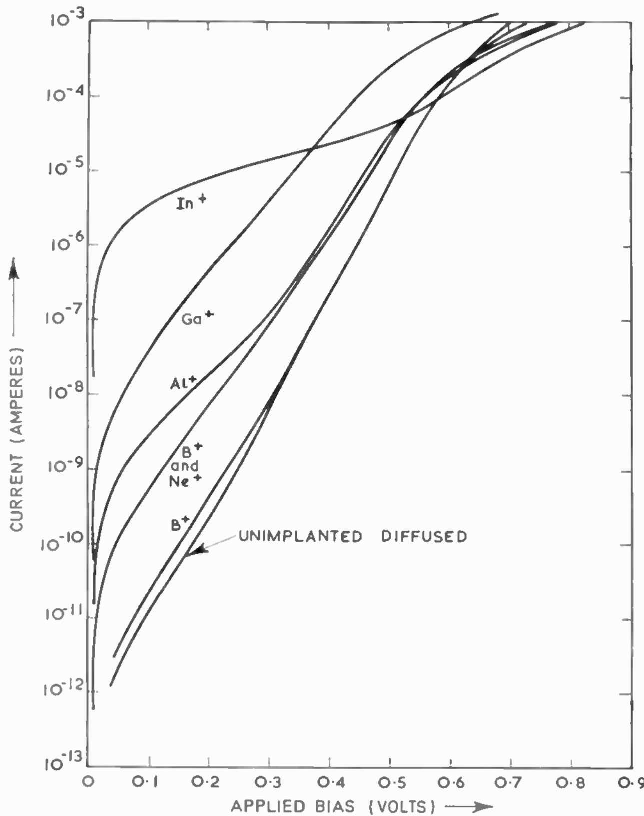


Fig. 9. Forward characteristics of boron, aluminium, gallium and indium implanted junctions (after Stephen and Grimshaw[22]).

Shannon and Ford[19] have measured the leakage of boron and phosphorus implanted junctions and their results for boron are in agreement with those of MacDougall *et al.* [20] They compare favourably with diffused boron junctions of the same area being about two to three times greater. When the annealing temperature was limited to 500°C a leakage current of 5×10^{-9} A/cm^2 was found for boron while a much larger current increasing linearly with dose was measured for phosphorus implants. Recently Pickar *et al.*[21] have reported making silicon vidicon targets by boron implantation and they obtained a total leakage current less than 10 nA (at −12V) for the array of 800 000 8 μm diameter diodes and a lifetime in the space charge region of over 200 μs. This result is as good or better than can be achieved by diffusion and it is a very critical test which confirms the uniformity and freedom from defect generation which is attributed to this doping method.

Stephen and Grimshaw[22] have studied the p-n junctions formed by implanting, separately, the four acceptor impurities and comparing these junctions with boron diffused p-n junctions on the same chip (Fig. 9). The annealing temperature was limited to 500°C to simulate m.o.s. implantation technology conditions. The boron results agree with those already discussed and the forward characteristics of boron implanted and diffused junctions were very similar with two exponential regions corresponding to the recombination current component followed by the diffusion component (Fig. 9). The large increase in leakage current for phosphorus implants observed by Shannon and Ford[19] at low annealing temperatures was also found for the three heavier acceptor ions Al, Ga and In. Their forward characteristics (Fig. 9) are no longer exponential and the reverse leakage currents are 2 to 3 orders of magnitude larger. These effects are most probably associated with the residual damage remaining in these partially annealed implants. It has been reported by Crowder and Title[23] and others that the damage profile associated with an implant is *shallower*, i.e. it has a shorter range than the implant profile. Thus, in an implanted junction, the concentration of damage should not be high at the metallurgical junction or in the space charge region beyond it and normal junction behaviour is to be expected. The experimental results for boron suggest that the concentration of damage is low but that for the heavier ions, with only partial annealing, there are sufficient damage centres in the junction to disturb its behaviour. The nature of the electrical centres produced by the damage is not clearly understood but it is expected that some form of recombination-generation action occurs, possibly involving more than one donor or acceptor level. In the case of the heavier ions, therefore, the annealing temperature must be high enough to allow the implanted silicon to recover, thus reducing the number of damage centres and at the same time increasing the fraction of electrically active implanted atoms. The annealing temperature must exceed 650°C for silicon.

2.5 *Masking against the Ion Beam*

As the whole area of the semiconductor wafer is scanned by the ion beam during implantation, it is necessary to define the areas, to be doped, by in-contact masking. It is possible to use out-of-contact masks for certain applications but for microelectronic devices in-contact masking is essential. The methods employed for preparing the masks are very similar to those used for masking against thermal diffusion. The ideal mask for ion implantation should have the following characteristics:

(i) it must be opaque to the ion beam without being too thick (the denser the medium the better is its stopping power),

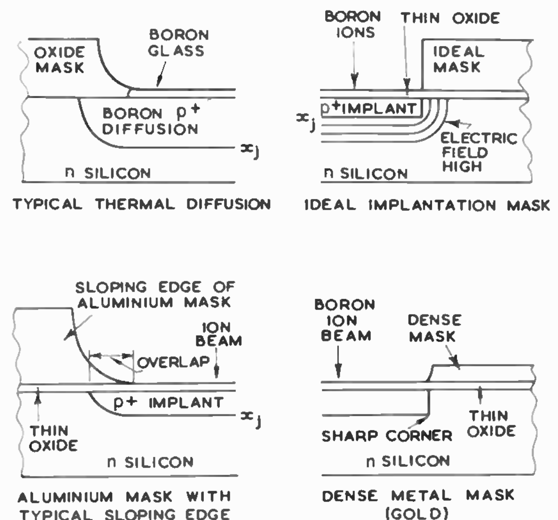


Fig. 10. Cross-sections of a diffusion and three implantation masks.

World Radio History

(ii) it must be easy to apply, photoengrave and remove,

(iii) it is not usually necessary for it to withstand high temperatures as it can be removed before annealing in most cases, and

(iv) the mask must not contaminate or damage the surface it is masking.

Figure 10 shows in cross-section, a typical thermal diffusion mask and three implantation masks. Two important differences between the end of a diffusion and implantation can be seen by comparing the upper two diagrams. The diffused junction edge is protected by the oxide mask due to sidewards diffusion whereas in the implantation case the junction is coincident with the mask edge. This has distinct advantages for the implanted m.o.s. transistor (Sect. 3.4), and it must be taken into account when cutting contact holes to avoid the metallization short circuiting the junction. The ideal implantation mask is approximated to by a dense metal mask. Many masks such as aluminium have sloping edges with the result that the ions penetrate the thin edge of the mask and produce an overlap. Figure 11 is a scanning electron micrograph of a typical engraved aluminium mask with a sloping edge. The serrated appearance of the edge is very common.

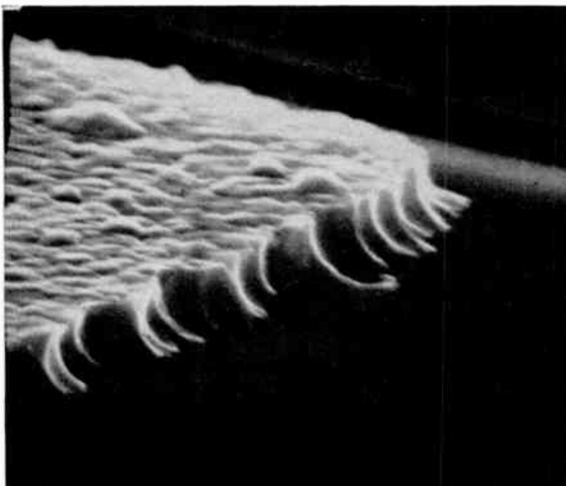


Fig. 11. Scanning electron micrograph of the edge of an aluminium mask about 1·4 μm thick. The sloping nature of the edge is clearly shown and the serrated effect is typical of etched aluminium layers.

There are no special problems in choosing masking materials. SiO_2, Si_3N_4, molybdenum and silicon, which are used for diffusion, have been successfully used for implantation together with aluminium, gold and photoresist. The thickness of the masking material is dependent on the ion, its energy and the stopping power of the masking material. Typical mask thicknesses used in practice vary between 0·1 and 2 μm. The ranges of 40 keV boron and phosphorus ions in aluminium and SiO_2 are approximately 20% shorter than those in silicon. The use of metal masks has the added advantage that the charge and energy deposited by the ion beam is conducted away from the surface of the wafer.

Wolf *et al.*[24] have combined three advanced techniques to produce a junction field-effect microwave switch made to replace a p-i-n diode in microwave switching applications. A very fine geometry is essential for this device, and a thin gold mask was used to define the area to be implanted. The apertures in the gold mask were formed by ion etching through a photoresist pattern previously exposed by electron beam. This approach indicates a probable evolutionary trend in masking for ion implantation when the dimensions required are smaller than those obtainable by optical printing methods.

2.6 *Implantation through Oxide Layers*

Insulating layers of silicon oxide or nitride are used in the fabrication of many devices for surface passivation and masking against the ion beam. It is important to consider the behaviour of these insulating layers when charged ions are deposited in or through them.

When a high energy ion beam passes through an insulating layer, it produces ionization in the insulator and makes it sufficiently conducting to prevent it charging up and failing by dielectric breakdown. Experiments have shown that thin oxide layers have been implanted by boron or phosphorus ions show a shift in their capacitance/voltage characteristics due possibly to charged states formed in the oxide. The oxide anneals rapidly above 300°C (Shannon and Ford[19], Glotin *et al.*[25]) and its electrical properties recover satisfactory for use as the gate insulator in a metal-oxide-silicon transistor (Aubuchon[26], MacDougall *et al.*[27]). No oxide problems have been experienced when implanting to align the gate of m.o.s.t., (Bower and Dill[5], Bower *et al.*[6], Shannon *et al.*[28]).

The position is different for thick oxide (oxide thickness greater than ion range) as the deposited charge cannot leak away through the oxide, as in the thin oxide case. It therefore accumulates at the surface and has to move *laterally* to the nearest 'sink' such as a thin oxide area, a contact hole or an aluminium conductor. The exact mechanism is not understood at present but experience has shown that permanent damage to the oxide does not occur in integrated circuits as the 'sinks' are very numerous and closely spaced. The charging of the oxide can be prevented by flooding the target with low energy electrons from a hot filament placed nearby (Bower[29]), but this complicates the measurement of ion dose.

2.7 *Summary*

The technique for implanting impurities in silicon have been described, and some of the electrical properties of boron and phosphorus layers discussed. It has been shown that ion implantation is an efficient doping method which is compatible with silicon planar technology as all the other processing steps are very similar. As will be indicated later, diffused and implanted layers can be combined easily in one device. Some of the advantages offered by implantation to the silicon technology are:

(i) wide range of dopants,

(ii) control over profile independent of temperature,

(iii) precise control over dose (impurity density),

World Radio History

(iv) precise control over lateral dimensions (by in-contact masking),

(v) uniformity of impurity concentration over the wafer,

(vi) reproducibility, from wafer to wafer, and from batch to batch,

(vii) low processing temperatures are feasible thereby reducing considerably the diffusion of unwanted impurities, maintaining carrier lifetime and restricting the propagation of crystal damage due dislocations etc. which arise at high temperatures,

(viii) implantation profile is independent of dislocation density,

(ix) clean environment with no exposure to very high temperatures and the risk of surface contamination and attack.

3. Applications of Ion Implantation

The areas in which it is technically feasible for ion implantation to make a positive contribution to the silicon planar technology are:

(i) controlled surface doping,

(ii) buried layers,

(iii) abrupt junctions,

(iv) precision alignment,

(v) doping of imperfect semiconductor material and

(vi) doping of high resistivity silicon.

Each of these topics will be discussed with appropriate examples.

3.1 *Controlled Surface Doping*

The dopant ions can be implanted uniformly over the wafer surface and the total number of ions deposited per unit area can be measured accurately over a very wide range of surface concentrations (10^{11} to 10^{16} impurities/cm^2). This degree of control is unique to ion implantation and makes it a very attractive technique.

3.1.1 Diffusion source

Roughan *et al.*[30] have described an implanted diffusion source and they give three potential advantages:

(i) it can be introduced at low temperatures, thereby preventing thermal diffusion,

(ii) the amount of diffusant can be more precisely controlled and

(iii) dopants can be used which are difficult if not impossible to introduce by thermal methods (for example, nitrogen).

They discussed results for boron, nitrogen and indium implantations in silicon. The boron implants had essentially bulk properties after a 800°C anneal while the nitrogen implants produced 1% electrical activity at 900°C and 100% at 1200°C. The diffusion constant for nitrogen was found to be $7{\cdot}6\times10^{-13}$ cm^2/s at 1100°C which compares with the diffusion constant for boron of about $3{\cdot}3\times10^{-13}$ cm^2/s at 1100°C. The activation energy of the nitrogen donor level was between 33 and 44 meV. In the case of indium, the depth profiles show the experimental peaks deeper than the calculations by the L.S.S. theory and with pronounced channelled tails. The boron and indium implants were thermally diffused in and the experimental junction depths were significantly deeper than those calculated from diffusion theory.

MacRae[16] has said that the production of low-power complementary m.o.s.t. structures has been difficult using thermal diffusion for the p-type 'tubs', in the silicon substrate, for the n-channel m.o.s.t.s. This difficulty has been largely overcome by implanting controlled amounts of boron which are subsequently annealed and thermally diffused in while growing the gate oxide (see also Dill *et al.*[88]).

3.1.2 High-value resistors

MacDougall *et al.*[20] have described making high value resistors by implanting boron through oxide windows and annealing at 950°C. The sheet resistances were shown to be inversely proportional to the dose and varied from 0·8 kΩ to 11 kΩ/square. The temperature coefficient of resistance (t.c.r.) changed from 800 to 4000 parts/10^6 (Fig. 12). The higher value resistors were non-linear with increasing applied voltage due to pinch off. The implantation of boron and phosphorus into one resistor was tried with the object of producing a temperature compensated resistor. The conductivity σ of a layer produced in this manner is given by

$$\sigma = q(N_A - N_D)\mu$$

where N_D and N_A are the donor and acceptor concentrations respectively and the mobility μ is an inverse function of $N_D + N_A$, being low for large impurity concentrations. These parameters are less sensitive to temperature changes when N_D and N_A are large, but it is then difficult to maintain close control over the difference $N_A - N_D$ (Fig. 12).

A direct comparison between boron-diffused and boron-implanted resistors in a linear integrated circuit was made by MacDougall *et al.*[20] The saving in area was very apparent and the leakage current of the implanted resistors to the substrate was of the order of 2 nA. The process was stated to be compatible with high β bipolar transistor technology.

Oosthoek *et al.*[31] have studied the variation of the temperature coefficient of resistance for boron and gallium implantations into silicon with annealing temperature. The t.c.r. was less than 1000 parts/10^6 for sheet resistances between 1 kΩ and 30 kΩ per square. They concluded that the low t.c.r. was a result of the high positive coefficient of charge carrier density produced by the lattice disorder from the implantation, just balancing the negative coefficient of mobility caused by lattice scattering.

Rosendal[32] has studied boron-implanted resistors for a range of doses and two ion energies. He reported that sheet resistances between 50 Ω and 50 kΩ per square were possible and that the uniformity was improved to a degree where the photo-lithography and etching tolerances dominate. The t.c.r. could be controlled by annealing in the range 400°C to 600°C for ion doses exceeding 10^{14} ions/cm^2. He drew the same conclusions as Oosthoek *et al.*[31].

World Radio History

Fig. 12. Variation of sheet resistance with dose and applied voltage and t.c.r. with temperature (after MacDougall *et al.*[57]).

The use of damage to deliberately reduce the mobility in implanted resistors has been reported by Nicholas and Ford.[33] They implanted neon into an existing boron-implanted resistor and reduced the mobility by a factor of 10 and obtained sheet resistances of the order of 100 kΩ/square with improved linearity.

Dill *et al.*[34] have incorporated implanted resistors in a 64-bit m.o.s.t. shift register in place of the usual enhancement m.o.s.t. loads. Figure 13 shows two cells of the register with the large pull-up resistors (2 kΩ/square). The gates are 5 μm long. The circuit operates normally at 10 MHz and can be made to work at 30 MHz by increasing the supply voltage and consequently the power dissipation.

Hodges *et al.*[35] have described an alternative approach to the implanted resistor. They used ion implantation to alter the doping immediately below a rhodium p-silicon Schottky diode. The reverse leakage current increases with the doping level. The diode can be used as a non-linear resistor as the reverse current does not saturate but increases in a non-ohmic manner with voltage. The saving in area achieved by using this type of resistor is over 40 times when compared to a diffused resistor. The t.c.r. is 3% per deg C.

There is no published information on the long term stability of ion implanted resistors particularly those annealed at the lower temperatures. There is no evidence to suggest that they will drift over a period of time under normal operating conditions.

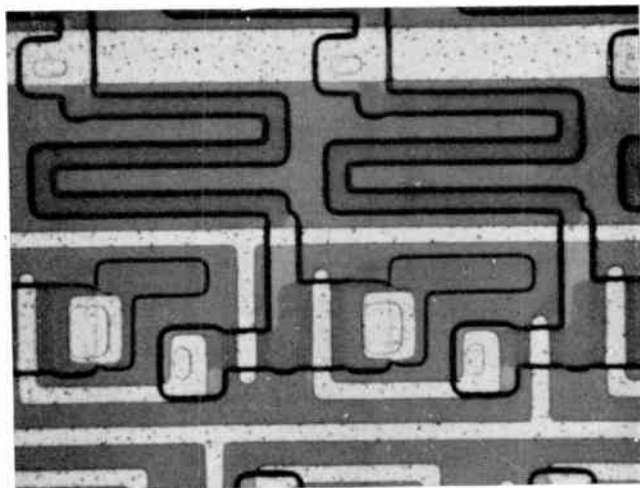


Fig. 13. Load resistors in a m.o.s.t. shift register (after Dill *et al.*[34]). The gates are about 5 μm wide.

World Radio History

3.1.3 Channel regions of metal oxide silicon transistors

The channel region of a m.o.s.t. can be implanted with a low dose of boron or phosphorus ions as appropriate for several purposes:

(i) To increase the impurity concentration in the channel region when working with a high resistivity substrate. Shannon *et al.*[28] used a boron implantation into a p^- substrate to increase the doping level in the channel region of a n-channel u.h.f. m.o.s.t. The increased doping prevents punch-through from drain to source while a large depleted volume can be formed around the drain to reduce the output capacitance. (Fig. 14).

(ii) Shannon and Ford[19] and Aubuchon[26] have altered the threshold voltages of p-channel devices in a positive direction by implanting small doses of boron through the gate oxide. An enhancement p-channel device can be altered to a depletion device. The carrier mobility in the doped channel tends to increase towards the bulk mobility. Aubuchon concluded that for the devices he studied, the only adverse effect of the implantation was to double the $1/f$ noise figure.

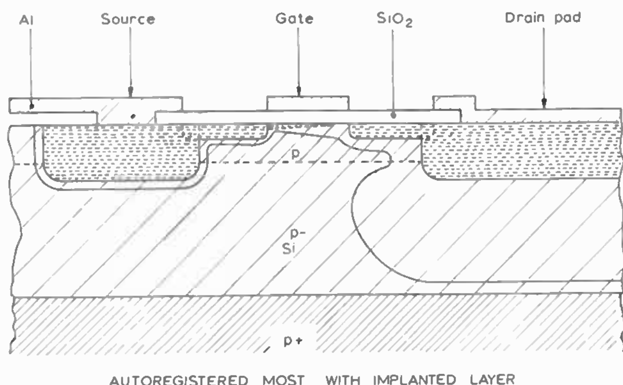


Fig. 14. Diagram of a n-channel microwave m.o.s.t. with doped channel (after Shannon *et al.*[28]).

(iii) MacDougall *et al.*[27] have outlined the advantages of changing V_T to make m.o.s. arrays compatible with TTL voltage levels and combining this with depletion load m.o.s.t.s on the same chip to give increased circuit speed; a factor of two for the same power dissipating was claimed. The depletion load m.o.s.t. is faster than a resistor load and occupies far less area.

3.1.4 Large area applications

(i) Solar cells: Burrill *et al.*[36] have described the implantation of phosphorus to produce n^+ on p-solar cells, 1 cm × 2 cm in area. The conversion efficiency is equal to that of diffused solar cells and is higher at the short wavelengths.

(ii) Nuclear particle detectors: As already mentioned, ion implentation was first applied to the fabrication of nuclear particle detectors.

The attractions of implantation for these devices are:

(i) shallow junctions producing thin 'windows' with dead volumes,

(ii) low processing temperatures and

(iii) high uniformity over a large area.

Successful detectors have been made by many workers, Dearnaley,[2] Dearnaley *et al.*,[37] Kalbitzer *et al.*,[38] Sebillotte *et al.*[39] and Howes and Knill,[40] in both silicon and germanium.

Position sensitive nuclear particle detectors utilizing the uniformity of the implanted layers have been made by Kalbitzer *et al.*[41] who obtained a 1 mm resolution in 15 cm for an X-Y detector and a 1% resolution for a linear detector 60 mm × 5 mm. Laegsgaard *et al.*[42] have reported making similar detectors with a positional resolution of 0·2 mm in 8 cm while Owen and Awcock[43] at Harwell have produced an X-Y detector with a resolution of 1 mm in 2 cm.

These results with very large area-devices are clear evidence of the high degree of uniformity obtainable with ion implantation when making resistive layers.

3.2 *Buried Layers*

3.2.1 Varactor diodes

These are designed to have a rapid change of capacitance with applied reverse voltage for tuning purposes. The required sensitivity is derived from

$$m = -\frac{d(\log C)}{d(\log V)}$$

where m is 0·5 for an abrupt p-n junction.

Two groups have reported improved values of m by implanting tailored phosphorus profiles below Schottky barriers. Brook and Whitehead[44] obtained $m = 2{\cdot}1$ for a series of summed implants while MacRae[16] obtained $m = 2{\cdot}5$ with a spread of less than 3% which was considered to be due to variations in the background doping of the epitaxial layer and not the implant.

3.2.2 Impatt diodes

MacRae[16] and later Seidel *et al.*[45] have used a tailored boron implantation to produce a double drift structure for an impatt diode, i.e. p^+-p-n-n^+ where the p-region is implanted and the p^+ contact diffused after the implantation. A power output of 1 W at 50 GHz with 12% efficiency has been obtained. Other structures operating at 100 GHz have been made as well as high frequency trapatt diodes.

3.2.3 Isolating layers

Schwuttke and Brack[46] have reported implanting 2 MeV nitrogen and oxygen ions into silicon to form a buried insulating layer of nitride or oxide. Stephen *et al.*[47] studied the electrical properties of a 3 MeV nitrogen implanted layer and found the electrical activity for a 900°C anneal to be less than 1% and the p-n junctions to be very leaky. Schwuttke *et al.*[48] also found that high-energy protons can produce a high-resistance layer in silicon which is stable up to 400°C.

3.2.4 Narrow-base bipolar transistors

The bipolar transistor has an important role to play at microwave frequencies (over 1 GHz). The main manufacturing problems are (i) to obtain narrow base

World Radio History

widths (< 0·2 μm) with low-base resistances ($R'_{bb} < 20\ \Omega$) and (ii) shallow highly-doped emitters with very abrupt emitter base junctions.

This is one application of ion implantation that is important to device technologists. Kerr and Large[49] have reported on accurate base profiles and Beale[50], in a recent review, has said that transistors with diffused emitters and implanted base regions have given good current gain ($h_{fe} = 100$) and that the intrinsic base resistance can be quite low. Much of the work on bipolar transistors in general has been empirical but computer simulation is now being used to predict performance and some earlier unorthodox ideas which were discarded, because they were not possible by diffusion, are now being tried out.[50]

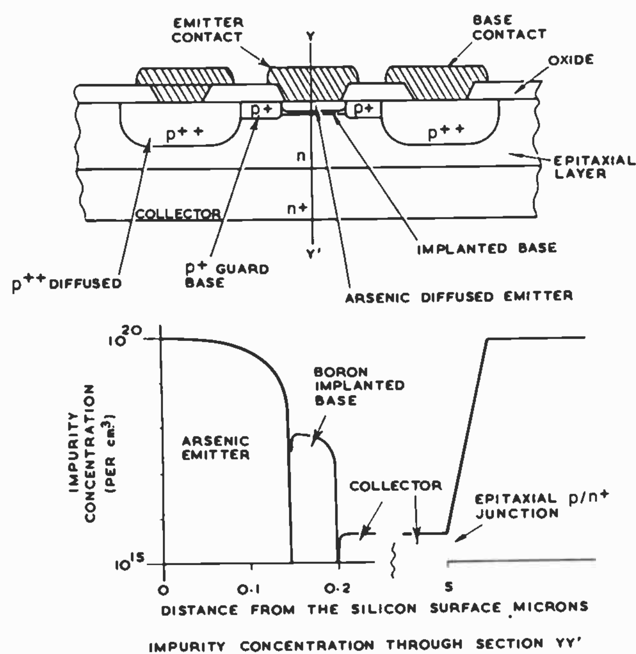


Fig. 15. Cross-section of n-p-n transistor with ion implanted base (after Fujinuma *et al.*[51]).

Fujinuma *et al.*[51] have reported on low-noise implanted base n-p-n transistors (Fig. 15). It is seen that the base region is surrounded by a diffused guard ring which helps contacting and lowers the extrinsic base resistance. The base region is implanted with boron *before* the emitter is diffused from a special solid to solid diffusion source using arsenic-doped germano-silicate glass which gives a very steep emitter profile with a high surface doping.[52] The base implantation is annealed as the silicon wafer is heated to 900°C for the emitter diffusion. The resultant basewidth is between 0·05 and 0·1 μm. A multi-layer metal contacting system is used to avoid the alloying effect experienced with aluminium which can penetrate the shallow layers. This device has a minimum available gain of 8 dB at 4 GHz with a total base resistance of 20 Ω and a noise figure of 4 dB at 4 GHz.

Ikeda *et al.*[53] have made n-p-n transistors by combinations of diffuson and implantation. They concluded that the gain (h_{fe}) of the n-p-n transistors with boron implanted bases were lower than the diffused transistors probably due to defects which reduce minority carrier lifetime in the base even after a 900°C anneal. The defects may have been affected by the high phosphorus concentration in the emitter or a highly damaged region associated with it. The leakage currents between emitter and collector were larger and this may be due to localized diffusion of phosphorus through defects produced during the boron implantation. In the case of the implanted emitter transistors, the 'tail' of the phosphorus implantation (Fig. 3) may compensate for the impurity in the base region and allow punch-through to occur from emitter to collector.

These results are in contrast to those reported by Fujinuma *et al.*[51] for the combination of base implantation followed by the emitter diffusion which was not tried by Ikeda but was later reported by Tokuyama *et al.*[54] from the same laboratory. They did not obtain the same successful results as Fujinuma *et al.* In the Fujinuma *et al.* case, the majority of the defects introduced by the boron implantation should be annealed out during the temperature rise preceding the emitter diffusion with the result that no enhanced diffusion should occur. The use of the germano-silicate glass for the emitter arsenic diffusion has been reported *not* to produce the 'emitter dip' effect as observed in the case of phosphorus emitter diffusions.[52] It is probable that the combination of first implanting the base and then diffusing arsenic into the emitter has been responsible for the precise control necessary over the basewidth. MacRae[87] has reported briefly on n-p-n transistors with implanted arsenic emitters and boron implanted bases, The β was in excess of 200 and very uniform over the wafer.

3.3 *Abrupt Junctions*

3.3.1 Impatt diode structure

Ying *et al.*[55] have reported on a p^+-n-n^+ diode in which the p^+ junction is made by implanting boron. A power output of 1·5 W at 11·7 GHz with an efficiency of 7% has been obtained. The diode has an improved performance over a similar diffused one because:

(i) it operates cooler as the junction is nearer the heat sink,

(ii) signal/noise ratio is better and

(iii) the implantation follows the surface profile of the epitaxial layer.

3.3.2 Avalanche photodiodes

Sherwell *et al.*[56] have studied avalanche multiplication in photosensitive diodes made by implanting phosphorus in to p-type silicon to form n^+-p junctions. The photosensitive area was delineated by etching a mesa. The junction has to be free of microplasma break-down for useful avalanche multiplication to occur and the diodes were tested for the presence of microplasmas. A high yield (90%) of useful devices was obtained showing a one order increase in the signal/noise ratio. This contrasts with a yield of 10% by diffusion. They also observed that the minority carrier lifetimes were longer in the implanted diodes.

World Radio History

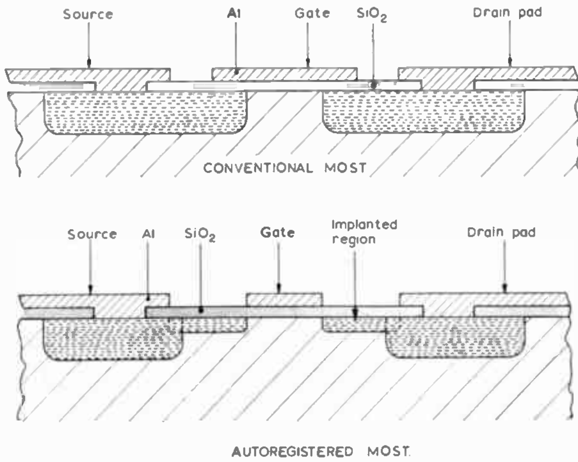


Fig. 16. Cross-sections of a conventional and an ion implanted metal oxide silicon transistor.

3.4 *Precision Alignment of Metal Oxide Silicon Transistors*

One of the most direct applications of ion implantation is the self-alignment or autoregistration of the metal gate electrode of the m.o.s.t. to the source and drain regions beneath the oxide layer. The method was first described by Bower and Dill.[5,6] Figure 16 shows the cross-section of two m.o.s.t.s, one made by diffusion methods and the other by ion implantation. The precision alignment of the gate electrode to the source and drain regions reduces the overlap capacitance between the drain and the gate by over an order of magnitude in the implanted transistor.[28] This reduction in the 'Miller' capacitance allows the implanted transistor to switch over 2·5 times faster than a conventional device and operate as a stable amplifier at frequencies up to 1 GHz without neutralization.

One of the first methods used by Shannon *et al.*[28] for making an aligned gate m.o.s. transistor is shown in Fig. 17. A dual gate, common source m.o.s.t. made by this method is shown in the scanning electron micrograph (Fig. 18). The implanted areas are shown up, darker, by voltage contrast and the offset in the gate lead is due to slight misalignment of the second aluminium engraving step. This basic approach is unsuitable for m.o.s. integrated circuits as it does not use a thick field oxide to cover the areas of silicon between the transistors and allow conductors to pass over without inducing m.o.s. action and causing short circuits. Figure 19 shows a typical p-channel transistor used in a m.o.s. array. The aperture in the thick oxide defines the area of the transistor. The diffused source and drain contacts are on either side, and well clear, of the aluminium gate electrode, which is about 5 μm wide and 1 μm thick. The gate oxide which is about 0·12 μm thick, is penetrated by the boron ions which dope the silicon between the gate and the source and drain. If implanted resistors are included (Fig. 13) then the tracks of the resistors must have thin oxide covering them to allow the implanted ions to dope the silicon.

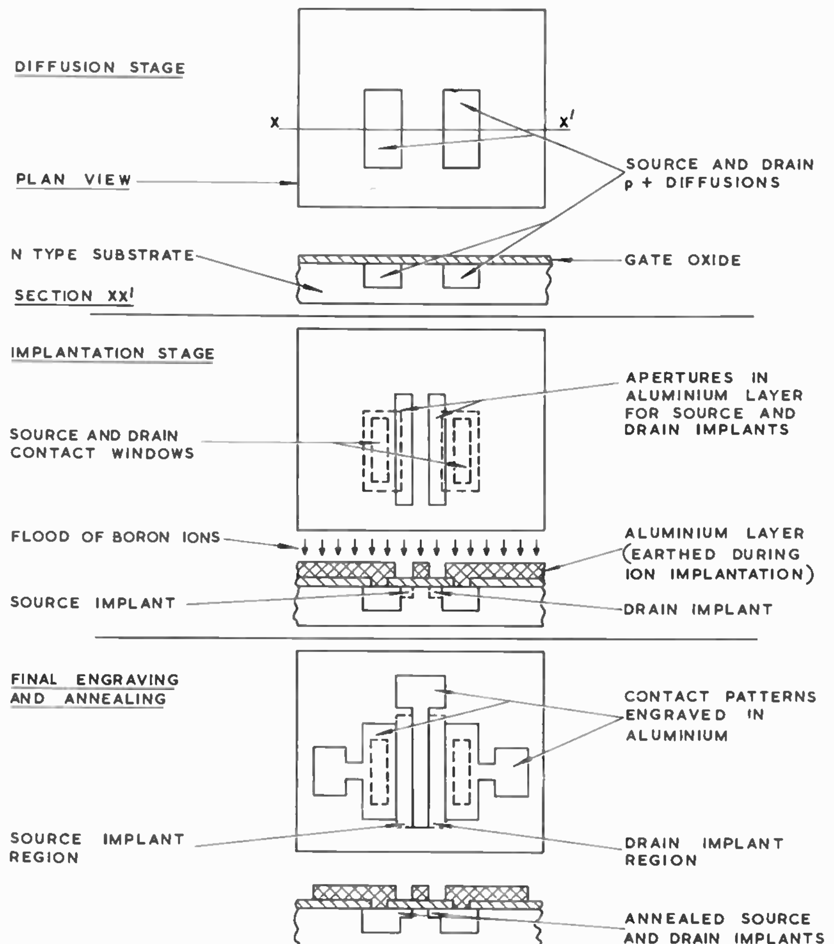


Fig. 17. Basic steps in the fabrication of an ion implanted m.o.s.t.

World Radio History

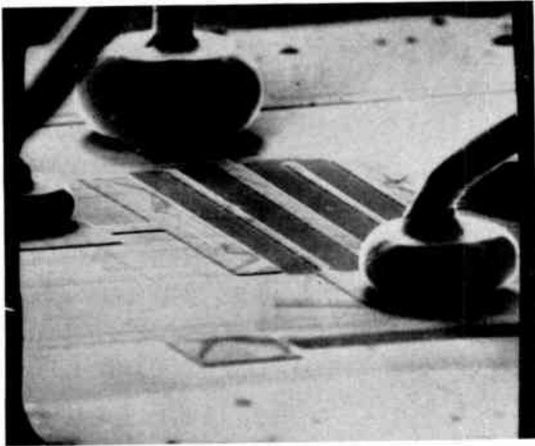


Fig. 18. Scanning electron micrograph of dual ion implanted m.o.s.t. There is a common source in the centre with the gates and drains on either side. The implanted regions are showing up by voltage contrast.

Alternative methods have been tried to replace the thermal diffusion of the source and drain contacts. MacDougall and Manchester[57] have given the details of an all-implanted m.o.s.t. using a molybdenum gate metal and thick oxide masking elsewhere. The use of molybdenum allows the annealing temperature to be raised to between 800°C to 900°C to fully anneal the boron implant. A disadvantage of this process is that an aluminium metallization layer is required to make ohmic contacts to the source and drain regions.

Josephy[58] and Nishimatsu *et al.*[59] have taken this work a step further by replacing the molybdenum gate by a silicon gate in p-channel m.o.s.t.s by combining implantation with silicon gate technology. The gate was aligned by the implantation, assuming no sidewards diffusion during annealing, and the advantage of the low threshold voltage associated with the silicon gate was retained. It was claimed that there was reduced contamination of the silicon in the source and drain regions as the oxide on either side of the gates does not have to be removed for usual thermal diffusion as the boron was implanted through the thin gate insulator. The annealing temperature can be chosen over a wide range to suit the circuit conditions as the restrictions imposed by the use of aluminium or molybdenum are not present.

The details have been given recently by Maloney[60] of an all-implanted p-channel m.o.s. process which has been successfully used to make m.o.s. arrays incorporating high-value resistors, diodes, self-aligned m.o.s. transistors and lateral bipolar transistors. It uses two implantations, one to replace the diffusion step and the second one for the self alignment of the m.o.s.t.s, chemically deposited field oxide and aluminium metallization. This gate insulator is a SiO_2–Si_3N_4 sandwich. A test chip with a ripple-through oscillator employing 54 m.o.s. inverters (with resistor loads) and a bistable flip-flop feed back circuit was used to test the process; 96% of the oscillators on a wafer were operative and 90% of these were ±4·8% of the mean oscillator frequency. Test resistors showed a ±3% variation across a wafer and a ±5% variation from wafer to wafer. Some of this variation was traced to changes in the dielectric thickness. The very promising results indicate that ion implantation can be used as the sole doping method capable for giving a high yield and good uniformity.

Lepselter *et al.*[61] have combined Schottky barrier techniques with implantation by replacing the diffused source and drain contact regions of a p-channel m.o.s.t. with platinum silicide contacts. A double boron implantation (150 keV ($1{\cdot}5 \times 10^{14}/cm^2$) and 50 keV ($1 \times 10^{14}/cm^2$)) was used to connect the contacts to the edge of the titanium-gold gate electrode. No difficulty was experienced in contacting the implant with the platinum-silicon Schottky barrier layers despite them being very thin.

The direct benefits obtained by implanting the gate of a m.o.s.t. to achieve self-alignment are:

(i) minimal overlap capacitance leading to an increase in the speed of operation,

(ii) short gate lengths giving reduced carrier transit time,

(iii) easing of photolithographic tolerances as precise overlaps are not required and

(iv) n- and p-channel devices can be produced in the same way and on the same chip.

To these benefits must be added the possibilities provided by the control of threshold voltage permitting depletion m.o.s.t. load devices to be made on the same chip as enhancement devices.[27] High-value resistors, already discussed, can be incorporated on the chip with enhancement m.o.s.t.s with no extra processing.[34]

Before leaving the ion-implanted m.o.s.t. it should be mentioned that several complex implanted m.o.s. devices have been reported in the literature.

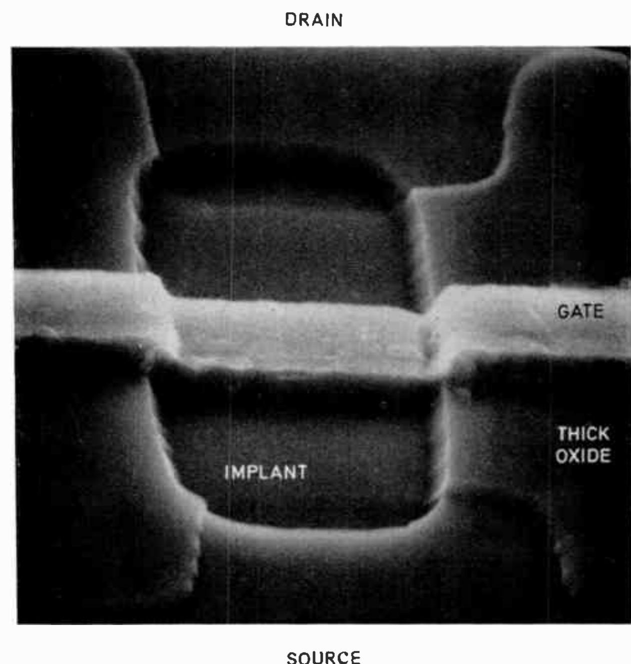


Fig. 19. A typical p-channel m.o.s.t. used in a m.o.s. array. The thick oxide and aluminium layers are over 1 μm thick and the diffused source and drain contact regions can be seen on either side of the gate. The gate length is about 5 μm.

World Radio History

Burt[62,63] has produced a tetrode structure which has a good u.h.f. performance up to 900 MHz. A power gain of 20 dB at 300 MHz is typical for this device.

Bower[64] with Dill *et al.*[34,88] have reported several complex m.o.s. circuits incorporating ion implantation. Among these devices are:

(i) Dual 64-bit shift register with resistor loads which can operate up to 30 MHz (details shown in Fig. 13).

(ii) 10- and 50-channel multiplexers, complete with address registers, which operate between 100 Hz and 10 MHz. The reduction in overlap capacitance reduces the switching noise generated in the series switching transistors.

(iii) 2048-bit 'read only' memory complete with addressing circuits and operating at voltage levels compatible with TTL circuitry. The access time is 100 ns. The data in the store are determined by a special photoresist mask which prevents implantation wherever a '0' is to be stored in the memory,

(iv) All-ion-implanted m.o.s. operational amplifier with a gain of 1000 times and the capability of operating at 4 K.

3.5 *Doping of Imperfect Semiconductor Materials*

One of the factors influencing the control of thermal diffusion is the perfection of the crystal structure of the semiconductor. Many semiconductor crystals contain a large number of defects such as dislocations, grain boundaries, twins, vacancy clusters etc., which influence the diffusion of an impurity. For example, diffusion is very fast along grain boundaries when compared with the diffusion in a crystal. The result of diffusing an impurity into an imperfect crystal is to produce a non-planar impurity concentration and a junction formed this way has poor electrical properties. The range distribution of high energy ions implanted in a substrate are determined by the ion, its energy and the elements in the substrate. Defects such as dislocations and grain boundaries have no effect on the profile of the implanted ions, but gross defects such as voids, cracks or inclusions will disturb the profile.

The single crystals of many Group III-V and Group II-VI semiconductors contain a large number of defects and implantation is a suitable method for introducing impurities without interference from these defects. The same situation exists in many thin semiconducting films produced by various deposition techniques, as the films may be either monocrystalline with many defects, polycrystalline or even amorphous.

There are no published papers describing this application of ion implantation.

3.6 *Doping of High-resistivity Silicon*

High-resistivity silicon apart from having a very low impurity concentration also has a low defect density and a long carrier lifetime. If silicon of this quality is heated to diffusion temperatures for any purpose, these properties are severely degraded by an increase in the number of dislocations and by the rapid diffusion of unwanted interstitial impurities such as copper which form deep traps and drastically reduce the carrier lifetime. The effect is more pronounced the higher the temperature.

By using ion implantation, the silicon can be processed at much lower temperatures which help maintain the original properties of the bulk silicon. There are two main classes of devices which require high-resistivity defect-free silicon. They are:

(i) Semiconductor radiation detectors which depend on the collection of electrons and holes produced by the incident radiation. Examples are photodiodes, X-ray and nuclear particle detectors.[2]

(ii) p-n junction structures based on the Read diode for producing microwave oscillation by avalanche processes, as discussed earlier in Sections 3.2.2 and 3.3.1.

4. Compound Semiconductors

The application of ion implantation to compound semiconductors is not as advanced as it is for silicon. From the device aspect, the main activity is concerned with the production of n and p layers and the formation of p-n junctions. Table 2 summarizes the present position of the ion implantation of GaAs. Further information can be obtained from a comprehensive review by Allen.[65]

One of the main problems with compound semiconductors is making ohmic contacts to the semiconductor. This is particularly important in the case of n-type GaAs because of several important applications in the microwave field (e.g. Gunn diodes). Ion implanted contacts to n-type GaAs have been studied by the group at Hughes Aircraft Corporation (Hunsperger *et al.*[66], Dunlap *et al.*[67]). They used sulphur, tin and argon ions implanted at room temperature and found that a high resistance ohmic contact was obtained provided the implant was not annealed above 200°C. They conclude that the damage rather than the electrical activity of the particular ion in the GaAs is responsible for the ohmic effect. Allen[65] states that if this method can be extended to other materials it should be a useful aid in the assessment of implanted layers.

Lindley *et al.*[68] have used proton bombardment to produce a semi-insulating guard ring around a platinum Schottky barrier photodiode made in n-type GaAs. These devices operate with a gain of 100 and have gain-bandwidth products greater than 50 GHz.

GaAs, CdS and ZnO have been investigated by Shifrin *et al.*[69,70] as piezoelectric crystals for acoustic microwave devices. The requirement is for a thin conducting layer on an insulating substrate to allow interaction between the shallow surface wave and the electric field. Ion implantation is being tried to produce these layers particularly in ZnO which offers the highest performance according to computer calculations.

Foyt, Lindley and Donnelly[71] at the Lincoln Laboratories have made a photovoltaic detector in InSb by a proton implantation. The p-type InSb was covered with

World Radio History

Table 2 Details of ion implantation in gallium arsenide

Ion	Experimental details and results
(i)	*Gallium arsenide*
H	$10^{13}H^+$ ions/cm² implanted in n and p GaAs reduced carrier concentration to < 10^{11}/cm³.[75]
S	n-p diodes formed in p substrate by 400 and 800 keV implants.[4] 30 keV implant produced n type activity before annealing but no activity afterwards. Another implant produced n layer ($\rho \sim 10^4$ ohms/square) which was removed on annealing to 700°C and reappeared again after annealing above 800°C.[67,77]
A	30 keV $1{\cdot}1 \times 10^{16}$ A^+/cm² implant without annealing produced high resistance ohmic contact to n-type GaAs.[67]
Zn	80 keV implant formed p-n diode with abrupt junction. Light emitted at $\lambda = 0{\cdot}92$ µm.[30,78] Implant formed p-n junction with large series resistance.[79] 20 keV implant formed p-layer after 400°C anneal with $\rho = 10^3$ ohms/square.[80] 70 keV implant formed p-layer with $\rho = 598$ Ω/square and $\mu = 45{\cdot}8$ cm²/V s.[66] 20 keV implant formed p-layer with $\rho = 10^4$ Ω/square and $\mu = 10$ cm²/V s.[67] 20 keV implant formed p-layer with 1% electrical activity after 600°C anneal with $\mu = 10$ cm²/V s.[81] 84 keV implant formed p-layer after 600°C anneal.[77]
Se	400 keV implant at 500°C formed n-layer with 50% electrical activity after 800°C anneal.[76]
Cd	20 keV implant formed p-layer, with $\rho = 2$ kΩ/square and $\mu = 40$ cm²/V s.[66] 20 keV implant produced p-i-n diodes.[81] 60 keV implant annealed at 800°C formed p-layer with 100% electrical activity for dose < 10^{14} ions/cm².[77]
Te	20 keV implants at 400°C formed a n-layer. A *C-V* plot indicated a graded junction.[80]
Sn	35 keV implant at 400°C gave n-layer after 600°C anneal.[66]
(ii)	*Semi-insulating gallium arsenide*
C	70 keV implant formed p-layer with 2% electrical activity and $\mu = 250$ cm²/V s.[82]
Si	10+30+50 keV multiple implant formed n layer with $\mu = 2600$ cm²/V s.[83] 70 keV implant, formed n layer after 700°C anneal with 10% electrical activity and $\mu = 2700$ cm²/V s.[82]
S	70 keV implant formed n layer with low electrical activity and $\mu = 3600$ cm²/V s.[82]
Cd	Cr doped GaAs when implanted with 5×10^{14} ions/cm² and annealed at 800°C gave 20% electrical activity with $\mu = 20$ cm²/V s.[70]
Zn Se Kr Xe	These four ions are reported to produce conducting layers in 10^8 Ω-cm GaAs. The energy was < 50 keV and dose about 6×10^{13} ions/cm².[84]

1500 Å of SiO_2 and masked against the protons by a thick layer of photoresist. Gold contacts were later made through windows in the oxide layer to the n-regions produced by the proton bombardment. Good *I/V* characteristics were obtained and the *C/V* characteristics suggested a slightly graded profile. The peak detectivity of 10^{11} cm Hz$^{\frac{1}{2}}$ W^{-1} was measured at 4·9 µm. The quantum efficiency at this wave-length was 35%.

The high resistivity layer produced in p-type ZnTe by proton implantation has been used by Donnelly *et al.*[72] to produce an electroluminescent metal-insulator-semiconductor structure. Gold was used for the metal contacts. When the ZnTe was biased positively with respect to the gold electrode on the insulating ZnTe, electrons were injected into the p-region and produced electroluminescence at 300K and 77K in the green and red with quantum efficiencies up to 0·3%.

Four-by-four planar mosaic photovoltaic infra-red diode detectors in InAs and InSb have been described by McNally and King[73] and McNally.[74] Sulphur and zinc ions were implanted through an in-contact oxide mask to produce n-regions in the p-type InAs and InSb. Surface layers of SiO_2, Al_2O_2 or MgO have been studied for surface passivation and anti-reflexion coatings. The detectivity (D^*) (500K, 2π) values are:

5×10^8 cm Hz$^{\frac{1}{2}}$ W^{-1} for InAs and

5×10^9 cm Hz$^{\frac{1}{2}}$ W^{-1} for InSb.

Uniformity values of 4% between individual elements in an array was observed.

There are various features of silicon planar technology which are directly applicable to compound semiconductors when ion implantation is used for doping. There is no restriction on the dopant as the masking material is used at room temperature and it is not exposed to the chemical attack which occurs in many cases at diffusion temperatures. The annealing of the implantation can be done at temperatures lower than diffusion temperatures and for much shorter times, avoiding the degradation produced by elevated temperatures. Many workers have found it beneficial to use a thin layer of SiO_2 on the compound semiconductor before implantation and certainly before annealing. The layer 'seals' the surface and prevents the escape of volatile components of the semiconductor and the implanted species. It can be readily removed by buffered hydrofluoric acid.

A great majority of the work on compound semiconductors is still at the research stage. It is confidently anticipated that as this work proceeds and expands, implantation will be incorporated as one of the standard doping methods, along with epitaxy and diffusion.

5. Production Equipment for Ion Implantation

The successful application of ion implantation to the *commercial* production of ion-implanted semiconductor devices depends on suitable production equipment being available. Most of the work described in this review has been done on ion accelerators designed as research tools which can handle only a few semiconductor wafers in a day.

Freeman *et al.*[85] have stated that these accelerators cannot be used for full-scale industrial assessments or be adapted to meet larger scale requirements. The absence of such facilities makes it difficult to demonstrate the benefits of uniformity and reproducibility afforded by ion implantation doping. A new implantation facility has been designed and built at Harwell to meet a wide range of industrial doping requirements and provide for the implantation of wafers on a production scale to permit the technical and economic assessments to be made under near realistic conditions.[85] The machine is primarily intended for repetitive doping operations on a

World Radio History

large batch scale but it can also be used for the experimental doping of individual samples. A very wide range of dopants can be used in the machine without any chance of contamination of the wafers. The time required to dope eighty 5 cm diameter wafers to a dose of 10^{15} ions/cm^2 requires 80 minutes for a 4% uniformity and about 4 hours for 0·5% uniformity.

Recently some uniformity measurements were made in collaboration with Mullard Research Laboratories[86] on boron resistors implanted in the Harwell machine. The uniformity was better than 2·5% over each wafer (Fig. 20) and < 5% from wafer to wafer, including the photolithographic tolerances. The implantation uniformity was deduced to be better than 1%.

Machines for the commercial implantation of semiconductors are also being built in the United States, France and Japan. This activity is a clear indication of a growing commercial awareness of the importance of ion implantation in present and future semiconductor fabrication processes.

6. The Future

In this paper ion implantation has been represented as a *complementary*, and perhaps an alternative technique to thermal diffusion for the doping of semiconductors, which is completely compatible with the concepts and requirements of the silicon planar technology. It offers measured control, with good uniformity and reproducibility over the number and position of the dopant atoms and lower processing temperatures can be used. The electrical properties of implanted layers in silicon are suitable for such critical devices as m.o.s. arrays, impatt diodes, bipolar transistors and vidicon targets. The work reported in this paper has all been published and there can be no doubt that there is considerable activity taking place at present in a number of laboratories which in a year's time may well necessitate a completely new review.

The future of ion implantation can be considered from two rather divergent aspects, firstly the potential offered by the technique to semiconductor fabrication and secondly the commercial acceptance and success of implantation on the production line. It is a stimulating mental exercise to allow one's imagination to 'run riot' over the possibilities offered by combining ion implantation with other associated techniques such as ion sputtering, electron beam writing and annealing, along with chemical vapour deposition and conventional evaporation. The ultimate aim is to combine these processes under computer control in an environment where the problems of contamination from the atmosphere and wet chemistry no longer apply. Clearly such a computer will use a program generated by computer-aided-design (c.a.d.) techniques. The considerable advantages of ion implantation can be obtained on a more modest plane by applying it to the devices, mentioned above or even the 'simple' high value implanted resistor or surface doping application.

The commercial prospects of ion implantation for semiconductor doping are entirely dependent on the future profitability and capital investment in the semiconductor industry. Production facilities based on diffusion have been established at considerable capital expenditure and their main task is now to operate efficiently and produce a return on the capital invested. Any small change proposed in a production schedule, which may upset the smooth flow, even if it offers an advantage, is resisted. However, for an organization setting up to manufacture semiconductors or establishing a new production line, (for m.o.s. arrays as an example) the incorporation of ion implantation is not so difficult.

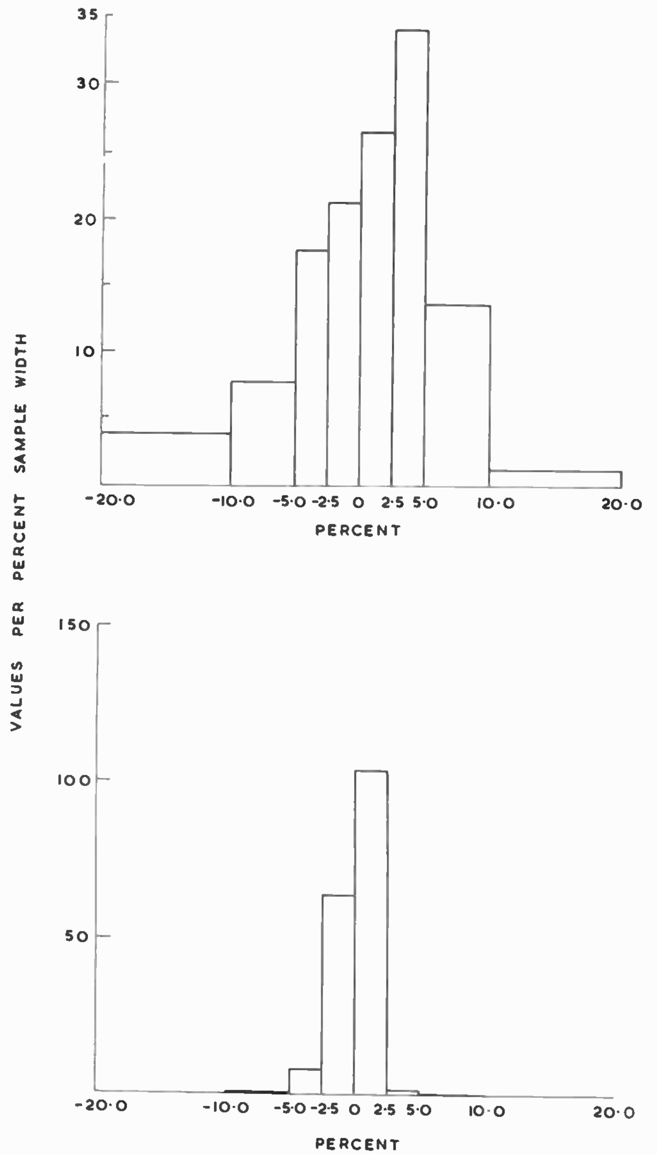


Fig. 20. Histograms showing the spread in sheet resistance for two boron resistor implantations. The upper histogram is for an implant annealed at 500°C and the lower one for an implant annealed at 900°C with precautions being taken to avoid a temperature rise during implantation (after Nicholas *et al.*[86]).

In conclusion, it is appropriate to repeat the remarks made by Beale[50] when addressing the European Ion Implantation conference at Reading in 1970. He compared the present state of ion implantation to that of epitaxy about 10 years ago. Its advantages were not clearly seen at the time but with use these gradually came to light. He considered that the development of

World Radio History

machinery and 'know-how' for implantation had gone critical, as there was no long term technical or economic barrier to its widespread introduction. This was a view shared by MacRae[16] as is shown by this quotation from his paper: 'The results of these electrical experiments are very encouraging. They indicate that the implantation process is an efficient doping technique and there is no fundamental reason to preclude its use in device fabrication'. The difficulty is to predict the time scale and using the example of epitaxy. Beale expects that 'in 10 years time the majority of *new* devices introduced onto the market will incorporate ion implantation'.

7. Acknowledgments

The author is grateful for the discussions he had with Dr. A. MacRae and Dr. K. Pickar of Bell Laboratories, Dr. R. Bower and Dr. Hans Dill of Hughes Aircraft Corp., Dr. John MacDougall of Sprague-Mostek, Dr. Hofker of Philips, Dr. K. Rosendal of the Technical University, Denmark, Mr. Eirug Davies of Rome Air Force Center, Bedford, Mass., Mr. J. Beale, Mr. J. Shannon, Dr. K. Nicholas and Mr. R. Ford at Mullard Research Laboratories, Mr. R. Knott, Mr. R. Josephy and Mr. D. Burt of Hirst Research Centre (G.E.C.), Mr. N. Sigournay and Mr. C. Maloney of Smiths Industries, Dr. J. A. Grimshaw of the Royal Military College of Science and his colleagues at Harwell, in particular, Mr. K. Kandiah and Dr. B. J. Smith.

8. References

1. Shockley, W., 'Forming semiconductive devices by ionic bombardment', U.S. Patent No. 2,787,564, 28th October 1954.
2. Dearnaley, G., 'Ion bombardment and implantation', *Reports on Progress in Physics*, 32, No. 4, pp. 405–91, August 1969.
3. Kellett, C. M., King, W. J. and Leith, F. A., 'High energy implantation of materials', Scientific Report No. 1 (Ion Physics Corp.), AD635267, May 1966.
4. Leith, F. A., King, W. J. and McNally, P., 'High energy ion implantation of materials', Final Report (Ion Physics Corp.), AD651313, January 1967.
5. Bower, R. W. and Dill, H. G., 'Insulated gate field effect transistors fabricated using the gate as source-drain mask', Paper presented at Int. Electron Devices Washington Meeting, 28th October 1966.
6. Bower, R. W., Dill, H. G., Aubuchon, K. G. and Thompson, S. A., 'M.i.s. field effect transistors formed by gate masked ion implantations', *I.E.E.E. Trans. Electron Devices*, **ED15**, No. 10, pp. 757–61, October 1968.
7. Morgan, R. and Greenhalgh, K. R., 'Ion implantation—a bibliography'. A.E.R.E. Harwell Bibliography, Bib. 176, 1972.
8. Freeman, J. H., Caldecourt, L. R., Done, K. C. W. and Francis, R. J., 'An industrial scale ion implantation facility', Proc. European Conf. Ion Implantation, pp. 36–40, 1970 (Peter Peregrinus Ltd., London).
9. Lindhard, J., Scharff, M. and Schiøtt, H. E., 'Range concepts and heavy ion ranges, (Notes on atomic collisions II)', *Kgl. Danske Vidensk. Selsk. Mat. Fys. Medd.*, 33, No. 14, pp. 1–39, 1963 (in English).
10. Johnson, W. S. and Gibbons, J. F., 'Projected Range Statistics in Semiconductors' (Stanford University Bookstore, 1969).
11. Smith, B. J., 'Projected ion range data for silicon planar technology', AERE R6660, 1971, A.E.R.E. Harwell.
12. Blamires, N. G., Matthews, M. D. and Nelson, R. S., 'Improved profiles of electrical activity in boron implanted silicon', *Physics Letters*, **28A**, No. 3, p. 178, 15th November 1968.
13. Shannon, J. M., Ford, R. A. and Gard, G. A., 'Annealing characteristics of highly doped ion implanted phosphorus layers in silicon', *Radiation Effects*, **6**, Nos. 3 and 4, pp. 217–23, 1970.
14. Blamires, N. G., Private communication, 1970.
15. Davies, D. E., 'Post annealing conductive behaviour of implanted layers in silicon', *Appl. Phys. Lett.*, **14**, p. 227, 1969.
16. MacRae, A. U., 'Device fabrication by ion implantation', *Radiation Effects*, 7, Nos. 1 and 2, pp. 59–64, January 1971.
17. Picker, K. and Dalton, J. V., 'Lifetime effects in ion implanted silicon', *Radiation Effects*, **6**, Nos. 1 and 2, pp. 89–94, November 1970.
18. Davies, D. E. and Roosild, S. A., 'Minority carrier lifetime in ion implanted and annealed silicon', *Appl. Phys. Lett.*, **17**, pp. 107–9, 1970.
19. Shannon, J. M. and Ford, R. A., 'Properties of implanted layers and oxides used for mos integrated circuits', Paper J.1, 1st Ion Implantation Conference, Thousand Oaks, 1970 (Private communication).
20. MacDougall, J. D., Manchester, K. E. and Roughan, P. E., 'High value implanted resistors for microcircuits', *Proc. I.E.E.E.*, **57**, No. 9, pp. 1538–42, September 1969.
21. Pickar, K., Dalton, J. V., Seidel, H. D. and Matthews, J. R., 'Electrical properties of silicon diode array camera targets made by boron ion implantation', *Appl. Phys. Lett.*, **19**, No. 2, p. 43, 15th July 1971.
22. Stephen, J. and Grimshaw, J. A., 'The electrical behaviour of abrupt ion implanted and diffused p^+–n junctions', *Radiation Effects*, 7, Nos. 1 and 2, pp. 73–86, January 1971.
23. Crowder, B. L. and Title, R. S., 'The distribution of damage produced by ion implantation of silicon at room temperature', *Radiation Effects*, **6**, Nos. 1 and 2, pp. 63–76, 1970.
24. Wolf, E. D., Bauer, L. O., Bower, R. W., Garvin, H. L. and Buckley, C. R., 'Electron beam and ion beam fabricated microwave switch', *I.E.E.E. Trans. Electron Devices*, **ED-17**, No. 6, p. 446, 1970.
25. Glotin, P., Bernard, J. and Monfret, A., 'Implications of ion implantation technology on ion implanted active devices in silicon', *Radiation Effects*, 7, Nos. 1 and 2, pp. 65–72, 1971.
26. Aubuchon, K. G., 'The use of ion implantation to set the threshold voltage of m.o.s. transistors', Proc. Int. Conf. Properties and Use of M.I.S. Structure, Grenoble, pp. 575–90, 1969.
27. MacDougall, J. D., Manchester, K. and Palmer, P. B., 'Ion implantation offers a bagful of benefits for mos', *Electronics*, 43, No. 13, pp. 86–90, June 1970.
28. Shannon, J. M., Stephen, J. and Freeman, J. H., 'MOS frequency soars with ion-implanted layers', *Electronics*, 42, No. 3, pp. 96–100, 3rd February 1969.
29. Bower, R. W., 'Method of irradiating dielectric-coated semiconductor bodies with low energy electrons', U.S. Patent No. 3,507,709, 1970.
30. Roughan, P. E., MacDougall, J. D., Clark, A. H., Manchester, K. E. and Anderson, F. W., 'Introduction of diffusion sources by ion implantation', Paper presented to Electrochemical Society, Boston Meeting, May 1968.
31. Oosthoek, D. P., den Boer, J. A. and Hofker, W. K., 'The thermal properties of high-value gallium and boron-implanted resistors in silicon', Proc. European Conf. Ion Implantation, pp. 88–91, 1970 (Peter Peregrinus Ltd., London).
32. Rosendal, K., 'Ion implanted planar resistors', *Radiation Effects*, 7, Nos. 1 and 2, pp. 95–100, January 1971.
33. Nicholas, K. H. and Ford, R. A., 'New techniques for improving high value ion implanted resistors', Proc. 2nd Int. Conf. Ion Implantation, Garmisch, pp. 357–61, 1971 (Springer-Verlag, Berlin).

World Radio History

34. Dill, H. G., Bower, R. W. and Toombs, T. M., 'Ion-implanted mos technology', *Radiation Effects*, 7, No. 1 and 2, pp. 45–58, 1971.

35. Hodges, D. A., Lepselter, M. P., MacDonald, R. W., MacRae, A. U. and Waggener, H. A., 'Low power bipolar transistor memory cells', *I.E.E.E. J. Solid State Circuits*, SC-4, No. 5, pp. 280–84, October 1969.

36. Burrill, J. T., King, W. J., Harrison, S. and McNally, P., 'Ion implantation as a production technique', *I.E.E.E. Trans. Electron Devices*, ED-14, No. 1, pp. 10–17, 1967.

37. Dearnaley, G., Hardacre, A. G. and Rogers, B. D., 'Ion implanted germanium particle detectors', *Nuclear Insts & Methods*, 71, pp. 86–92, 1969.

38. Kalbitzer, S., Bader, R., Herzer, H. and Bethge, K., 'The fabrication of high quality silicon junction detectors by low energy ion implantation', *Z. f. Physik*, 203, pp. 117–8, 1967.

39. Sebillotte, Ph., Siffert, P. and Coche, A., 'Characteristics and window thickness of ion implanted semiconductor detectors', *I.E.E.E. Trans. Nuclear Science*, NS-17, No. 1, pp. 24–32, February 1970.

40. Howes, J. H. and Knill, G.. 'The characteristics of large area ion implanted p–n junctions for nuclear radiation detectors', Proc. European Conf. Ion Implantation, pp. 97–101 (Peter Peregrinus Ltd., London, 1970).

41. Kalbitzer, S., Bader, R., Melzer, W. and Stumpfi, W., 'An X–Y position sensitive detector', *Nuclear Insts & Methods*, 54, pp. 323–24, 1967.

42. Laegsgaard, E., Martin, F. W. and Gibsen, W. M., 'Position sensitive detectors made by ion implantation in silicon', *Nuclear Insts & Methods*, 60, No. 1, pp. 24–6, 1968.

43. Owen, R. B. and Awcock, M. L., 'One and two-dimensional position sensing semiconductor detectors', *I.E.E.E. Trans. Nuclear Science*, NS-15, No. 3, pp. 290–303, 1968.

44. Brook, P. and Whitehead, C. S., 'Hyperabrupt junctions in Au–Si Schottky diodes by ion implantation', *Electronics Letters*, 4, No. 16, pp. 335–7, 1968.

45. Seidel, T. E., Davis, R. E. and Iglesias, D. E., 'Double-drift-region ion implanted millimeter wave impatt diodes', *Proc. I.E.E.E.*, 59, No. 8, p. 1222, August 1971.

46. Schwuttke, G. H. and Brack, K., 'Annealing of high energy ion implantation damage in single crystal silicon', *Trans. Metallurg. Soc. AIME*, 245, No. 3, pp. 475–482, 1969.

47. Stephen, J., Smith, B. J., Hinder, G., Marshall, D. C. and Wittam, E. M., 'Some observations on high energy nitrogen implantations in silicon', Proc. 2nd Int. Conf. Ion Implantation, Garmisch, pp. 489–4 (Springer-Verlag, Berlin, 1971).

48. Schwuttke, G. H., Brack, K., Gorey, E. F., Kahan, A. and Lowe, L. F., 'Resistivity and annealing properties of implanted Si : H^+, *Radiation Effects*, 6, No. 2, pp. 103–6, 1970.

49. Kerr, J. A. and Large, L. N., 'Semiconductor devices made by ion implantation', Proc. Int. Conf. on Applications of Ion Beams to Semiconductor Technology, Grenoble, pp. 601–18, 1967.

50. Beale, J. R. A., 'Applications to semiconductor devices', Proc. European Conf. Ion Implantation, Reading, pp. 81–7 (Peter Peregrinus Ltd., London, 1970).

51. Fujinuma, K., Sakamoto, T., Abe, T., Sato, K. and Ohmura, Y., 'Silicon microwave transistor made by new techniques', Proc. 1st Conference on Solid State Devices, Tokyo, 1969. (Supplement to *J. Jap. Soc. Appl. Phys.*, 39, pp. 71–7, 1970.)

52. Abe, T., Sato, K., Konaka, M. and Miyazaki, A., 'Impurity diffusion into silicon using Ge-doped SiO_2 films', Extended Abstract No. 184, p. 483, Electrochem. Soc. Fall Meeting, Detroit, 1969.

53. Ikeda, T., Tokuyama, T. and Tsuchimuto, T., 'Silicon n–p–n structures by ion implantation', Extended Abstract No. 170, Electrochemical Society Fall Meeting, Detroit, p. 447, 1969.

54. Tokuyama, T., Ikeda, T. and Tsuchimoto, T., 'Junction structures and electrical properties of silicon n-p-n transistors fabricated by various combinations of diffusion and ion implantation', Proc. 4th Microelectronics Congress, Munich, 1970, p. 36 (Oldenburg Verlag).

55. Ying, R. S., Mankarious, R. G., English, D. L., Bower, R. W. and Coerver, L. E., 'Characteristics of ion implanted impatt oscillators', *I.E.E.E. J. Solid State Circuits*, SC-3, No. 3, pp. 225–31, September 1968.

56. Sherwell, R. J., Raines, J. A. and Large, L. N., 'Avalanche multiplication in photosensitive junctions formed by ion implantation', Proc. Int. Conf. Ion Implantation, Grenoble, pp. 641–6, 1967.

57. MacDougall, J. D. and Manchester, K. E., 'Implanted components in microcircuits', Paper presented at the National Electronics Conference, Chicago, December 1969.

58. Josephy, R., Private communication (Hirst Research Centre, G.E.C.), 1970.

59. Nishimatsu, S., Natsuaki, N., Warabisako, T. and Tokuyama, T., 'Ion implanted mosfet', Proc. 2nd Conf. on Solid State Devices, Tokyo, 1970. (Supplement to *J. Jap. Soc. Appl. Phys.*, 40, pp. 29–35, 1971.)

60. Maloney, C. P. C., Private communication (Smiths Industries Ltd.), 1971.

61. Lepselter, M. P., MacRae, A. U. and MacDonald, R. W., 'SB–IGFET. II, an ion implanted igfet using Schottky barriers', *Proc. I.E.E.E.*, 57, No. 5, pp. 812–3, 1969.

62. Burt, D., 'The design and characterization of high frequency m.o.s. transistor utilising ion-implantation techniques', Proc. Int. Conf. Properties and Use of M.I.S. Structures, Grenoble, pp. 605–7, 1969.

63. Burt, D., Private communication (Hirst Research Centre, G.E.C.), 1971.

64. Bower, R. W., 'Applications of ion implantation doping to the planar m.o.s. integrated circuit technology', NEREM Record, p. 120, 6th November 1969.

65. Allen, R. M., 'Implantation of compound semiconductors', Proc. European Conf. Ion Implantation, Reading, pp. 127–34, (Peter Peregrinus Ltd., London, 1970).

66. Hunsperger, R. G., Dunlap, H. L. and Marsh, O. J., 'Development of ion implantation techniques for microelectronics' NASA Report N69–24439, October 1968, (Hughes Research Labs.), (NASA Scientific and Technical Information Facility, P.O. Box 33, College Park, Maryland 20740).

67. Dunlap, H. L., Hunsperger, R. G. and Marsh, O. J., 'Develop, ment of ion implantation techniques for microelectronics'- NASA Report N70–17314, October 1969, (Hughes Research Labs.), (NASA Scientific and Technical Information Facility, P.O. Box 33, College Park, Maryland 20740).

68. Lindley, W. T., Phelan, R. J., Wolfe, C. M. and Foyt, A. G., 'GaAs Schottky barrier avalanche photodiodes', *Appl. Phys. Lett.*, 14, No. 6, pp. 197–9, 15th March 1969.

69. Shifrin, G. A., Zanio, K. R., Jamba, D. M., Jones, W. R., Marsh, O. J. and Wilson, R. G., 'Selective doping of piezo-electric crystals by ion implantation', Semi-annual Report (Hughes Research Labs.) AD693154, 1969.

70. Shifrin, G. A., Jamba, D. M., Jones, W. R., Marsh, O. J., Wauk, M. T. and Wilson, R. G., 'Selective doping of piezo-electric crystals by ion implantation', Technical Report No. 2 (Hughes Research Labs.) AD702778, December 1969.

71. Foyt, A. G., Lindley, W. T. and Donnelly, J. P., 'n–p junction photodetectors in InSb fabricated by proton bombardment', *Appl. Phys. Lett.*, 16, pp. 335–7, May 1970.

72. Donnelly, J. P., Foyt, A. G., Hinkley, E. D., Lindley, W. T. and Dimmock, J. O., 'Type conversion and p–n junctions in n–CdTe produced by ion implantation', *Appl. Phys. Lett.*, 12, No. 9, pp. 303–5, 1968.

73. McNally, P. J. and King, W. J., 'Ion implanted mosaic infrared detectors', Paper presented at 16th National Infrared Information Symposium, Fort Monmouth, May 1968.

World Radio History

74. McNally, P. J., 'Ion implantation in InAs and InSb', *Radiation Effects*, 6, No. 2, pp. 149–53, 1970.
75. Foyt, A. G., Lindley, W. T., Wolfe, C. M. and Donnelly, J. P., 'Isolation of junction devices in GaAs using proton bombardment', *Solid State Electronics*, 12, No. 4, 209–14, 1969.
76. Foyt, A. G., Donnelly, J. P. and Lindley, W. T., 'Efficient doping of GaAs by Se^+ ion implantation', *Appl. Phys. Lett.*, 14, pp. 372–4, 1969.
77. Hunsperger, R. G. and Marsh, O. J., 'Electrical properties of Cd, Zn and S ion-implanted layers in GaAs', *Radiation Effects*, 6, Nos. 3 and 4, pp. 263–8, 1970.
78. Manchester, K. E., 'Getting dopants on the beam is a long term project', *Electronics*, 40, pp. 116–7, 13th November 1967.
79. Schroeder, J. B. and Dieselman, H. D., 'Implantation of zinc into gallium arsenide', *Proc. I.E.E.E.*, 55, pp. 125, January 1967.
80. Mayer, J. W., Marsh, O. J., Mankarious, R. and Bower, R., 'Zn and Te implantations into GaAs', *J. Appl. Phys.*, 38, pp. 1975–6, 1967.
81. Hunsperger, K. G. and Marsh, O., 'Anneal behaviour of defects in ion-implanted GaAs diodes', *Metallurgical Trans.*, 1, p. 603, March 1970.
82. Sansbury, J. D. and Gibbons, J. F., 'Properties of ion implanted sulphur and carbon in GaAs', *Radiation Effects*, 6, Nos. 3 and 4, pp. 269–78, 1970.
83. Sansbury, J. D. and Gibbons, J. F., 'Conductivity and Hall mobility of ion-implanted silicon in semi-insulating gallium arsenide', *Appl. Phys. Lett.*, 14, No. 10, pp. 311–12, May 1969.
84. Zelevinskaya, V. M., Kachurin, G. A., Pridachin, N. B. and Smirnov, L. S., 'Low resistivity layers produced by ion bombardment in semi-insulating gallium-arsenide', *Soviet Physics—Semiconductors*, 4, No. 2, pp. 258–60, 1970.
85. Freeman, J. H., 'Implantation machines', Proc. European Conf. Ion Implantation, pp. 1–18 (Peter Peregrinus Ltd., London, 1970).
86. Nicholas, K. H., Goldsmith, B. J., Freeman, J. H., Gard, G. A., Stephen, J. and Smith, B. J., 'The evaluation of the Harwell–Lintott industrial ion implantation machine by making silicon planar resistors', *J. Phys.*, E, 5, pp. 309–10, 1972.
87. MacRae, A. U., 'Recent advances in ion implanted junction device technology', Proc. 2nd Int. Conf. Ion Implantation Garmisch, pp. 329–34 (Springer-Verlag, Berlin, 1971).
88. Dill, H. G., Toombs, T. N. and Bauer, L. O., 'Recent advances in ion implanted m.o.s. technology', Proc. 2nd Int. Conf. Ion Implantation Garmisch, pp. 315–28 (Springer-Verlag, Berlin, 1971).

Manuscript received by the Institution on 24th February 1972.
(Paper No. 1453/CC132)

© The Institution of Electronic and Radio Engineers, 1972

STANDARD FREQUENCY TRANSMISSIONS—May 1972

(Communication from the National Physical Laboratory)

May 1972	Deviation from nominal frequency in parts in 10^{10} (24-hour mean centred on 0300 UT)			Relative phase readings in microseconds N.P.L.—Station (Readings at 1500 UT)		May 1972	Deviation from nominal frequency in parts in 10^{10} (24-hour mean centred on 0300 UT)			Relative phase readings in microseconds N.P.L.—Station (Readings at 1500 UT)	
	GBR 16 kHz	MSF 60 kHz	Droitwich 200 kHz	GBR 16 kHz	†MSF 60 kHz		GBR 16 kHz	MSF 60 kHz	Droitwich 200 kHz	GBR 16 kHz	†MSF 60 kHz
1	—	0	0	—	650·0	17	—	0	+0·1	—	647·0
2	—	+0·2	0	—	657·0	18	—	0	0	—	646·8
3	—	+0·1	0	—	656·4	19	—	0	0	—	647·1
4	—	0	+0·1	—	658·8	20	—	0	0	—	647·0
5	—	0	0	—	659·1	21	—	+0·1	0	—	646·4
6	—	+0·1	0	—	658·5	22	—	—	0	—	—
7	—	+0·1	0	—	657·1	23	—	—	0	—	—
8	—	+0·1	0	—	657·5	24	—	—	0	—	—
9	—	+0·1	0	—	656·5	25	—	—	0	—	—
10	—	0	0	—	656·3	26	—	—	0	—	—
11	—	+0·1	+0·1	—	655·3	27	—	—	0	—	—
12	—	0	−0·1	—	655·0	28	—	—	0	—	—
13	—	+0·1	0	—	654·5	29	—	—	0	—	—
14	—	+0·1	0	—	653·6	30	—	—	0	—	—
15	—	+0·2	0	—	651·9	31	—	—	0	—	
16	—	+0·1	0	—	651·2						

All measurements in terms of H.P. Caesium Standard No. 334, which agrees with the N.P.L. Caesium Standard to 1 part in 10^{11}.

† Relative to AT Scale; $(AT_{NPL} - \text{Station}) = +468{\cdot}6$ at 1500 UT 31st December 1968.

The GBR and MSF Transmitters at Rugby are off the air for maintenance until the end of July.

LETTERS

The History of Positive Feedback

In his most interesting article on 'The History of Positive Feedback', Professor Tucker says of the Autodyne 'that it was only used for receiving telegraph signals (which was all it was suited for)'.

A few months ago I dismantled the last of several regenerative receivers, designed and built by myself around 1931, which were all used for broadcast reception. A total of five valves was used, the first an autodyne frequency changer feeding a steerable i.f. amplifier and with a second detector fitted with switchable super-regenerative circuits. The receivers were very sensitive and easy to operate and it was possible to listen to American broadcasting on most days. When ionospheric conditions permitted, the police cars touring around the streets of Chicago and New York were a part of my regular listening.

MICHAEL LEESTON-SMITH
T.Eng.(CEI), (*Associate*).

9 Sandpits Road,
Petersham,
Richmond-on-Thames,
Surrey.
20th March 1972.

In Professor Tucker's paper the word 'autodyne' is chiefly used to describe a self-oscillating detector which is intended for c.w. reception; that is, so that an audio-frequency beat note is produced. One sentence, (p. 76, Section 5, 2nd para) however, suggests that it applies also to the case of a much higher beat frequency.

I do not myself remember 'autodyne' being commonly used for a self-oscillating detector. I fancy its use was dying at the time when I was first getting seriously interested in radio (about 1925). What I do remember is that 'autodyne' was used for a single-valve frequency changer in a superheterodyne; it was usually called an autodyne frequency changer.

Mr. Leeston Smith's receiver employed an arrangement which was sometimes adopted to prevent radiation of the quench frequency.

W. T. COCKING
C.Eng., F.I.E.E.

29 Arundel Avenue,
East Ewell,
Surrey.
18th April 1972.

I am grateful to Mr. Cocking for adding his experiences of early superheterodyne receivers to Mr. Leeston Smith's interesting recollections. It seems that the term 'autodyne' originally used for a self-oscillating c.w. heterodyne device, became transferred to a single-valve frequency-changer performing almost identically the same function.

D. G. TUCKER
D.Sc., C.Eng., F.I.E.E., F.I.E.R.E.

Department of Electronic & Electrical Engineering,
The University of Birmingham,
P.O. Box 303,
Birmingham B15 2TT.
2nd May 1972.

Professor Tucker's paper throws light both on the technical background to the technique of positive feedback and on the patent litigation which often follows an invention. In connection with the latter some observations by Professor E. H. Armstrong following the final Supreme Court decision add an interesting postscript which may well be new to many readers.

The American magazine *Electronics* of June 1934 had said:

'The amount of money that has gone into this fight must run to several millions of dollars; so far as the art was concerned, wasted, gone to attorneys and patent lawyers instead of being reinvested in further research to the benefit of the art . . . So far as recognition goes, both de Forest and Armstrong are appreciated as inventors of the first rank—the only regret is that their energies could not have been spent exclusively in invention and not futilely dissipated in litigation'.

Armstrong's reply in a letter to W. R. Maclaurin (quoted in the latter's book 'Invention and Innovation in the Radio Industry', Macmillan, New York, 1949) was as follows:

'There is one comment that might be made with respect to the editorial in *Electronics* and that is in respect to the words "not futilely dissipated in litigation". The whole proceeding may appear on the surface to have been a futile one but a man's destiny sometimes moves in a very strange way. Because of the circumstance that a patent attorney on the other side made a statement in the regenerative circuit case that wasn't true, and because I had the burden of setting up apparatus to prove what the truth was, I accidentally ran into the phenomenon of super-regeneration. The sale of that invention a year later was to net me more of a return than the sale of the regenerative circuit and the superheterodyne combined. It was that invention (super-regeneration) which furnished, in fact, the resources by which I was able to continue my investigation of the problem of static that was to lead to the development of frequency modulation. I would never have made the super-regenerative invention had I not been engaged in the regenerative circuit litigation'.

Perhaps Armstrong's experience will provide a ray of encouragement to the many other inventors whose expectations may seem to have been dashed to pieces in the law courts!

F. W. SHARP
C.Eng., F.I.E.R.E.

65 Priory Road,
Cheam,
Sutton, Surrey.
12th May 1972.

Letters commenting on published papers or putting forward other matters of technical interest are invited by the Editor for consideration for publication. Longer letters will normally be regarded as 'short contributions' and will be dealt with in the same manner as submitted papers with an accelerated refereeing process.

World Radio History

U.D.C. 621.317.361: 621.374

Digital Methods of Frequency Measurement: A Comparison

J. D. MARTIN, M.Sc.(Eng.), C.Eng., M.I.E.E.*

SUMMARY

Some well known and some novel methods of digital frequency measurement are compared, with particular emphasis on their dynamic performance. A simple uniform analysis helps to show the inter-relationships between these measurement systems.

* *School of Electrical Engineering, University of Bath, Claverton Down, Bath BA2 7AY.*

1. Introduction

Frequency measurement is often required in signal processing for control, instrumentation and communication purposes. In communications, frequency tends to be a time-varying quantity and estimates of instantaneous-frequency have been made by using the phase-shifting properties of analogue networks although digital circuits are rapidly being applied. In control and instrumentation applications, frequency tends to be regarded as a slowly varying parameter, and digital methods of frequency measurement are today preferred. There are several methods of measuring frequency using common digital techniques, and it is not always obvious which of these will give the desired result, or indeed whether there is anything at all to choose between them.

A digital frequency measuring device takes in a pulse-rate or digital frequency signal as input, and produces a digital-code parallel-output signal representing an estimate of the input frequency.

In this present discussion, it is hoped to survey, describe, and compare several of the common and some not-so-common methods of frequency measurement using digital techniques. Each will be considered from the points of view of resolution, and tracking rate. Some idea may be acquired of relative circuit complexity, but since this depends largely upon the type of digital circuits to be used no universal judgments can be made.

2. Terminology and Comparison Techniques

In order to compare the different circuits, they must be put into a generalized form, which may appear somewhat peculiar to someone familiar with a particular circuit in its conventional form. On the schematic diagrams, signals are labelled according to their type, while when considering the functions performed by the circuit elements, signals are referred to only by corresponding problem variables. Thus, a typical pulse-rate or digital frequency signal will be written as $(X)_r$ on a schematic diagram. When using the numerical value of this signal, it will be referred to simply as $x(t)$ or $X(s)$, or just x. Counter contents are expressed by a problem variable such as a, $(a < 1)$. On the schematic diagram this would be shown as $(A)_c$—a digital-code signal in parallel form. The capacity of the counter would be represented by an integer N say, so that the actual count held by the counter at any time would be aN. Signals written as B on a schematic diagram are interpreted as standard logical control signals of the normal Boolean or switching algebra type. A standard block diagram convention is adopted for circuit elements, which will become clear as various examples are taken.

Estimation of pulse-rate or frequency is subject to several kinds of error. First, the pulses themselves may not be spaced equally in time, giving rise to small variations in period, i.e. a fluctuating instantaneous-frequency. Secondly, the measuring circuit has a finite resolution. Thirdly, the measuring circuit suffers from tracking errors when operating on a varying frequency. Errors of the first kind will not be considered further unless they are introduced by synchronizing an arbitrary pulse sequence to a reference clock signal. The discussion

World Radio History

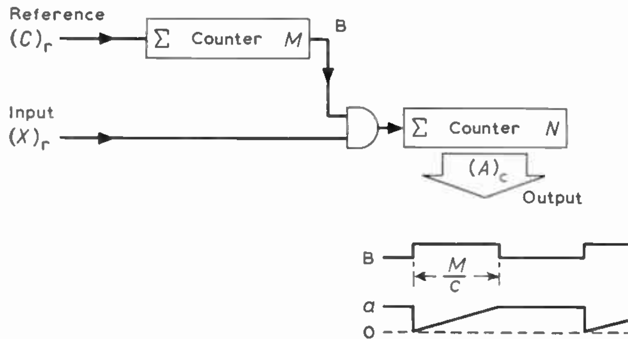


Fig. 1. Gated-counter.

will be limited to errors of the second and third kinds which enable one to compare the different kinds of digital frequency measurement circuit.

The *resolution* of a measuring device will be defined as the change in input frequency which will increment the digital-code output by its least significant digit. *Probable error* or *uncertainty range* is a fractional measure of resolution and is the resolution divided by the actual frequency value. An estimate of small perturbation *tracking performance* is obtained by considering the action when the input frequency varies by an amount equal to its resolution, in a time equal to the measurement time of the device. A large-signal tracking performance index is the time to settle after a full-scale input step of frequency.

3. Counter Systems

Measuring devices in this class assess the frequency of the input pulse-rate by counting the number of pulses occurring in a given time. The resolution is directly determined by the time allowed for the count, and hence the actual number of pulses accumulated, since the uncertainty amounts always to one pulse.

The ubiquitous gated-counter is the prime example of this kind of instrument and is shown in Fig. 1. The input rate $(X)_r$ is compared indirectly with the clock reference rate $(C)_r$, by means of two counters of capacity M and N respectively. The gating time is determined by the clock, and is constant at

$$T = \frac{M}{c}. \qquad \text{......(1)}$$

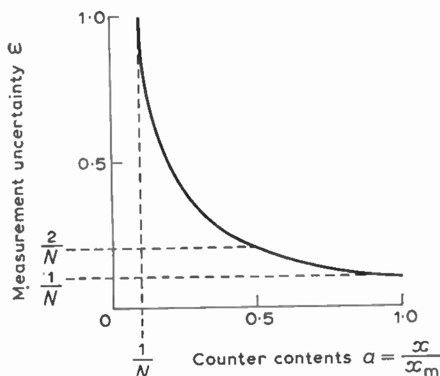


Fig. 2. Measurement uncertainty for gated-counter.

The counter N accumulates the measuring count, and:

$$a = \frac{\bar{x}.M}{c.N} = \frac{\bar{x}}{x_m}. \qquad \text{......(2)}$$

It may be noted that $a < 1$, $\bar{x}$ is a short-term average over T seconds and x_m is the full-scale input frequency.

The resolution of this counter is readily seen to be:

$$\rho = \frac{c}{M} = \frac{1}{T} \qquad \text{......(3)}$$

and the error or uncertainty in measurement is given by:

$$\varepsilon = \frac{\rho}{x} = \frac{x_m}{x.N}. \qquad \text{......(4)}$$

Although the figure of $1/N$ is usually quoted for this error, it is as well to point out by reference to Fig. 2 that the error does rise rapidly at low input frequencies.

Consider now the tracking performance of this circuit, as the input frequency varies linearly with time. A reasonable estimate is that the input signal must not change by more than its resolution during one complete gating period.

Then

$$\dot{x} \leqslant \frac{x_m^2}{N^2}. \qquad \text{......(5)}$$

It is immediately clear that this counter is not very suitable for dynamic situations, and this is the reason why other, more complicated forms of circuit have been investigated.

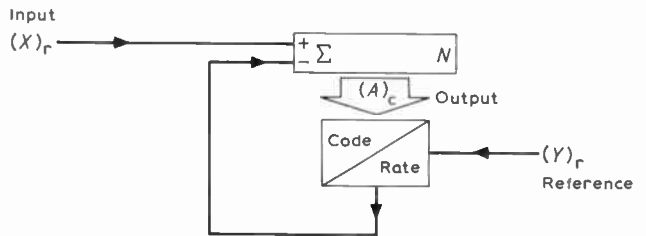


Fig. 3. Closed-loop counter.

The closed-loop counter (Fig. 3) is one attempt at making the basic counter track an input rate-of-change of frequency.[1, 2] The output of the counter is converted into pulse-rate form and subtracted from the input pulse-rate, thus forming a closed loop which settles to the condition where the fed-back pulse-rate and the input pulse-rate are equal. The situation as analysed by the author in another paper,[2] gives the following effective transfer function of this loop as:

$$A(s) = \frac{X(s)}{N} \cdot \frac{1}{s + y/N}. \qquad \text{......(6)}$$

The performance to dynamic inputs is therefore that of a first-order lag circuit with time-constant N/y. The static condition is given by:

$$a = \frac{x}{y} = \frac{x}{x_m}. \qquad \text{......(7)}$$

If the input rate x is fluctuating, the output a is an exponentially smoothed estimate of mean frequency.

World Radio History

The static resolution and error are given by:

$$\rho = \frac{x_m}{N} \qquad \ldots\ldots(8)$$

$$\varepsilon = \frac{x_m}{x.N}. \qquad \ldots\ldots(9)$$

It is possible to change the range of this instrument by including a divider circuit in the loop. A divider of K would reduce the range by K, but by introducing a further integration in the loop would change the dynamic performance.

The tracking performance suffers from the usual characteristics of a first-order lag, in that there is a constant error of $N\dot{x}/x_m^2$ to a ramp input of slope $\dot{x}$, after the initial transient. There will be a further error due to the discrete nature of this device, which can be estimated by considering that the input frequency x changes by an amount ρ, when the counter should be able to track the change by incrementing by $1/N$. Since the input and feedback reference have a random phase, a time of up to N/x_m may elapse before the counter increments.

Thus

$$\dot{x} < \frac{2x_m^2}{N^2} \text{ average.} \qquad \ldots\ldots(10)$$

In practice there will be further aberrations due to unfortunate phase relationships between input and reference, and irregularities in the fed-back pulse-rate if a binary-rate-multiplier is used.[3]

Of further interest is the settling time to a large input step of frequency. Consider the effect of a large step of size x_1. Then the time t_1 for the loop to settle to the new state within a tolerance equal to the resolution, is

$$t_1 = \frac{N}{x_m}.\log_e\left\{\frac{N.x_1}{x_m}\right\}. \qquad \ldots\ldots(11)$$

Notice that the N/x_m term is the settling time of the gated-counter, which is regarded as standard. A comparison will be made later (see Section 7), but it should be noted that the full-scale settling time is considerably increased.

A system having similar characteristics but different hardware is that used by Vincent and Rowles.[4] Instead of pulse-rate feedback, this device uses a digital-code subtraction from a counter contents to produce the exponential smoothing effect. In other respects this system can be classified with the closed-loop counter.

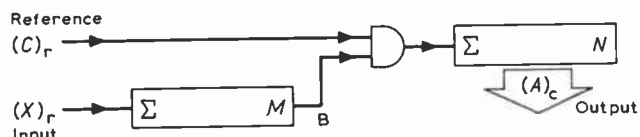


Fig. 4. Period measurement.

4. Period Measurement

The most rapid estimate that can be made of the frequency of a regularly spaced series of pulses, is one based on measurement of the period of the pulse-rate. Period is quite easy to measure, but must be inverted to give the frequency, and hence has not been used very often in the past because of the relatively high complexity of the inverting circuitry. However, modern developments in integrated-circuit fabrication indicate that such objections will not be so important in the near future. Sinha *et al.*,[5] describe a feasible device.

Consider first the problems in measuring period, assuming that the conversion to frequency indication can be carried out without further errors. Figure 4 shows the basic circuit, which is very similar to that for normal counting measurement (Fig. 1). The gating time is now variable, and

$$T = \frac{M}{x}. \qquad \ldots\ldots(12)$$

The count accumulated is given by

$$a = \frac{c.M}{x.N} = \frac{x_m}{x.N}. \qquad \ldots\ldots(13)$$

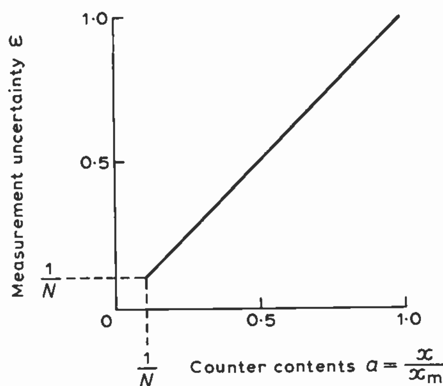


Fig. 5. Measurement uncertainty for period measurement.

This reading is the result of averaging M periods of the input frequency. Since the resolution of the counter is $1/N$, the error or uncertainty in period estimate is

$$\varepsilon_p = \frac{1}{a.N}. \qquad \ldots\ldots(14)$$

The corresponding error for the frequency estimate is

$$\varepsilon_f = \frac{a.N}{1} = \frac{x}{x_m} \quad \text{where} \quad x_m = c.M. \qquad \ldots\ldots(15)$$

Because this error rises to ridiculous heights at large frequencies, only the lowest fraction of the measuring scale can be used (Fig. 5). Further comments are made in Section 7, where the various schemes are compared.

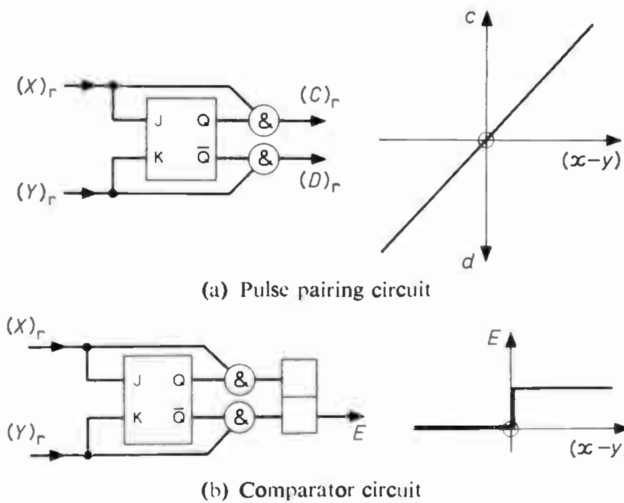
Period measurement is useful in two situations:

(i) For very low frequencies, < 100 Hz say.

(ii) For rapid assessment of frequency in control applications.

For case (i), the error problem highlighted in equation (15) is circumvented by arranging the counter capacity M to vary with range, so that the gating time is approximately constant as ranges are changed. In control applications, particularly in speed measurement where frequencies down to zero are required with high acquisition rate, a rotating transducer may be designed to add a

World Radio History



(a) Pulse pairing circuit

(b) Comparator circuit

Fig. 6. Pulse-rate comparator.

bias frequency to the signal, as described by Sinha *et al.*[5] If the original frequency is represented by y, then the input to the measuring device is actually

$$x = ky + z. \qquad \ldots\ldots(16)$$

Suppose that y is to have a resolution ρ; then this may be related to the counter parameters by taking the maximum value of ε_f in equation (15):

$$x^2 < k\rho x_m. \qquad \ldots\ldots(17)$$

It may be noted that z only enters into the picture to the extent that it indirectly influences x_m and x. There will no doubt be other means of adding a bias frequency to an arbitrary input pulse-rate signal, than the opto-mechanical arrangement used by Sinha *et al.*

Processing the period measurement to derive frequency measurement, is carried out in the case quoted above by a novel application of conventional digital processing techniques. The resolution achieved by this type of processing can be made better than that of the original period measurement, so the overall errors are not significantly increased. Several other approaches are possible, and one employing pulse-rate techniques could be used, but is inherently very slow, giving a conversion time of about 1 second.

5. Frequency Comparators

A frequency or rate comparator compares an unknown frequency with a standard frequency by comparing periodic times. The basic comparator operates by subtraction, pairing pulses such that for the difference $(x-y)$ and $x > y$, one pulse of $(X)_r$ must be suppressed for each pulse of $(Y)_r$. Similarly for $x < y$, the reverse must be true. Such a circuit has been used before for this purpose, and has been called a 'pairing' circuit or a backlash circuit.[1,3,6] Figure 6(a) shows one form of such a circuit, which is self-explanatory except that the bistable must be of a type which will resolve the race hazard between the input signal and the output of the bistable. Many of the integrated-circuit master-slave flip-flops currently available will do this nicely. The flip-flop shown operates on the back edge of an input pulse to the J or K inputs.

The input/output characteristic of this subtractor is also shown in Fig. 6(a). In principle the output rates $(C)_r$ and $(D)_r$ will also be regular if the inputs are both regular, but there will be some small irregularity in practice since it is necessary to guard against coincidence in the input pulses, and this is most conveniently done by synchronizing $(X)_r$ and $(Y)_r$ to two different phases of a higher frequency clock.

Consider now the performance of such an ideal comparator, in terms of time for a decision and resolution. Suppose that we wish to determine whether the two inputs x and y are similar, with an uncertainty ε, the condition is

$$\frac{x-y}{y} \leqslant \varepsilon. \qquad \ldots\ldots(18)$$

Maximum time for a decision is

$$T = \frac{1}{\varepsilon y}. \qquad \ldots\ldots(19)$$

The average time will be half this, but since the phase of the two signals is never known in the unclocked case, the maximum time must always be allowed for a reliable decision.

There is a further parameter, that of resolution. Suppose that the logic circuits which make up the comparator are capable of resolving a time interval τ, such that the result is uncertain if two pulses occur at parallel inputs within this time interval. Then the detection time is virtually unaltered if $\tau \ll 1/y$, but there is an uncertainty zone where the result is in question. The uncertainty zone is given by

$$2y^2\tau. \qquad \ldots\ldots(20)$$

Although this sounds a drastic restriction on the comparator, yet in practice events tend to work together in favour of good resolution. This coincidence problem is well known in the pulse field.

The resolution problem may be circumvented in the comparator by first synchronizing the input pulse-rates to two separate phases of a higher frequency clock. The performance of such a circuit is illustrated in Fig. 7.

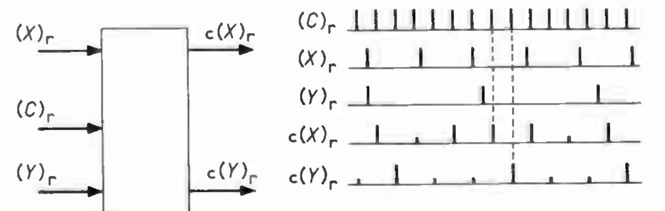


Fig. 7. Clock-synchronizing action.

Clearly the time for a definite decision in a subsequent comparator will now be extended by an additional time interval $1/c$, which is negligible. The comparator action is now ideal, but the resolution problem has been relegated to the clock synchronizing circuit: which can be an advantage in that the synchronizer can be carefully designed without considering any other part of the system.

Although the comparator is comparing frequencies on the basis of period times, the decision time is very long.

World Radio History

This is to take account of arbitrary phase relationships between the two signals. For the case where a period of the unknown frequency x starts off at the same time as that of the reference frequency y, a firm decision can be made within one period time. Thus, if the reference is arranged to always start off at an x channel pulse, then the decision time can be reduced to $1/y$. Thomas[7] makes use of a counting type of comparator which employs this principle.

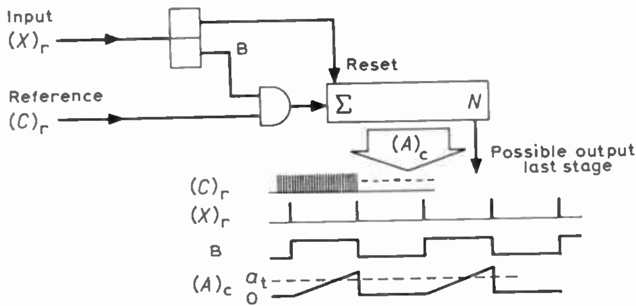


Fig. 8. Counting comparator principle.

The counting comparator (Fig. 8), uses a reference frequency which is higher than the unknown, and is continuously accumulated in a counter. The counter is reset by the unknown signal pulses. The counter contents are represented by the variable a, and a threshold of a_t is set into the counter, perhaps by an overflow output. At the end of an input signal period, the counter contents are

$$a = \frac{c}{xN}. \qquad ......(21)$$

An output occurs if $a \geqslant a_t$, or

$$x \leqslant \frac{c}{a_t . N}. \qquad ......(22)$$

The effective reference is $y = c/n$ say, so that the comparison is

$$x \leqslant \frac{c}{n}. \qquad ......(23)$$

Such a device has a time resolution of $\rho = 1/c$, which is an improvement over the simple comparator, but of course the improvement is only gained by using a correspondingly higher frequency clock reference. If the reference source could be instantaneously phase-locked to the input signal at the beginning of a period, then the resolution of the circuit would be that of the logic elements themselves. This would be tantamount to providing a precision time reference, in the form of a monostable circuit when $n = 1$.

6. Comparator Systems

Frequency comparators similar to that described in the previous section do not seem to have been used very extensively in the art of digital frequency measurement. However, some interesting systems are possible somewhat akin to the several types of analogue-digital converter (a.d.c.) which are in occasional use.[8]

6.1. *Period-loop Counter*

This device is described by Thomas[7] under a different name, but this title is used for comparison with the other systems already discussed. The counting type of comparator is used as a basic element, but is included in a closed loop so as to provide a digital-code type of output signal (Fig. 9).

The input pulse-rate period is compared with the fed-back pulse-rate d, by the counting comparator with an overflow count of N. Thus, at balance

$$d = Nx \qquad ......(24)$$

and overflows alternate with every other period of x. If the input rate is decreased, then the overflows become consistent, and the main accumulator of capacity M climbs steadily at a rate x. Feedback action reduces the rate d until balance is again obtained.

At balance, the counter contents are

$$a = \frac{KNx}{y}. \qquad ......(25)$$

The maximum input rate that can be accommodated is

$$x_m = \frac{y}{KN}. \qquad ......(26)$$

Since many pulses are counted from the d pulse-rate, there is an effective averaging action within the feedback loop which is beneficial if the code/rate converter is of the traditional b.r.m. type. The input signal is not averaged at all.

Considering the resolution of this system, one contribution is made by each of the counters, so adding them together as a first approximation to the total order of uncertainty:

$$\varepsilon = \frac{x_m}{xM} + \frac{1}{N}. \qquad ......(27)$$

The first term is due to the main accumulator, and the second to the comparator. Normally the second term could be made negligible.

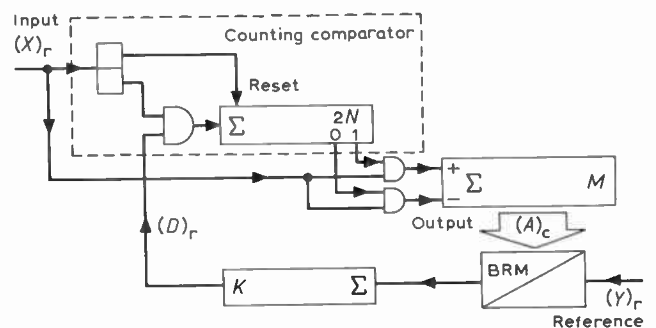


Fig. 9. Period-loop counter.

For an estimate of maximum input tracking rate, suppose that the input rate changes by the amount of its resolution in one input signal period. Then the maximum input tracking rate is given by

$$\dot{x} = \frac{x^2}{N} + \frac{x . x_m}{N}. \qquad ......(28)$$

World Radio History

For a step input, supposing initial conditions of x_1, a_1, and final conditions of x_2, a_2, the main accumulator fills linearly, yielding an overall time of

$$T_{21} = \left\{1 - \frac{x_1}{x_2}\right\} \frac{M}{x_m}. \qquad \ldots\ldots(29)$$

These relationships will be compared in Section 7 with those for the other systems.

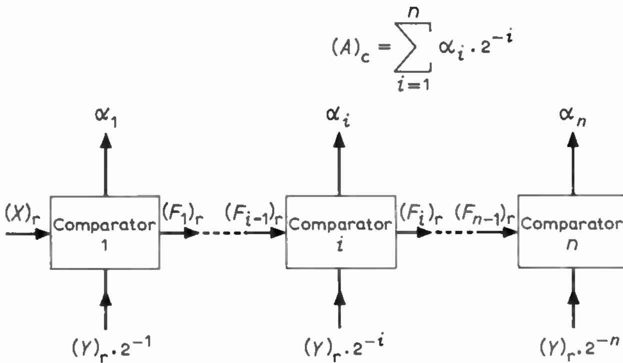


Fig. 10. Sequential-approximation converter.

6.2. *Sequential-approximation Converter*

There are several useful additions which might be made to the basic comparator as described in Section 5.

A flip-flop operated from the two rate outputs will indicate the sign of the difference (Fig. 6(b)). One important feature of this form of comparator is that the transition from one state to another is logically distinct, there is no flutter at the decision boundary.

The comparator error output may also have several different forms. The error is available in signed form, treating $(C)_r$ as positive and $(D)_r$ as negative, or the modulus may be formed by taking the logical operation $(C+D)$. Various combinations of $(X)_r$, $(Y)_r$, $(C)_r$, $(D)_r$, may be employed, switched forward by the state of the sign flip-flop E, and will allow considerable flexibility in the use of this comparator.

This rate comparator may be used in a number of arrangements to improve the performance of pulse-rate to digital-code converters, in a manner analogous to the use of an analogue voltage comparator in conventional a.d.c.s.

A number of the comparator/subtractor circuits may be cascaded after the style of a sequential-approximation a.d.c.[8, 9] with the difference that the reference is conveniently scaled down along the cascade, instead of the error being amplified, and the comparators do not place any limitation on the magnitudes of the compared quantities. The only limitation is one of time, and source stability. Figure 10 shows the form of this sequential-approximation converter, for the case where a binary code is produced.

The scaled down versions of the basic reference frequency y, can be readily obtained from a conventional frequency divider, in either a binary or binary coded decimal form. A comparison is then made between the input frequency or rate, and the scaled down references one at a time. When steady conditions obtain, the digital-code output of the cascade is given by $a = x/y$. The general structure is shown in Fig. 10, where the code output of stage i, α_i is 1 if $f_{i-1} > y.2^{-i}$. The cascaded quantities $(F)_r$ determine the form of code. If a pure binary code is required, then

$$\left.\begin{aligned} f_i &= f_{i-1} \quad \text{if} \quad \alpha_i = 0, \\ f_i &= (f_{i-1} - y.2^{-i}) \quad \text{if} \quad \alpha_i = 1. \end{aligned}\right\} \qquad \ldots\ldots(30)$$

Figure 11(a) shows the form of this transfer characteristic, and the code structure which obtains. The performance is similar to that of the successive-approximation a.d.c., where standards are tested against the unknown, and subtracted from it if the difference is positive.

Use of a different algorithm enables a Gray code to develop. Here in one form

$$f_i = |f_{i-1} - y.2^{-i}| \qquad \ldots\ldots(31)$$

and a transfer characteristic as shown in Fig. 11(b) obtains. The code is actually an inverted form of Gray code, but this may readily be converted to the normal form. Yet another approach gives the transfer characteristic of Fig. 11(c) using the algorithm

$$f_i = |f_{i-1}| - y.2^{-i}. \qquad \ldots\ldots(32)$$

This latter characteristic is that used in the asynchronous sequential-approximation a.d.c. or cyclic one-bit encoder and is only a small variation on that of Fig. 11(b). Changes between these three different transfer characteristics may be effected by very simple alterations in gating structure between stages.

Discrimination is very good with these converters, there being no uncertainty except for hysteresis at the least-significant digit. The Gray code gives a very smooth settling sequence, but the binary code sequence while avoiding critical races between successive digits, allows large spikes in the output code as settling takes place.

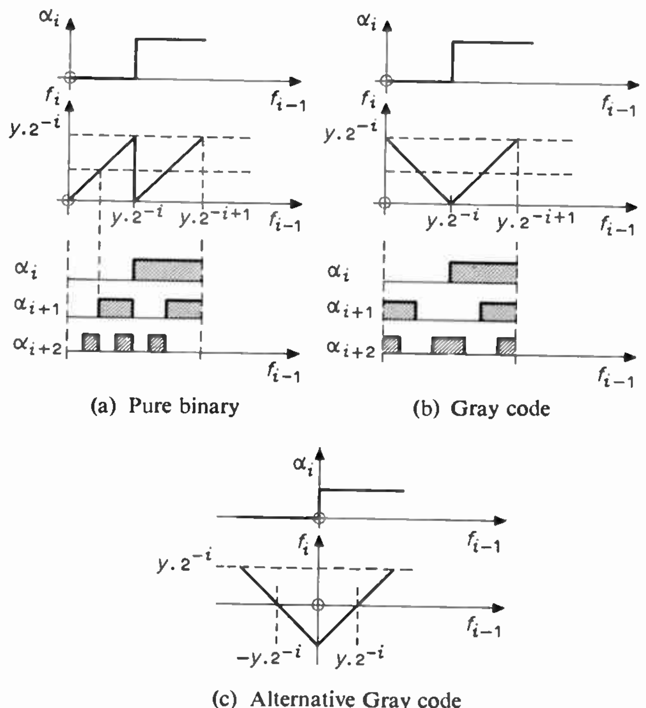


Fig. 11. Comparator transfer characteristics.

World Radio History

Thus, while the binary or b.c.d. converter would be suitable for use with a visual display where short-lived discontinuities would not be noticed, a Gray code converter would be necessary for continuous measurement of frequency changes.

A b.c.d. sequence can be obtained, but division of the reference is not quite so convenient. While it is easy to obtain a 1, 1/2, 1/4, 1/8 sequence of reference frequencies within one decade, an independent divider must be used to provide the 1/10 reference needed to drive the next lower decade.

The decision time needed in the worst case by the ith stage is $2^i/y$, so that the maximum time required by the whole converter will be

$$\sum_{i=1}^{n} \frac{2^i}{y}$$

which is approximately $2^{n+1}/y$. Although this is twice as long as the time needed by the gated counter, this converter is a tracking type, and the ith stage will settle within $2^{i+1}/y$ s. Taking a specific example, consider a 10-bit converter, and a reference rate $y=1$ MHz. The gated counter sampling time is therefore approximately 1 ms, and the worst settling time for this sequential-approximation converter is about 2 ms. However, the latter will settle to 6-bit accuracy within approximately 0·12 ms, and to 8-bit accuracy within 0·5 ms.

In practice the settling time may be expected to be shorter than the figures mentioned above, since the error signal is propagated quite quickly down the cascade. For instance consider the case where full-scale input is applied, when the converter is initially cleared. In the zero input condition, assuming pure binary connexion, all stages will be set to 0 and each one will be passing on its own input. When the signal is applied, it will propagate instantaneously down the cascade, and will set the stages to 1, starting at the least-significant end. As time progresses, the propagated signal is progressively reduced as the stages change state, and the cascade settles to a stable condition. A result is thus obtained quickly, but the reliable result can only be guaranteed after the maximum decision time deduced above. The mechanism of settling is rather complex, and it is difficult to forecast a condition where the maximum decision time obtains in practice. The practical settling time should certainly be between $2^n/y$ and $2^{n+1}/y$, for full accuracy. Experimental results indicate that this limit can be exceeded, when some stages change twice in a settling sequence. This is thought to be due to irregularities in the input pulse-rate, produced by clock synchronizing action.

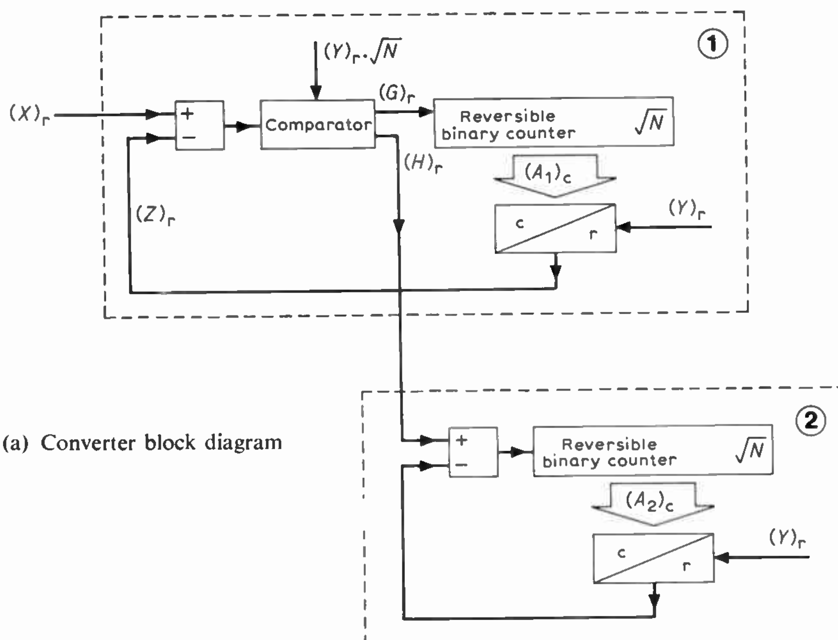
In comparison with the gated counter, the sequential-approximation converter appears slightly more complex, although not more so than the gated counter with storage. However it does provide good tracking performance, and lends itself to a modular construction. Since the converter employs a counter of order n to generate the reference rates, then it would appear to be feasible to construct a general purpose frequency-meter/timer using the sequential-approximation technique.

6.3. *Section Counter*

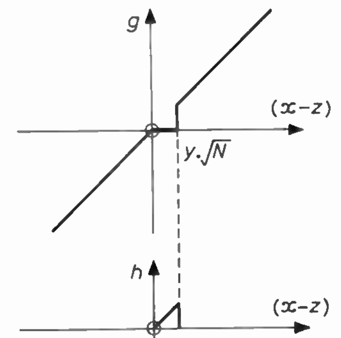
In Section 3, the closed-loop counter is described, and is shown to have an exponential settling characteristic, with time-constant N/y. Thus, the response to a full-scale step change of input is that the counter contents reach full-scale to within a fraction $1/N$, within

$$\frac{N}{y}.\ln(N) \text{ seconds.} \qquad \ldots\ldots(33)$$

For normal values of N this represents a considerable increase over the gated counter time. The code/rate converter shown in Fig. 3 is usually a binary rate-multiplier, which produces an irregular pulse-rate. Consequently, the settling time estimated above is liable in practice to unpredictable variations of the order of



(a) Converter block diagram



(b) Comparator characteristics

Fig. 12. Section-counter converter.

World Radio History

1·5 times.[3] Backlash units like that in Fig. 6(a) may be included in the loop after the subtractor, to iron out these fluctuations to a certain extent.

It appears feasible to divide this counter loop into sections, and achieve separate balance for each section. For example consider division into two, shown in Fig. 12. Balance is obtained in loop 1 over a most-significant range $\sqrt{N}$, and when the loop error falls below $y/\sqrt{N}$, the comparator circuit immobilizes counter 1 and allows counter loop 2 to balance the least-significant range $\sqrt{N}$. Subsequent changes in input are balanced by counter 2 and/or counter 1, depending on their magnitude.

Using equation (33) above, it is possible to estimate the overall conversion time.

Balance for counter 1:

$$t_1 = \frac{\sqrt{N}}{y} \cdot \frac{1}{2} \ln (N). \qquad \ldots\ldots(34)$$

Comparator operation:

$$t_2 = \frac{\sqrt{N}}{y}. \qquad \ldots\ldots(35)$$

Balance for counter 2:

$$t_3 = \frac{N}{y} \cdot \frac{1}{2} . \ln (N). \qquad \ldots\ldots(36)$$

Thus, for normal values of $N(> 100)$, t_3 is the main contributor to the overall conversion time for a step input, which is

$$t = \frac{N}{y} \frac{1}{2} . \ln (N). \qquad \ldots\ldots(37)$$

Following this with a more general argument, if the counter is divided into m sections, then the overall conversion time will be improved by a factor of about $1/m$, compared to the conversion time of the one section counter. This estimate must be treated with care owing to fluctuations produced by the rate-multipliers and because this simple analysis becomes more inaccurate as the number of stages in any particular loop becomes small and exponential behaviour is less evident. As the number of sections increase, the theoretical limiting conversion time is that of the gated counter.

This section-counter modification to the usual form of closed-loop rate/code converter does improve the time response with only a modest increase in hardware. It would seem to be most useful for high resolution converters, and lends itself to a modular construction in b.c.d. units. The precise performance of such converters would need a more careful analysis than that presented here.

7. Comparison of Measurement Systems

The different systems described in the previous sections are now compared by means of Table 1, and the sketch graphs in Figs. 13, 14, 15. The nomenclature used in the table is that used in the schematic diagrams of the systems. Comparison is made of the static uncertainty ϵ, the small-signal tracking rate $\dot{x}$, and the large-signal response time to a full-scale step, T_s. These parameters are normalized so as to aid comparison, making use of the maximum range x_m, and the counter capacity N.

Consider first of all the static uncertainty ϵ, Fig. 13. The period measurement comes off badly in this comparison (curve (c)), and only the lowest part of the measurement scale can be used. There is little difference between the other systems on this basis.

Figure 14 shows the comparison between the maximum allowable small-signal tracking rates. A high tracking rate is often an advantage, although this is not the case for averaged measurements. The best tracking performance is offered by the period measurement technique,

Table 1 Comparison of digital frequency measuring systems

measurement system	maximum range x_m	static uncertainty ε (Fig. 13)	small signal tracking $\dot{x}$ (Fig. 14)	large-signal response time T_s (Fig. 15)
gated-counter (Fig. 1) (a)	$\frac{c.N}{M}$	$\frac{x_m}{xN}$	$\frac{x_m^2}{N^2}$	$\frac{N}{x_m}$
closed-loop counter (Fig. 3) (b)	y	$\frac{x_m}{xN}$	$\frac{2x_m^2}{N^2}$ average	$\frac{N}{x_m} . \ln (N)$
period measurement (Fig. 4) (c)	cM	$\frac{x}{x_m}$	$\frac{x^3}{x_m M}$	$\frac{1}{x_m}$ $(M=1)$
period-loop counter (Fig. 9) (d)	$\frac{y}{KN}$	$\frac{x_m}{xM} + \frac{1}{N}$	$\frac{x^2}{N} + \frac{x.x_m}{M}$	$\frac{M}{x_m}$
sequential-approximation converter (Fig. 10) (e)	y	$\frac{x_m}{xN}$	$\frac{x_m^2}{N^2}$ average	$\frac{2N}{x_m}$
section counter (Fig. 12) (f)	y	$\frac{x_m}{xN}$	$\frac{2x_m^2}{N^2}$ average	$\frac{N}{mx_m} . \ln (N)$

World Radio History

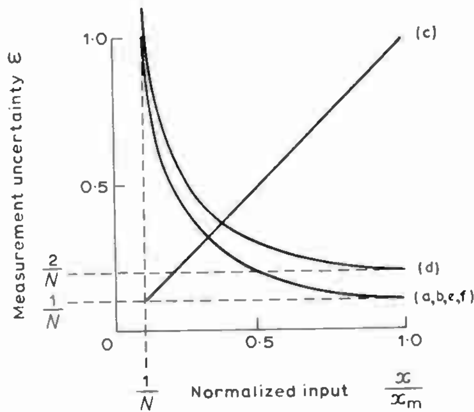


Fig. 13. Comparison of static measurement uncertainty.

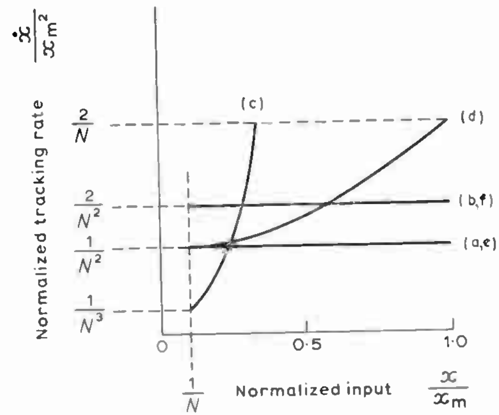


Fig. 14. Comparison of small-signal tracking rate.

since this allows a measurement of frequency to be made within one period. Notice however that the conversion time from period to frequency measurement has been ignored, and would add a restricting limit to the allowable tracking rate. For low frequencies, a conversion time of a few microseconds would not add appreciably to the overall measurement time. The tracking rates for the closed-loop counter, sequential-approximation converter, and section counter, are quoted as average figures, since due to the nature of these devices their performance depends on the phase relationship between the input and reference. The worst tracking rate should not be less than half the average figure.

The large-signal response time (Fig. 15), is basically the response time to a full-scale step of the input frequency. In the case of the exponential response systems (closed-loop counter, and section counter), the response time is that taken to settle to full-scale within an interval equal to one scale increment. This result is almost meaningless for the period measurement, since first, this technique is never used at its full-scale, and secondly the conversion time from period to frequency indication has again been neglected. In the case of the section counter, the parameter m is chosen for the graph such that each section of the counter is a single decade.

8. Conclusion

A comparison has been made of several different techniques for digitally measuring frequency. Some of these are well known, and others are attempts to improve the dynamic performance of the basic gated-counter system. It appears that so far as these particular systems are concerned, the gated-counter represents the optimum for full-scale input step response; period measurement being excluded. For small-signal tracking, Thomas's period-loop counter shows the most promising performance, with only a small degradation in static uncertainty, and also the optimum large-signal response. Systems making use of unlocked frequency comparators tend to be slow because of the fortuitous nature of the phase relationships between unknown and reference. Period measurement represents the fastest technique for frequency measurement, but depends on a digital-code type of processing to convert the period measurement to frequency. Only a limited portion of the total scale can be utilized.

In dynamic applications the tracking types of converter would often be preferred to the gated counter, since they behave better with slewing inputs whose rates exceed the accurate tracking rate of the measurement system.

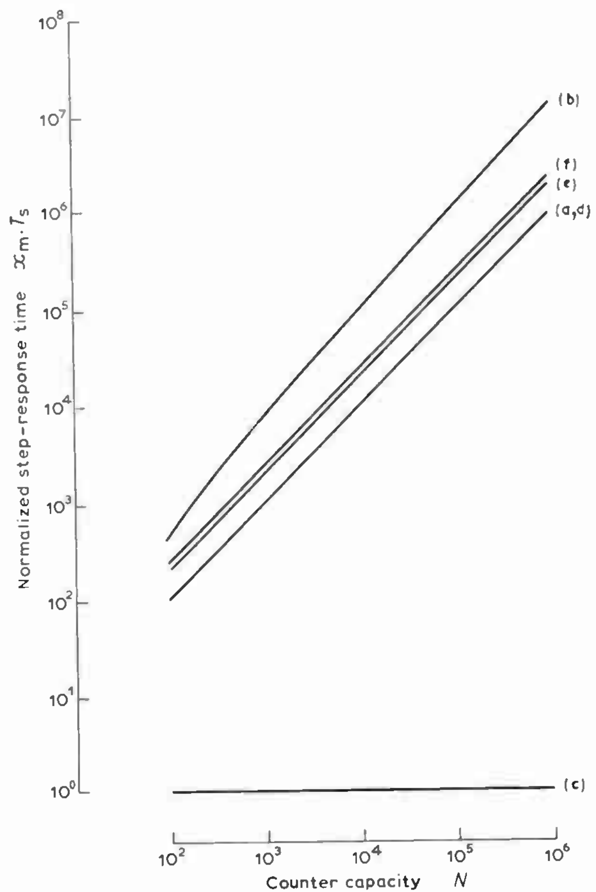


Fig. 15. Comparison of large-signal response times.
Key to curves in Figs. 13, 14 and 15.

(a) gated counter
(b) closed-loop counter
(c) period measurement
(d) period-loop counter
(e) sequential-approximation converter
(f) section counter

World Radio History

9. Acknowledgments

The author would like to thank Professor A. J. Eales, Head of the School of Electrical Engineering, University of Bath, for the use of facilities both in experimental work and preparation of this paper. Mr. D. A. Horne constructed and tested a small sequential-approximation converter, and contributed some interesting ideas.

The contents of Sections 6.2 and 6.3 of this paper were presented in a different form at the I.E.R.E. Conference on 'Digital Methods of Measurement' at Canterbury, July 1969.[9]

10. References

1. Wood, P., 'Frequency meter with continuous digital presentation', *J. Brit. Instn Radio Engrs.* **26**, pp. 109–13, August 1963.
2. Martin, J. D., 'Signal processing and computation using pulse-rate techniques', *The Radio and Electronic Engineer*, **38**, pp. 329–44, December 1969.
3. Lundh, Y., 'Digital techniques for small computations', *J. Brit. I.R.E.*, **19**, pp. 429–49, January 1959.
4. Vincent, C. H. and Rowles, J. B., 'A digital linear ratemeter', *Nuclear Instruments and Methods*, **22**, pp. 201–20, 1963.
5. Sinha, N. K., Szabados, B., and Di Cenzo, C. D., 'New high-precision digital tachometer', *Electronics Letters*, **7**, pp. 174–6, 22nd April 1971.
6. Rey, R. J., 'Digital rate synthesis for frequency measurement and control', *Proc. Inst. Radio Engrs*, **47**, pp. 2106–12, December 1959.
7. Thomas, T. H., 'Rapid pulse-frequency to number converter', *Electronics Letters*, **7**, pp. 345–6, 17th June 1971.
8. Hoeschele, D. F., 'Analog-Digital and Digital-Analog Conversion Techniques' (Wiley, New York, 1968).
9. Martin, J. D., 'Pulse-rate to digital-code tracking converters', Proc. I.E.R.E. Conf. on Digital Methods of Measurement, Canterbury, July 1969, pp. 95–104 (I.E.R.E. Conference Proceedings No. 15).

Manuscript first received by the Institution on 19th January 1972 and in final form on 7th March 1972. (Paper No. 1454/IC64.)

© The Institution of Electronic and Radio Engineers, 1972

World Radio History

U.D.C. 681.38:621.374

Threshold Logic Network Synthesis with Specific Threshold-gate Sensitivities

S. L. HURST,
M.Sc.(Eng.), Ph.D., C.Eng., M.I.E.R.E.*

SUMMARY
The sensitivity characteristics of threshold logic gates have been shown to be dependent upon the Chow parameter values of the gates. This information is here extended to enable threshold circuit synthesis to be undertaken using gates of chosen sensitivity, a requirement akin in Boolean synthesis to using Boolean gates with chosen fan-in limitations.

* *School of Electrical Engineering, University of Bath, Claverton Down, Bath BA2 7AY.*

1. Introduction

Threshold logic gates are binary gates obeying the following input/output relationships, instead of the more familiar Boolean relationships of AND and OR, NAND and NOR, etc.:

$$\text{Gate output} = 1 \text{ iff } \sum_{i=1}^{n} (a_i x_i) \geqslant t_1$$

$$= 0 \text{ iff } \sum_{i=1}^{n} (a_i x_i) \leqslant t_2$$

where x_i = independent binary gate inputs of value 0 or 1, i = 1 to n,

a_i = real-number 'weight' associated with each respective input x_i,

t_1 = real-number upper gate threshold,

t_2 = real-number lower gate threshold, $t_2 \leqslant t_1$.

This output relationship may be expressed in the form:

$$\langle a_1 x_1 + a_2 x_2 + \ldots + a_n x_n \rangle_{t_1 : t_2},$$

where normal arithmetic products and summations, not Boolean, hold within the $\langle\ \rangle$ brackets.

The power of threshold logic gates and the advantages of their use in comparison with normal Boolean gates will be found described in several sources[1-4], although the full impact awaits the forthcoming commercial availability of such gates. Circuit synthesis techniques using threshold gates are also now well established, relying very heavily upon published weight-threshold classification tables, or 'Chow parameters', to identify all possible threshold functions of up to 7 variables,[1, 5] and to give the appropriate minimum-integer realizing weights a_i and gate thresholds t_1, t_2.

As threshold gates act by appropriately weighting the input signals, summing these weighted input quantities, and equating this summation against the preset internal gate threshold, tolerance considerations of the weight and threshold values are of paramount importance. It has, however, been shown that the maximum design tolerance on inputs and thresholds of all threshold gates may be related very simply to the Chow parameters,[6] being inversely proportional to the sum of all the $|a_i|$ values. It has further been suggested that allowable maximum gate tolerance, δ_{max}, may be taken as:

$$\delta_{max} = \pm \left\{ \frac{0 \cdot 5}{\sum_{i=0}^{n} |a_i|} \right\} \times 100\% . \qquad \ldots\ldots(1)$$

The value of the numerator used in this expression is empirical, and in practice may be pessimistic. However, irrespective of the value used, a valid comparison between sensitivities of different gates may be made.

2. Threshold Gate Sensitivities

The Chow parameter tabulations constitute the most compact classification procedure for all possible threshold functions. Their limitation is that they are not readily available and published for functions of more than seven input variables, but this has little real engineering disadvantage—the majority of Boolean gates used rarely have more than four inputs available.

World Radio History

Each Chow parameter classification covers a number of different possible threshold gates, corresponding to permuting the Chow parameter values and taking positive and negative values for each a_i.[2,4] The sensitivity of all these possible gates per classification, however, remains unaltered, as given by equation (1) above.

Hence the published Chow parameter tables may be extended to list the above δ_{max} parameter in addition to the existing $|b_i|$ and minimum-integer $|a_i|$ values. For 3, 4 and 5 input variables, the tabulations therefore become as given in Table 1.

Table 1. Chow parameter tabulations and gate tolerance, δ_{max}, for all 3, 4 and 5-variable threshold functions

n	$\|b_i\|$						$\|a_i\|$						$\pm\delta_{max}$
$\leqslant 3$	8	0	0	0			1	0	0	0			50%
	6	2	2	2			2	1	1	1			10%
	4	4	4	0			1	1	1	0			16·7%
$\leqslant 4$	16	0	0	0	0		1	0	0	0	0		50%
	14	2	2	2	2		3	1	1	1	1		7·1%
	12	4	4	4	0		2	1	1	1	0		10%
	10	6	6	2	2		3	2	2	1	1		5·6%
	8	8	8	0	0		1	1	1	0	0		16·7%
	8	8	4	4	4		2	2	1	1	1		7·1%
	6	6	6	6	6		1	1	1	1	1		10%
$\leqslant 5$	32	0	0	0	0	0	1	0	0	0	0	0	50%
	30	2	2	2	2	2	4	1	1	1	1	1	5·6%
	28	4	4	4	4	0	3	1	1	1	1	0	7·1%
	26	6	6	6	2	2	5	2	2	2	1	1	3·8%
	24	8	8	4	4	4	4	2	2	1	1	1	4·6%
	24	8	8	8	0	0	2	1	1	1	0	0	10%
	22	10	10	6	2	2	5	3	3	2	1	1	3·3%
	22	10	6	6	6	6	3	2	1	1	1	1	5·6%
	22	12	12	4	4	0	3	2	2	1	1	0	5·6%
	20	12	8	8	4	4	4	3	2	2	1	1	3·8%
	20	8	8	8	8	8	2	1	1	1	1	1	7·1%
	18	14	14	2	2	2	4	3	3	1	1	1	3·8%
	18	14	10	6	6	2	5	4	3	2	2	1	2·9%
	18	10	10	10	6	6	3	2	2	2	1	1	4·6%
	16	16	16	0	0	0	1	1	1	0	0	0	16·7%
	16	16	12	4	4	4	3	3	2	1	1	1	4·6%
	16	16	8	8	8	0	2	2	1	1	1	0	7·1%
	16	12	12	8	8	4	4	3	3	2	2	1	3·3%
	14	14	14	6	6	6	2	2	2	1	1	1	5·6%
	14	14	10	10	10	2	3	3	2	2	2	1	3·8%
	12	12	12	12	12	0	1	1	1	1	1	0	10%

From these results, ignoring the trivial cases where $f(x_1, \ldots, x_n)$ is always equal to 0 or 1, we may extract the following information:

(i) Threshold functions of exactly one variable: only one threshold gate is available, namely, a gate with a weighting and threshold of $\langle 1\rangle_{1:0}$, and which has a tolerance value δ_{max} of $\pm 50\%$.

(ii) Threshold functions of exactly two variables: two gates are available, with weightings and thresholds of $\langle 1, 1\rangle_{1:0}$ and $\langle 1, 1\rangle_{2:1}$; both have a tolerance value δ_{max} of $\pm 16{\cdot}7\%$.

(iii) Threshold functions of exactly three variables: the gate specifications available, with their tolerance value δ_{max} are:

$$\left.\begin{array}{lll}\langle 2, 1, 1\rangle_{2:1} & \text{and} & \langle 2, 1, 1\rangle_{3:2} \\ \langle 1, 1, 1\rangle_{1:0} & \text{and} & \langle 1, 1, 1\rangle_{3:2}\end{array}\right\}\delta_{max} = \pm 10\%$$

and

$$\langle 1, 1, 1\rangle_{2:1},\ \delta_{max} = \pm 16{\cdot}7\%$$

(iv) Threshold functions of exactly four variables: the gate specifications available, with their tolerance value δ_{max} are:

$$\left.\begin{array}{lll}\langle 3, 1, 1, 1\rangle_{3:2} & \text{and} & \langle 3, 1, 1, 1\rangle_{4:3} \\ \langle 1, 1, 1, 1\rangle_{1:0} & \text{and} & \langle 1, 1, 1, 1\rangle_{4:3}\end{array}\right\}\delta_{max} = \pm 7{\cdot}1\%$$

$$\langle 2, 1, 1, 1\rangle_{3:2},\ \delta_{max} = \pm 10\%$$

$$\left.\begin{array}{lll}\langle 3, 2, 2, 1\rangle_{4:3} & \text{and} & \langle 3, 2, 2, 1\rangle_{5:4} \\ \langle 3, 2, 1, 1\rangle_{3:2} & \text{and} & \langle 3, 2, 1, 1\rangle_{5:4} \\ \langle 2, 2, 1, 1\rangle_{2:1} & \text{and} & \langle 2, 2, 1, 1\rangle_{5:4}\end{array}\right\}\delta_{max} = \pm 5{\cdot}6\%$$

$$\left.\begin{array}{lll}\langle 2, 2, 1, 1\rangle_{3:2} & \text{and} & \langle 2, 2, 1, 1\rangle_{4:3} \\ \langle 2, 1, 1, 1\rangle_{2:1} & \text{and} & \langle 2, 1, 1, 1\rangle_{4:3}\end{array}\right\}\delta_{max} = \pm 7{\cdot}1\%$$

and

$$\langle 1, 1, 1, 1\rangle_{2:1} \text{ and } \langle 1, 1, 1, 1\rangle_{3:2},\ \delta_{max} = \pm 10\%$$

(v) Threshold functions of exactly five variables: forty-eight gate specifications of exactly five variables are available, compared with the nine for the $n = 4$ case detailed immediately above. Reading from Table 1, it will be found that these forty-eight gates range from $\langle 4, 1, 1, 1, 1\rangle_{4:3}$ and $\langle 4, 1, 1, 1, 1\rangle_{5:4}$, of $\delta_{max} = \pm 5{\cdot}6\%$, through to $\langle 1, 1, 1, 1, 1\rangle_{3:2}$, of $\delta_{max} = \pm 10\%$. The poorest tolerance 5-variable gates have $\delta_{max} = \pm 2{\cdot}9\%$, whilst the best tolerance gate is the latter function with $\delta_{max} = \pm 10\%$.

If the same calculations are carried out for 6-variable functions (not listed in Table 1), it will be found that the poorest and the best maximum-tolerance six-input gates have δ_{max} values $\pm 1{\cdot}5\%$ and $\pm 7{\cdot}1\%$, respectively.

To this information may be added similar information relating to the tolerance of 'universal' threshold gates, that is, gates with a certain chosen combination of inputs and input weightings which are capable of realizing a whole series of threshold functions, for example, all possible functions of four or fewer input variables. Details of proposed universal gate specifications have been published,[7,8] leading to the formulation of the following gate specifications:

$\langle 1, 1, 1, 1\rangle_{2:1}$ for all possible threshold functions of 3 or fewer variables,

$\langle 2, 1, 1, 1, 1, 1, 1\rangle_{4:3}$ for all possible threshold functions of 4 or fewer variables,

and

$\langle 3, 2, 2, 2, 1, 1, 1, 1, 1, 1, 1\rangle_{8:7}$ for all possible threshold functions of 5 or fewer variables.

World Radio History

Table 2 summarizes all this information on gate sensitivities. Of interest from this Table is the result that the tolerance of the universal gates for $\leqslant n$ variables, $n = 3$, 4, and 5, is never inferior to the worst-tolerance threshold function of exactly n variables. This is rather remarkable, as the universal gates always have more than n inputs provided, so as to make them of universal usage, and also their specifications were developed with no consideration for tolerance and sensitivity criteria.

3. Multi-level synthesis of Linearly-separable Functions

Without any restriction of gate sensitivity, the threshold logic realization of linearly-separable functions is entirely straightforward, utilizing the Chow parameter $|b_i|$ and $|a_i|$ tables as necessary.[1, 2, 5] However, when some sensitivity limitation is applied, the single-gate realization of the given function may have to become a multi-level realization in some threshold-OR or threshold-AND form.

The present art of integrated-circuit technology readily allows the matching of resistor values on any one chip to be made within, say, $\pm 5\%$. Matching within $\pm 2\%$ or less can be provided, but with correspondingly poorer yields in production-line acceptance checks. (These percentages are in the matching of resistor values with each other per chip, and not their absolute ohmic values; very fortunately it is the relative values of the resistors that is of significance in the integrated circuit realizations of Amodi[9] and similar circuits rather than their absolute values, and thus this is in line with i.c. fabrication performance.)

Using these figures, which possibly err on the pessimistic side in view of continuing improvements in i.c. manufacturing techniques, from Table 2 it is seen that:

(i) there is no problem in the fabrication of any i.c. threshold gate of up to four variables, or the universal gate for $n \leqslant 4$;

(ii) the better-tolerance $n = 5$ gates present no problem, but the worst-tolerance $n = 5$ gates and the universal $n \leqslant 5$ gate are becoming difficult propositions;

(iii) best-tolerance $n = 6$ gates (a minority) are also possible, but the majority are more difficult propositions.

For the purpose of the immediately following exercise it will be assumed that δ_{max} shall not be less than $\pm 5\%$. Hence one can specify any $n \leqslant 4$ gate, including the universal $n \leqslant 4$ gate without restriction, but only a decreasing percentage of the $n = 5$ and $n = 6$ gates.

Suppose the function:

$$(x_1, \ldots, x_5) = [\bar{x}_2(x_1+\bar{x}_3+x_4+\bar{x}_5)+ +x_4(x_1+\bar{x}_5)+x_1\bar{x}_3\bar{x}_5]$$

is to be realized in threshold-gate form.

Using the standard Chow parameter tabulations, it will be found that this function is linearly-separable, with $|a_i|$ values of 4, 3, 3, 2, 2, 1, giving a single-gate realization of

$$f(x_1, \ldots, x_5) = \langle 2x_1+4\bar{x}_2+\bar{x}_3+3x_4+2\bar{x}_5\rangle_{5:4}$$

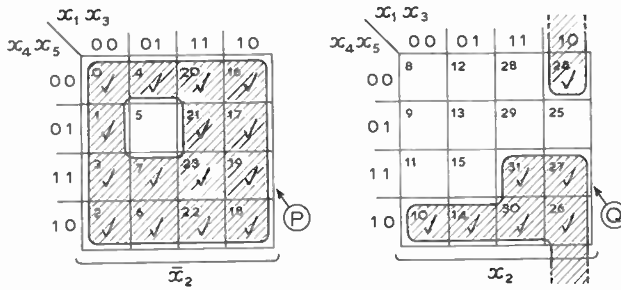
From Table 1, this gate has a tolerance δ_{max} of $\pm 3{\cdot}3\%$, which is outside our arbitrary limit. Therefore a multi-level realization must be sought.

As the function is a linearly-separable function, it may be decomposed about any input variable x_i, the two resulting terms always themselves being linearly-separable functions.[4, 10] Making the arbitrary choice of decomposition about $\bar{x}_2$, x_2, the given function may be plotted on a pair of Karnaugh maps, as shown in Fig. 1(a).

Table 2. Gate tolerance statistics for standard threshold gates

Function			Tolerance $\pm\delta_{max}$
Description	Chow $\|a_i\|$ parameters	Threshold realizations	
3 input variables, worst tolerance	2, 1, 1, 1	$\langle 2,1,1\rangle_{2:1}$ through to $\langle 1,1,1\rangle_{3:2}$	10%
3 input variables, best tolerance	1, 1, 1, 0	$\langle 1,1,1\rangle_{2:1}$	16·7%
4 input variables, worst tolerance	3, 2, 2, 1, 1	$\langle 3,2,2,1\rangle_{4:3}$ through to $\langle 2,2,1,1\rangle_{5:4}$	5·6%
4 input variables, best tolerance	2, 1, 1, 1, 0 and 1, 1, 1, 1, 1	$\langle 2,1,1,1\rangle_{3:2}$, $\langle 1,1,1,1\rangle_{2:1}$, $\langle 1,1,1,1\rangle_{3:2}$	10%
5 input variables, worst tolerance	5, 4, 3, 2, 2, 1	$\langle 5,4,3,2,2\rangle_{8:7}$ through to $\langle 4,3,2,2,1\rangle_{9:8}$	2·9%
5 input variables, best tolerance	1, 1, 1, 1, 1, 0	$\langle 1,1,1,1,1\rangle_{3:2}$	10%
6 input variables worst tolerance	8, 7, 6, 5, 4, 3, 2	$\langle 8,7,6,5,4,3\rangle_{16:15}$ through to $\langle 7,6,5,4,3,2\rangle_{18:17}$	1·5%
6 input variables, best tolerance	1, 1, 1, 1, 1, 1, 1	$\langle 1,1,1,1,1,1\rangle_{3:2}$ and $\langle 1,1,1,1,1,1\rangle_{4:3}$	7·1%
Universal gate for $\leqslant 3$ variables	1, 1, 1, 1, 1	$\langle 1,1,1,1\rangle_{2:1}$	10%
Universal gate for $\leqslant 4$ variables	2, 1, 1, 1, 1, 1, 1, 1	$\langle 2,1,1,1,1,1,1\rangle_{4:3}$	5·6%
Universal gate for $\leqslant 5$ variables	3, 2, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1	$\langle 3,2,2,2,1,1,1,1,1,1,1\rangle_{8:7}$	2·9%

World Radio History



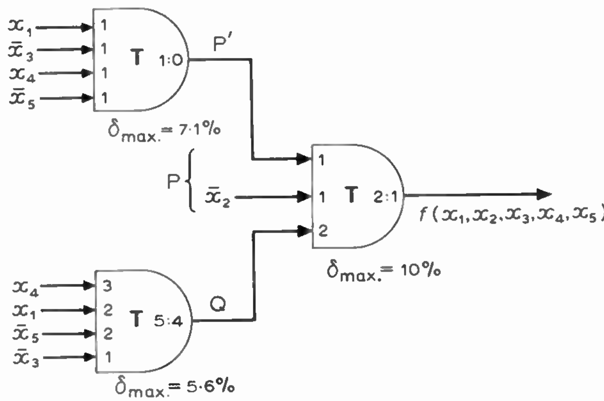
(a) Karnaugh-map plot of function

$P = \bar{x}_2(x_1 + \bar{x}_3 + x_4 + \bar{x}_5),$

$= \bar{x}_2[\langle x_1 + \bar{x}_3 + x_4 + \bar{x}_5\rangle_{1:0}]$

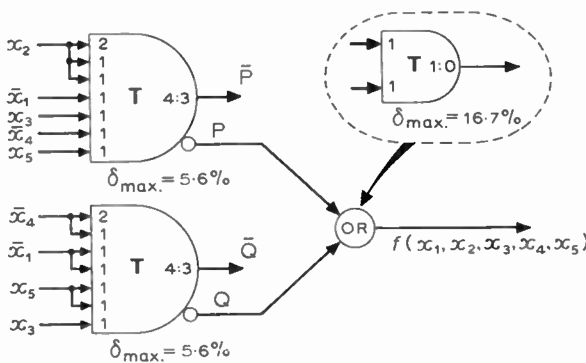
$Q = [x_4(x_1 + \bar{x}_5) + x_1\bar{x}_3\bar{x}_5],$

$= \langle 2x_1 + \bar{x}_3 + 3x_4 + 2\bar{x}_5\rangle_{5:4}$



(b) Two-level realization

$P' = \langle x_1 + \bar{x}_3 + x_4 + \bar{x}_5\rangle_{1:0}$



(c) Two-level realization using universal $n \leqslant 4$ threshold gates

$P = \bar{x}_2[\langle x_1 + \bar{x}_3 + x_4 + \bar{x}_5\rangle_{1:0}],$

$= \langle 4\bar{x}_2 + x_1 + \bar{x}_3 + x_4 + \bar{x}_5\rangle_{5:4},$

$\equiv \langle 4x_2 + \bar{x}_1 + x_3 + \bar{x}_4 + x_5\rangle_{4:3}$

$Q = \langle 2x_1 + \bar{x}_3 + 3x_4 + 2\bar{x}_5\rangle_{5:4},$

$\equiv \langle 2\bar{x}_1 + x_3 + 3\bar{x}_4 + 2x_5\rangle_{4:3}$

Fig. 1. Multi-level realization of linearly-separable function $[\bar{x}_2(x_1 + \bar{x}_3 + x_4 + \bar{x}_5) + x_4(x_1 + \bar{x}_5) + x_1\bar{x}_3\bar{x}_5]$, gate tolerances $\geqslant 5\%$.

Each of these four-variable maps may now be separately realized by appropriate 4-variable threshold gates, and the result combined by appropriate AND–OR connectives. Figure 1(b) illustrates one possible solution. An alternative solution using two universal $n \leqslant 4$ gates plus a two-input OR is shown in Fig. 1(c). Both of these example solutions provide a realization with a worst gate tolerance δ_{max} of $\pm 5{\cdot}6\%$.

Alternative realizations for this given function can of course be made possible by choosing some other decomposition and cover for the minterms of the function.

4. Multi-level Synthesis of Non-linearly-separable Functions

As an example of multi-level synthesis of a non-linearly-separable function with specific sensitivity constraints, consider the non-linearly-separable five-variable function

$$f(x_1, \ldots, x_5) = [x_1x_2\bar{x}_3\bar{x}_5 + \bar{x}_1\bar{x}_3x_5 + \bar{x}_2x_3x_5 + x_3x_4x_5 + \bar{x}_3x_4x_5]$$

Assume that a solution with a very high gate-tolerance throughout, say, not less than $\pm 10\%$ in any gate is required. This will preclude the use of the universal $n \leqslant 4$ gate, and only the following threshold-gate specifications, taken from Table 1 will be found useful:

(i) $\langle 1\rangle_{1:0}$, $\delta_{max} = \pm 50\%$

(ii) $\langle 2, 1, 1, 1\rangle_{3:2}$, $\delta_{max} = \pm 10\%$

(iii) $\langle 2, 1, 1\rangle_{2:1 \text{ or } 3:2}$, $\delta_{max} = \pm 10\%$

(iv) $\langle 1, 1, 1\rangle_{1:0 \text{ or } 3:2}$, $\delta_{max} = \pm 10\%$

(v) $\langle 1, 1, 1\rangle_{2:1}$, $\delta_{max} = \pm 16{\cdot}7\%$

(vi) $\langle 1, 1\rangle_{1:0 \text{ or } 2:1}$, $\delta_{max} = \pm 16{\cdot}7\%$ and

(vii) $\langle 1, 1, 1, 1\rangle_{2:1 \text{ or } 3:2}$, $\delta_{max} = \pm 10\%$

Let the given function be plotted, say, on a $\bar{x}_5, x_5$ pair of Karnaugh maps, as shown in Fig. 2(a). (*Note:* the choice of $\bar{x}_5$, x_5 is arbitrary; any other choice may be tried.)

All threshold gate specifications are characterized by producing unique patterns on Karnaugh map layouts. These may readily be plotted when required, and have been catalogued for immediate reference for all possible 3 and 4-variable functions.[11] It may therefore readily be shown that a choice of four of the available threshold gate specifications listed above will cover the given function, as shown in Fig. 2(b). These covers therefore are:

Cover A: $\bar{x}_5.[x_1x_2\bar{x}_3],$

$= \bar{x}_5.[\langle x_1 + x_2 + \bar{x}_3\rangle_{3:2}].$

$\nwarrow \delta_{max} = \pm 10\%.$

Cover B: $\bar{x}_5.[\bar{x}_3x_4],$

$= \bar{x}_5.[\langle \bar{x}_3 + x_4\rangle_{2:1}].$

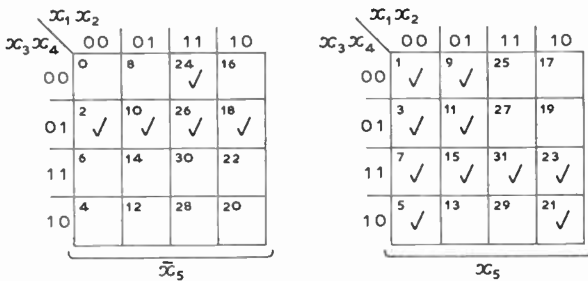
$\nwarrow \delta_{max} = \pm 16{\cdot}7\%.$

Cover C: $x_5.[\bar{x}_1(\bar{x}_3 + x_4)],$

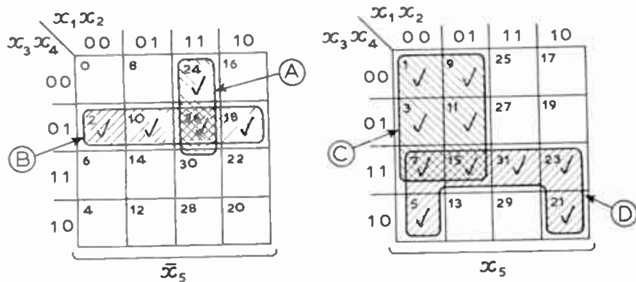
$= x_5.[\langle 2\bar{x}_1 + \bar{x}_3 + x_4\rangle_{3:2}].$

$\nwarrow \delta_{max} = \pm 10\%.$

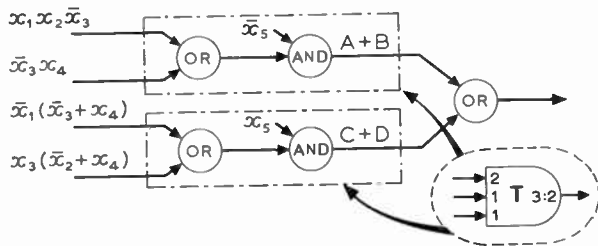
World Radio History



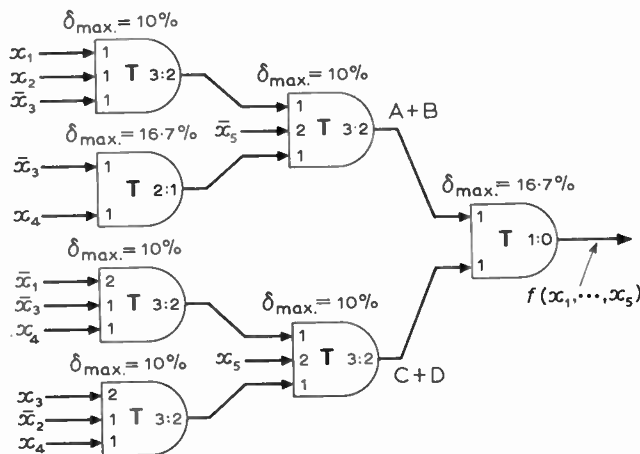
(a) Karnaugh-map plot of given function.



(b) Cover of given function with four permitted threshold gates.



(c) Schematic of final realization.



(d) Final realization using permitted gates.

Fig. 2. Multi-level realization of non-linearly-separable function $[x_1 x_2 \bar{x}_3 \bar{x}_5 + \bar{x}_1 \bar{x}_3 x_5 + \bar{x}_2 x_3 x_5 + x_3 x_4 x_5 + \bar{x}_3 x_4 \bar{x}_5]$, gate tolerances $\geqslant 10\%$.

$$\text{Cover D:} \quad x_5 . [x_3(\bar{x}_2 + x_4)],$$
$$= x_5 . [\langle 2x_3 + \bar{x}_2 + x_4 \rangle_{3:2}].$$
$$\nwarrow \delta_{max} = \pm 10\%.$$

These four gates of permitted sensitivity must now be combined together, as shown schematically in Fig. 2(c), giving a final three-level realization shown in Fig. 2(d).

A relaxation of the original tolerance specification from $\pm 10\%$ to $\pm 5\%$ would readily result in the saving of at least two gates. For example, the final $\langle 1, 1 \rangle_{1:0}$ OR-gate and either the upper or the lower preceding gate could be combined into a single $\langle 3, 2, 1, 1 \rangle_{3:2}$ gate, and also the [A+B] cover could be combined into a single $\langle 3, 2, 1, 1 \rangle_{5:4}$ gate.

It may be concluded that if an extremely high gate tolerance is specified, the number of gates necessary to realize any given function may be appreciably greater than if smaller-tolerance gates are permitted. This of course is closely analogous to fan-in considerations in normal Boolean gate synthesis—if only a small fan-in availability is present, then a greater number of gates will be required compared with the situation where higher fan-in capabilities are available.

5. Further Considerations

In the examples considered above, decomposition of the given binary function into an appropriate threshold-gate realization was readily performed, as the number of input variables present was not excessive. Indeed, for up to six variables the techniques illustrated with supporting Karnaugh-map plotting provide a ready procedure, particularly with the availability of appropriate published information.[11]

Where a large number of input variables is involved, say $n \geqslant 7$, the situation is not so clear, as indeed is the case in normal Boolean synthesis. The Chow parameters of Winder[5] classifying all possible threshold functions of $n \leqslant 7$ represents the upper limit of currently catalogued information for threshold working, and in view of the exponential increase in tabulation lengths with n this limit is likely to remain. Further, in both the Boolean and the threshold field it is unlikely that general-purpose gates with a high fan-in capability will become readily and economically available.[7, 8]

Thus in both fields of realization, a method of decomposing any given function of a large number of variables into some multi-level realization using available gates, becomes necessary. At present this is largely an intuitive process, which yields good if not necessarily optimum solutions. To replace such intuitive work by more formal methods undoubtedly involves computer-aided-design methods, in view of the volume of data to be handled when n is large. Current work in this field on decomposition and matrix operators [12, 13] may well be extended to computer-aid the design of threshold networks, with or without the sensitivity constraints introduced above as desired.

On the question of hazards and propagation delays in multi-level threshold logic networks, the situation is again very similar to that existing in the Boolean field.

World Radio History

The propagation delay of certain threshold logic gates has been shown to be comparable with that of Boolean gates,[9] and hence no aggravation of this problem should arise with the adoption of threshold logic realizations.

6. Conclusions

Previously available information on threshold logic network synthesis, combined with the recently available information on threshold gate sensitivity performance, means that threshold network synthesis can now be undertaken incorporating gates of any chosen or available sensitivity performance.

Hence all the major design problems of engineering a given digital system in threshold-gate rather than Boolean-gate form would appear to be now appreciated, practical system realization awaiting only the commercial availability of such gates.

In connexion with the latter point, the following factors may be worthy of consideration:

(i) The number of threshold gates required to realize any given digital system is never greater than the number of Boolean gates necessary; on the average the adoption of threshold gates show a saving of some two or three to one in the total gate population per system;[1, 2, 3]

(ii) there would appear to be no problem in producing threshold gates, of performance comparable with Boolean gates, with present-day i.c. technology;[9]

and finally,

(iii) the manufacturing costs, reliability and repair time of a modern logic system are largely becoming a function of the number of i.c. packages per system, irrespective of package complexity.

These factors combined would suggest that general-purpose threshold logic gates should now be considered as a viable commercial proposition, their adoption providing a valuable reduction in the gate population per complex system.

6. References

1. Hurst, S. L., 'Introduction to threshold logic: a survey of present theory and practice', *The Radio and Electronic Engineer*, **37**, pp. 339–51, June 1969.
2. Lewis, P. M., 'Practical guide to threshold logic', *Electronic Design*, **22**, pp. 66–8, 25th October 1967.
3. Winder, R. O., 'Threshold logic will cut costs, especially with boost from LSI', *Electronics*, **41**, No. 11, pp. 94–103, 27th May 1968.
4. Hurst, S. L., 'Threshold Logic: An Engineering Survey' (Mills & Boon, London, 1971).
5. Winder, R. O., 'Threshold functions through $n = 7$', Scientific Report No. 7, Air Force Cambridge Research Laboratory, Cambridge, Mass., U.S.A., Contract No. AF19(604)–8432, October 1964.
6. Hurst, S. L., 'Sensitivity performance of threshold-logic gates using Chow parameters', *Electronics Letters*, 7, pp. 371–3, 1st July 1971.
7. Cohen, S. and Winder, R. O., 'Threshold gate building blocks', *Inst. Elect. Electronics Engrs Trans. on Computers*, **C-18**, pp. 816–23, June 1964.
8. Hurst, S. L., 'Universal threshold gates for 3-, 4- and 5-variable linearly-separable (threshold) functions', *Electronics Letters*, **6**, pp. 324–6, 28th May 1970.
9. Amodi, J. J., *et al.*, 'An integrated threshold gate', 1967 International Solid-State Circuits Conference, Digest of Technical Papers, University of Pennsylvania, pp. 114, 115.
10. Dertouzos, M. L., 'Threshold Logic: A Synthesis Approach', M.I.T. Research Monograph No. 32, 1965.
11. Hurst, S. L., 'Synthesis of threshold logic networks using Karnaugh-mapping techniques', *Proc. Instn Elect. Engrs*, To be published.
12. Edwards, C. R., 'Partitioning of Boolean functions by variable complementation', *Electronics Letters*, **8**, pp. 138–40, 9th March 1972.
13. Edwards, C. R., 'Some Boolean matrix operators', *ibid.*, **8**, pp. 113–5, 9th March 1972.

Manuscript first received by the Institution on 23rd February 1972 and in final form on 4th April 1972. (Paper No. 1445/CC133.)

© The Institution of Electronic and Radio Engineers, 1972

World Radio History