# THE RADIO AND ELECTRONIC ENGINEER

**The Journal of the Institution of Electronic and Radio Engineers**

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 31 FEBRUARY 1966 NUMBER 2


## INTERNATIONAL ASSOCIATIONS OF ENGINEERS

THE pattern of association between engineering institutions in Great Britain which has been established during the past three years, culminating in the Council of Engineering Institutions, must sooner or later logically find extension by similar groupings on an international scale. Indeed there has been for several years a number of associations having an international basis, but the effectiveness of these organizations has perhaps been limited in the past to their immediate Continents.

In Europe there are two associations: E.U.S.E.C. and F.E.A.N.I. The Conference of Representatives from Engineering Societies in Western Europe and the United States of America (usually referred to as E.U.S.E.C.) was founded in 1948 under United Kingdom initiative and has a somewhat broader based membership than the European Federation of National Associations of Engineers (F.E.A.N.I.), which was founded on the initiative of the French Union of Engineering Societies in 1951 and now has seventeen member countries with a combined membership estimated at 650,000. Great Britain was elected to F.E.A.N.I. in 1965 and representation is jointly through the Council of Engineering Institutions and the Engineers' Guild: the broad range of interests of the Federation embraces both the corporate and individual interests of engineers.

These two international associations have so far tended to emphasize different aims. For instance, E.U.S.E.C. has put forward useful definitions of 'professional engineers' and 'engineering technicians', and the problems of adopting a basic code of professional conduct have been the subject of much active consultation. F.E.A.N.I. for its part has been concerned with the registration of engineers and the equivalence of qualifications in different countries, while it also considers problems associated with the education and training of technicians as distinct from the training of engineers. Both bodies co-operate with international organizations, E.U.S.E.C. having liaison with U.N.E.S.C.O., while F.E.A.N.I. has rather naturally tended to have links with European organizations such as O.E.C.D., the European Economic Community, the Council of Europe and Euratom. E.U.S.E.C. and F.E.A.N.I. have a common membership of fifteen nations and the Joint Committee set up in 1963 will facilitate further discussions on the possibility of establishing an international organization of engineers.

There are several other international groupings of engineering associations and these, because their membership overlaps with E.U.S.E.C. and F.E.A.N.I., provide a useful basis for eventual links on an almost world-wide basis. The United States and Canada are members of the Pan-American Federation of Engineering Societies (U.P.A.D.I.) while Canada and Great Britain are both represented on the Conference of Engineering Institutions of the British Commonwealth which also includes representatives from Australia, Ceylon, India, New Zealand and Pakistan.

The first major meeting of C.E.I.B.C. to be held since the formation of C.E.I. will take place in London in May of this year and it will undoubtedly be a meeting of considerable importance. There is a vital need to establish common ground within the Commonwealth at least on the matter of recognition of qualifications in view of the requirements of legal registration of engineers in countries such as Canada and New Zealand. The tour which the Secretary of the Institution of Electronic and Radio Engineers, Mr. Graham D. Clifford, is currently undertaking in India, Australia, New Zealand and Canada will include discussions on the greater co-operation within the Commonwealth between electronic engineers and their place in Commonwealth and International societies. His discussions in Israel, Singapore and the United States will also provide useful exchanges of opinions on wider associations.

One important point which must be borne in mind in furthering international associations of engineers was stressed by the immediate past President of F.E.A.N.I., Professor Siegfried Balke—the idea of unity must not be confused with that of uniformity.

G. D. C.



World Radio History

# INSTITUTION NOTICES

## New Year Honours

The Council has sent its congratulations to the following members, whose appointments or promotions in the Most Excellent Order of the British Empire have appeared in Her Majesty's New Year Honours Lists:

Professor M. R. Gavin (Member) on his promotion to be a Commander of the Order (Professor Gavin was, until September last, professor of electronic engineering at the University College of North Wales, Bangor; during this period he was the External Examiner for degrees in Electrical Engineering at the University of Hong Kong. He is now Principal of Chelsea Polytechnic, and Vice-Chancellor Designate of the University of Chelsea.)

Commander Sydney Robert Hack (Associate Member) on his appointment as an Officer of the Military Division. (Commander Hack is on the staff of the Director-General, Dockyards and Maintenance, Telecommunications Section, Ministry of Defence.)

Mr. W. D. Hatcher (Associate Member) on his appointment as an Ordinary Member. (Mr. Hatcher is Senior Assistant, Maintenance Co-ordination, Television, British Broadcasting Corporation.

Mr. L. N. Karpovich (Associate) on his appointment as an Ordinary Member. (Mr. Karpovich is a Police Radio Communications Officer in the Hong Kong Police Force.)

Mr. T. J. H. Stanley (Graduate) on his appointment to Ordinary Member. (Mr. Stanley is a Supervisory Engineer at the Broadcasting Division of the Ministry of Culture, Singapore.)

## Conference on Non-Destructive Testing

The Fifth International Conference on Non-Destructive Testing is being sponsored by the Canadian Council for Nondestructive Technology and will be held at the Mount Royal Hotel, Montreal, from 21st to 26th May 1967.

An invitation has been received for British authors to submit papers and copies of 'Notes for the Guidance of Authors' can be obtained from the Secretary of the British National Committee for Non-Destructive Testing, Redfields Home Farmhouse, Church Crookham, Aldershot, Hampshire. Synopses, not exceeding 400 words, are required in Canada by 30th April 1966 and the final papers by 30th November 1966. It is requested that authors send a copy of their synopsis to the British National Committee so that the Committee can co-ordinate the British contribution to this important conference.

The Institution is represented on B.N.C.N.D.T. by Dr. A. Nemet (Member).

## Conference on Electronic Engineering in Oceanography

The Institution's Conference on Electronic Engineering in Oceanography will be held during the week of 12th to 19th September, 1966 at the University of Southampton. (This is a week earlier than was announced in the November 1965 *Journal*.)

The Organizing Committee can still accept offers of papers and further information may be obtained on application to the Secretary of the Programme and Papers Committee, I.E.R.E., 8–9 Bedford Square, London, W.C.1. Registration details and outline programme of the conference will be published in an early issue of *The Radio and Electronic Engineer*.

## International Conference on Electronic Switching

An International Conference on Electronic Switching as applied to telecommunications engineering is being held in Paris from 28th March to 2nd April, under the organization of the Société Française des Electroniciens et des Radioélectriciens. The Institution of Electrical Engineers and the I.E.R.E. were invited to combine to produce a substantial British contribution to the Conference which will comprise some 40 papers; the Institution's representative on the British committee has been Mr. J. Powell, M.Sc. (Associate Member).

The Conference will deal with the following broad topics:

System Design; Connection Networks for Spatial-Switching Systems; Control Components; Programming of Electronic Systems; Time Division Electronic Switching Systems; Special Devices for Switching Systems; and Applications of Computers.

Requests for further information should be addressed to Colloque International de Commutation Electronique, 16 Rue de Presles, Paris 15e, France.

## Members' Annual Subscriptions

Members are advised that as from 1st April 1966 the following subscription rates will be payable:

| | | | |
|---|---|---|---|
| Members, Companions | £12 | Graduates 35 and over | £9 |
| | | Graduates 25–34 | £7 |
| Associate | | Graduates under 25 | £6 |
| Members | £10 10s. | Students 25 and over | £5 |
| Associates | £10 | Students under 25 | £3 10s. |

It should be noted that these subscription rates will apply for members resident in all parts of the world. Details of the new rates, including the amounts payable in local currency in the various parts of the world, are being sent to members, together with a letter from the President, explaining the Council's decision to increase the rates.



World Radio History

U.D.C. 621.382.232.3.011.4

# Varactor Diode Measurements

*By*

Professor
F. J. HYDE, D.Sc.C.Eng.
(*Member*),†

S. DEVAL, M.Sc.†

AND

C. TOKER, Ph.D.‡

*Reprinted from the Proceedings of the Joint I.E.R.E.–I.E.E. Symposium on 'Microwave Applications of Semiconductors' held in London from 30th June to 2nd July, 1965.*

**Summary:** A comparison has been made of the 'relative impedance' and 'transmission' techniques of measuring the small-signal parameters of varactor diodes. Relative-impedance measurements have been made at u.h.f. and X-band and transmission measurements at X-band.

It has been shown that losses associated with the matching circuit used in relative-impedance measurements can give rise to a degraded diode specification. Their effect is less if reduced-height waveguide is used at X-band. The transmission method leads to the most favourable specification for a diode.

The measurements support the representation of the diode by a simple equivalent circuit to a first order.

## 1. Introduction

Two typical cross-sections of varactor diodes designed for use at frequencies up to about 10 Gc/s are shown in Fig. 1. The thickness of the semiconductor region is only a small fraction of a wavelength so that it may be represented by the lumped $R$ and $C$ shown on the right-hand side of Fig. 2(a), with good approximation. For parametric amplifier applications diodes are not markedly driven into conduction so that $C$ and $R$ may be identified, respectively, with the small-signal depletion-layer capacitance of the p-n junction and the bulk resistance of the p- and n-regions on either side of the junction (internal contact resistance is also included in $R$). The resistivity of the p- and n-regions of most varactor diodes is high enough for skin effect to be negligible[1] up to several

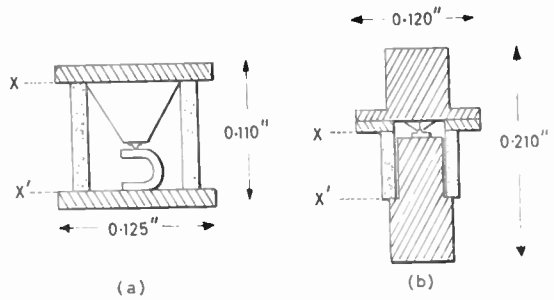



Fig. 1. Typical cross-sections of microwave varactor diodes.

gigacycles per second. Hence $C$ and $R$ can be considered independent of frequency in the u.h.f. and lower microwave regions although $R$ should rise and $C$ fall as skin effect increases in importance. There is some evidence[2, 3] that both $C$ and $R$ may be influenced by surface effects in the v.h.f. and lower u.h.f. bands. The width of the depletion layer is dependent on the junction voltage, so that both $C$ and $R$ are functions of applied voltage. The dependence of $C$ on applied voltage is a first-order effect, and is the reason for the practical usefulness of varactor diodes. For parametric amplifier diodes the dependence of $R$ on applied voltage is expected to be a second-order effect.

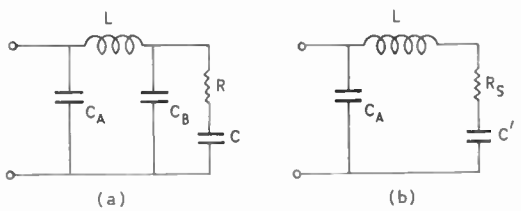



Fig. 2. Lumped small-signal equivalent circuits of a varactor diode: (a) general form (b) simplified form, in which $R_s \simeq R/(C_B/C + 1)^2$, $C' \simeq C + C_B$.

It is obvious from Fig. 1 that the energy-storage capability of the capsule, within which the semiconductor is housed, can only be represented rigorously on a distributed basis. For practical purposes this is too complicated so that the simple approximate representation of energy storage by the lumped elements $L$, $C_A$ and $C_B$ in Fig. 2(a) is used.[4] These play an important part in determining the gain-bandwidth and noise performance of a parametric amplifier. They must be specified in relation to well-defined planes of mounting, such as those shown as X and X′ in Fig. 1.

Several different methods of measuring some or all of the elements in the equivalent circuit of Fig. 2

† Electronics Division, School of Engineering Science, University College of North Wales, Bangor.
‡ Now with the Middle East Technical University, Ankara, Turkey.



World Radio History

have been described. To date the most useful of these are the 'relative impedance' methods devised by Harrison[5] and Houlding[6] and augmented by Mavaddat,[7] and the 'transmission' method described by DeLoach.[8]

In this paper, which is not exhaustive, a comparison is made between these different methods with a view to showing the significance of the effects of losses in the measuring equipment. Theoretical investigation of such effects has been carried out in some detail by Hyde and Smith.[9] Uhlir[10] has also drawn attention to the effect of loss in a special case.

## 2. Capacitance Measurements in the H.F. Band

Measurement of diode capacitance in the microwave region is difficult. In the h.f. band it is much simpler; at such frequencies the reactance of $L$ is negligible and $1/\omega C \gg R$, so that the measured capacitance, $C_m(V)$, at an applied bias voltage $V$ is $C_s+C(V)$; here $C_s = C_A+C_B$ is the total stray capacitance and $C(V)$ is the junction capacitance. A transformer ratio-arm bridge can conveniently be used for such measurements. Care has to be taken in the design of the jig which supports the diode to ensure that there is no significant extra capacitance added.

$C_s$ can be measured for a capsule from which the semiconductor has been omitted. A check on this value and on the 'law' of the diode can conveniently be made using an unpublished method devised by Smith.[11] This will be outlined here. For both alloy-type and simple diffused-junction diodes the depletion-layer capacitance should depend on bias voltage as follows,[12] to a good approximation,

$$C = \frac{C(0)}{\left(1-\frac{V}{\phi}\right)^{1/n}} \qquad \text{.....(1)}$$

Here $C(0)$ is the capacitance at zero bias and $\phi$ is the diffusion potential. For an abrupt junction $n = 2$ and for a linearly-graded junction, to which a diffused junction approximates, $n = 3$. Hence

$$C_m = C_s + \frac{C(0)}{\left(1-\frac{V}{\phi}\right)^{1/n}} \qquad \text{.....(2)}$$

This may be differentiated to give

$$\frac{dC_m}{dV} = \frac{C(0)}{\phi n}\left(1-\frac{V}{\phi}\right)^{-\left(\frac{n+1}{n}\right)} \qquad \text{.....(3)}$$

Since eqn. (2) can be written in the form

$$\left(\frac{C_m-C_s}{C(0)}\right)^{n+1} = \left(1-\frac{V}{\phi}\right)^{-\left(\frac{n+1}{n}\right)} \qquad \text{.....(4)}$$

it follows, on combining eqns. (3) and (4) and taking logarithms, that

$$\log\left(\frac{dC_m}{dV}\right) = (n+1)\log(C_m-C_s) - n\log(C(0)) - \log(\phi^n) \qquad \text{.....(5)}$$

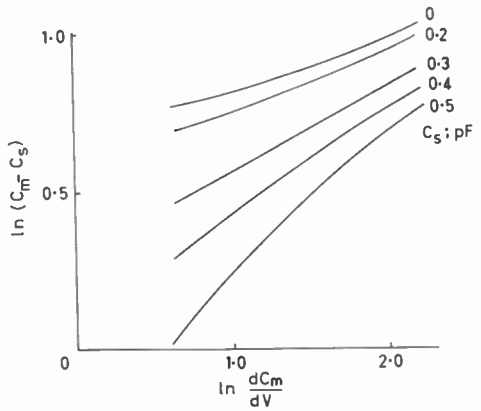



Fig. 3. Plot of high-frequency capacitance data for a silicon diffused-junction diode.

A family of curves of $\log(C_m-C_s)$ versus $\log(dC_m/dV)$ can be plotted for a range of values of $C_s$ around the estimated value. Only for the true value of $C_s$ should a straight line be obtained, provided that the assumed-capacitance-voltage relationship of eqn. (1) is valid. A typical family of curves is shown in Fig. 3 for a silicon diffused-junction diode. A straight line is obtained for $C_s = 0{\cdot}3$ pF. From the slope it is found that $n = 2{\cdot}8$ and from eqn. (1) we find $\phi = 0{\cdot}6$ eV. It should be noted that there is a range of values for $C_s$ over which the departure from linearity is small, so that precise determination of $C_s$, $n$ and $\phi$ is not possible.

## 3. Relative Impedance Measurements

### 3.1. *Theoretical Considerations*

This class of measurements, which gives information on the semiconductor region only, is based on the Weissfloch transformer theorem.[13] In relation to the assessment of varactor diodes this may be paraphrased as follows: By augmenting the reactances of the discontinuities represented by $L$, $C_A$ and $C_B$, $C$ (as defined by a standard bias voltage—which is usually zero, so that $C = C(0)$), with lossless reactive elements, it is possible to define a plane in an input waveguide or coaxial line at which the impedance $Z_{in}$ is given by

$$Z_{in} = Z_0\left(1+j\frac{\Delta X}{R}\right) = Z_0(1+j\Delta Q) \qquad \text{.....(6)}$$

Here $Z_0$ is the characteristic impedance of the waveguide or coaxial system being used and $\Delta X$ is the change in junction reactance corresponding to a



World Radio History

change in capacitance $\Delta C$ produced by changing the bias from the standard value by a known amount $\Delta V$. Equation (6) defines the unit resistance circle on a Smith chart. The change in reactance, $\Delta X$, is measured 'relative to $R$' as $\Delta Q$. It follows that

$$R = \frac{\Delta C}{\Delta Q \omega C(C+\Delta C)} \qquad \ldots\ldots(7)$$

To utilize this formula it is necessary to know the actual dependence of $C$ on $V$, and it is assumed that this can be obtained in the manner described in the previous section. Harrison's, Houlding's and Mavaddat's techniques are all based on eqns. (6) and (7).†

It has been shown theoretically by Hyde and Smith[9] that the $Q$-factor of the semiconductor region of varactors as measured by relative impedance techniques may be lower than the true value because of the effect of losses in the components used to match $R$ to the impedance of the measuring system at the standard bias voltage.

In a typical experimental arrangement at X-band the diode is mounted between the centres of the broad faces of the standard rectangular waveguide, whose height is preferably reduced from the normal dimension to that of the diode at the place where the diode is located, using 'reflectionless' tapers. A choke is incorporated at either the top or bottom of the guide to permit the application of bias. Choke design is described in Appendix 1. Behind the diode there is an adjustable 'short-circuit' and in front an E-H tuner preceded by a slotted waveguide section used as a standing-wave indicator, a precision attenuator, a level-setting attenuator, a wavemeter, an isolator and the klystron supplying the small-signal power. The usual method of making measurements is as follows: with the standard bias voltage applied, minimum v.s.w.r. in the slotted section is obtained by adjusting the short-circuit position behind the diode; matching is completed by adjusting the E-H tuner; the bias is then changed and the v.s.w.r. measured. (The convention used here is that v.s.w.r. > 1.)

† Harrison considered in addition the effect of replacing the impedance of the semiconductor by a short-circuit. In relation to the plot defined by eqn. (6) a point is then defined on the periphery of the Smith chart, whose normalized reaction is $Q = 1/\omega CR$, which is the $Q$-factor of the semiconductor region at the standard bias voltage. Mechanical replacement of the semiconductor region by a short-circuit is not satisfactory in the microwave region. In materials such as silicon and germanium, however, a near-short-circuit can in principle be created by passing a forward current of some tens of mA through the diode. The junction admittance then becomes large and $R$ is reduced towards zero by conductivity-modulation. In practice the near-short-circuit is not good enough for accurate evaluation, so that this technique has not been employed in the work described in this paper. In gallium arsenide varactors the diffusion length of injected carriers is in any event too small for the effect to be utilized.

### 3.2. *The Practical Effect of Losses in the Measuring System*

#### 3.2.1. X-band measurements

The effect of losses in the measuring system is illustrated in Fig. 4. This shows the variation of $\Delta Q$, for a range of values of $\Delta V$, as a function of the position of the short-circuit behind the diode, when this was systematically set to different positions before matching was accomplished by adjusting the E-H tuner. Full-height waveguide was used with the frequency (9 Gc/s) such that the combination of diode and supporting post had inductive reactance. It is to be presumed that the maximum value of $\Delta Q$ obtained for a given $\Delta V$ is the best approximation to the true $\Delta Q$ of the diode, although the smaller the diode resistance the greater will be the effect of losses elsewhere. These are thought to arise mainly in the transformed values at the diode of the non-zero resistances of the adjustable 'short-circuits' behind the diode and in the arms of the E-H tuner.

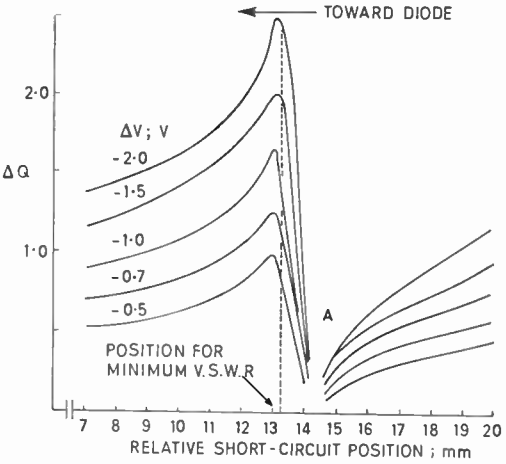



Fig. 4. Dependence of $\Delta Q$ on the position of the short-circuit behind the diode. $\Delta V$ is the parameter.

The general shape of the curves can be explained as follows. When the 'short-circuit' behind the diode is located $\lambda_g/2$ from it, where $\lambda_g$ is the guide wavelength, a near-short-circuit is produced across the diode. Changes in diode bias voltage then produce no change in $\Delta Q$ (region A). When the short-circuit is slightly closer to the diode it has capacitive reactance; at the maximum of the $\Delta Q$ curves this is in approximate resonance with the inductive reactance of the diode and post. For other positions of this short-circuit the effects of transformed loss associated with it and with the short-circuits in the E-H tuner have a greater effect.

At upper X-band frequencies the maximum values of $\Delta Q$ are less than those obtained at lower X-band,



World Radio History

although the difference between the short-circuit positions to give minimum and maximum $\Delta Q$ does not change much. At the higher frequencies the impedance locus lies inside the unit resistance circle on the Smith chart, showing that there are significant losses in the measuring system.

When the standard application of Harrison's technique is employed, in which minimum v.s.w.r. is first obtained by adjustment of the short-circuit behind the diode (the E-H tuner being off-tune), the values of $\Delta Q$ subsequently measured are less than the maximum which are possible. In some cases the differences can be large.

The procedure for obtaining the maximum $\Delta Q$ position is as follows. First the standard matching procedure is carried out. A large forward current is passed through the diode and the v.s.w.r. due to the resulting mismatch measured. The standard bias voltage is then reapplied, the short-circuit position is changed slightly and the E-H tuner re-adjusted for a match. The v.s.w.r. is again measured when a large forward current is passed. The procedure is repeated until the highest v.s.w.r. is obtained in this condition. The short-circuit position is then such that maximum $\Delta Q$ is obtained in normal measurements.

3.2.2. U.h.f. measurements

A coaxial measuring system of 50 ohms characteristic impedance has been developed in which a single adjustable short-circuit is used in conjunction with an L-section matching circuit. The effect of loss associated with a single short-circuit can be more readily interpreted than when several short-circuits are involved.

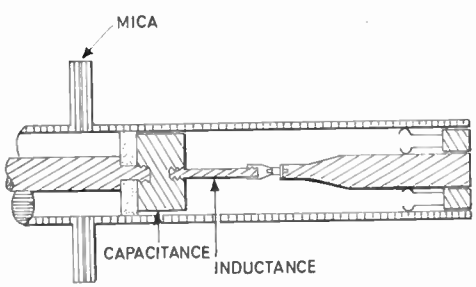
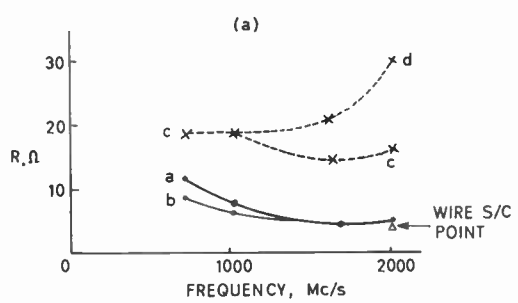


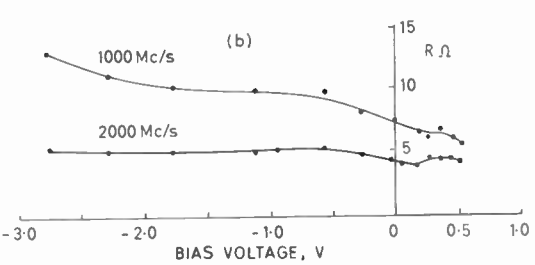

Fig. 5. Diode mount for relative impedance measurements at u.h.f.: spring finger short circuit.

The diode mount and matching elements are shown in Fig. 5. The apparatus is completed by standard equipment for the measurement of standing wave ratio; this is connected on the left. The diode is mounted in a section of central conductor, behind which an adjustable short-circuit is located. This makes sliding contact to both inner and outer conductors by tapered phosphor-bronze spring fingers. The purpose of the adjustable short-circuited length of line behind the diode is to provide a variable inductance which can be adjusted so that it is approximately in resonance with the effective reactance of the diode. In front of the diode two short lengths of line, the first of high characteristic impedance and the second of low characteristic impedance, behave to a close approximation as the 'lossless' inductance and capacitance respectively of an L-C matching section. They are demountable and are connected together and to the line by screw threads. A range of individual L and C components (50 of each) was constructed so that 'matching' could be effected for different diodes over a range of frequencies. V.s.w.r.'s realized in practice were better than 1·05. The varying length of different L and C combinations was accommodated by bodily moving the diode and short-circuit behind it. On the left of the adjustable section a 1200-pF blocking capacitor is incorporated in the outer conductor, with the inner conductor continuous; bias is applied across this capacitor.

Typical apparent variations of $R$ with frequency, obtained with the apparatus as described above and calculated using eqn. (7) are shown as the continuous curves 'a' in Figs. 6(a) and 7(a), for a gallium arsenide and a silicon diode respectively. The standard technique for matching at the standard bias voltage




(a) Dependence of calculated 'diode' resistance on frequency for matching with: a, b lumped L, C components; c adjustable line and stub; d two stubs.

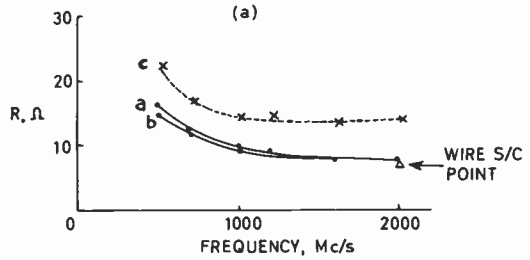



(b) Dependence of calculated 'diode' resistance on bias voltage at 1000 and 2000 Mc/s.

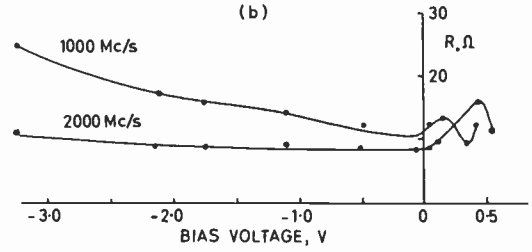
Fig. 6. Gallium arsenide diode.



World Radio History




(a) Dependence of calculated diode resistance on frequency for matching with: a, b lumped L, C components, c adjustable line and stub; d two stubs.




(b) Dependence of calculated diode resistance on bias voltage at 1000 and 2000 Mc/s.

Fig. 7. Silicon diode.

is shown by the triangular points in Figs. 6(a) and 7(a). These were obtained using an improved type of short circuit shown in Fig. 8.† In this design the diode was mounted directly on the back-plate, so that mechanical tuning was not possible. In consequence matching could only be effected at the highest frequencies. The short-circuit to the outer conductor is provided by a wire A which makes a biting contact under pressure from the bevelled surface of the moveable inner member B. For readjustment the wire is released, after removing C and L, by movement of the sleeve C to the left. The triangular points lie a little below those obtained using the spring-finger short-circuit.

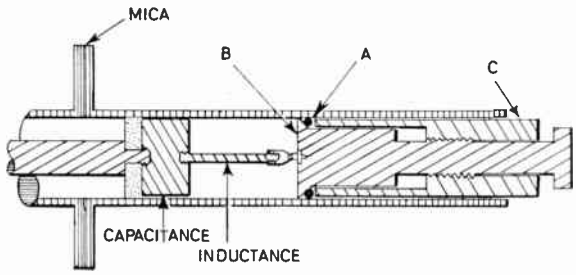



Fig. 8. Diode mount for relative impedance measurements at u.h.f.: wire short-circuit.

was used; i.e. short-circuit adjustment for minimum v.s.w.r., followed by adjustment of the matching elements in front of the diode. It is shown in Appendix 2 that the rise in resistance with decreasing frequency is due to the transformed resistance, at the diode, of the resistance of the 'short-circuit'. This is greater as the inductance is larger and hence the greater the fractional wavelength of the short-circuited line.

As at X-band it was found that larger values of $\Delta Q$ and hence smaller values of $R$ can be obtained if the short-circuit is located closer to the diode than the resonant position. Optimum curves obtained by doing this are shown as 'b' in Figs. 6(a) and 7(a); it will be noted that although the values of $R$ at the lower frequencies are reduced, the asymptotic high frequency value is not affected. If the length of short-circuited line behind the diode is reduced too much the values of $L$ and $C$ required for matching are increased. If the lengths of the corresponding sections become comparable with $\lambda/4$ the transformation is affected and matching may become impossible.

It could be assumed that the asymptotic value of $R$ is the real diode resistance. However, it is possible that distributed losses are increasing at the higher frequencies, thereby masking any continuing fall in transformed short-circuit loss. Evidence that the short-circuit loss at 2000 Mc/s is still slightly effective

The effect of the additional losses introduced when the C and L sections described above were replaced by an adjustable length of line and single-stub are shown by the broken curves (c) in Figs. 6(a) and 7(a). The apparent value of $R$ is increased over the whole range. The effect of double-stub tuning is shown by curve (d) in Fig. 6(a). Considerable extra loss is introduced at high frequencies.

In Figs. 6(b) and 7(b) the corresponding apparent dependence of $R$ on bias voltage is shown for each diode at the 'safe' frequency, for zero bias, of 2000 Mc/s and also at 1000 Mc/s. At reverse bias the smaller values of diode capacitance require a longer length of short-circuited line behind the diode to give resonance and it is to be expected therefore that the apparent value of $R$ will increase as the reverse bias is increased. The effect is only significant at 1000 Mc/s. Under forward bias the converse is true. For the gallium arsenide diode the value of $R$ remains substantially constant at 2000 Mc/s and the values obtained at 1000 Mc/s converge towards the 2000 Mc/s value. It is very likely therefore that the diode resistance is approximately 4 ohms. The anomalous behaviour of the silicon diode under forward bias is not yet understood.

† The idea for this type of short-circuit was suggested by Dr. A. Fairweather of the British Post Office Research Station.



World Radio History

### 4. Transmission Measurements

Measurements based on the transmission of power past a diode were first described by DeLoach.[8] A requirement is that the frequency of measurement can be chosen so that there is an effective series resonance produced between suitable mounting planes. For the diodes shown in Fig. 1 this resonance occurs in X-band for the planes shown as X and X′. Since the distance between these planes is much less than the full height of standard X-band waveguide it is necessary to transform down from the standard height on either side to the diode dimensions using reflectionless tapers. In the present work (cosine)$^2$ tapers as analysed by Bolinder[14] have been used. V.s.w.r.'s better than 1·05 have been obtained over the major part of X-band.

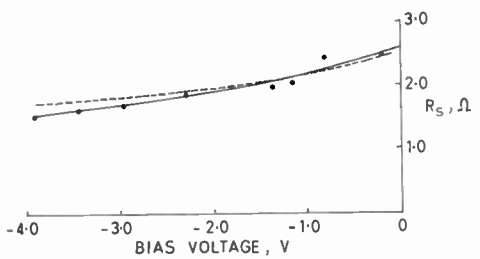



Fig. 9. Dependence of $R_s$ on bias voltage for a silicon diode. Experimental curve ———; Synthetic curve – – – – –, based on $R = 4{\cdot}1\ \Omega$, $C_B = 0{\cdot}15$ pF.

A typical experimental arrangement comprises a taper mount as above, with a choke on one face so that bias can be applied, a matched load (crystal detector) on the output side, and on the input side a precision attenuator preceded by a level-setting attenuator, a wavemeter and a small-signal power source. Facility also has to be provided for the measurement of the available power to the diode.

From an analysis made using the equivalent circuit of Fig. 2, it follows that when either the bias voltage or the frequency is adjusted so that there is minimum power delivered to the load,

$$R_s = \frac{R}{\omega^2 C_B^2 R^2 + \left(\frac{C_B}{C}+1\right)^2} \simeq \frac{R}{\left(\frac{C_B}{C}+1\right)^2} = \frac{Z_0}{2(\sqrt{T}-1)} \qquad \text{.....(8)}$$

Here $Z_0$ is the characteristic impedance (resistance) of the reduced height waveguide and $T$ is defined as the ratio of available power to power absorbed in the load. The condition is effectively one of series resonance of the elements $L$, $C_B$, $C$ and $R$ of Fig. 2(a) so that $R_s$ is the resonant series resistance of these combined elements as shown in the simplified equivalent circuit of Fig. 2(b): the shunting effect of the reactance of $C_A$ for practical purposes is negligible. The neglect of $\omega^2 C_B^2 R^2$ relative to $(C_B/C+1)^2$ in eqn. (8) is also justifiable. A point of some difficulty is the attribution of a value to $Z_0$. DeLoach[8] used the expression due to Schelkunoff for the impedance presented to a small diameter wire connected between the broad walls of a waveguide of height $b \ll \lambda_g/4$;

$$Z_0 = 754\frac{b}{a}\frac{1}{\sqrt{1-\left(\frac{\lambda_g}{2a}\right)^2}} \qquad \text{.....(9)}$$

$a$ is the width of the guide. For the diodes investigated by him this choice is not too inappropriate. For those shown in Fig. 1 it cannot be as good an approximation. Because of the complexity of the situation it has, however, been used here. In Fig. 9 a typical variation of $R_s$ with bias voltage is shown as the continuous curve associated with the experimental points. This was obtained from the expression on the right-hand side of eqn. (8). A synthesized broken curve is also shown. This is calculated from $R_s = R/\{(C_B/C)+1\}^2$, with values of $R$ and $C_B$ chosen to give a good fit.

DeLoach also showed that the transmission method can lead to a microwave determination of junction capacitance. The frequencies $\omega_1$ and $\omega_2$ on either side of the resonant frequency at which the power in the load is increased by 3 dB above the minimum are measured. Adapting DeLoach's result to suit the equivalent circuit used here, in which $C_B$ is incorporated, it can be shown that

$$C + C_B \simeq \frac{\omega_1 - \omega_2}{\omega_1 \omega_2} \cdot \frac{2}{Z_0} \cdot (\sqrt{T}-1)\left(1-\frac{2}{T}\right)^{\frac{1}{2}} \qquad \text{.....(10)}$$

In Fig. 10 the dependence of $C+C_B$ on bias voltage is shown by the continuous curve linking the experimental points; the diode is the same for which Fig. 9 was obtained. For comparison the corresponding broken curve for $C+C_A+C_B$ at h.f. is also shown.

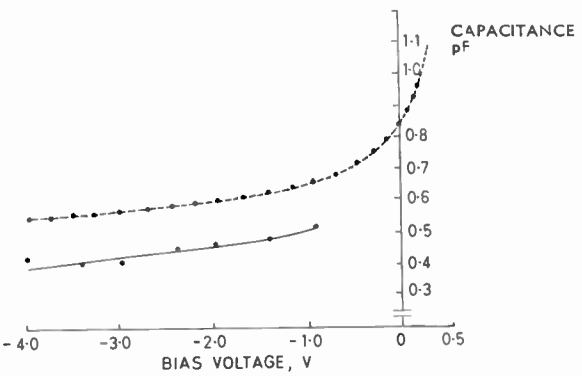



Fig. 10. Dependence of capacitance on bias voltage for a silicon diode. X-band measurements ———; H.f. measurements, – – – – –.



World Radio History

This is seen to be approximately 0·15 pF higher at all points: the difference may be attributed to $C_A$.

### 5. Comparison of Data

By comparing the results obtained by the different techniques which have been described it is possible to assess their relative usefulness and accuracy. In Table 1 some representative data are shown for silicon diodes A, B and C and a gallium arsenide diode D. All diodes have diffused junctions and the encapsulation of Fig. 1(b). In each case the total stray capacitance $\simeq 0{\cdot}3$ pF; by fitting $C_B$ and $R$ values to the data from X-band transmission measurements it is found that $C_A \simeq C_B \simeq 0{\cdot}15$ pF.

In view of the anomalous behaviour of the silicon diodes under forward bias the value of $R$ given is the minimum value, which was obtained near zero bias. There is only one clear deduction to be made from these results. This is that the values of resistance obtained by the transmission method are the lowest. The reason may be that no external matching elements are involved, but it could be that an incorrect value for waveguide impedance has been used.

It is interesting to note the close agreement between the values of $R$ obtained for the gallium arsenide diode by Harrison's method at u.h.f. and the X-band transmission method. As was pointed out in Section 3.2.2, the 2000 Mc/s measurement of $R$ for this diode at zero and forward bias was practically constant, as would be expected. The only significant difference between the silicon and gallium arsenide diodes was that the latter had a considerably smaller value of junction capacitance. Values of $C(0)$ obtained by h.f. measurements are shown in the fifth column. In the next column values of 'cut-off frequency' at zero bias are given: they are calculated using the values of $R$ obtained by the transmission method. The cut-off frequency is defined as the frequency at which the $Q$-factor of the semiconductor region is unity. Hence

$$f_{c0}(0) = \frac{1}{2\pi C(0)R} \qquad \text{.....(11)}$$

Corresponding values at $-6$ V bias, which are often quoted by manufacturers, are about twice the values at zero bias.

It should be noted that if the measured values of $R_s$ and $C$, as defined by eqns. (8) and (10) respectively, are used to define a cut-off frequency its value will be

$$f'_{c0}(0) = \frac{1}{2\pi C(0)R} \cdot \frac{C(0)+C_B}{C(0)} = f_{c0}(0)\frac{C(0)+C_B}{C(0)} \qquad \text{.....(12)}$$

i.e. a false, high value will be obtained. For parametric amplifier applications, however, the cut-off frequency is not a sufficient specification of the usefulness of a diode. A commonly-used[15] figure of merit is the product $\gamma Q_s$; here $Q_s$ is the $Q$-factor of the semiconductor region at the signal frequency, which is proportional to the cut-off frequency, and $2\gamma$ is the first coefficient of *junction* capacitance expressed as a Fourier series involving only cosine terms. If the product $\gamma Q_s$ is formed either on the basis of actual $C$ and $R$ values or on apparent $C$ and $R$ values obtained by transmission measurements, the same value is obtained.

The final column in Table 1 indicates the frequencies at which self-resonance occurred in reduced-height X-band waveguide and the corresponding bias voltages.

In Table 2 data are presented for two silicon diodes having the encapsulation of Fig. 1(a). A comparison is made between values of $R$ obtained in both full- and reduced-height X-band waveguide by Harrison's $\Delta Q$ method and by DeLoach's

**Table 1**

Collected data for diodes having the form of Fig. 1(b).

| Diode | Resistance, $R$; Ω | | | $C(0)$; pF | $f_{c0}$ at 0V; Gc/s (Transmission method) | $f_0$; Gc/s / Bias; V |
|---|---|---|---|---|---|---|
| | Harrison's $\Delta Q$ method | | | | | |
| | X-band in reduced-height waveguide | U.H.F. (2000 Mc/s) | X-band transmission method | | | |
| A | 5·7 | 7·1 | 5·2 | 0·63 | 48 | 8·9 / −1·0 |
| B | 7·4 | 7·8 | 5·3 | 0·56 | 54 | 9·15 / −2·0 |
| C | 5·3 | 6·1 | 4·1 | 0·58 | 67 | 9·0 / −1·0 |
| D | 6·4 | 4·0 | 3·7 | 0·38 | 113 | 9·1 / 0 |



World Radio History

**Table 2**

Collected data for diodes having the form of Fig. 1(a)

| Diode | Resistance $R$; Ω | | | $C(0)$; pF | $f_{c0}$ at 0V; Gc/s (Transmission method) | $f_0$; Gc/s / Bias; V |
|---|---|---|---|---|---|---|
| | Harrison's $\Delta Q$ method | | X-band transmission method | | | |
| | Full-height X-band waveguide | Reduced-height X-band waveguide | | | | |
| E | 8·0 | 6·1 | 6·8 | 0·40 | 58 | 8·9 / 0 |
| F | 7·0 | 5·0 | 4·4 | 0·51 | 70 | 8·9 / − 1·0 |

transmission method in the same reduced-height waveguide mount.

The values obtained using Harrison's method in full-height waveguide, in which it is necessary to support the diode by an inductive post, are greater than those obtained in reduced height waveguide. It is to be inferred that there are external losses associated with the modified transformation due to this post. For the order of cut-off frequency involved here it can be concluded that the accuracy of measurement in reduced-height X-band waveguide is of the order of ± 10–20%, considering both Tables 1 and 2.

## 6. Conclusion

The relative-impedance and transmission methods of measuring varactor diode parameters have been compared. It has been found that the transmission method due to DeLoach, as applied at X-band to diodes having cut-off frequencies up to about 100 Gc/s at zero-bias, gives the lowest values of diode resistance and, correspondingly, the highest cut-off frequencies. The accuracy of determination of junction capacitance and diode resistance at X-band is of the order of ± 10% at best. The simple equivalent circuit of Fig. 2(a) appears to fit in with experimental data to a first order and there is correlation between capacitance measurements made in the h.f. and X-bands.

## 7. Acknowledgments

The authors wish to acknowledge the benefit of discussions and the provision of diodes by Mr. J. D. Pearson and Dr. A. H. Benny of Ferranti Ltd., Wythenshawe, and Mr. D. D. Jones and Mr. K. Wilson, of Associated Semiconductor Manufacturers, Hirst Research Centre, Wembley. Mr. Wilson also kindly provided a copy of an unpublished report describing the application of the transmission method at X-band. The mechanical work was performed by Mr. K. Beardsell and Mr. R. Jones.

This research is supported by a Research Grant provided by the Science Research Council.

## 8. References

1. K. E. Mortenson, 'Alloyed, thin-base diode capacitors for parametric amplification', *J. Appl. Phys.*, **30**, p. 1542, 1959.
2. O. E. Sawyer, 'Surface-dependent losses in variable reactance diodes', *J. Appl. Phys.*, **30**, p. 1689, 1959.
3. S. T. Eng and R. Solomon, 'Frequency dependence of the equivalent series resistance for a germanium parametric amplifier diode', *Proc. Inst. Radio Engrs*, **48**, p. 358, 1960.
4. T. Hyltin, 'Varactor packaging and paramp performance', *Microwaves*, **1**, p. 12, 1962.
5. R. I. Harrison, 'Parametric diode $Q$ measurements', *Microwave J.*, **3**, p. 43, 1960.
6. N. Houlding, 'Measurement of varactor quality', *Microwave J.*, **3**, p. 40, 1960.
7. R. Mavaddat, 'Varactor diode Q-factor measurements', *J. Electronics Control*, **15**, p. 51, 1963.
8. B. C. DeLoach, 'A new microwave measurement technique to characterize diodes and an 800 Gc cut-off frequency varactor at zero volts bias', *Trans. I.E.E.E. on Microwave Theory and Techniques*, MTT-12, p. 15, 1964.
9. F. J. Hyde and R. B. Smith, 'Effect of losses in varactor-diode, impedance measurements', *Proc. Instn Elect. Engrs*, **111**, p. 471, 1964.
10. A. Uhlir, 'Apparent frequency dependence of series resistance of varactor diodes', *Proc. I.E.E.E.*, **51**, p. 1246, 1963.
11. R. B. Smith, 'Study of Varactor Parameters', Ph.D. Thesis, University College of North Wales, 1964.
12. J. M. Early, 'Design theory of junction transistors', *Bell Syst. Tech. J.*, **32**, p. 1271, 1953.
13. H. M. Barlow and A. L. Cullen, 'Microwave Measurements', p. 205. (Constable, London, 1950.)
14. F. Bolinder, 'Fourier transforms in the theory of inhomogeneous transmission lines', *Trans. Royal Inst. Technology Stockholm*, Nr. 48, 1951.
15. L. A. Blackwell, 'Semiconductor-diode Parametric Amplifiers', (Prentice-Hall, Englewood Cliffs, N.J., 1961).
16. C. G. Montgomery, R. H. Dicke and E. M. Purcell, 'Principles of Microwave Circuits', p. 196, Radiation Laboratory Series, Volume 8, (McGraw-Hill, New York, 1948).

## 9. Appendix 1: Choke Design

One terminal (end piece) of the varactor diode can make direct contact to the waveguide. The other must be insulated from it for d.c., so that bias can be applied. Ideally an r.f. short-circuit must be provided



World Radio History

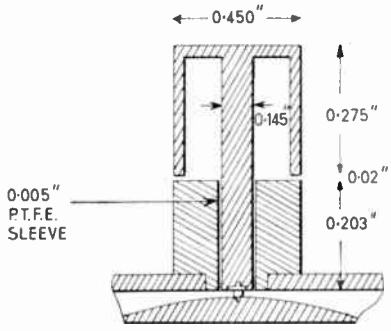



Fig. 11. Detail of diode mount for reduced-height X-band waveguide showing choke design.

at this terminal to coincide with the chosen plane of measurement. Although Harrison's method and the determination of resistance by the transmission method are not in principle affected if the choke possesses reactance, it is nevertheless desirable that this should approach zero. Determination of diode capacitance by the transmission method depends on this. Simple coaxial chokes made of low impedance line, one half-wavelength long and topped by a capacitive plate are effective, but have a relatively narrow bandwidth of about 5%.

A wide-band choke, which is similar to that described by Wilson (see Sect. 7), was used for most of the work. The design is shown in Fig. 11 for diodes having the form of Fig. 1(b). It is based on the choke joint described by Montgomery, Dicke and Purcell,[16] and comprises two sections of line, each approximately a quarter wavelength long; the top one has high characteristic impedance and the bottom one low characteristic impedance. The short-circuit at the top is transformed to an equivalent short circuit at the bottom. Transmission measurements made with an inductive post replacing the diode showed that the bandwidth was of the order of 30%. A similar choke was made for diodes having the form of Fig. 1(a).

## 10. Appendix 2: Apparent Frequency-Dependence of the Series Resistance of a Varactor Diode

We consider a diode mounted coaxially in the central conductor of a coaxial line, which is terminated by an adjustable 'short-circuit'. The length of this short-circuited line is $l$, its phase constant $\beta = 2\pi/\lambda_g$, and its characteristic impedance $Z_0$. It is assumed that the 'short-circuit' has a resistance $r$ and that the line itself is lossless. It follows from standard transmission line theory that the impedance $Z_l$ presented at the diode by this line is

$$Z_l = Z_0 \frac{\left(\frac{r}{Z_0}\right)(1+\tan^2\beta l)+j\left(1-\left(\frac{r}{Z_0}\right)^2\right)\tan\beta l}{1+\left(\frac{r}{Z_0}\right)^2\tan^2\beta l} \qquad \ldots\ldots(13)$$

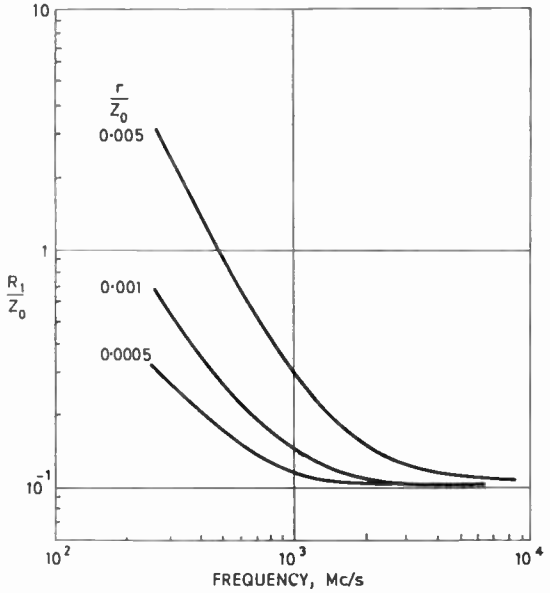



Fig. 12. Apparent diode resistance, due to 'short-circuit' loss.

For the measuring system described in Section 3.2.2, the reactive part of this, $X_l$, resonates with the effective diode reactance and therefore determines $\beta l$. It follows that

$$\tan\beta l = \frac{\left\{1-\left(\frac{r}{Z_0}\right)^2\right\} \pm \left\{\left[1-\left(\frac{r}{Z_0}\right)^2\right]^2 - 4\left(\frac{r}{Z_0}\right)^2\left(\frac{X_l}{Z_0}\right)^2\right\}^{\frac{1}{2}}}{2\left(\frac{X_l}{X_0}\right)\left(\frac{r}{Z_0}\right)^2} \qquad \ldots\ldots(14)$$

Practical values of $r/Z_0$ are such that $(r/Z_0)^2$ can be neglected in comparison with unity. Also $4(rX_l/Z_0{}^2)^2$ is small compared with unity so that $\{1-4(rX_l/Z_0{}^2)^2\}$ can be represented by the first two terms of the binomial expansion. Then, taking the alternative negative sign in the numerator it follows that

$$\tan\beta l \simeq \frac{X_l}{Z_0} \qquad \ldots\ldots(15)$$

Hence the transformed resistance at the diode is given by

$$R_l \simeq r\left\{1+\left(\frac{X}{Z_0}\right)^2\right\} \qquad \ldots\ldots(16$$

Theoretical curves showing the apparent variation with frequency of normalized diode resistance are presented in Fig. 12. It has been assumed that $R = 5\Omega$, $C = 0{\cdot}5$ pF, $Z_0 = 50\Omega$; the values of the parameter $r/Z_0$ used are 0·0005, 0·001 and 0·005.

*Manuscript first received by the Institution on 8th April 1965 and in final form on 10th May 1965. (Paper No. 1022).*

© The Institution of Electronic and Radio Engineers, 1966



World Radio History

U.D.C.621.382.2 : 621.396.67

# Tunnel Diodes Integrated with Microwave Antenna Systems

*By*

Professor
Dr. H. H. MEINKE†

*Reprinted from the Proceedings of the Joint I.E.R.E.–I.E.E. Symposium on 'Microwave Applications of Semiconductors' held in London from 30th June to 2nd July 1965.*

**Summary:** The applications of tunnel diodes to antenna systems make use of devices as (1) negative conductances, (2) as pumped elements, and (3) as rectifiers.

## 1. Introduction

Many problems of antenna technique cannot be solved by the help of purely passive antennae. These include, for example, the broadband radiator of small height, the highly directive antenna of small size, and dipole arrays without mutual coupling between adjacent dipoles. If we can tolerate relatively poor noise performance it is probable that many of these problems can be solved by combining passive antennae with active elements as tunnel diodes, varactor diodes or transistors. This paper is confined to a consideration of the problems of combining tunnel diodes with antennae.

The low power capability of present tunnel diodes limits the practical use of tunnel diode antennae to receiving antennae. But it cannot be excluded that in the future a high-power element will be discovered having similar characteristic curves, which would make possible the use of such active antennae for transmission. The behaviour of a transmitting antenna can already be measured if the amplitudes are restricted to low values.

A tunnel diode can be used in combination with antennae as a negative conductance, as a frequency converter, as a pumped element with auxiliary frequency input, or as a rectifier. The number of possible combinations of tunnel diodes with passive antenna elements is, of course, infinite and it is not feasible to give a complete survey of all possible useful forms of active antennae. Only a few examples will be explained here to illustrate new effects. Most of the research on active antennae in the past was carried out by the Research and Technology Division of the U.S. Air Force and its research contractors. Our own research was sponsored by this Division.

## 2. Tunnel Diodes as Negative Conductances

Figure 1 shows the application of a tunnel diode as negative conductance to control the input admittance of the antenna. The antenna consists of a main radiator feeding a coaxial output. It is well known that a second radiator parallel to the main radiator and grounded at one end influences the admittance of the main radiator at the feeding point as shown in Fig. 2, curve I (without tunnel diode). The admittance curve in the complex plane forms a loop and thus produces a broadband characteristic over a limited frequency region. If a tunnel diode is mounted between

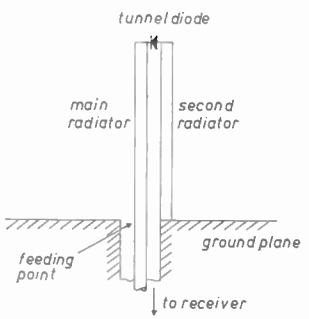



**Fig. 1.** Folded radiator with tunnel diode.

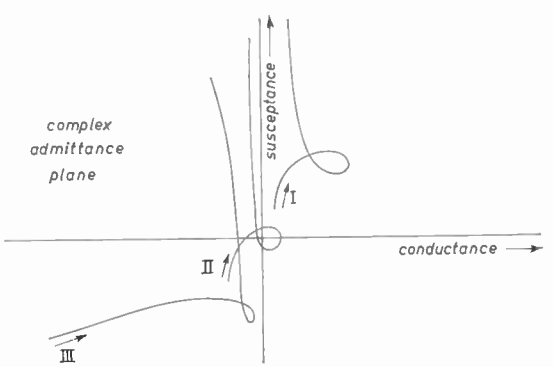



**Fig. 2.** Input admittance of the radiator of Fig. 1. The arrows indicate increasing frequency.
Curve I: Diode conductance $G_D = 0$.
Curve II: $G_D = -0{\cdot}006$ mho.
Curve III: $G_D = -0{\cdot}01$ mho.

† Institute for High Frequency Techniques, Technical University, Munich, Germany.



World Radio History

the ends of the two radiators the impedance at the feeding point is changed depending on the value of the negative conductance; in addition the tunnel diode may be shunted by a reactance which, at microwave frequencies, is in all cases at least the diode capacitance. By proper choice of the conductance and the shunt reactance of the diode, many different admittance curves in the complex plane may be achieved. For example, one may produce a loop around the zero point or a curve along the imaginary axis (Fig. 2, curve II) or curves in the negative part of the complex plane (curve III). Thus a tunnel diode can produce frequency-dependent antenna-admittance curves of a form which cannot be produced by passive antennae. This offers new possibilities especially by combining passive broadband two-port networks at the feeding point. In addition to these advantages amplification of the received signal takes place if the antenna is connected through a suitable input admittance to the receiver.

Figure 3 shows a radiator which is divided into two parts and these parts are connected by a tunnel diode. The tunnel diode may be shunted by a capacitance or by an inductance, thus producing a complex admittance with negative conductance of general form. The most recent branch of antenna research studies the behaviour of radiators interrupted by concentrated admittances and the use of tunnel diodes extends this research to negative real parts of these admittances. In this case not only the amplification and the admittance behaviour at the feeding point causes interest, but also the changed current distribution on the radiator, which may influence the radiation diagram, may cause directivity and lower the mutual coupling between adjacent radiators in antenna arrays. We studied the radiator of Fig. 3 in which the upper part has a length $b$. Inside this upper part a short-circuited line is mounted which acts as a choke to bring the bias to the diode by help of a separate wire through the interior of the lower part of the radiator. This short-circuited line also acts as a reactance in parallel with the diode. By moving the short circuit and by proper choice of the length $b$ and of the conductance of the diode, the load $Z$ at the end of the lower part of the radiator can be adjusted to any complex value. We especially studied the case $Z = -Z_L$, in which $Z$ is negatively matched to the characteristic impedance $Z_L$ of the lower radiator part and thus in the transmitting case produces a wave propagating along the radiator towards the feeding point. This is a quite new and very simple current distribution unknown on purely passive radiators. The radiator input impedance is clearly $-Z_L$ and the mutual coupling between equal parallel radiators is very low.

In all combinations with tunnel diodes the danger of self-excitation presents a hard problem. There is low-frequency self-excitation because a tunnel diode needs extra wires, blocking capacitors and chokes to supply the bias. This d.c. circuit introduces low self-excitation frequencies which must be attenuated carefully to avoid the effect. A further self-excitation at microwave frequencies is generated by the capacitance and the inductance of the diode itself. A stabilizing two-terminal network using micro-circuit techniques with very low inductance has been developed and the tunnel-diode is shunted by this network to attenuate microwave self-excitation. Of course, if this self-exciting frequency is very near to the intended operating frequency it may be impossible to design such stabilizing network without a heavy sacrifice in efficiency of the antenna at the operating frequency. The antenna system can also have resonances and self-excitation at many other frequencies; the input impedance of the receiver is a part of the load circuit of the diode and participates in the occurence of such resonances. Every combination of antenna and receiver must therefore be studied carefully from the point of view of self-excitation. In our experimental systems the tunnel diode is therefore always shunted by an ordinary rectifying diode with separate d.c. circuit to indicate any alternating voltage

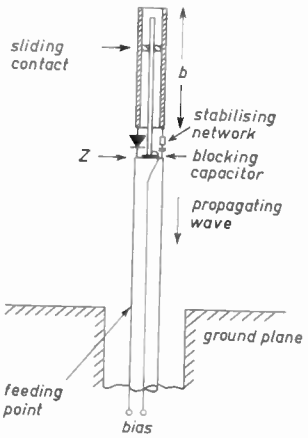



**Fig. 3.** Radiator in two parts joined by a tunnel-diode arrangement.

existing across the tunnel diode in the case of self-excitation. It is our opinion that a tunnel diode as a negative conductance in broadband antennae will probably not come into extended use on account of this danger of mostly unpredictable self-excitation.

### 3. Auxiliary Frequency Pumped Applications

A very interesting modification of a tunnel diode admittance in antennae seems to be a tunnel diode pumped by an auxiliary frequency as in Fig. 4. The antenna system then may be represented by a tunnel



World Radio History

diode and a network parallel to the terminals 1 and 2 of the diode. The network forms a frequency-dependent complex admittance $Y(f)$ parallel to the diode. In the receiving case the receiver is coupled to the network, the input impedance of the receiver being

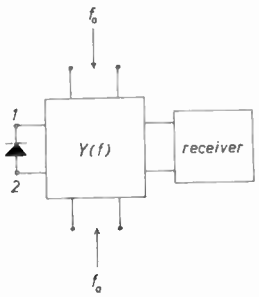



Fig. 4. Receiving network with tunnel diode, pumped by an auxiliary frequency $f_a$.

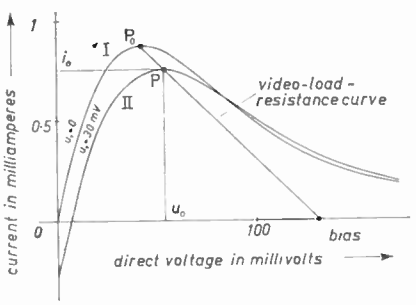



**Fig. 5.**

Curve I: Current-voltage characteristic of a tunnel diode.

Curve II: Rectifier current $i_0$ depending on the d.c.-voltage $u_0$ between the terminals of the diode for constant amplitude $u_1$ of the pump voltage.

a part of the network. The network is fed by the signal (operating frequency $f_0$), which is generated by the received electromagnetic wave and by an auxiliary generator (frequency $f_a$), the internal impedances of these two sources being included in the network.

In Fig. 5 curve I shows the current-voltage characteristic of the diode and curve II the rectified current $i_0$ dependent on the direct voltage $u_0$ for a given voltage amplitude $u_1$ of the auxiliary frequency. The derivative of curve II

$$s_0 = \frac{\partial i_0}{\partial u_0} \qquad \ldots\ldots(1)$$

at the operating point $P$ is the primary admittance of the tunnel diode in the circuit of Fig. 4. By changing $u_0$ we can change $s_0$ from positive to negative values including the value zero. When pumping the diode by $f_a$, the non-linearity of the diode generates additional periodic currents with combination frequencies $f_a \pm f_0$, $2f_a \pm f_0$, and so on. These currents flowing through the admittance $Y$ of the network generate voltages at these frequencies and these voltages are found between the terminals of the diode. By backward mixing in the diode these combination frequencies together with the pump frequency excite additional currents in the diode at the operating frequency $f_0$. The diode thereby forms parallel to its primary admittance $s_0$ additional admittances at the operating frequency which may be written in the form of infinite series

$$-\frac{s_1^2}{s_0+Y(f_a+f_0)} - \frac{s_1^2}{s_0+Y^*(f_a-f_0)} + \ldots \qquad \ldots\ldots(2)$$

where

$$s_1 = 2\frac{\partial i_0}{\partial u_1}$$

is the converter transconductance, $Y(f_a+f_0)$ is the complex admittance of the network at the combination frequency $f_a+f_0$, and $Y^*(f_a-f_0)$ is the conjugate of the complex admittance of the network at the combination frequency $f_a-f_0$.

This admittance of a pumped diode in combination with a frequency-dependent network offers an abundance of new forms of complex admittance curves beyond the possibilities of passive networks. As an example, Fig. 6 shows this admittance made up of the sum of the admittances represented by eqns. (1) and (2) for a very simple network, consisting of two resonant circuits with equal $C$ and different $L$. One resonance frequency is set below $f_a$, the other resonance frequency is equally above $f_a$. The diode admittance resembles that of two coupled circuits, but inverted

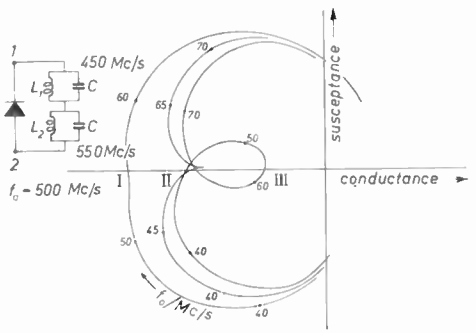



**Fig. 6.** Complex admittance of a tunnel diode pumped by a frequency $f_a$ = 500 Mc/s, shunted by two resonant circuits with resonances at 450 Mc/s and 550 Mc/s. The values at the curves show the operating frequencies $f_0$ along the admittance curves. The three curves correspond to different values of $s_0$, decreasing from curve I to curve III. The form of curve II is very similar to the input admittance of a filter with two critically coupled resonance circuits, but inverted into the negative half of the complex plane.



World Radio History

into the negative half of the complex plane. The admittance curve depends considerably on the value of $s_0$, thus imitating circuit coupling, which is critical, below critical or above critical. Such an admittance in combination with passive coupled networks can give compensation over a very broad band. In antennae a symmetrical push-pull circuit may be used at the auxiliary frequency to avoid radiation at the frequency $f_a$. In Fig. 4 the receiver may also be tuned to one of the combination frequencies, for example $f_a - f_0$, so that the antenna system acts as a frequency converter. The internal admittance of this converter at the frequency $f_a - f_0$ has similar forms as in eqns. (1) and (2) and produces similar curves to those of Fig. 6.

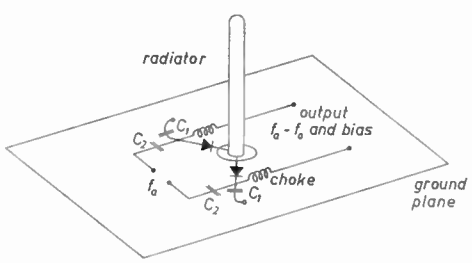



Fig. 7. Radiator with two-tunnel diodes pumped by $f_a$ in a push-pull circuit. One end of $C_1$ is grounded.

Figure 7 shows a schematic sketch of a simple radiator with two tunnel diodes pumped by $f_a$ in a push-pull circuit and the output at the frequency $f_a - f_0$. The very loose coupling of $f_a$ to the diodes is achieved by the capacitors $C_1$ and $C_2$ as a capacitive voltage divider, where $C_1$ has a high and $C_2$ has a low value.

### 4. Tunnel Diodes as Rectifiers

Useful results may be obtained with tunnel diodes as rectifiers in low-level video detection directly integrated with a waveguide-antenna system. A tunnel diode with a characteristic as in Fig. 5, curve I, is driven by a bias of 0·13 V with a video-load resistance of 100 ohms.

This gives an operating point $P_0$ near the current maximum where the curvature of the characteristic is very pronounced and the rectifying process correspondingly intensive. According to Fig. 8 this gives a nearly linear detection for h.f.-voltages in the range between 30 and 300 mV, while ordinary diodes show a linear detection above 1 V only. Furthermore, the intrinsic admittance is very low on account of the small derivative of the characteristic in the operating point $P_0$. The intrinsic admittance is the real admittance generated by the voltage current characteristic across the barrier layer. By shifting the bias to more positive values, the intrinsic admittance will become

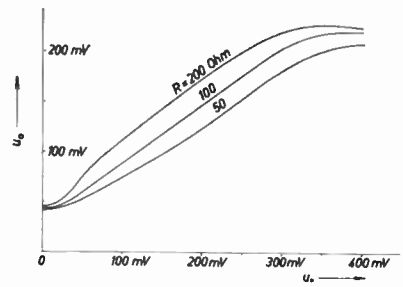



Fig. 8. Rectifier voltage $u_0$ depending on h.f.-voltage amplitude $u_1$.

slightly negative and thus can even compensate losses in the series resistance of the diode and in the input circuit. Thus we obtain good linearity for extremely low input power. The sensitivity of this rectifier is about 40 dB better than for ordinary diodes and not far from the sensitivity of a superheterodyne receiver using ordinary diodes. This opens new possibilities for simple low-level video detection, especially for reception of pulsed signals in very broad frequency bands.

An important feature of tunnel-diode rectification is its stability against excessive h.f.-overloads. As the curvature of the characteristic at the operating point is negative, the direct current decreases and the diode becomes cooler with increasing r.f.-voltage, while an ordinary diode becomes increasingly hot in this case. While a tunnel diode as video detector is usually operated at a standard level of about $10^{-9}$ W h.f. power, it can withstand 30 mW before burn-out. This allows a power overload of more than 70 dB. But this burn-out will never happen because the intrinsic admittance of the diode will increase considerably at high h.f. amplitudes, and this causes a severe mismatch between antenna and receiver input at high power levels. The high power, therefore, can never enter the diode if there is a suitable matching network between antenna and diode.

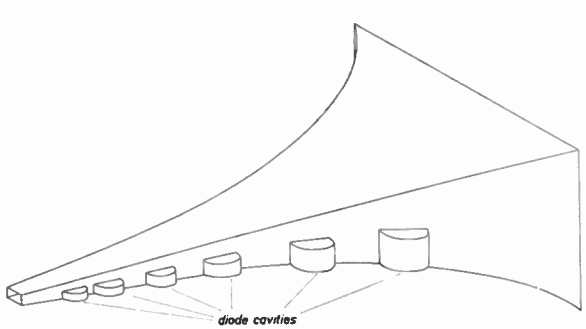



Fig. 9. Broadband-exponential horn-receiver with six tunnel diodes.



World Radio History

We constructed a broadband receiver as in Fig. 9 for a frequency range from 1·3 to 5·4 Gc/s consisting of a horn antenna and an exponentially tapered waveguide. There are six tunnel diode rectifiers for six different overlapping frequency ranges within the band mentioned above. The selectivity of the six diode circuits is first given by the fact that the waveguide has a cut-off frequency depending on the width of the guide at that point, where the diode is coupled to the waveguide. Secondly, every diode is located in a cavity as in Fig. 10. The cavity is open at one side and this opening is a part of the waveguide wall. There is strong coupling between cavity and waveguide and therefore the resonance of such a cavity is very broad. This gives a very simple broadband receiver with high sensitivity, especially for pulsed signals. A burn-out of the diodes by receiving higher power from nearby transmitter stations, which is very dangerous for ordinary receivers, is nearly impossible in tunnel-diode receivers because of the automatic mismatch of the input at high amplitudes.

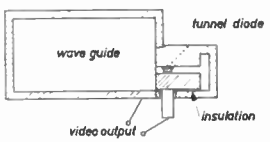



**Fig. 10.** Tunnel-diode cavity coupled with waveguide.

*Manuscript first received by the Institution on 26th April 1965. (Paper No. 1023.)*

© The Institution of Electronic and Radio Engineers, 1966

# I.F.A.C. Congress 1966

As announced in the January *Journal*, the Third Congress of the International Federation of Automatic Control is to be held in London from 20th to 25th June, 1966. The Duke of Edinburgh has graciously consented to be Patron of the Congress, which is being supported by the Ministry of Technology. The Institution has been associated with the planning of the Conference through its representatives on the U.K. Automation Council (Mr. A. St. Johnston and Mr. W. Renwick).

It is expected that more than 2000 specialists on automatic control from thirty-five countries will be attending the Congress. Two hundred and eighty-eight papers will be presented on subjects which include systems engineering, real-time digital computers, automatic control and medicine, adaptive learning and pattern recognition, computer-aided design and micro-electronics.

Survey papers, contributed by distinguished scientists, will be presented at the plenary sessions of the Congress. Plenary sessions will be followed by several technical sessions, held in parallel, in which papers covering specialized aspects of the subject will be presented and discussed. Tutorial sessions on optimal control, computer control, statistical identification procedures and computer languages, will precede the appropriate technical sessions. Informal colloquia, which will also be held in conjunction with the technical sessions, will cover terminology, education, standards and evaluation, artificial languages, reliability and other subjects.

Several members of the Institution are reading papers at the Congress, among them Dr. H. J. Meyerhoff (Huntec Ltd., Toronto), Professor H. H. Rosenbrock (Manchester College of Science and Technology) and Mr. D. Shaw (United Steel).

Simultaneous interpretation facilities into English, French, German and Russian will be available at the plenary sessions. Plenary sessions of the Congress will be held at Central Hall, Westminster. Technical sessions, tutorial sessions and the informal colloquia will be held at Church House, Westminster, and in other buildings in the neighbourhood.

An extensive social and ladies' programme is being organized, of which the most notable event will be a Conversazione at the Royal Festival Hall on the evening of 20th June. A programme of tours will take place in the week following the Congress, combining visits to factories to see installations of automation and advanced automatic control systems with visits to places of historic and cultural interest.

Further information, including copies of the advance programme, may be obtained from the Institution.



World Radio History

U.D.C.621.396.965

# A Phase-switched Radar System Giving Improved Control of Directional Pattern

*By*

ROBERT H. MacPHIE, Ph.D.†

**Summary:** This paper is a mathematical analysis of a new type of phase-switched radar which uses four antennae, two for transmitting and two for receiving. By phase-switching the pulses of one of the transmitting antennae and by multiplicative detection of the outputs from the two receiving antennae one can obtain a two-way power pattern proportional to the product of the four directivity (field strength) patterns of the system's four antennae. Furthermore, by appropriate phase and amplitude switching of a single array's element currents it is possible to simulate four coincident arrays whose product pattern can be obtained. The greater flexibility of this system (four pattern functions instead of one) is demonstrated by examples of improved pattern synthesis. Comparison is made with a conventional array of the same size whose pattern beam-width is the same as that of the phase-switched system; the phase-switched pattern has lower side-lobes and four times as many nulls. The effect of uncorrelated element noise and of a second target (either coherent or incoherent) is also considered. It is shown that for any degree of coherence one can synthesize a better pattern with the phase-switched system.

**List of Symbols**

| | |
|---|---|
| $t$ | time |
| $\omega_0$ | angular frequency of the r.f. carrier of the radar |
| $c$ | velocity of light |
| $\beta$ | phase constant or wave number |
| $\theta$ | bearing angle |
| $u = \beta \sin \theta$ | bearing coordinate |
| $r$ | range coordinate |
| $r_{max}$ | maximum range of the radar |
| $\tau$ | pulse repetition period, $\left(\tau = \frac{2r_{max}}{c}\right)$ |
| $l$ | element spacing of array |
| $v$ | Dolph-Chebyshev pattern variable |
| $\mathrm{Re}\{\ldots\}$ | indicates 'The real part of . . .' |
| $\mathrm{E}\{\ldots\}$ | indicates 'The mathematical expectation of . . .' |
| $A^*$ | indicates the complex conjugate of $A$ |
| $\left(\frac{S}{N}\right)$ | signal/noise ratio |

## 1. Introduction

It has been known for some time that the performance of receiving antenna systems can be improved by multiplicative, cross-correlation, or phase-switching techniques.[1, 2, 3] Instead of using a single antenna with square-law detector, the multiplicative system uses two antennae whose terminal voltages are effectively multiplied together. Associated with this voltage product is a system power pattern which is proportional to the product of the directivity patterns of the two antennae used.‡ This paper describes a new type of radar system which uses phase-switching techniques that are natural extensions of those mentioned above for passive receiving systems. Since the radar is transmitting as well as receiving, it is possible to use two antennae for transmitting and two for receiving, four antennae in all.

Section 2 contains a description of the basic system which uses four physically separate antennae. The transmitting antennae are fed from the same transmitter but the pulses sent to one of the antennae are phase-switched in synchronism with the pulse repetition period. After striking the target these pulses return and are received by the remaining two antennae whose terminal voltages are multiplied together. The video output will have a fluctuating component due to the phase-switching in the transmitting section. By making appropriate use of delay lines and a polarity switching device, this fluctuating component can be detected and the effective two-way power pattern associated with the output can be written as

$$\mathrm{Re}\{A_1(u)A_2^*(u)\}\ \mathrm{Re}\{B_1(u)B_2^*(u)\}$$

where $A_1(u)$, $A_2(u)$, $B_1(u)$, and $B_2(u)$, are the directivity patterns of the system's four antennae. In contrast to

† Electrical Engineering Department, University of Waterloo, Ontario, Canada.

‡ This paper is a revised version of 'A New Type of Cross-Correlation Radar System' which was presented at the Symposium on Signal Processing in Radar and Sonar Directional Systems, held at the University of Birmingham, in July 1964, and published in the *Proceedings* of that Symposium.



World Radio History

this, the two-way power pattern of a conventional radar, employing a single antenna with directivity pattern $C(u)$, is given by

$$|C(u)|^4$$

Comparing these two power patterns, one sees that the phase-switched system has, in a sense, four times as many 'degrees of freedom' as the conventional system since in the former there are four independent directivity patterns instead of the single pattern of the conventional system.

In a recent paper, Cooper[5] describes a system which also uses four antennae. The Cooper system uses two orthogonally polarized signals in lieu of two phase-switched signals to obtain an effective two-way pattern which is proportional to product of the patterns of the system's four elements.

In Section 3 it is shown that by phase and amplitude switching of the currents of a *single array's elements* one can effectively simulate *four coincident arrays* and obtain their product pattern just as with physically separate antennae. Two examples of pattern synthesis for coincident phase-switched linear arrays are presented in Section 4. In both cases the number of nulls of the phase-switched pattern is four times that of the pattern of a conventional array with the same number of elements. When, for purpose of comparison, the phase-switched pattern's beam-width is set equal to that of the conventional pattern, then the new system's side-lobe level is significantly reduced.

In Section 5 an analysis of the effect of antenna element noise (uncorrelated between elements) is presented and a comparison between the conventional and phase-switched systems is made. When one assumes that both systems are fed with the same average power the analysis shows that there is a slight improvement with the phase-switched system for low noise levels and a slight degradation for high noise levels.

The effect of a second interfering target of arbitrary amplitude, phase, and direction of arrival with respect to the primary target, is also considered. It is shown that in the limit of complete incoherence the rejection of this interference is governed by the two-way power patterns of the radars and hence the phase-switched system is superior since its pattern has lower and more numerous side-lobes. However, if the targets are coherent, the rejection of interference is shown to be due primarily to the *cross-product* pattern which in the case of the conventional array is $C(u)^2$ and for the phase-switched array is

$$[A_1(u)+A_2(u)][B_1(u)+B_2(u)] = A_+(u)\,B_+(u)$$

The latter system, with two pattern functions to vary instead of one, is again superior and an example of an improved cross-product pattern for coherent targets is given in Section 6. Furthermore, it turns out that not only is the cross-product pattern better (lower side-lobes with same beam-width) but the associated two-way power pattern is also better with lower side-lobes and a narrower main-lobe. Consequently, by proper design a phase-switched system can be made superior to the conventional system for all degrees of target coherence.

## 2. Operation of the Phase-Switched Radar System

A schematic diagram of the system is shown in Fig. 1 where it will be noted that four antennae are used, two for transmitting—$A_1$ and $A_2$—and two for receiving—$B_1$ and $B_2$. The pulsed transmitter power is divided into two equal parts which are fed to the transmitting antennae. In the transmission line feeding antenna $A_2$ is inserted a phase-switch acting in synchronism with a pulser from the transmitter and causing the pulses fed to antenna $A_2$ to be alternately in-phase and out-of-phase with the pulses fed to antenna $A_1$. The complex far zone fields of the transmitting antennae, except for a constant factor, are given by

$$A_1(t,u) = \sum_{k=-\infty}^{\infty} P(t-k\tau)A_1(u)\exp\{j(\omega_0 t+\gamma_k)\} \quad \ldots\ldots(1)$$

$$A_2(t,u) = \sum_{k=-\infty}^{\infty} P(t-k\tau)A_2(u)(-1)^k \exp\{j(\omega_0 t+\gamma_k)\} \quad \ldots\ldots(2)$$

where

$P(t)$ is the radar pulse,

$A_i(u)$ is the directivity pattern of antenna $i$, $i=1, 2$.

$\gamma_k$ is the phase of the r.f. carrier for the $k$th pulse.

Note the switching factor $(-1)^k$, which distinguishes the field of antenna $A_2$ from that of antenna $A_1$.

In the presence of a single point target at range $r$, this double series of pulses, after a delay of

$$t_r = \frac{2r}{c}$$

will return and be received by antennas $B_1$ and $B_2$ whose output voltages can be written as

$$B_i(t) = [A_1(t-t_r,u)+A_2(t-t_r,u)]B_i(u)$$
$$= \sum_{k=-\infty}^{\infty} P(t-t_r-k\tau)[A_1+(-1)^k A_2]\times$$
$$\times\, B_i \exp\{j[\omega_0(t-t_r)+\gamma_k]\} \quad \ldots\ldots(3)$$

where to simplify the notation, we have suppressed the $u$ variable. If these two voltages are fed into a multiplier, followed by a low-pass filter, one obtains

$$m(t) = \mathrm{Re}\{B_1(t)B_2^*(t)\} \quad \ldots\ldots(4)$$

This output is then passed through a delay line network which forms



World Radio History

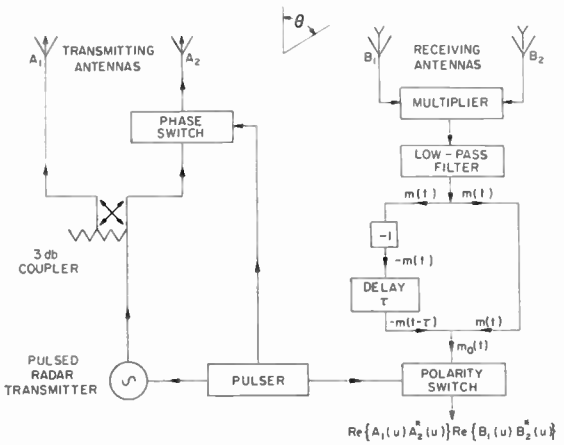



Fig. 1. Phase-switched radar system.

$$m_0(t) = m(t) - m(t-\tau) \qquad \text{.....(5)}$$

$$= \mathrm{Re}\{B_1(t)B_2^*(t) - B_1(t-\tau)B_2^*(t-\tau)\} \qquad \text{.....(6)}$$

Substituting (3) into (6) we obtain

$$m_0(t) = \mathrm{Re}\left\{\sum_{k=-\infty}^{\infty} |P[t-t_r-k\tau]|^2 |A_1+(-1)^k A_2|^2 \times \right.$$
$$\times B_1 B_2^* - |P[t-t_r-(k+1)\tau]|^2 \times$$
$$\left. \times |A_1+(-1)^{k+1}A_2|^2 B_1 B_2^* \right\} \qquad \text{.....(7)}$$

$$= \mathrm{Re}\left\{\sum_{k=-\infty}^{\infty} |P(t-t_r-k\tau)|^2 B_1 B_2^* \times \right.$$
$$\times [|A_1|^2 + |A_2|^2 + (-1)^k 2\,\mathrm{Re}\{A_1 A_2^*\} -$$
$$\left. -|A_1|^2 - |A_2|^2 + (-1)^k 2\,\mathrm{Re}\{A_1 A_2^*\}] \right\}$$
$$\text{.....(8)}$$

$$= 4 \sum_{k=-\infty}^{\infty} |P(t-t_r-k\tau)|^2 (-1)^k \times$$
$$\times \mathrm{Re}\{A_1 A_2^*\}\,\mathrm{Re}\{B_1 B_2^*\} \qquad \text{.....(9)}$$

Again, note the phase-switching factor $(-1)^k$, which causes the output $m_0(t)$ to alternate in polarity from pulse to pulse. However, $m_0(t)$ can be rectified by feeding it into a polarity switching device, acting in synchronism with the pulser of the transmitter; one obtains

$$\bar{m}_0(t) = (-1)^k m_0(t) \qquad \text{.....(10)}$$

on the $k$th pulse (see Fig. 1). Clearly this will yield our required output which can be written, except for a constant factor, as

$$\bar{m}_0(t,u) = \sum_{k=-\infty}^{\infty} |P(t-t_r-k\tau)|^2\,\mathrm{Re}\{A_1(u)A_2^*(u)\} \times$$
$$\times \mathrm{Re}\{B_1(u)B_2^*(u)\} \qquad \text{.....(11)}$$

$$= \sum_{k=-\infty}^{\infty} |P(t-t_r-k\tau)|^2 R(u) \qquad \text{.....(12)}$$

In (11) above it has been tacitly assumed that the directivity patterns are referred to some fixed reference point located on the common axis of all four linear antennae. If the distance from this point to antennae $A_1$, $A_2$, $B_1$ and $B_2$ are respectively $l_1$, $l_2$, $l_1$ and $l_2$, then the patterns become

$$A_i(u) = A_{i0}(u)\exp\{jl_i u\}$$
$$B_i(u) = B_{i0}(u)\exp\{jl_i u\} \qquad i = 1,2$$

where $A_{i0}(u)$ or $B_{i0}(u)$ would be the pattern (usually real) if the antenna in question were located at the reference point.

Consequently we may rewrite (11), in the case of real patterns $A_{i0}(u)$ and $B_{i0}(u)$, as follows:

$$\bar{m}_0(t,u) = \sum_{k=-\infty}^{\infty} |P(t-\tau_r-k\tau)|^2 A_{10}(u)A_{20}(u) \times$$
$$\times \cos[(l_1-l_2)u] \times B_{10}(u)B_{20}(u) \times$$
$$\times \cos[(l_1-l_2)u] \qquad \text{.....(11a)}$$

Note the two interferometer factors, $\cos[(l_1-l_2)u]$ and $\cos[(l_1-l_2)u]$, which depend on the mutual separation of antennae $A_1$ and $A_2$ and of $B_1$ and $B_2$; such factors might be usefully employed in overall pattern synthesis.

Now with a conventional radar system, where quite often the signals are transmitted and received on the same antenna (isolated by means of a duplexer), the analogous output is

$$\bar{m}_c(t,u) = \sum_{k=-\infty}^{\infty} |P(t-t_r-k\tau)|^2 |C(u)|^4 \qquad (13)$$

where $C(u)$ is the antenna directivity pattern and we have assumed that square-law detection is used. Comparing eqns. (11) and (13) we see that, mathematically, the latter is just a degenerate case of the former, which occurs when

$$A_1(u) = A_2(u) = B_1(u) = B_2(u) = C(u) \qquad \text{.....(14)}$$

Clearly, a phase-switched system is preferable since its four distinct patterns can be specified independently; in a sense, this system has four times as many 'degrees of freedom' as the conventional one. In Section 4 we will consider the implications of this in the design of linear array radar patterns.

### 3. Phase-switching with a Single Aperture

In some applications it is convenient to use a single aperture in lieu of four separate apertures. In effect we would have four coincident antennae each with a distinct pattern. A diagram of the single-aperture system for the case of a linear array is shown in Fig. 2. The aperture weighting network will vary the amplitude and phase of the currents fed to and from the array elements in synchronism with the pulse repetition period. If the four array patterns are $A_1(u)$, $A_2(u)$, $B_1(u)$, $B_2(u)$, then on transmit the aperture weighting network produces the patterns



World Radio History

$$A_+(u) = A_1(u)+A_2(u) = \sum_{k=-M}^{M} (a_{1k}+a_{2k}) \exp\{jklu\} \quad \text{.....(15)}$$

or

$$A_-(u) = A_1(u)-A_2(u) = \sum_{k=-M}^{M} (a_{1k}-a_{2k}) \exp\{jklu\} \quad \text{.....(16)}$$

and on receive it gives either

$$B_+(u) = B_1(u)+B_2(u) = \sum_{k=-M}^{M} (b_{1k}+b_{2k}) \exp\{jklu\} \quad \text{.....(17)}$$

or

$$B_-(u) = B_1(u)-B_2(u) = \sum_{k=-M}^{M} (b_{1k}-b_{2k}) \exp\{jklu\} \quad \text{.....(18)}$$

Note that, in effect, the currents of antennae $A_2$ and $B_2$ are either in phase or in phase-opposition with the corresponding current of antennae $A_1$ and $B_1$.

There are four possible transmit-receive product pattern combinations, namely,

$$(A_+, B_+), (A_+, B_-), (A_-, B_+), (A_-, B_-)$$

In operation, the radar, during a sequence of four pulses, will use each of the above combinations before repeating the first combination on the fifth pulse. Thus the video output of the square-law detector, in the presence of a single point target, will be a series of pulses whose amplitudes vary with period $4\tau$. This fluctuating output is fed into the delay-line network containing polarity switches which is shown in Fig. 2. In the Appendix it is demonstrated that the output of this network is proportional to

$$R(u) = \mathrm{Re}\{A_1(u)A_2^*(u)\}\ \mathrm{Re}\{B_1(u)B_2^*(u)\} \quad \text{.....(19)}$$

which, of course, is the four-antenna product pattern we have been seeking. The obvious virtue of this method of pattern multiplication is that the conventional square-law detector is used. The method's drawback is the necessity of phase and amplitude switching each of the array elements both on transmit and on receive. In the following section we will demonstrate the advantages of the phase-switched single-aperture in the synthesis of linear array radar patterns. Comparison will be made with a single-aperture conventional radar using square-law detection.

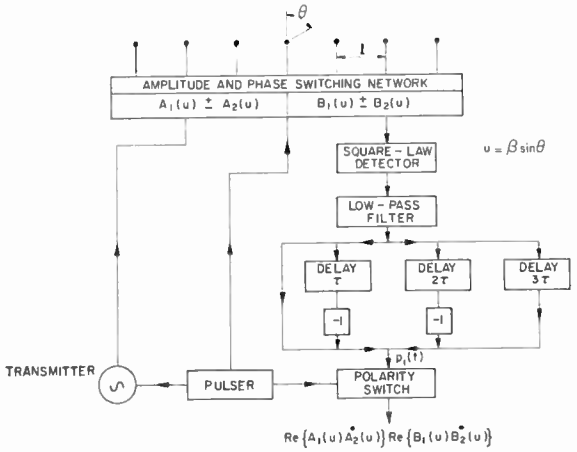



**Fig. 2.** Phase-switched radar system using a single aperture.

## 4. Quadrupling the Number of Nulls of the Radar Pattern

Let us consider a conventional array of $2M+1$ uniformly weighted isotropic elements with uniform spacing $l$. The directivity pattern of the array is

$$C(u) = \sum_{k=-M}^{M} \exp\{jklu\} = \frac{\sin\left[(2M+1)\frac{lu}{2}\right]}{\sin\frac{lu}{2}} \quad \text{.....(20)}$$

and the corresponding two-way radar pattern is

$$|C(u)|^4 = \left|\frac{\sin\left[(2M+1)\frac{lu}{2}\right]}{\sin\frac{lu}{2}}\right|^4 \quad \text{.....(21)}$$

Such a pattern has $4(2M) = 8M$ nulls but only $2M$ of them are distinct; each null or zero is of multiplicity four. It is reasonable to suspect that a better pattern could be obtained if the nulls were all *distinct* and if they were *equally spaced.* Such a pattern can be synthesized by the single-aperture phase-switched radar system. In order to compare this pattern with the conventional pattern given by (21) we will contrive to let both patterns have the same beam-width measured to the first null. For the particular case of a seven-element array the four patterns of the new radar system which will achieve this are as follows:

$$A_1(u) = 1+1{\cdot}842\cos[l(u+\tfrac{3}{2}\Delta u)]+ +1{\cdot}396\cos[2l(u+\tfrac{3}{2}\Delta u)]+ +0{\cdot}740\cos[3l(u+\tfrac{3}{2}\Delta u)] \quad \text{.....(22)}$$

$$A_2(u) = A_1(u-\Delta u)$$
$$B_1(u) = A_1(u-2\Delta u) \quad \text{.....(23)}$$
$$B_2(u) = A_1(u-3\Delta u)$$

where

$$l\Delta u = \frac{10\pi}{161} = 0{\cdot}1951$$

The four patterns, each with distinct nulls, are shown in Fig. 3, along with the directivity pattern of the conventional linear array of seven elements,



World Radio History

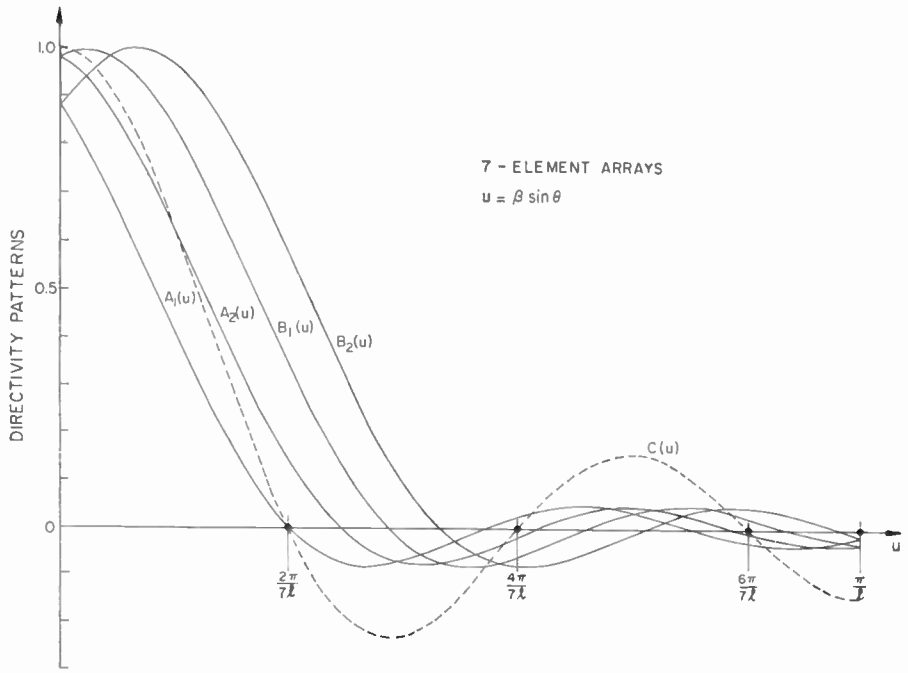



**Fig. 3.** Factor patterns of phase-switched and conventional radars for equispaced nulls, and equal main-lobes measured to first null.

$$C(u) = \frac{\sin 7\,\frac{lu}{2}}{\sin \frac{lu}{2}} \qquad \text{.....(24)}$$

In Fig. 4 the phase-switched radar pattern is compared with the conventional radar pattern; because of the large amplitude variation, a decibel plot is used. Note that both patterns have the same beam-width measured to the first null. However, the side-lobes of the phase-switched pattern are 7 dB or more below those of the conventional pattern and are more than four times as numerous (more precisely there are twenty-three phase-switched side-lobes and five conventional side-lobes). In the region farthest away from the main lobes the improvement in side-lobe level—in excess of 30 dB—is quite remarkable!

The author has shown elsewhere[4] that the optimum Dolph-Chebyshev array design can be extended to the two-way patterns of radar. When compared to the pattern of the conventional Chebyshev array the phase-switched Chebyshev two-way pattern has side-lobes which are about 16 dB lower when both patterns have the same beam-width measured to the first null. For arrays of $2M+1$ elements, the conventional pattern, and the phase-switched pattern (with four times as many nulls) are

$$|C(v)|^4 = |T_{2M}(v)|^4 \quad \text{(conventional)} \qquad \text{.....(25)}$$

$$R(v) = T_{8M}(v) \quad \text{(phase-switched)} \qquad \text{.....(26)}$$

where

$$v = v_0 \cos \frac{lu}{2}$$

with $v_0$ a constant determined by the side-lobe lever. The improvement in the side-lobe level of the phase-switched pattern is shown in Fig. 5 as a function of the conventional array two-way side-lobe level. The curve is for an element spacing of $l = \lambda/2$ and for large arrays, $M \geq 5$.

## 5. Noise and Interference

### 5.1. *Analysis of the Effect of Antenna Element Noise*

We assume that at each element there is a random Gaussian noise current which is independent of the signal and of the noise current at any other element. We let the current at the $k$th element be $N_k(t)$ whose average power is $\overline{N}$, independent of $k$. This noise will be present for both conventional and phase-switched arrays.

#### 5.1.1. Conventional array

In the presence of noise the output of the conventional system, using a single antenna and square-law detection, is

$$v_c(t) = \sum_{q=-\infty}^{\infty} \left( [\,|P(t-t_r-q\tau)|^2\,|C(u)|^4 + \right.$$
$$+ 2\,\mathrm{Re}\,\{P(t-t_r-q\tau)\,\exp[\mathrm{j}(\omega_0 t_r - \gamma_q)] \times$$
$$\left. \times\, C^2(u) N^*(t)\} \right) + |N(t)|^2 \qquad \text{.....(27)}$$



World Radio History

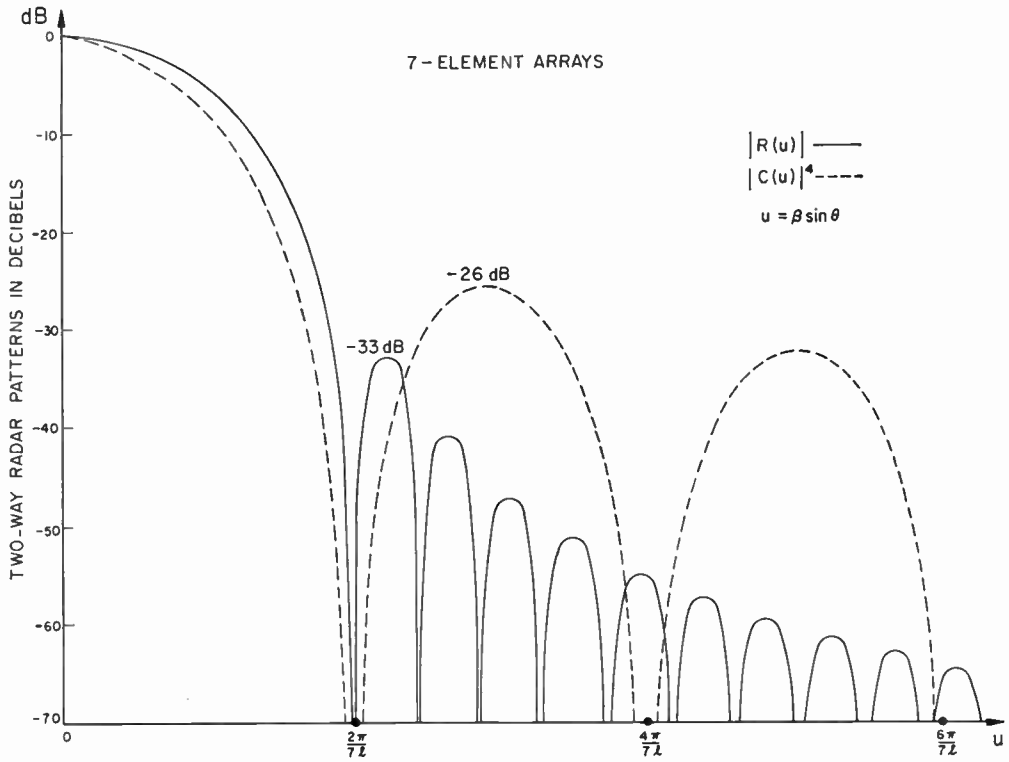



**Fig. 4.** Phase-switched and conventional radar patterns with equispaced side-lobes.

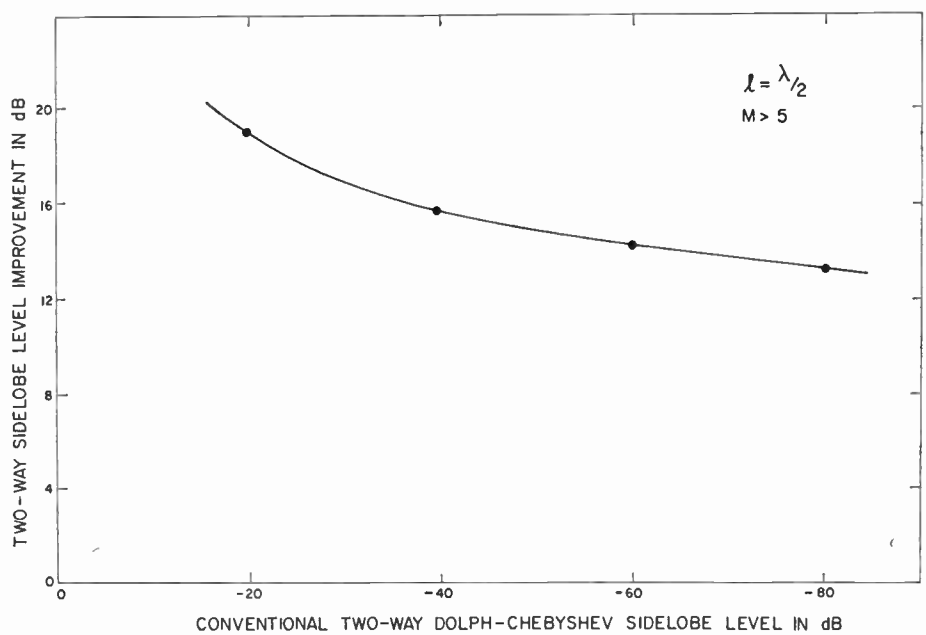



**Fig. 5.** Side-lobe level improvement of the phase-switched radar patterns as a function of the conventional Dolph-Chebyshev side-lobe level.



World Radio History

where

$$N(t) = \sum_{k=-M}^{M} c_k N_k(t) \qquad \text{.....(28)}$$

and $c_k$ is the $k$th element's weighting coefficient.

It is not difficult to show that if $t = t_r$ and $P(0) = 1$, then the expected value and variance of the output are respectively

$$E\{v_c(t_r)\} = |C(u)|^4 + \bar{N}\bar{C} \qquad \text{.....(29)}$$

and

$$\sigma_c^2(t_r) = 2|C(u)|^4\bar{N}\bar{C} + \bar{N}^2\bar{C}^2 \qquad \text{.....(30)}$$

where

$$\bar{C} = \sum_{k=-M}^{M} |c_k|^2 \qquad \text{.....(31)}$$

and

$$\bar{N} = E\{|N_k(t)|^2\} \qquad k = -M, \ldots, M \qquad \text{.....(32)}$$

5.1.2. Phase-switched array

The signal-plus-noise output can be written as

$$\begin{aligned} v_p(t) = \sum_{q=-\infty}^{\infty} \Big[ & 16\,\mathrm{Re}\{A_1 A_2^*\}\,\mathrm{Re}\{B_1 B_2^*\} \times \\ & \times |P(t-t_r-q\tau)|^2 + 2\,\mathrm{Re}\{A_+ B_+ \exp[j\gamma_q] N_+^*(t) \\ & - A_+ B_- \exp[j\gamma_{q-1}] N_-(t-\tau-q\tau) - \\ & - A_- B_+ \exp[j\gamma_{q-2}] \times N_+^*(t-2\tau-q\tau) + \\ & + A_- B_- \exp[j\gamma_{q-3}] N_-^*(t-3\tau-q\tau)\exp(-\omega_0 t_r) \\ & \times P(t-t_r-q\tau)\} \Big] + |N_+(t)|^2 - |N_-(t-\tau)|^2 \\ & - |N_+(t-2\tau)|^2 + |N_-(t-3\tau)|^2 \end{aligned} \qquad \text{.....(33)}$$

where

$$N_\pm(t) = \sum_{k=-M}^{M} (b_{1k} \pm b_{2k}) N_k(t) \qquad \text{.....(34)}$$

and, for simplicity, we have suppressed the pattern variable; thus $A_+(u)$ has become simply $A_+$, etc.

One can show that the expected value and the variance of this output for $t = t_r$, and $u = 0$, are given as follows:

$$E\{v_p(t_r)\} = 16\,\mathrm{Re}\{A_1(0)A_2^*(0)\}\,\mathrm{Re}\{B_1(0)B_2^*(0)\} \qquad \text{.....(35)}$$

$$= 16A_1(0)A_2(0)B_1(0)B_2(0) \quad \text{(for real patterns)} \qquad \text{.....(35a)}$$

$$\sigma_p^2(t_r) = 2[(\bar{B}_+^2 + \bar{B}_-^2)\bar{N}^2 + |A_1(0) + A_2(0)|^2 \times |B_1(0) + B_2(0)|^2 \bar{B}_+ \bar{N}] \qquad \text{.....(36)}$$

where

$$\bar{B}_\pm = \sum_{k=-M}^{M} |b_{1k} \pm b_{2k}|^2 \qquad \text{.....(37)}$$

5.1.3. Signal/noise ratios

If we define the output signal/noise ratio as

$$\left(\frac{S}{N}\right) = \frac{E\{v_s(t_r)\}^2}{\sigma^2(t_r)} \qquad \text{.....(38)}$$

where $v_s(t)$ is the signal component of the output voltage, then the signal/noise ratios for the conventional and phase-switched systems for $u = 0$, $t = t_r$, are respectively

$$\left(\frac{S}{N}\right)_c = \frac{|C(0)|^8}{2|C(0)|^4\bar{C}\bar{N} + \bar{C}^2\bar{N}^2} \qquad \text{.....(39)}$$

$$\left(\frac{S}{N}\right)_p = 128\,\frac{[A_1(0)B_1(0)A_2(0)B_2(0)]^2}{A_+^2(0)B_+^2(0)\bar{B}_+\,\bar{N} + (\bar{B}_+^2 + \bar{B}_-^2)\bar{N}^2} \qquad \text{.....(40)}$$

if the patterns are all real as is usually the case.

Now in comparing the phase-switched and conventional radars from the point of view of signal/noise ratio we must, to be realistic, assume that both systems are fed by transmitters of equal power. However, due to the amplitude and phase-switching of the new system, the power radiated by it will vary, being greater for the $A_+$ mode than the $A_-$ mode. We will assume that the *average* of the powers radiated by the two modes is the same as that of the conventional radar; thus,

$$\int |C(u)|^2\,du = \tfrac{1}{2}\left[\int |A_+(u)|^2\,du + \int |A_-(u)|^2\,du\right] \qquad \text{.....(41)}$$

$$= \int [|A_1(u)|^2 + |A_2(u)|^2]\,du \qquad \text{.....(42)}$$

But, by Parseval's theorem, this is equal to

$$\sum_{k=-M}^{M} |c_k|^2 = \sum_{k=-M}^{M} \{|a_{1k}|^2 + |a_{2k}|^2\} \qquad \text{.....(43)}$$

if the element spacing is $n\dfrac{\lambda}{2}$, $n = 1, 2, 3, \ldots$. In simpler notation (43) can be written as

$$\bar{C} = \bar{A}_{11} + \bar{A}_{22} \qquad \text{.....(44)}$$

In a similar fashion we can reason that the ratio between the conventional and phase-switched receiving patterns should be

$$\int |C(u)|^2\,du = \tfrac{1}{2}\left[\int |B_+(u)|^2\,du + \int |B_-(u)|^2\,du\right] \qquad \text{.....(45)}$$

$$= \int [|B_1(u)|^2 + |B_2(u)|^2]\,du \qquad \text{.....(46)}$$

or

$$\sum_{k=-M}^{M} |c_k|^2 = \sum_{k=-M}^{M} \{|b_{1k}|^2 + |b_{2k}|^2\} \qquad \text{.....(47)}$$

Again we can write (47) in simpler notation as

$$\bar{C} = \bar{B}_{11} + \bar{B}_{22} \qquad \text{.....(48)}$$

5.1.4. The seven-element arrays

It should be remarked that by equating the total radiated powers of the two systems the average value of the phase-switched power density radiated in the $u = 0$ direction (direction of the product pattern's maximum) will in general differ from that radiated in



World Radio History

the same direction by the conventional system. A similar condition obtains for the receiving half of both systems—the responses to identical returning pulses may differ. Consequently the signal output from each system will differ in magnitude even when the total input power is the same. The important question, however, is whether the phase-switched system's *signal/noise* ratio is improved, a question which we will now answer for the case of the seven-element arrays of Section 4.

For the particular case of the seven-element arrays, whose patterns are shown in Fig. 3, it is convenient to let the element currents of the conventional array be 1 ampere, and hence (43) and (47) become

$$7 = \sum_{k=-3}^{3} (|a_{1k}|^2 + |a_{2k}|^2) \qquad \text{.....(49)}$$

$$7 = \sum_{k=-3}^{3} (|b_{1k}|^2 + |b_{2k}|^2) \qquad \text{.....(50)}$$

In order to satisfy these equations the current amplitudes must be rescaled from the values given in eqn. (22); specifically all currents must be decreased by 5·81%. Thus, the first pattern becomes

$$A_1(u) = 0{\cdot}9419 + 1{\cdot}735 \cos\left[lu + \frac{15\pi}{161}\right] + 1{\cdot}315 \cos\left[2\left(lu + \frac{15\pi}{161}\right)\right] + 0{\cdot}6972 \cos\left[3\left(lu + \frac{15\pi}{161}\right)\right] \qquad \text{.....(22a)}$$

The other three patterns are scaled similarly and when all are substituted into (40), the signal/noise ratio becomes

$$\left(\frac{S}{N}\right)_p = \frac{213{\cdot}0}{\overline{N} + {\cdot}00232\overline{N}^2} \qquad \text{.....(40a}$$

Similarly, for conventional array currents of 1 ampere, (39) becomes

$$\left(\frac{S}{N}\right)_c = \frac{171{\cdot}5}{\overline{N} + {\cdot}00146\overline{N}^2} \qquad \text{.....(39a)}$$

The improvement in signal/noise ratio can be written as

$$\frac{\left(\frac{S}{N}\right)_p}{\left(\frac{S}{N}\right)_c} = 1{\cdot}24\,\frac{1 + 0{\cdot}00146\overline{N}}{1 + 0{\cdot}00232\overline{N}} \qquad \text{.....(51)}$$

and is plotted in decibels as a function of the original signal/noise ratio $\left(\frac{S}{N}\right)_c$ (also in decibels) in Fig. 6. Note that for high original signal/noise levels there is a slight improvement with the new system whereas for low levels there is a slight degradation. However, we must conclude that the two systems have essentially the same performance with respect to uncorrelated element noise.

### 5.2. *Analysis of Interference due to a Single Point Target*

Let us assume that in addition to the desired target of unit amplitude located in the main beam of the pattern at $u = 0$, there is a secondary target of arbitrary amplitude $N$, phase $\xi$, and direction $u$. We will

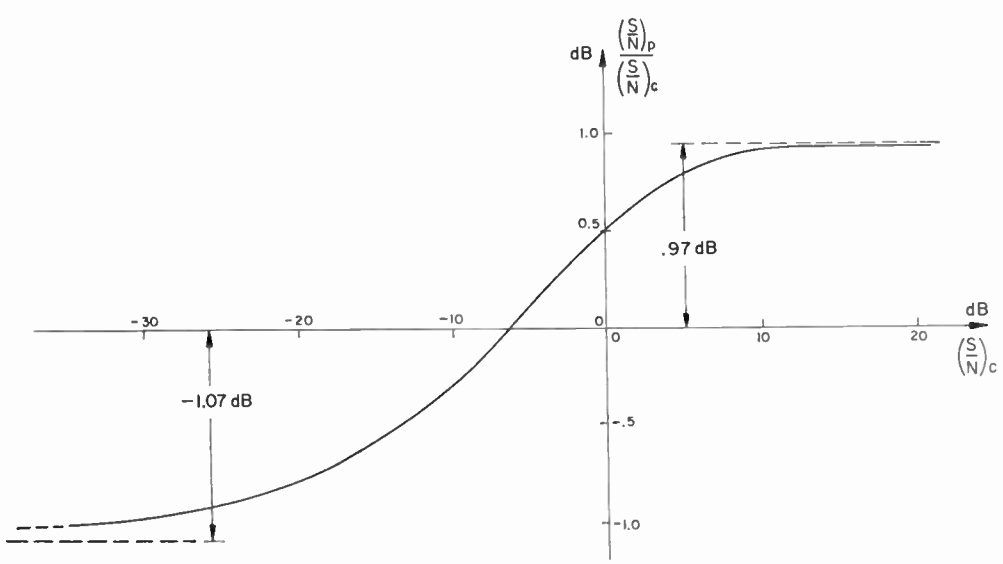



**Fig. 6.** Comparison of output signal-to-noise ratios for seven-element arrays with uncorrelated element noise.



World Radio History

now consider the effect of this interference on the conventional and phase-switched systems.

5.2.1. Conventional system

The square-law output of the radar can be written as

$$v_c = |C(0)|^4 + 2\,\mathrm{Re}\,\{C^2(0)C^2(u)N\exp(j\xi) + N^2|C(u)|^4 \quad \ldots\ldots(52)$$

Here we have suppressed the range variable, tacitly assuming the targets to be at the same range. Inspection of (52) shows that only the first term is due to signal alone; the second is a cross-product of signal and interference and the third is due to interference alone. Let us define the interference function of the radar to be the ratio of the interference components to the signal component in the system output. Thus for a real pattern the interference function is

$$I_c(u, N, \xi) = 2\bar{C}^2(u)N\cos\xi + N^2\bar{C}^4(u) \quad \ldots\ldots(53)$$

where

$$\bar{C}(u) = \frac{C(u)}{C(0)} \quad \ldots\ldots(54)$$

In practice, the phase angle $\xi$ may be time-independent or may vary randomly over all possible values from 0 to $2\pi$. We say, in the former case, that the interference is coherent and in the latter that it is incoherent.

Incoherent interference: since the statistical average or expectation of $\cos\xi$ is

$$\mathrm{E}\{\cos\xi\} = 0$$

the interference function for incoherent interference becomes

$$I_c(u, N, \xi) = N^2\bar{C}^4(u) \quad \text{(incoherent)} \ldots\ldots(55)$$

which is proportional to the interference power and the two-way power pattern of the radar. Consequently, a radar whose pattern has low side-lobes will tend to reject incoherent interference, a not unexpected conclusion.

Coherent interference: let us consider the worst possible case of coherent interference—$\cos\xi = 1$. The interference function then becomes

$$I_c(u, N, 0) = 2N\bar{C}^2(u) + N^2\bar{C}^4(u) \quad \text{(coherent)} \ldots\ldots(56)$$

and in this case depends not only on the two-way power pattern $\bar{C}^4(u)$ but also on the 'cross-product pattern' $\bar{C}^2(u)$. We also note that due to the cross-product term the interference is no longer linear in power, and since in many cases the cross-product term will dominate, the interference will be proportional to the square root of the power of the interfering signal. However, even in the coherent case, a radar pattern with low side-lobes will have a good interference function with a low response to interference from outside the main lobe of the pattern.

5.2.2. Phase-switched system

The system output for real patterns can be written as

$$\begin{aligned} V_p = 16\Big[ & A_1(0)B_1(0)A_2(0)B_2(0) + \\ & + \frac{N}{2}\{A_1(0)B_1(0)A_2(u)B_2(u) + \\ & + A_1(0)B_2(0)A_2(u)B_1(u) + \\ & + B_1(0)A_2(0)A_1(u)B_2(u) + \\ & + A_2(0)B_2(0)A_1(u)B_1(u)\}\cos\xi + \\ & + N^2A_1(u)B_1(u)A_2(u)B_2(u)\Big] \quad \ldots\ldots(57) \end{aligned}$$

Note that if $A_1 = A_2 = B_1 = B_2 = C$ this reduces to the conventional expression given by eqn. (52). Again we define the interference function as the ratio of interference to signal components; we obtain

$$\begin{aligned} I_p(u, N, \xi) = {} & \tfrac{1}{2}[\bar{A}_1(u) + \bar{A}_2(u)] \times \\ & \times [\bar{B}_1(u) + \bar{B}_2(u)]N\cos\xi + \\ & + N^2\bar{A}_1(u)\bar{B}_1(u)\bar{A}_2(u)\bar{B}_2(u) \quad \ldots\ldots(58) \end{aligned}$$

where

$$\bar{A}_i(u) = \frac{A_i(u)}{A_i(0)}, \qquad \bar{B}_i(u) = \frac{B_i(u)}{B_i(0)}, \qquad i = 1, 2 \quad \ldots\ldots(59)$$

We distinguish the two limiting cases of coherent and incoherent interference as with the conventional radar.

Incoherent interference: the equation corresponding to (55) is

$$I_p(u, N, \xi) = N^2\bar{A}_1(u)\bar{B}_1(u)\bar{A}_2(u)\bar{B}_2(u) \quad \text{(incoherent)} \ldots\ldots(60)$$

Just as in the case of conventional radar, the function is proportional to the two-way power pattern of the system and to the power of the interfering signal.

Coherent interference: for the worst case of coherent interference ($\xi = 0$ or $\pi$) we obtain

$$\begin{aligned} I_p(u, N, 0 \text{ or } \pi) = {} & \tfrac{1}{2}\big|[A_1(u) + A_2(u)] \times \\ & \times [B_1(u) + B_2(u)]\big|N + \\ & + N^2\big|\bar{A}_1(u)\bar{B}_1(u)\bar{A}_2(u)\bar{B}_2(u)\big| \\ & \text{(coherent)} \quad \ldots\ldots(61) \end{aligned}$$

Note that the magnitude of the two components *add* for the worst case. This will occur sometimes for $\xi = 0$ and sometimes for $\xi = \pi$.

5.2.3. The seven-element arrays

We will now compare the interference rejection of the two systems which use the seven-element linear arrays whose patterns are shown in Fig. 3. Since we have shown that the rejection of incoherent interference depends only on the two-way power pattern of either system we can conclude immediately that the phase-switched system has a superior interference function simply by comparing the two-way power patterns shown in Fig. 4. For the more complicated



World Radio History

case of coherent interference, it is necessary to compare the interference functions of the two systems for several values of the amplitude of the interfering signal. Thus in Fig. 7(a) the interference functions of both systems have been plotted for $N = 0{\cdot}1$. It will be observed that the side-lobes of the phase-switched interference functions are lower and more than twice as numerous as those of the conventional system. However, the conventional system has a narrower main lobe. Similar interference functions can be obtained for $N = 1$ and $N = 10$ and are shown in Figs. 7(b) and 7(c).

From these curves we reach the important conclusion that it is *not* the two-way power pattern of the system that is important in rejecting coherent interference but it is the *cross-product pattern*. The cross-product pattern for the conventional and the phase-switched systems are given below:

$$2C^2(u) \qquad \text{(conventional)}$$

$$\tfrac{1}{2}[A_1(u)+A_2(u)][B_1(u)+B_2(u)] = \tfrac{1}{2}A_+(u)B_+(u) \qquad \text{(phase switched)} \qquad \text{.....(62)}$$

Note that the conventional pattern is simply the square of the directivity pattern $C(u)$ of the conventional array, whereas the pattern of the phase-switched system is the *product* of the transmitting pattern $A_+(u)$ and the receiving pattern $B_+(u)$ of the phase-switched system. We infer that in synthesizing such patterns the new system will have twice as many degrees of freedom as will the old and in general will yield better patterns.

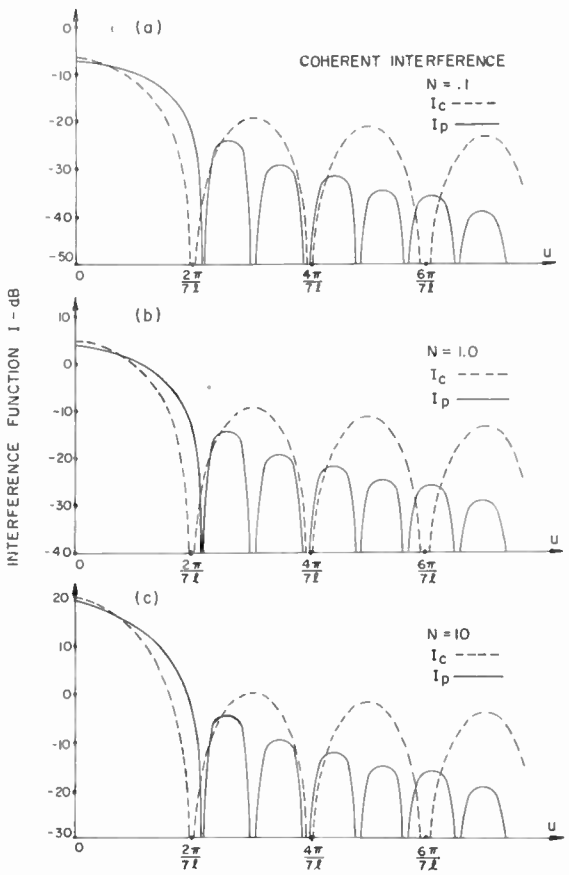



Fig. 7. Interference function for coherent interference when: (a) $N = 0{\cdot}1$, (b) $N = 1{\cdot}0$, (c) $N = 10$.

### 6. An Improved Pattern for Coherent Targets

If it were known that the target distribution was coherent, it would be logical to synthesize a good cross-product pattern, and if possible simultaneously produce a good two-way power pattern. We will now do this by letting the phase-switched cross-product pattern $A_+(u)B_+(u)$ have the same beam-width as that of the conventional pattern but require that the former pattern's nulls be equispaced and hence produce lower side-lobes.

For the seven-element array we must let

$$A_+(u) = 1+1{\cdot}8833\cos[l(u+\Delta_1 u)]+ +1{\cdot}5389\cos[2l(u+\Delta_1 u)]+ +0{\cdot}9546\cos[3l(u+\Delta_1 u)] \qquad \text{.....(63)}$$

$$B_+(u) = A_+(u-2\Delta_1 u) \qquad \text{.....(64)}$$

where

$$l\Delta_1 u = \frac{5\pi}{77}$$

If we choose

$$A_1(u) = A_+(u+\tfrac{1}{2}\Delta_1 u) \qquad \text{.....(65)}$$

$$B_2(u) = B_+(u-\tfrac{1}{2}\Delta_1 u) = A_+(u-\tfrac{5}{2}\Delta_1 u) \qquad \text{.....(66)}$$

then

$$A_2(u) = 2A_+(u)-A_1(u) \qquad \text{.....(67)}$$

$$B_1(u) = 2B_+(u)-B_2(u) \qquad \text{.....(68)}$$

The cross-product pattern that results is shown in Fig. 8, along with that of the conventional seven-element uniform array. We see that the improvement in side-lobe level varies from 3 dB for the first side-lobes to about 10 dB for those farthest away from the main lobe. This also indicates the improvement in the ability of the new system to reject a coherent interfering signal when the desired signal is incident at $u = 0$.

The regular two-way patterns of the two systems are shown in Fig. 9 where we see that not only does the phase-switched pattern still have lower side-lobes but its main-lobe is about 11·3% narrower. Since this pattern is not only improved for coherent targets but is also better in the case of incoherent targets, we can expect a better performance from this phase-switched system for any degree of target coherence.



World Radio History

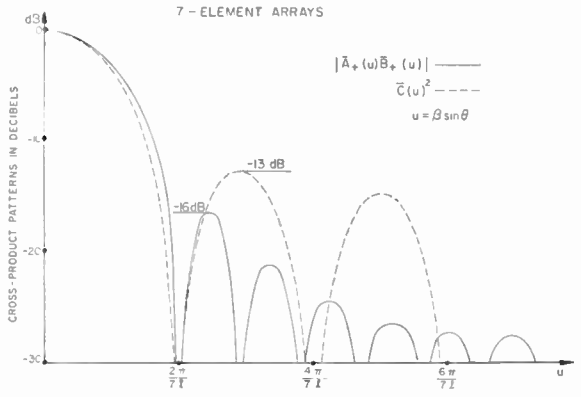



Fig. 8. Graph of cross-product radar patterns showing improved phase-switched pattern with lower side-lobes.

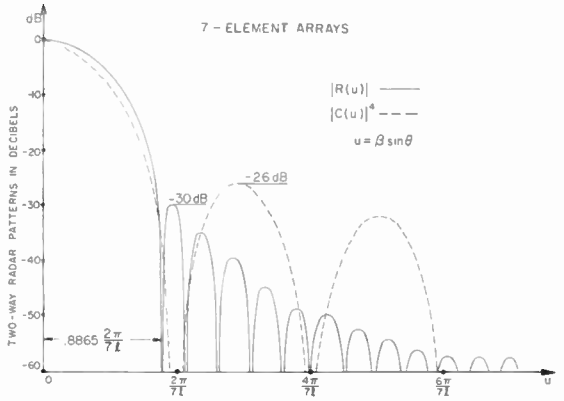



Fig. 9. Graph of two-way power patterns showing narrower main-lobe and lower side-lobes of the phase-switched pattern.

## 7. Conclusions

It has been demonstrated that the new four-antenna phase-switched radar has several advantages when comparison is made with a conventional one-antenna radar; indeed from the pattern-synthesis viewpoint the latter may be thought of as a degenerate form of the new system which occurs when all four of the phase-switched patterns are identical. Consequently, the new system should prove useful in many practical applications where increased directivity is needed. A curious aspect of the phase-switched radar is that its average response is zero to both element noise and external noise sources (if the latter are uncorrelated with the transmitted pulses). The new system is 'blind', on the average, to everything but its own returning pulses.

Finally it should be remarked that a simplification of the system can be effected by using a single transmitting antenna and two receiving antennae whose outputs are multiplied in the usual fashion. The resulting two-way pattern, with $A_1 = A_2 = A$, is now

$$R(u) = |A(u)|^2 \operatorname{Re}\{B_1(u)B_2^*(u)\}$$

On the other hand it is possible, by a somewhat more involved process of phase-switching, to obtain the following output:

$$R(u) = \operatorname{Re}\{A_1(u)A_2^*(u)B_1(u)B_2^*(u)\}$$

However, the pattern obtained with the system described in this paper,

$$R(u) = \operatorname{Re}\{A_1(u)A_2^*(u)\} \operatorname{Re}\{B_1(u)B_2^*(u)\}$$

is probably sufficient for most applications.

## 8. Acknowledgment

This work was initiated at the University of Illinois where the author was a graduate student under the guidance of Professor Georges A. Deschamps. The work was sponsored by the Aeronautical Systems Division, Wright-Patterson Air Force Base, Ohio, through contract no. AF33(657)-10474.

The work was completed at the University of Waterloo with the aid of Research Grant No. A-2176 of the National Research Council of Canada.

## 9. References

1. V. G. Welsby and D. G. Tucker, 'Multiplicative receiving arrays', *J. Brit. I.R.E.*, **19**, pp. 369–82, June 1959.
2. I. W. Linder, 'Resolution characteristics of correlation arrays', *J. Res. Nat. Bur. Stand.*, **65D**, pp. 245–52, May–June 1961.
3. B. Y. Mills and A. G. Little, 'A high resolution aerial system of a new type', *Australian J. Phys.*, **6**, pp. 272–8, September 1953.
4. R. H. MacPhie, 'Optimum cross-correlation radar system', *Proc. Inst. Radio Engrs*, **50**, No. 12, pp. 2508–9, December 1962.
5. D. C. Cooper, 'The use of an effective transmission pattern to improve the angular resolution of within-pulse sector-scanning radar or sonar systems', *The Radio and Electronic Engineer*, **28**, No. 3, pp. 153–60, September 1964.

## 10. Appendix

If we consider a single target at the range $r$ the output of the square-law detector of the single-aperture system will be a series of pulses of varying amplitudes with a period of fluctuation of $4\tau$. If we let

$$\begin{aligned}\alpha(t) &= |A_1+A_2|^2\,|B_1+B_2|^2\,|P(t)|^2\\ \beta(t) &= |A_1+A_2|^2\,|B_1-B_2|^2\,|P(t)|^2\\ \gamma(t) &= |A_1-A_2|^2\,|B_1+B_2|^2\,|P(t)|^2\\ \delta(t) &= |A_1-A_2|^2\,|B_1-B_2|^2\,|P(t)|^2\end{aligned} \qquad \ldots\ldots(69)$$

where $A_i = A_i(u)$, $B_i = B_i(u)$, for $i = 1, 2$, and $P(t)$ is the radar pulse, then the output of the square-law detector can be written as



World Radio History

$$p_0(t) = \sum_{k=-\infty}^{\infty} \{\alpha[t'-4k\tau]+\beta[t'-(4k+1)\tau]+ +\gamma[t'-(4k+2)\tau]+\delta[t'-(4k+3)\tau]\} \quad \text{.....(70)}$$

where

$$t' = t - \frac{2r}{c}$$

Now the delay-line network forms the following sum (except for a constant factor)

$$p_1(t) = p_0(t') - p_0(t'-\tau) - p_0(t'-2\tau) + p_0(t'-3\tau) \quad \text{.....(71)}$$

Substituting eqn. (70) in the above we obtain

$$p_2(t) = \sum_{k=-\infty}^{\infty} (-1)^k \bar{p}(t'-k\tau) = \sum_{k=-\infty}^{\infty} (-1)^k \bar{p}\left(t - \frac{2r}{c} - k\tau\right) \quad \text{.....(72)}$$

where

$$\bar{p}(t) = \alpha(t) - \beta(t) - \gamma(t) + \delta(t) \quad \text{.....(73)}$$

Finally, the delay-line network output is fed into a polarity switching device operating in synchronism with the pulse repetition frequency which 'rectifies' $p_2(t)$ to yield

$$p_3(t) = \sum_{k=-\infty}^{\infty} \bar{p}\left(t - \frac{2r}{c} - k\tau\right) \quad \text{.....(74)}$$

We will now show that $\bar{p}$ is proportional to the product of our four antennae patterns.

First we write

$$\alpha(t) = [|A_1|^2 + |A_2|^2 + 2\,\mathrm{Re}\{A_1 A_2^*\}] \times \times [|B_1|^2 + |B_2|^2 + 2\,\mathrm{Re}\{B_1 B_2^*\}]|P(t)|^2 = [A + A_{12}][B + B_{12}]|P(t)|^2 \quad \text{.....(75)}$$

where

$$A = |A_1|^2 + |A_2|^2, \qquad A_{12} = 2\,\mathrm{Re}\{A_1 A_2^*\}$$
$$B = |B_1|^2 + |B_2|^2, \qquad B_{12} = 2\,\mathrm{Re}\{B_1 B_2^*\} \quad \text{.....(76)}$$

Thus

$$\alpha(t) = [AB + AB_{12} + BA_{12} + A_{12}B_{12}]|P(t)|^2 \quad \text{.....(77)}$$

Similarly

$$-\beta(t) = [-AB + AB_{12} - BA_{12} + A_{12}B_{12}]|P(t)|^2$$
$$-\gamma(t) = [-AB - AB_{12} + BA_{12} + A_{12}B_{12}]|P(t)|^2$$
$$\delta(t) = [AB - AB_{12} - BA_{12} + A_{12}B_{12}]|P(t)|^2 \quad \text{.....(78)}$$

From the above it is easy to see that

$$\bar{p}(t) = \alpha(t) - \beta(t) - \gamma(t) + \delta(t) \quad \text{.....(79)}$$
$$= 4A_{12}B_{12}|P(t)|^2$$
$$= 16\,\mathrm{Re}\{A_1 A_2^*\}\,\mathrm{Re}\{B_1 B_2^*\}|P(t)|^2 \quad \text{.....(80)}$$

which is the desired result.

*Manuscript first received by the Institution on 8th June 1965, and in final form on 18th August 1965. (Paper No. 1024/RNA48.)*

© The Institution of Electronic and Radio Engineers, 1966



World Radio History

U.D.C.621.382

# Intrinsic Frequency Limitations for Semiconductor Microwave Devices

*By*

H. V. SHURMER, M.Sc., Ph.D.†

*Reprinted from the Proceedings of the Joint I.E.R.E.–I.E.E. Symposium on 'Microwave Applications of Semiconductors' held in London from 30th June to 2nd July 1965.*

**Summary:** The upper frequency of operation for a semiconductor microwave device is usually considered in terms of the equivalent circuit parameters associated with the immediate neighbourhood of the active element, together with any stray reactances of the package. Whilst it is generally recognized that in the semiconductor contribution both the resistive and reactive terms may have a frequency dependence, the exact nature of this is difficult to determine experimentally and its importance is generally obscure.

This paper seeks to give a brief survey of the various mechanisms which can conceivably introduce frequency-dependent effects and attempts to gauge which of them is likely to be important in devices of present interest —crystal valves, varactors, tunnel/backward diodes and p-i-n switches. The mechanisms discussed include such items as skin effect, carrier/phonon collisions, dielectric relaxation, carrier lifetime, transit time, etc.

**List of Symbols**

| | |
|---|---|
| $B$ | bandwidth |
| $c$ | velocity of light |
| $C_B$ | capacitance of barrier layer |
| $C_I$ | capacitance of intrinsic layer |
| $d$ | thickness of intrinsic layer |
| $D$ | electronic diffusion constant |
| $e$ | electronic charge |
| $E$ | field strength |
| $f$ | frequency |
| $J$ | current density |
| $k$ | Boltzmann's constant |
| $K$ | a constant |
| $L$ | diffusion length |
| $m$ | electronic mass |
| $n$ | carrier density |
| $P$ | ionization probability |
| $q$ | charge density |
| $Q$ | quality factor |
| $r$ | series resistance |
| $R_B$ | resistance of barrier layer |
| $R_I$ | resistance of intrinsic layer |
| $R_N$ | negative resistance |
| $S$ | thickness of depletion layer |
| $S_0$ | thickness of depletion layer at zero applied voltage |
| $t$ | time |
| $T$ | temperature, deg K |
| $v_T$ | thermal velocity |
| $V$ | instantaneous applied voltage |
| $V_s$ | static voltage ('built-in' plus bias voltage) |
| $V_0$ | amplitude of alternating voltage |
| $w$ | width of wafer |
| $\Delta W$ | depth of ionic trap |
| $\delta$ | skin depth |
| $\varepsilon$ | permittivity |
| $\varepsilon_r$ | relative permittivity |
| $\lambda$ | wavelength |
| $\rho$ | resistivity |
| $\sigma$ | conductivity |
| $\tau_m$ | mean free time |
| $\tau_p$ | lifetime of holes |
| $\tau_s$ | switching time |
| $\tau_t$ | transit time |
| $\mu$ | permeability |
| $\mu_e$ | electronic mobility |
| $\omega$ | angular frequency |

† Associated Electrical Industries Ltd., Central Research Laboratory, Rugby, Warwickshire.



World Radio History

## 1. Introduction

Semiconductor microwave devices mainly depend on the property of a potential barrier to behave as a voltage variable resistance or capacitance, one or other of these being made to predominate, according to the circuit function required. The usual circuit representation and cut-off frequency $f_c$ for each device is shown in Table 1.

**Table 1**

| DEVICE | EQUIVALENT CIRCUIT | CUT-OFF FREQUENCY |
|---|---|---|
| CRYSTAL MIXER | FORWARD VOLTAGE: $r$; REVERSE VOLTAGE: $r$, $C_B$ | $f_c = \frac{1}{2\pi C_B r}$ |
| CRYSTAL DETECTOR | $r$, $R_B$, $C_B$ | $f_c = \frac{1}{2\pi C_B \sqrt{(R_B r)}}$ |
| VARACTOR | $r$, $C_B$ | $f_c = \frac{1}{2\pi C_B r}$ |
| TUNNEL DIODE | $-R_N$, $C_B$, $r$, $L$ | $f_c = \frac{1}{2\pi C_B R_N}\sqrt{\left(\frac{R_N}{r} - 1\right)}$ |
| P-I-N SWITCHING DIODE | LOW FREQUENCIES, ZERO BIAS: $R_B$, $R_I$, $C_B$, $C_I$; HIGH FREQUENCIES, ZERO BIAS: $R_I$, $C_I$; ALL FREQUENCIES, FORWARD BIAS: $R_I'$, $R_I' \ll R_I$ | |

As the above modes of characterization do not express the non-linear qualities upon which device operation essentially depends, their usefulness is mainly restricted in the one case to affording guidance on matching properties and in the other to indicating that part of the frequency spectrum above which efficiency decreases rapidly.

The term 'cut-off frequency' enjoys widespread use, but its meaning varies somewhat with each device, so that it is not legitimate to attempt any basic comparison in the devices. For instance, the criterion used with crystal mixers is that frequency at which the back-to-front resistance ratio becomes small, whilst for detectors it is that at which half of the absorbed power is dissipated in the spreading resistance. For varactors, the corresponding criterion is when equal voltages appear across the barrier and the series resistance. The resistive cut-off for tunnel diodes is however rather more absolute in being the maximum frequency at which the device will exhibit negative resistance.

In designing and developing microwave devices, the engineer should be aware not only of the first order effects which apply at low frequencies but also of second-order effects which are of a frequency dependent nature. This is particularly true when radically new materials or geometries are contemplated.

It is likely that one of the most significant of the frequency-dependent mechanisms is skin effect, giving an additional contribution to the series resistance of p-n junction devices at microwave frequencies. The trapping of charge carriers is probably of similar importance in relation to barrier capacitance, for point contact crystals and, possibly, varactor diodes also. Barrier resistance may be governed by transit time, which must be short compared to a microwave period for a good rectifier. Other mechanisms are undoubtedly operative too and we now attempt to make some assessment of the importance of each of them, taking first the effects which are most general in character.

## 2. Skin Effect

Table 2 gives the skin depths at 9·375 Gc/s for resistivities which are typical of material used in the various microwave semiconductor devices, calculated from the formula:

$$\delta = \frac{1}{\sqrt{(\pi f \mu \sigma)}} = \frac{1}{\pi}\sqrt{\left(\frac{\lambda \rho}{120}\right)}$$

**Table 2**

Skin depths at 9·375 Gc/s for various semiconductor devices

| Device | Resistivity (ohm-cm) | Skin depth (microns) |
|---|---|---|
| p-i-n switch | | |
| zero bias | 500 | 11600 |
| forward | 0·2 | 232 |
| Varactor | 0·10 | 164 |
| | 0·05 | 116 |
| Point contact crystal | 0·03 | 90 |
| | 0·01 | 52 |
| Tunnel diode or | 0·003 | 29 |
| substrate for | 0·001 | 16 |
| epitaxial device | 0·0003 | 9 |
| Tungsten | $5{\cdot}50 \times 10^{-6}$ | 1·22 |
| Copper | $1{\cdot}72 \times 10^{-6}$ | 0·68 |

The possible importance of skin effect in increasing the effective series resistance mainly concerns the region in the immediate vicinity of the metal/semiconductor boundary adjacent to the active area of each device, but an exact solution for this situation appears to be lacking.

The problem has been considered by Torrey and Whitmer[1] in relation to point contact crystals, using a simplified model of two equi-diameter semi-infinite cylinders of metal and semi-conductor in end-to-end contact. The variation in power dissipation over the



World Radio History

contact face was estimated to be less than 0·5% at 10 cm wavelength and it was concluded that skin effect is entirely negligible at any microwave frequency for all point contact rectifiers.

For the model described, the skin effect was shown to be an increasing function of the product of conductivity and contact area. With some of the present devices this product is however much larger than for point contact crystals and skin effect may well not always be negligible. Varactors, for instance, have an effective area greater than crystals by two orders of magnitude, although they are one order less in conductivity, whilst forward biased p-i-n diodes are not much lower in conductivity than varactors but are about four orders of magnitude greater in area than crystals. Backward diodes, although similar in area to crystals, have conductivities about one order of magnitude higher. The greatest effects would however be expected with tunnel diodes, having even higher conductivities and areas comparable with varactors.

If skin effects are in fact appreciable for conventional geometries they will certainly impose a restriction on attempts to reduce spreading resistance by epitaxial techniques, since the success of these depends on a parallel current flow in the immediate vicinity of the contact. Further, a more realistic model for the conventional case would tend to accentuate skin effects. It is clear that this particular subject deserves much closer study.

### 3. Effect on Carrier/Phonon Collisions at Very High Frequencies

In a semiconducting material the drift velocity of electrons (or holes) for a given electric field is set by the scattering of electron waves through interference with phonons. Quantum mechanical considerations show that the effect of phonons can be accurately derived if they are represented by a set of phantom particles, with which the electrons make nearly elastic collisions. For applied frequencies higher than the collision frequency the usual charge carrier process through a semiconductor will clearly be modified.

Taking, as a rough approximation, the electron effective mass as equal to that of a free electron and defining a thermal velocity $v_T$ in accordance with the principle of equipartition of energy, i.e.

$$v_T = \left(\frac{3kT}{m}\right)^{\frac{1}{2}}$$

gives the thermal velocity of an electron at room temperature as $v_T \simeq 10^5$ m/s.

The electronic mobility is expressed in terms of the charge-to-mass ratio and mean free time as[2]

$$\mu_e \simeq \frac{e\tau_m}{m}$$

For p-type silicon, with a hole mobility of 0·025 $m^2/Vs$, $\tau_m \simeq 1{\cdot}4 \times 10^{-13}$ s. Whilst this mean free time is much shorter than a microwave period, for germanium of electron mobility 0·360 $m^2/Vs$, $\tau_m \simeq 2 \times 10^{-12}$ s, which is slightly longer than the half period of 1 mm waves. The highest mobility known at present is that for electrons in indium antimonide, which is 8·0 $m^2/Vs$ at 77°K for carrier concentrations appropriate to semiconductor microwave devices, i.e. about $10^{23}$ atoms/$m^3$. Although this material is not widely used at present, it has a value for $\tau_m$ of about $4{\cdot}5 \times 10^{-11}$ s, which corresponds to a frequency in the 3-cm band. It is possible therefore that the usual laws of conduction in semiconductors may be modified at microwave frequencies as a result of there being fewer carrier/phonon collisions per period.

In the absence of collisions, the equation of motion (for electrons) may be written as

$$m\ddot{x} = Ee$$

Putting

$$E = E_0 \exp(j\omega t)$$

and current density

$$J = ne\dot{x}$$

the result is obtained

$$J = \frac{1}{j\omega}\frac{ne^2}{m}E_0 \exp(j\omega t) \qquad \text{.....(1)}$$

It will be observed that the factor $m/ne^2$ in eqn. (1) is equivalent to an inductance and arises from the inertia of the charge carriers. That this is of importance in the microwave range of frequencies has recently been demonstrated by Champlin *et al.*,[3] who have shown this effect to have a considerable influence on low temperature measurements at 24 Gc/s of conductivity and permittivity for both n-type silicon and p-type germanium.

### 4. Dielectric Relaxation Time

The dielectric relaxation time $\tau_d$ is usually defined in relation to the dispersion of a local excess of charge introduced into a dielectric medium and is the time constant $\varepsilon/\sigma$ in the decay formula

$$q = q_0 \exp(-\sigma/\varepsilon . t)$$

Although in devices of present interest the above situation is not strictly appropriate, the ratio $\varepsilon/\sigma$ is nevertheless of importance. For example, this determines the basic $Q$-factor at a given frequency, which is of interest for p-i-n diodes (see Sect. 8), since the smaller $\tau_d$ is, the greater is the intrinsic bandwidth. A low value of $\tau_d$ is also of importance in varactor diodes, as discussed below.

Consider a varactor having an asymmetrical abrupt p-n junction, heavily doped on one side and more lightly on the other. The depletion layer thickness is



World Radio History

then given by the formula

$$S = \frac{2\varepsilon}{en}[V_s - V]^{\frac{1}{2}} \qquad \text{.....(2)}$$

where $n$ is the carrier density on the more lightly doped side.

We shall determine the response of the depletion layer to an incremental voltage $\delta V$ which is suddenly applied across the device. If $w$ is the thickness of the semiconductor wafer, for a one-dimensional representation, an incremental field $\delta E$ due to $\delta V$ is initially set up throughout the wafer and is given by

$$\delta E = \frac{\delta V}{w}$$

The velocity with which the carriers at the edge of the depletion layer begin to respond to this field is

$$v_0 = \frac{\mu_e \delta V}{w}$$

Let $\delta S$ be the change in $S$ corresponding to $\delta V$ when equilibrium has been reached. The velocity will fall nearly exponentially to zero as this condition is approached and the time constant is obtained by assuming that the initial velocity is maintained, i.e.

$$t = \frac{w}{\mu_e} \cdot \frac{\delta S}{\delta V} \simeq \frac{w}{\mu_e} \cdot \frac{\mathrm{d}S}{\mathrm{d}V}$$

Hence, from eqn. (2)

$$t = \frac{w}{S} \frac{\varepsilon}{\mu_e en} = \frac{w}{S} \cdot \frac{\varepsilon}{\sigma} = \frac{w}{S} \cdot \tau_d$$

It should be noted that this same result can also be derived from the equivalent electrical circuit.

For silicon varactors $\varepsilon = 12 \times 8{\cdot}85 \times 10^{-12}$ farad/m and $\sigma \simeq 1000$ mho/m, so that $t \simeq 10^{-13}$ s for $S \simeq w$.

Thus, the varactor depletion layer can respond to a small increment of voltage, suddenly applied, in a time which is short compared to the period of microwaves.

It is interesting to note that with p-i-n devices the conductivity is much lower and the depletion layer boundary could not respond in a time short compared to microwave periods, although this is not necessary for their operation.

In applications such as parametric amplifiers and harmonic generators, a relatively large driving voltage is applied to a varactor and it is relevant to consider how rapidly the depletion layer can respond to such excitation. If the drift velocity corresponding to the time derivative of the depletion layer formula becomes of the same order as the thermal velocity of the charge carriers, the assumption of virtually instantaneous response will not be justified.

For the case of a varactor driven from a high-level sinusoidal voltage source the general solution is complicated when allowance is made for the resistance drop across the neutral bulk of the semiconductor behind the depletion layer. However, assuming that the amplitude of the driving voltage is made equal to the total bias voltage $V_s$, a rough solution may readily be obtained for the maximum velocity of the depletion layer boundary, by neglecting the resistance drop, and is given by

$$\left(\frac{\mathrm{d}S}{\mathrm{d}t}\right)_{\max} = \omega\left(\frac{\varepsilon V_s}{2en}\right)^{\frac{1}{2}}$$

For a typical high frequency varactor driven at 10 Gc/s, assuming $V_s \simeq 5$ volts, $S_0 \simeq 10^{-7}$ m, the above expression gives $(\mathrm{d}S/\mathrm{d}t)_{\max} \simeq 10^4$ m/s, which is less than $V_T$ by an order of magnitude.

### 5. Minority Carrier Lifetime Effects

The possibility of minority carrier lifetime effects is obviously to be considered in any discussion of frequency limitations, so far as p-n junction devices are concerned. However, of the devices of present interest, we note that crystal valves involve essentially single carrier transport mechanisms and tunnelling devices, depending on quantum mechanical processes, are inherently extremely fast and do not involve minority carriers. Varactors, too, only involve minority carrier action in so far as this contributes extra capacitance to that provided by the depletion layer and, although this may be of importance in some applications, the effect is still somewhat obscure and is outside the scope of this review.

The device which it is relevant to discuss here, which depends principally on minority carrier injection into a near-intrinsic region, is the p-i-n diode. However, as this device does not change its state at a microwave frequency our attention must be confined to the limit imposed by carrier lifetime on the modulation frequency.

In a p-i-n diode the transition from a high to a low impedance state is accomplished by the injection of carriers into the near-intrinsic region, which must not exceed the order of a diffusion length in thickness, whilst being sufficient to sustain the maximum inverse voltage if power handling capability is required. The formula for diffusion length, representing the distance over which the charge density falls to 1/e of its value at the injecting contact is[2]

$$L = \sqrt{(D\tau_p)}$$

where $D$ = electronic diffusion constant and $\tau_p$ = hole lifetime.

Taking $D = 0{\cdot}0065$ m$^2$/s for holes in silicon and $\tau_p = 4$ μs, the above formula gives $L \simeq 1{\cdot}6 \times 10^{-4}$ m —which represents a typical thickness for $d$, the width of the intrinsic layer. For satisfactory switching into the low impedance state, the switching time $\tau_s$ will



World Radio History

be of the order which is required for carriers to diffuse across the intrinsic layer, i.e. $\tau_s = \tau_p$.

If switching back into the high impedance state is accomplished simply by removing the biasing potential the diode impedance would be expected to return to its original high state in a time which is of the order of the hole lifetime, i.e. a few microseconds. The speed at which the high impedance state is recovered may be increased by applying reverse bias, $V_R$. One might then naïvely expect the time $t$ taken for holes to be swept right through the intrinsic layer to be given by

$$t = \frac{d^2}{\mu_e V_R}$$

which, for $V_R = 100$ V, is about $6 \times 10^{-9}$ s. The observed result is, however, very different and, even with a reverse voltage of this order, the time required to restore the high impedance state exceeds $10^{-7}$ s. The reason for this is probably that in most p-i-n diodes the electric field does not extend completely across the i-region, so that recombination is still an effective mechanism. This relatively long 'switch-off' time, together with the similar time required to switch the device to its low impedance condition, limits the upper modulation frequency to about 100 kc/s, dependent somewhat on the permissible distortion of the modulation waveform.

### 6. The Trapping of Charge Carriers

In p-n junction devices, direct recombination or generation of a hole-electron pair is a very unlikely event because of the large energy difference between valence and conduction bands and this process mainly takes place via fixed sites which are associated with impurity atoms or defects in the crystal lattice. Similar sites can trap charge carriers and then release an electron or hole after a time interval. This process can occur in single or double carrier systems. There is evidence with point contact crystals, particularly those employing silicon, that strong relaxation effects occur as a result of such a mechanism.

The lifetime of carriers may be expressed in terms of the probability $P$ per unit time for the ionization of an impurity atom for which a theoretically derived formula is[1]

$$P = 1{\cdot}5 \times 10^{11} \exp\frac{\Delta E}{kT}\ \mathrm{s}^{-1} \qquad \ldots\ldots(3)$$

Substituting into eqn. (3) the value for $\Delta E$ which corresponds to the depth of an acceptor site below the valence band in silicon (0·08 electron volt) gives $P \simeq 6{\cdot}6 \times 10^9\ \mathrm{s}^{-1}$. It may be shown that an ionization probability of this order will cause the capacitance of point contact crystals to decrease with frequency in the microwave range. This is indeed borne out by experience, the conversion loss of mixers and the sensitivity of detectors both falling less rapidly with frequency than would be expected in the absence of relaxation effects. It should be added that the capture cross-sections of impurities are now known to vary considerably in magnitude and eqn. (3) should therefore be regarded with some caution.

### 7. Transit Time

A crystal rectifier depends on the thermal diffusion of charge carriers across the barrier depletion layer and will only function if this time is short compared with the period of the signal frequency. To estimate the transit time $\tau_t$ the diffusion formula $\tau_t = d^2/6D$ may be used, taking the barrier thickness as 200 Å, a value estimated from capacitance and area measurements. The result is $\tau_t \simeq 10^{-14}$ s, which is small compared to all microwave periods and gives the crystal valve its unique property as a rectifier of microwaves, other devices with suitably low capacitance (e.g. space-charge-limited diodes) being unable to compete on transit time.

### 8. Basic $Q$-factor and Bandwidth

In p-i-n switching diodes the performance may be degraded at high frequencies by a restriction of a rather special character. This relates to the bandwidth over which a high switching ratio may be maintained, the limitation arising in the high impedance state.

The basic $Q$-factor for a semiconducting material is given by the usual formula $Q = \omega CR$ for a resistance and capacitance in parallel, where these relate to a unit cube of the material. Thus

$$Q = \frac{\varepsilon_r \rho}{60\lambda}$$

where $\rho$ = resistivity in ohm-metres

and $\lambda$ = wavelength in metres.

Considering an element of the semiconductor as a parallel circuit tuned by external inductance to resonance at the wavelength $\lambda$, the bandwidth over which the impedance exceeds $\sqrt{2}$ times its value at resonance is given approximately by [4]

$$B = \frac{c}{\lambda Q} = \frac{60c}{\varepsilon_r \rho}\ \mathrm{c/s}$$

where $c$ is the velocity of light in m/s.

If the resistivity for silicon p-i-n diode material is taken as 5 ohm-metres, for the unbiased state, and $\varepsilon_r = 12$, the following result is obtained:

$$Q = \frac{1{\cdot}0}{\lambda} \quad \text{and} \quad B = 300\ \mathrm{Mc/s}$$

Although the bandwidth as defined above is inversely proportional to resistivity it may still be possible to improve the frequency performance of a



World Radio History

### 2.1. *Assessment of Device Failure Data*

Although much information relating to an e.m. device may be available from the manufacturer, this information will often be in terms of a specific set of environmental conditions. If time permits, it is to the user's benefit to establish, either by observation of working equipment or by testing under the actual conditions encountered in field use, the various ways that failures occur and associated data such as the overall constant failure rate and when the onset of the wear-out period can be expected. Each faulty part should be carefully examined and the failure mechanism determined, for only by this means will it be clear whether the failure occurred because of high stresses, mechanical or electrical, resulting from its mode of use, or because the device failed to give its promised performance.

If life testing is used to obtain this information it is most important that the mechanical and electrical stresses encountered by the mechanism simulate very closely those which would be encountered in field use; if not, failures may be obtained which, because they would not occur in normal service, will unduly bias the subsequent test-result analysis, or vice versa.

Furthermore, although for most practical purposes it is necessary to accelerate life testing, it must not be assumed that this can be taken to the point of continuously cycling an e.m. device through its sequence of normal operation. For example, a ratio of 1 : 1 or even 2 : 1 should be arranged for the rest to operating times of sliding contacts such as a contact wiper rotating over banks of fixed contacts.

### 2.2. *Failure Data and Basic System Design*

When discussing reliability there are a number of terms used which are virtually self explanatory: mean life, mean-time-between-failures, mean-time-to-failure, and mean-time-to-first-failure. Which are the relevant terms for any particular system depends largely on the system's operating conditions, e.g. high reliability, short life, or high reliability, long life, and ease of access by maintenance staff. It is, therefore, important for the systems engineer to decide which are applicable at an early stage, the decision being based on the knowledge that either the system will not be maintained or it will be maintained by fault repair, and on the overall life requirements.

### 2.3. *Maintenance Organization*

Electromechanical devices are designed to function correctly as long as their mechanical adjustments remain within prescribed limits. While it is possible with electronic equipment to apply marginal testing to reveal component drift, it is very difficult and in many cases impossible to detect, except by visual inspection, that mechanical adjustments are reaching their permitted limits. With the exception of a few fully-sealed e.m. devices such as miniature relays, which are easily replaced when faulty, it is possible to bring many e.m. devices back into use by applying appropriate re-adjustments.

These factors, together with those considered in the preceding section, indicate clearly that for a system employing e.m. devices the maintenance organization set up will play a large part in establishing the overall reliability.

No hard and fast maintenance organization can be laid down which is applicable to all types of systems; each maintenance method adopted must take account of the particular problems and conditions associated with the equipment to be maintained. However, the following basic factors will materially affect the resulting system reliability.

(*a*) The training and experience of maintenance personnel.

(*b*) Conditions under which maintenance is carried out.

(*c*) The types and availability of test equipment and tools, and their calibration.

(*d*) Methods of recording maintenance work performed.

When few e.m. devices are used in a complex and large installation, they tend to be left to the tender mercies of anyone who happens to show any interest. Electromechanical devices may look relatively simple and easy to 'adjust' so that they are apparently operating correctly, but if long, frequent and reliable operation is required they must be treated with considerable respect: the adjustments, and cleaning and lubricating instructions must be strictly adhered to. This calls for adequate training, whether it be for the maintenance of comparatively simple devices such as relays or for 200-outlet 2-motion selectors with their many closely related adjustments, and for test equipment and tools to be supplied which are appropriate for the work to be carried out.

If concrete information about the reliability of the devices is to be obtained, strict control must be exercised over the recording of all maintenance work. Such paper work is not wasted: it will pinpoint weaknesses in the e.m. devices, or in their method of use, and will also yield data important in the selection of e.m. devices for future requirements.

While, as stated above, each maintenance organization must be virtually tailor-made for each system the following principles are offered for consideration; they have been applied very effectively to maintaining telephone switching-equipment reliability.



World Radio History

Readjustments of the moving parts of an e.m. device must disturb the relative positioning of the various mating surfaces; as soon as a device is returned to service a number of operations will be required before the surfaces bed down once more. It is found that a device is more liable to failure in the period of time immediately following a check and subsequent readjustment of a few simple adjustments.

Furthermore, due to the variation of telephone traffic with time, the switching equipment is connected in such a manner that some equipment carries more traffic than other similar equipment, while common-control equipment is usually very heavily worked.

These two factors are incompatible with a system of regular routine equipment overhauls. Basically, what is required is a procedure which gives attention to individual items as and when the need arises, and from these considerations the fault-history-card procedure was evolved. With this method each item or shelf of small items has a fault-history card on which all faults are recorded, the faults being classified as predictable or chance: predictable faults are those due to wear and tear, and are those which mainly govern the need for overhaul, while chance faults, such as a broken spring or dirty contact of an otherwise correctly adjusted spring-set, rarely have any bearing on the need for overhaul.

Following a predictable fault the mechanism is given a careful inspection to determine wear and tear, and the inspecting officer decides whether or not to overhaul. A chance fault calls for the minimum attention consistent with restoring the mechanism to service. Lubrication of the mechanism is, however, required at regular intervals and must be carried out with the least possible disturbance of parts. When lubrication is being carried out a decision must be made with respect to the accumulated dust and dirt in the mechanism—whether the disturbance created during its removal will cause more trouble than the long-term effect of leaving it in the mechanism.

**Table 1**

Rate of overhaul per 1000 items per year

| Item | 2-motion selectors | Uni-selectors | Average operations per year |
|---|---|---|---|
| Group and final selectors. | 4 | — | 22 000 |
| A-digit selectors (common equipment). | 44 | — | 265 000 |
| 1st code selectors. | 8 | — | 33 500 |
| Subscriber's uniselectors. | — | 3 | 2 700 |
| Director uniselectors (common equipment). | — | 31 | 420 000 |

The rate of overhaul of various types of e.m. switching mechanisms in telephone exchanges where this maintenance procedure has been in use for periods of up to nine years are given in Table 1. It should be noted that the basic mechanisms for a group, final, 1st code, or A-digit selector are the same, but the uniselector used in the director is a heavy-duty uniselector and cannot be directly compared with the subscriber's uniselector in the table.

As far as relays are concerned the above principles still apply, but it has been found by experience that, following a predictable fault on a relay, it is worth while to give all associated relays in the relay-set a general examination, particularly checking that spring follow, or block lift, is satisfactory.

### 2.4. *Electromechanical Switching Devices*

Electromechanical switching devices can be split broadly into two classes: those which can be used to choose a specific set of contacts from a considerable number of possible selections, e.g. uniselectors and 2-motion selectors; and those which, when operated, will always establish the same set of contacts or which, when released, may also always establish an alternative set of contacts, e.g. relays. The above general considerations apply equally to both categories.

## 3. Selectors or Rotary Switches

Selectors may be divided roughly into two classes: 2-motion selectors, in which a contact can be selected from a large number of contacts arranged in a matrix by operating the moving contact wiper along the appropriate coordinates of the matrix; and uniselectors, the function of which is to select one contact from several contacts arranged in a plane by simple rotation of the moving contact wiper. With both classes of device, multiple contacts can be established by arranging for several moving contact wipers to be moved together in parallel planes and for different sets of fixed contacts to be also arranged in parallel planes.

Before considering either class it should be pointed out that these mechanisms, particularly the uniselector type, can be electrically driven by a variety of control circuits. It is therefore important, if reliable operation over the full life of a device is to be obtained, that the control circuits used are those recommended by the manufacturer.

### 3.1. *Two-motion Selectors*

It is unlikely that semiconductor equipment designers will be particularly concerned with 2-motion selectors, which are widely used throughout the telecommunications field for switching telephone and telegraph traffic. Nevertheless, for completeness and



World Radio History

for comparative purposes the following information is provided.

The selector mechanism can be arranged for either the vertical or rotary mechanisms, or both, to be self-driven or driven by controlled pulses. To ensure reliable and consistent operation it is necessary for the energizing pulses to be of at least 30 ms each, with a break between pulses of at least 12 ms. If the driving action of the selector is to be stopped by recognition of a marking condition applied to one of the selector outputs it must be appreciated that the recognition period will be less than the 12 ms—it may be as short as 7 ms.

When a bank-outlet marking system is used to drive and stop the selector as it rotates around the bank, the moving wiper connected to the detecting circuit may be making and breaking electrical circuits. The current must be kept as low as possible, otherwise severe erosion of the bank contacts can occur, not only at the leading and trailing edges of the fixed bank contact but also along the track across the fixed contact due to wiper bounce and small particles deposited on the contact from the atmosphere. Whenever possible the circuit to be established through the sliding contacts of the mechanism should be switched through by relay contacts only after the wipers have come to rest on the required outlet.

In general, both the wipers and associated bank contacts of 2-motion selectors are of base metal : brass or nickel-silver bank contacts with nickel-silver wipers. As might be expected, these can lead to difficulties if low-voltage circuits are to be switched or if low-resistance connexions are required. Precious metal contacts can be used, but for the bank contacts this would be very expensive and plating might not last long; the wipers can be provided with precious metal tips. The problems associated with sliding contacts are very complex and it is a highly specialized subject. For those occasions when it is necessary to use a 2-motion e.m. switching device it is probably safest to use the manufacturers recommendations regarding its wiper-to-bank characteristics, but it is recommended that with base-metal contacts a minimum current of 5 mA should flow through the bank contact and wiper, since below this value the wiper to bank contact resistance may be unstable.

A 2-motion switching mechanism as used for telephone traffic switching would be expected to perform at least 100 000 operations without failure or need for major readjustments. No excessive wear of any part, excluding wipers, would be expected under $1 \times 10^6$ operations. Referring back to the figures in Table 1 it will be seen that for group, final, A-digit and 1st code selectors the average number of operations per overhaul ranges from $4 \times 10^6$ to $6 \times 10^6$.

### 3.2. *Uniselectors*

Uniselectors can be divided into two groups: those driven by a ratchet and pawl actuated by an electromagnet, and those driven by a small electric motor. Both types consist of three main sub-assemblies: the bank, the wiper assembly and the driving mechanism.

The wipers and banks are normally of nickel-silver but the banks can be plated with a noble metal if they are required for low-voltage applications. The wipers are arranged so that contact is established on both sides of each bank contact, giving improved contacting. Even so, at between 5 and 25 mA there may be up to 12 ohms maximum resistance from brush feed to bank contact, although in 99·9 per cent of the cases it will be less than 0·02 ohms. Below 5 mA the resistance will be unstable, and less than 5 mA is not recommended for base-metal contacts and wipers. There is also some wiper bounce which must be considered for some circuit arrangements, e.g. the non-bridging wipers of a motor uniselector bounce continuously while the wiper assembly is in motion, and bouncing can continue for up to 20 ms after wiper rotation has stopped—the bridging wipers are generally free from bounce as they move smoothly from contact to contact.

As already mentioned, the correct control circuits for these mechanisms are essential for reliable functioning. In general, for the ratchet-and-pawl mechanism using remotely-controlled driving pulses to the magnet, the pulses should be at least 30 ms long and the break between pulses at least 10 ms; this will give a speed of rotation of approximately 25 contacts/second. Under self-drive control, when the magnet is wired in series with the uniselector interrupter contacts, the wipers will rotate step by step at, typically, 60 contacts/second or 75 revolutions/minute. At this speed the wipers are stationary on each bank contact for approximately 10 ms.

For the motor-driven uniselector, one level, or arc, of bank contacts is used to determine the bank outlets at which the wipers will come to rest. A high-speed relay is used to control the cutting of the motor drive when the marked outlet is reached. As the speed of the motor-driven uniselector is between 170 and 230 contacts/second and the wiper system is locked in the required position with a mechanical latch, it is important that the locking is controlled by a correctly designed circuit using the high-speed relay developed for this purpose—the wiper, which must be a bridging one to limit bounce, sweeps over a contact in approximately 4 ms and only a small proportion of this time is available for the relay to function.

Assuming that lubrication is carried out as specified, the Post Office light-duty uniselector, including the wipers, would be expected to perform at least $1 \times 10^6$



World Radio History

half revolutions without failure or excessive mechanical wear; for a heavy-duty uniselector the figure would be $4 \times 10^6$ half revolutions, and for a miniature uniselector the figure would be $3 \times 10^6$ revolutions. These are, respectively, the equivalent of $25 \times 10^6$ steps, $100 \times 10^6$ steps and $108 \times 10^6$ steps.

The Post Office motor-driven uniselector, excluding the wipers, would be expected to perform $4 \times 10^6$ half revolutions without failure or excessive wear, and the speed of rotation would still be expected to lie in the range 170 to 230 contacts/second. The wipers and feeder brushes would be expected to perform at least $1 \times 10^6$ half revolutions.

For the ratchet-and-pawl type uniselectors it is essential that the interrupter contacts are fitted with spark quenches; normally CR quenches are used, as discussed later. The motor uniselector requires both the motor coils and the latch coil to be quenched; as the quenches used can influence the performance of the latching arrangement it is important to use the recommended values of C and R for the CR quenches.

### 4. Relays

Today there is a great variety of electromechanical relays available, covering a wide field of application. Some of them are highly specialized, but for many purposes a comparatively simple device is satisfactory: an electromagnet, an armature device for actuating the contact springs, and a contact spring-set. Many relays are manufactured so that a considerable choice of contacting arrangements is available for association with the basic electromagnet and armature assembly, but there is a much more limited choice of contact material.

From the point of view of the circuit designer it is principally the reliability of the mechanical portion and of the contacts themselves that is most important, with the emphasis on the latter. In view of the complex nature of the problems associated with electrical contacts they are discussed in a separate section below and some indication is given of methods by which some of the difficulties can be overcome and contact reliability improved.

Probably the best known relay in Great Britain is the Post Office 3000 type. This has been proved over many years of service under widely differing conditions. It can be provided for general-purpose switching, it can be fitted with appropriate slugs to provide slow-operating or slow-releasing features, it can be provided with special residual-screw arrangements to allow closely-controlled current operating characteristics. This relay is but one of many and each type presents its own problems if the best possible life with reliable contacting is to be achieved. It is not possible to do more here than to indicate one or two of the common difficulties and improvements to overcome them.

Although the mechanically-operated parts of a relay are few, there are several points at which wear can occur which will eventually cause contact failures. For example, many relays use a pin and stud arrangement for the link between the armature and the spring-set. Naturally, after long use wear occurs, leading to lack of spring follow when the relay is operated. This in turn leads to lack of contact pressure and, hence, in time to failure of the contacts.

In an effort to overcome this problem with the 3000 type relay the pin and stud arrangement was changed for a comb arrangement, the relay being known as the Post Office type 10 relay. This arrangement has given an increased mechanical life, of at least $100 \times 10^6$ operations, but, unfortunately, the contact bounce characteristics are in general inferior to those of their 3000 type equivalents due to the lack of the frictional forces present with the older arrangement which quickly damp out contact-spring oscillations. Furthermore, there is a tendency for armature rebound on release that, under certain light-load conditions, can lead to complete reoperation before the armature finally releases. The type 10 relay is therefore, limited to relays without slugs or timing requirements. It is used when the number of operations per year is expected to exceed 500 000, e.g. as storage relays in common control equipment.

Another redesign of special 3000 type relays used for pulsing applications has produced the Post Office type 19 relay. It is also a comb-operated relay in which a reverse-action type of spring-set operation has been used—this has greatly reduced contact bounce. It will give a much longer life than the particular 3000 type relays it will replace, and the time required for maintenance will be considerably reduced.

In the above, contact bounce has been referred to. It should be pointed out that contact bounce is often taken as synonymous with contact chatter but they are, in fact, two distinct characteristics. Contact bounce refers to one or more reopenings of contacts immediately after closure, usually following rebound of the armature or actuating device. Chatter is a more sustained and more rapid opening and closing of contacts caused by variations in the coil current and flux, or by mechanical vibration or both. Both bounce and chatter are equivalent to a substantially increased number of contact actuations, with the contacts opening and closing under abnormally severe conditions, as discussed later.

Every type of relay will have its own particular field of use, and to endeavour to give figures to indicate here the reliability that can be obtained with



World Radio History

each is impossible. However, the following will give some indication of what may be expected under normal circuit conditions. A large group of inter-connected 3000 type relays were arranged via their own contacts so that the failure of any one contact to be correctly established at the appropriate point in the circuit operation would bring the test to a halt and ring an alarm. It was found that with over 550 separate contacts involved, and with quenches provided where necessary, the number of contact operations per fault ranged from approximately $60 \times 10^6$ to $100 \times 10^6$, depending on the contact unit considered, i.e. make or break action in clean or dirty atmosphere. The major cause of contact failure was loss of block lift, due to both mechanical wear of the actuating mechanism and electrical wear of the contact points.

Before considering the difficulties associated with electrical contacts, it must be emphasized that the figures given by manufacturers for reliability of a specific relay will normally refer to a particular circuit application, and if the parameters are changed, apparently only slightly, a considerably different level of reliability may result.

## 5. Electrical Contacts

The physical characteristics of the contacts in any circuit should ensure that the load can be controlled reliably and economically throughout the working life of the equipment. Electrical loads on contacts can vary widely and, although there is some restriction to the variety of contacts available, switching devices usually cater for a range of load conditions; some devices offer a choice of contact materials to extend the loading capacities.

Ideally, a complete specification of the contact rating of a switching device should cover not only voltage, current and power, but also other factors such as the characteristics of suitable loads, the estimated life of the contacts under various load conditions, and, as far as they are known, any limitations due to operating features of the device itself. In practice, current load ratings are usually quoted in a generalized and abbreviated form, making it difficult to predict the performance accurately where there is a wide variation in circuit applications.

Ratings are often quoted without reference to the corresponding life which may be expected from the contacts and, conversely, contact-life figures are often given in isolation although a particular maximum load is probably implied. Load ratings for opening contacts are almost invariably quoted with reference to non-inductive loads. A maximum permissible current may be specified for a particular voltage or, alternatively, a permissible wattage loading may be given. Usually a note is added to the non-inductive or resistive load ratings recommending a suitable quench to be used for inductive loads: this gets round the problem of the great variety of factors involved. It must not, however, be assumed that with a suitably quenched inductive load the contact life will be that given for the non-inductive load—contact life is generally appreciably reduced.

In all types of e.m. switching devices the closing and opening of electrical circuits is involved, and there is a very wide variation of the contacting conditions. The contact points may touch momentarily or they may be semi-permanent; contact may be established under very light or very heavy pressures; the contact points may move in a direction normal to their surface to make and break the circuit, or they may slide over each other for short or long periods of time. Their study, therefore, involves aspects of mechanical contacting, microgeometry of surfaces, hardness, plastic and elastic deformation, friction, cold welding and sliding.

The contact area can be considered as nominally flat with a topography such that the actual areas in contact are very small compared with the total area; this leads to plastic and elastic deformation. Electrical erosion of the contact points takes place as a result of the sequence of opening and closing the contact points. In general, two physical processes are involved: (i) electrical discharges between the contact points, and (ii) formation and rupture of metallic bridges between the contact points.

Knowledge of many of the theoretical aspects of contact switching phenomena is still far from complete, but there is a fair degree of agreement over some practical aspects. One of the points often overlooked by design engineers is the importance of the phenomena associated with closing contacts: one naturally tends to think more often than not in terms of opening contacts interrupting appreciable currents through inductive circuit components. Admittedly, action taken to deal with the consequences of opening contacts often eliminates closure effects, but it is, nevertheless, as well to consider the two conditions separately.

Investigations of closing contacts show that as the contact points come very close together the electrical field may be sufficient to cause a cold metallic discharge to bridge the gap, prematurely completing the line circuit. This metallic bridge is usually ruptured by the consequent premature line current before contact closure is complete—hence the name 'pre-closure' for the effect. In practice, there may be time for a number of bridgings and ruptures to occur, each causing contact damage. Naturally, the greater the contact potential and the larger the capacitance of



World Radio History

**Table 2**
Comparison of the four principal contact materials

| Metal | Max. circuit current in amps (non-inductive load at 50 V) | Melting point °C | Boiling point °C | Remarks |
|---|---|---|---|---|
| Silver (Ag) | 0·3 | 961 | 1950 | General purpose material for low-current light-duty work. Relatively cheap. High electrical and thermal conductivity. Its low m.p. and b.p. make it prone to damage and errosion if rated current and voltage exceeded, especially when frequency of operations is high. |
| Platinum (Pt) | 1·0 | 1769 | 3910 | Hard. Has high m.p. and b.p. and better minimum arc-current limits than Ag. Better for heavier loads and higher frequency of operations. |
| Palladium (Pd) | 1·0 | 1552 | 2540 | Better than silver for hardness, m.p. and b.p., and minimum arc-current limits. Is not as good as Pt but is much cheaper. |
| Tungsten (W) | 3·0 | 3400 | 5400 | Very hard; has high m.p. and b.p., and very good minimum arc-current and voltage limits. Suitable for heavier loads and high frequency of operations. Requires heavy contact pressure and rubbing action to break surface film. |

the wiring connected to the battery-connected contact point, the worse the effects; any contact bounce or chatter present will further aggravate the situation.

For opening contacts the most severe form of contact damage arises from breaking currents above the minimum arc-current (see Sect. 6.5). Normally, the circuit current should be kept well below the minimum arc-current value; as this value refers to a circuit with a non-inductive load, a much greater margin must be given between the circuit current and minimum arc-current value when the load is inductive. In general terms, as the contacts open, contact is maintained at an extremely small area and the current density becomes very high. The temperature is raised above the metal's melting point, producing a molten metal bridge which is finally ruptured either mechanically or by boiling. The mode of electrical breakdown which follows the bridge rupture is highly complex, but for practical purposes it should be remembered that it can result in many sparks and arcs—often hundreds—being formed during each opening, with surge-current and voltage peaks which may exceed 10 amperes and 1000 or more volts, respectively.

In this country four main types of relay contact material are used: silver, platinum, palladium and tungsten. Some of their more important characteristics from a contacting point of view are given in Table 2. Atmospheric environments obviously will affect the performance of these materials, and should not be overlooked. To a lesser extent the contact-point shape, the contact spring and the method of actuation will also affect the reliability of the device, but are probably outside the sphere of electronic equipment designers. However, it should be pointed out that clean, smooth, contacts are less liable to sparking and arcing than those with irregular and roughened surfaces—thus, once damage to contacts has started it tends to accelerate with use.

## 6. Quenching of Electrical Contacts

Due to the nature of contact switching phenomena the selection of suitable quenches is very largely an empirical process. One of the difficulties encountered in literature on the subject is the general terminology describing contact switching, arcing and sparking, and the principal terms used are discussed in the following sections.

### 6.1. *Quench*

A quench is a component, or components, used to dissipate energy that might damage contacts or cause other difficulties during switching, and is taken to include devices for the suppression of surges and interference.

### 6.2. *Spark*

A spark is a discharge current of a transient nature across the air-gap between contacts, caused by a high voltage developed across the load controlled by the contact. The spark current is relatively small but nevertheless causes heating and erosion of the contact material. Conduction across the gap is by ionization of the air and vaporization of the contact metal. It can normally be assumed that spark breakdown will not occur unless the gap voltage exceeds 300 volts, the approximate minimum breakdown voltage of air. Spark extinction occurs when the gap voltage is insufficient to maintain the discharge current.



World Radio History

6.3. *Arcing*

An arc is a relatively large discharge current across the air-gap between contacts. For the duration of the arc the voltage stabilizes at approximately 15–20 volts. Heating of the contact material is greater than with sparking, and conduction in the gap is largely due to vaporized metal.

6.4. *Glow*

A characteristic violet-blue glow, sometimes termed a glow spark or discharge, may occur at contacts interrupting an inductive load circuit. Glow discharge in air requires a minimum maintaining voltage of approximately 300 volts.

6.5. *Minimum Arc-Current*

Each contact metal has, for a given applied voltage, a characteristic value of current which if exceeded will cause arcing to occur when contacts of that material open an electrical circuit. This value, termed the minimum arc-current, should not normally be exceeded at the working voltage.

A quench will usually be provided specifically to avoid contact damage, or surge-voltage breakdowns, or interference, but generally it will be of benefit for all three because of the close relationship between them. The basic aim will be to limit the voltage across opening contacts to less than 300 volts, the minimum spark breakdown voltage of air, and this will prevent all discharges except those due to re-closure effects at very minute contact openings. For some methods of connecting the quench this 300 volts will include a standing contact potential, such as the battery voltage.

A variety of quenches is available. The more usual ones are the capacitor plus series-resistor (CR) quench, the non-linear resistor (NL) quench, shunt-resistor quench, and the metal-rectifier or diode quench, the first two being the more popular.

6.6. *Capacitor-Resistor Quench*

The CR quench provides, in effect, a bypass for momentary continuation of the circuit current after it has been interrupted at the opening contact. The quenching action is essentially due to the capacitor: as far as opening contacts are concerned the resistor is not only unnecessary but, in fact, reduces the voltage-limiting action of the capacitor; a resistor is, however, essential on contact closure to limit the instantaneous current due to the capacitor charging or discharging via the contacts—otherwise welding might occur. The values of C and R for the quench may considerably influence the release time of an armature actuated by the load coil: the release time may be reduced or increased, or the armature be re-operated to a reverse oscillatory surge.

The factors involved in contact switching are very complex, and circuit analysis to determine the component C and R values still requires somewhat arbitrary assumptions. In practice, the C and R values are generally not too critical and may be determined by the following simple rules, but when conditions are particularly complex it is as well to carry out comparative tests using a range of C and R values based on those determined by the simple rules.

For inductive loads, provided the contact opening speed is not abnormally slow and is reasonably free from chatter, the current value in amps disconnected at the contacts may be taken as the minimum value of C in microfarads for satisfactory quenching, while twice this value is recommended to allow a margin for variation of working conditions. The corresponding value of R can generally be taken as approximately equal to the resistance of the load coil.

The CR quench may be connected across the coil or across the contact. Theoretically, the method chosen should not affect its performance, but in practice there are a number of reasons for connecting it across the contacts, and also for it to be arranged so that, on restoration of the circuit to its unoperated condition, a capacitor discharge path is provided.

For non-inductive loads the value of R will be less than for equivalent inductive loads, but too small a value will increase the risk of damage from capacitor discharge currents on closure of contacts.

6.7. *Non-Linear Resistor Quench*

The non-linear resistor has a marked non-linearity of voltage/current characteristic. They are formed from silicon-carbide, which has a rapidly decreasing resistance with increase of voltage and is therefore very suitable for suppression of surge voltages. In general, the NL resistor quench is less efficient than the CR quench—it will give approximately half the working life for a contact under comparable working conditions. Nevertheless, for general-purpose, light-load applications it usually extends contact life to that of the life of those parts subject to mechanical wear. NL resistor quenches will extend armature-release time and cannot be used where critical timing is required, but there is no possibility of a reverse surge as with CR quenches.

At the normal working voltage the NL resistor will pass a leak current, and it is, therefore, usually connected across the load coil, where its small size makes direct connexion across the coil very easy. However, because of their negative temperature coefficient, their use in places with high ambient temperatures should be avoided.



World Radio History

6.8. *Shunt-Resistor Quench*

The shunt-resistor quench is simple, stable and reliable, with a performance that can be accurately predicted. However, when the associated coil is energized the shunt-resistor quench also passes a relatively high current, the resistor current increases the load current for the contact, and the armature release time is considerably increased by the slugging effect of the resistor. This last disadvantage can be turned to good effect when a closely-controlled delayed contact opening is required, thereby precluding the use of other quenching techniques, since the armature release time can be fairly accurately calculated.

6.9. *Metal-Rectifier or Diode Quench*

A metal-rectifier or semiconductor diode quench is simple and robust, giving a performance comparable with that of the shunt-resistor quench, but without the additional load current for the contact. Armature release time is, however, considerably increased. The range of metal-rectifier and diode characteristics is wide and the actual device to be used in any situation must be carefully chosen. Diodes are generally more easy to use, and their characteristics do not vary as those of metal rectifiers do with age and with conditions of use.

## 7. Circuit Design Considerations

It will be apparent from the above that careful attention must be paid to the suitability of each e.m. switching device so that the advantages of using a specific device are not offset by some other consideration. The following examples illustrate the difficulties.

High-speed relays have platinum contacts, speedy actuation and small pulse distortion. It would thus appear that these contacts may be used to pulse onerous loads. In fact, under these conditions contact damage is likely as the moving contacts are light and subject to bounce, particularly on restoration. The resultant damage and metal transfer may form spikes and, as the gaps are small, the contacts may be short-circuited.

Similarly, a load well within the capabilities of an opening relay contact may cause severe damage when disconnected by a moving wiper. Contact erosion may also be increased by wiper bounce and the presence of abrasive particles. As damage to a bank contact will require expensive replacement of a complete bank assembly, a critical check should be made of any element in which it is proposed to disconnect load currents, of any magnitude whatsoever, by movement of a wiper. Where heavily-worked equipment or heavy electrical loads are involved, this method of control should be avoided; for all other applications a quench should be provided unless the number of operations is very small.

The distribution of common pulses generated by relays should be carefully watched: the loading of the relays can be inadvertently increased beyond the point at which the insertion of intermediate distribution relays is necessary to prevent contact damage.

When the voltage of a battery-connected contact exceeds 50 volts some restriction should be placed on the number of bank multiples connected in parallel to the contact to ensure that pre-closure effects are reduced to a minimum.

Circuit logic should be arranged so that there are no ineffective operations of loaded contacts; if these cannot be avoided it must be borne in mind when assessing the number of operations of the contacts to establish their expected life.

Slow actuation of contacts leads to increased contact damage as well as increasing the risk of contact chatter. Apart from providing a spark quench, keep the contact load current small, and, if it is necessary to switch a heavy load, insert a quick-acting device controlled by the slowly-actuated contacts. Do not overlook the possibility that a device thought to be reasonably quick in operation has been made slow-acting unintentionally by some feature of associated circuit elements in related parts of the general circuit logic.

While quenches will more often than not be fitted to e.m. switching devices to improve contact life and reliability, where such devices are associated with electronic equipment there will be many points at which the use of quenches must be considered if protection is to be given from excessive surge voltages. This also applies to interference which is likely to affect amplifiers, electronic toggle devices, sensitive pulse-detecting circuits, etc.

In many organizations the circuit designer may not be the person responsible for the selection of the specific e.m. device to be used. It is, therefore, essential for the circuit designer to estimate the expected number of operations required of any electrical contacts associated with the e.m. device and to indicate the expected loading conditions.

The actual choice of contact material will be limited for any particular device, but Table 3 gives examples of some standard Post Office devices and the choice of contact material available.

While all precautions may be taken during circuit and equipment design stages it is still likely that unsuitable choices have been made both from a mechanical and an electrical point of view. Mechanical wear and tear will only begin to show up after a period of time and, as stressed in Section 2.3, this



World Radio History

**Table 3**

Device application and contact materials available

| Device | Circuit application and design details | Contact materials available |
|---|---|---|
| Relay, type 10 | Long mechanical life. Local pulse distribution, etc. | Ag: general purpose material; current maximum 0·3 A.<br>Pt: heavily-worked circuits; heavy loads; current maximum 1 A. |
| Relay, type 12 | Subscribers' line circuits; lightly loaded and lightly worked. | Ag: general purpose material. |
| Relay, type 17 | Switching heavy loads at mains voltage (250 V). | Ag-Ni: extends load current ranges 2·5 A d.c., 5·0 A a.c. |
| Relay, type 19 | Long mechanical life. Specifically designed for responding to line pulse signals. | Ag: general purpose material.<br>Pt: heavy electrical loads, such as selector magnets. |
| Relay, type 3000 | General purpose with a wide range of applications; light and heavy loads; lightly and heavily worked. | Ag: general purpose material, current maximum 0·3 A.<br>Pd: increased reliability; heavier loads with current maximum 1 A. |
| Relay, high-speed | Critical circuit elements; only one contact point; small contact rubbing action. | Pt: increased reliability. |
| Uniselector, type 2 | Interrupter contacts. | W: general purpose material; heavy duty.<br>Pt: increased reliability and less onerous applications. |
| Uniselector, type 4 | Interrupter contacts: small rubbing action. | Pt: to increase reliability. |

is when the maintenance organization chosen is most important. Any errors with the electrical aspects, and particularly with the choice of contact arrangements and materials, can usually be found earlier on by a careful visual check of switching contacts. When making this check, which cannot replace careful assessment during the design stage, it is well to bear in mind the following points.

(*a*) Damage is not directly related to the size of the visible discharge between contacts. Pronounced blue-glow discharge may cause relatively slight damage while a hardly noticeable white-hot arc may cause severe damage.

**Table 4**

Identification of common forms of contact damage

| Form of contact damage | Probable cause |
|---|---|
| Severe erosion and burning. | Excessive current. |
| Severe erosion without burning. | Excessive number of operations. |
| Welding | Capacitive loads.<br>Incorrect CR quench.<br>Line discharges.<br>Low-resistance loads.<br>Slugged load-coils.<br>(Each of these conditions is appreciably worsened by contact bounce.) |
| Spike and crater tending to interlock. | Contact chatter, usually during slow release. |
| Spike, whisker or rod of contact material short-circuiting contacts (more likely with Pt). | Repetition rate of contact operation and/or load too onerous for the switching device. |

(*b*) Welding of contacts usually occurs without visible sparking or arcing.

(*c*) Contacts can suffer severe short-term erosion from oscillatory discharges in the ultra-violet light or r.f. range.

(*d*) Contact damage can often occur in a random manner, depending on complex physical factors.

(*e*) By careful examination of contact surfaces it is possible to detect at an early stage the formation of spikes and craters liable to cause interlocking faults.

Certain general forms of contact damage are usually associated with particular switching conditions and recognition of these may aid investigations. Some of the more common forms of damage are set out in Table 4.

## 8. Acknowledgment

Acknowledgment is made to the Engineer-in-Chief of the General Post Office for permission to publish this paper.

## 9. Bibliography

1. S. Rudeforth and F. V. Sanders, "The use of fault-history records for the maintenance of telephone switching mechanisms", *P.O. Elect. Engrs J.*, **58**, p. 19, April 1965.
2. J. S. Young, "The Post Office 2000 type selector: its development and mechanical details", *P.O.E.E. J.*, **28**, p. 249, January 1936.
3. R. F. Purves and A. Walker, "The 4000-type selector", *P.O.E.E. J.*, **51**, p. 168, October 1958.
4. J. O. Thompson, "The new Post Office standard uniselector", *P.O.E.E. J.*, **42**, p. 17, April 1949.
5. F. Haythornthwaite and D. J. Manning, "Improvements to the Post Office type 2 uniselector", *P.O.E.E. J.*, **57**, p. 156, October 1964.



World Radio History

6. D. J. Manning, "The Post Office type 4 uniselector", *P.O.E.E. J.*, **52**, p. 215, October 1959.
7. C. A. May, "The Post Office standard motor uniselector", *P.O.E.E. J.*, **46**, p. 79, July 1953.
8. R. Barker, "Design of relays for automatic telephone equipment circuits, with special reference to relay type 3000", *P.O.E.E. J.*, **26**, p. 15, April 1933.
9. C. W. Clack, "The Post Office 600 type relay", *P.O.E.E. J.*, **28**, p. 293, January 1936.
10. B. H. E. Rogers, "The Post Office type 10 relay", *P.O.E.E. J.*, **51**, p. 14, April 1958.
11. A. Knaggs, "A new pulsing relay—Post Office relay type 19", *P.O.E.E. J.*, **57**, p. 190, October 1964.
12. H. A. Turner and B. Scott, "A polarized relay of improved performance", *P.O.E.E. J.*, **43**, p. 85, July 1950.
13. G. Windred, "Electrical Contacts" (MacMillan, London, 1940).
14. "Electrical Contacts" (Johnson, Matthey, London, 1960).
15. R. L. Peek and H. N. Wagar, "Switching Relay Design" (Van Nostrand, New York, 1955).
16. A. Fairweather and J. Ingham, "Subsidence transients in circuits containing a non-linear resistor, with reference to the problem of spark-quenching", *J. Instn Elect. Engrs*, **88**, Part 1, p. 330, 1941.
17. F. Ashworth, W. Needham and R. W. Sillars, "Silicon carbide non-ohmic resistors", *J. Instn Elect. Engrs*, **93**, Part 1, p. 385, September 1946, and **93**, Part 1, p. 595, December 1946.
18. A. Fairweather, F. Lazenby and A. E. Parker, "Long-life, low-voltage contacts", *Proc. Instn Elect. Engrs*, **109**, Pt. B, Supplement No. 22, pp. 567–586, 1962. (I.E.E. Paper No. 3789E.)
19. A. Fairweather and E. J. Frost, "The design and testing of semi-permanent metallic contacts for use at low-voltages", *P.O.E.E. J.*, **53**, p. 26, April 1960.
20. H. W. Hermance and T. F. Egan, "Organic deposits on precious metal contacts", *Bell Syst. Tech. J.*, **37**, p. 739, 1958.


*Manuscript first received by the Institution on 14th May, 1965 and in final form on 22nd October, 1965 (Paper No. 1026.)*

© The Institution of Electronic and Radio Engineers, 1966

# I.E.R.E. GRADUATESHIP EXAMINATION, NOVEMBER 1965—PASS LISTS

The following candidates who sat the November 1965 examination at centres outside Great Britain and Ireland succeeded in the sections indicated. The examination, which was conducted at 75 centres throughout the world, attracted entries from 404 candidates. Of these 176 sat the examination at centres in Great Britain and Ireland and 228 sat the examination at centres overseas. The names of successful candidates resident in Great Britain and Ireland will be published in the January–February issue of the *Proceedings* of the I.E.R.E.

| | *Candidates appearing* | *Pass* | *Fail* | *Refer* |
|---|---|---|---|---|
| *Section A* | | | | |
| Great Britain | 108 | 55 | 44 | 9 |
| Overseas | 143 | 44 | 88 | 11 |
| *Section B* | | | | |
| Great Britain | 68 | 19 | 39 | 10 |
| Overseas | 85 | 20 | 54 | 11 |

## OVERSEAS

**The following candidates have now completed the Graduateship Examination and thus qualify for transfer or election to Graduate or a higher grade of membership.**

BHASKARAN, K. P. (S), *Calcutta*
CURTIS, T. D. W. (S), *Ottawa*
GOPALAKRISHNA, S., *Hyderabad*
KRISHNASWAMY, C. (S), *Calcutta*
LEE, Y. S. (S), *Hong Kong*
LEES, F. P. (S), *Salisbury*
LO LOKE YEE (S), *Singapore*
MOUSSAVI, A., *Teheran*
NARANG, I. K., *Dehradun, India*
NAYAK, N. M., *Stuttgart*
PANT, L. M., *Lucknow*
PILLAI, R. (S), *Bangalore*
PILLAY, R. (S), *Hyderabad*
RAJAPAKSE, K., *Colombo*
SHAHI, B. L. (S), *Shillong*
SHARMA, C. R., *Dehradun*
SOONAWALA, N. M. (S), *Delhi*
TOH PENG KUAN, *Singapore*
TRINATHARAO, P. (S), *Bangalore*
WHITELEY, R. C. (S), *Oranjemond—S.W.A.*

**The following candidates have now satisfied the requirements of Section A of the Graduateship Examination.**

ALEGE, R. A. (S), *Lagos*
AMBRE, A. B., *Bombay*
AROOZOO, E. P. (S), *Kuala Lumpur*
BELL, G. (S), *Hong Kong*
BHAKAR, A. S. (S), I.N.S. *Investigator, India*
BOERSEMA, D., *Delft, Holland*
CHAN SUN PO, *Hong Kong*
CHHIKARA, R. S. (S), *Delhi*
DAHIYA, D. S., *Delhi*
D'CRUZ, W. V. (S), *Bombay*
EADINGTON, J., *Wellington, New Zealand*
ELESHIN, K. (S), *Lagos*
FERNANDO, T. M., *Madras*
GADEPALLI, V. N. M. (S), *Hyderabad*
GUNDAPPA, T. R. (S), *Bangalore*
IGHARAO, M. O. U. (S), *Lagos*
KATOCH, D. S. (S), *Bombay*
KOHLI, M. S., *Delhi*
KRISHNA SARMA K., *Colombo*
KRISHNAN, D. (S), *Calcutta*
LIM SHO SHIAW (S), *Singapore*
MAK, A. (S), *Delft, Holland*
MUKERJEE, S. M., *Delhi*
MULTANI, N. S. (S), *Lucknow*
NDIBE, F. O. (S), *Lagos*
NUSS, A. (S), *Tel-Aviv*
OBI, L. M. (S), *Lagos*
OKONYIA, A. (S), *Lagos*
PAGE, M. G. (S), *Bombay*
PAMCHAPAKESEN, S. (S), *Bombay*
RANDEV, S. K. (S), *Delhi*
RENGASWAMI, R. *Madras*
SAGOO, A. S. (S), *Nairobi*
SAMEL, P. Y., *Bombay*
SANKARANARAYANA, A., *Delhi*
SANKARANARAYANAN, K. V. (S), *Delhi*
SINGH, GURMIT (S), *Delhi*
SUGATHADASA, A. G. (S), *Colombo*
TEO SOON LYE, (S), *Singapore*
THIRUNAVUKKARASU, S. (S), *Colombo*
WEINTRAUB, J., *Tel-Aviv*
WHITESIDE, R., *R.A.F. Muharraq*
WONG SOH SOON (S), *Singapore*
WONG WING-CHEUNG, *Hong Kong*

(S) Denotes a Registered Student.



World Radio History

U.D.C. 621.37/9.004.6

# The Reliability of Electronic Systems

*By*

J. C. CLULEY, M.Sc., C.Eng.†

*Originally presented at a Symposium on 'Engineering for Reliability in the Design of Semiconductor Equipment' held at Hatfield College of Technology on 13th–14th May 1965 under the aegis of the I.E.E. and the I.E.R.E.*

**Summary:** Starting from the agreed definition of reliability the use of probability theory enables the design engineer to predict the reliability of a complex system as a function of the reliability of its component parts. The characteristics of series and parallel grouping are given, together with some applications of hammock networks. The exponential survival law is derived, and the conditions under which it may be applied are discussed.

The use of redundancy to improve reliability is introduced, and examples are given of the effect of various schemes upon the reliability of a small electronic system, with and without maintenance. Some possible forms of passive redundancy are described, including block redundancy with majority voting, and quadded logic.

## 1. Introduction

The increased use of electronic equipment in critical situations where failure can cause loss of life, damage to expensive plant or equipment, or involve very costly replacement procedures, has directed much attention to the problem of the reliability of electronic apparatus. Typical of these situations are the electronic navigational and control systems of civil aircraft, electronic data-handling systems for air traffic control, on-line process controllers and satellite communications.

In all of these systems, cost and reliability are both design parameters, but their relative importance depends upon the application. In the case of satellite communications, or underwater telephone repeaters, there is no direct hazard to life and the main consideration is an economic one: to provide the required communication channels of prescribed performance for a minimum total cost over 20 years or so. In both of these examples the cost of replacing faulty equipment is very high due to the physical difficulty of launching another satellite, or locating and recovering a faulty repeater from the ocean bed. The actual cost of the faulty equipment is relatively insignificant. On top of this is the additional cost caused by loss of earnings while the channels are out of service.

Thus in both of these applications, high reliability is necessary to provide an economically competitive system. In other cases a minimum reliability may be demanded by the user, or a licensing authority. An example of this is the use of automatic blind landing equipment in civil aircraft. Here the Air Registration Board has laid down the maximum probability of a serious incident during the automatic phase of the landing, and the designers' task is then to satisfy this standard of reliability under the constraints of cost, size, weight, etc.

In most of the applications mentioned above, some degree of redundancy is required to achieve an adequate level of reliability, but it is nevertheless useful to start the design by considering a non-redundant or basic system. This is one in which no attempt has been made to duplicate any components or sub-systems, and for the purposes of computation we assume that any component fault will cause a system failure. The designer needs first of all to assess the reliability of the basic system before investigating how this may be improved by various means.

The reliability of a system is generally defined as the probability that it will operate within defined performance limits for a specified time under specified environmental conditions. Thus the reliabilities of various components of a system may be combined to give an overall performance figure in accordance with probability theory.

## 2. The Poisson Distribution and the Exponential Law

In the basic system, where a fault in any element will cause a system failure, the elements are said to be functionally in series, and the probability of the system working decreases as the number of components rises.

In terms of reliability, such a system is always less reliable than its least reliable part. If the reliability of each part is known, the system reliability can be computed by the product rule, so that if the reliabilities of the parts are $R_1$, $R_2$, $R_3$, etc. the system reliability is $R = R_1.R_2.R_3.R_4$ etc., and for an assembly of $n$ equally reliable parts of reliability $R_1$

† Department of Electronic and Electrical Engineering, University of Birmingham.



World Radio History

the system reliability is

$$R = R_1{}^n$$

However, the data available to the designer are generally not in the form of reliability figures but of the failure rates of various types of component, or for sub-systems, perhaps its reciprocal, the mean time between failures. These figures can be converted into reliability figures by using the Poisson probability distribution,[7] which gives the probability of a certain number of events occurring in terms of the number of events which are expected to occur as a result of previous experience. This is the situation in an electronic system, the event being the failure of a component during a certain period of test or operation. As a result of previous tests, or from data accumulated in the field, the average number of failures will be known, and the Poisson distribution may then be used to compute the probability that 0, 1, 2, etc. components will fail. For reliability studies we are generally concerned only with the probability of zero failures, i.e. fault-free operation.

If the probability of failure of any one of the $n$ similar components of a system during a prescribed interval is $p_f$, the expected number of failures is $a = np_f$.

Thus the Poisson series:

$$e^{-a}+ae^{-a}+\frac{a^2e^{-a}}{2!}+\frac{a^3e^{-a}}{3!}+-+\frac{a^me^{-a}}{m!}+-$$

contains term by term the probabilities:

$p_0 = e^{-a}$ = probability of 0 failures

$p_1 = ae^{-a}$ = probability of 1 failure

$p_2 = \frac{a^2e^{-a}}{2!}$ = probability of 2 failures

$p_m = \frac{a^me^{-a}}{m!}$ = probability of $m$ failures

The series must terminate at $m = n$, since this embraces all the possible number of failures, and must sum to 1. We are generally concerned only with the first term, since this gives the *reliability* of the unit, that is the probability that it will function without failure. We thus arrive at the Exponential Survival Law giving the reliability $R$ in terms of the expected number of failures $a$ as

$$R = e^{-a}$$

If the data are in the form of a *failure rate*, or failure probability per component per hour $p$, and the test lasts for $T$ hours, the expected number of failures per hour is $pn$, and during the entire test is

$$a = pnT.$$

Thus the reliability of $R = e^{-pnT}$.

It is sometimes convenient to express this in terms of the mean time between failures (m.t.b.f.) of the unit. Since we expect $pn$ failures per hour, the m.t.b.f. is the reciprocal of this, $M = 1/pn$, and we have

$$R = e^{-T/M}$$

Figure 1 shows a graph of reliability in terms of the ratio $x = T/M$ It is clear from this that in order to achieve a high degree of reliability the operating period must be much less than the mean time between failures.

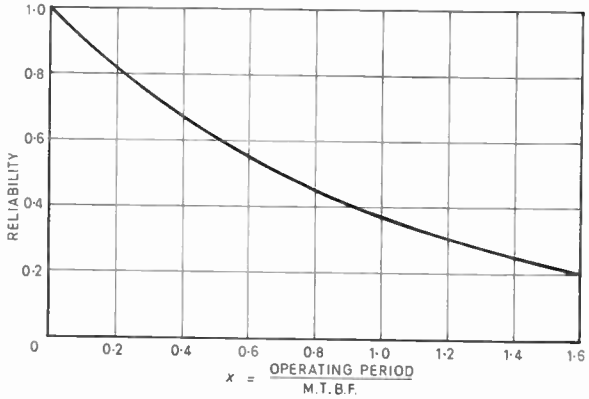



**Fig. 1.** Exponential failure law.

The value of $x$ is then small, say less than 1/10, and the first two terms of the exponential series afford a good approximation to the sum of the series.

The reliability may then be written as

$$R = 1-x$$
$$= 1-T/M$$

This may be seen intuitively since $T/M$ is the probability of a failure occurring during the interval $T$.

In the more general case of a system with various classes of components, for example $n_1$ each having a failure rate $p_1$, $n_2$ each having a failure rate $p_2$, etc., the total number of failures per hour is now

$$N = \sum p_1n_1$$

and the m.t.b.f. becomes

$$M = \frac{1}{N} = \frac{1}{\sum p_1n_1}$$

It may be necessary to subdivide one class of component into several sub-classes, each with its appropriate failure rate, if the operating conditions of the various sub-classes are significantly different. For example transistors used near their maximum electrical and thermal ratings may have a considerably greater failure rate than those used in low-power logic circuits.



World Radio History

If we consider the case of a satellite containing say 1000 transistors, with a failure rate of 0·02% per 1000 hours, or $2\times10^{-7}$ per hour, the total failure rate for the system is

$$10^3\times2\times10^{-7} = 2\times10^{-4} \text{ per hour}$$

Thus the m.t.b.f. is $\frac{1}{2\times10^{-4}} = 5000$ hours, if failure in other components are neglected. The reliability over 8000 hours, or nearly a year is thus

$$R = \exp\left(-\frac{8000}{5000}\right)$$
$$= 0{\cdot}202$$

Thus there is only a 20·2% probability that the system will operate successfully for a year. We may expect some reduction in the transistor failure rate assumed above if particularly reliable transistors are used well within their ratings, but if the other components are included in the calculation (perhaps 5000–8000 in all), the m.t.b.f. cannot be increased very much.

Since we would require a reliability of better than 80% over five years to make the satellite profitable, it is clear that this particular non-redundant system is not likely to be adequate.

### 3. Variation of Failure Rate with Time

The above calculations are based upon a failure rate which does not vary with time. This is a reasonable assumption for most electronic components, which do not have any clearly-defined wear-out mechanism, and life-tests on semiconductor devices generally show a roughly constant failure rate. However, there are exceptions to this; if for example a batch of transistors having a small leak in the sealing can were used, these could fail due to excessive collector leakage current after a relatively short working life, and the failure rate for this batch would then have a pronounced peak.

Fault records have been collected by the author's Department over several years for a medium-sized digital computer used regularly to assist air traffic controllers, and the distribution of times between failures is in good agreement with that expected on the basis of a constant failure rate for the components. Since the majority of failures are in semiconductor diodes or transistors, it has been concluded that an assumption of a constant failure rate over a period of years is a reasonable basis for predicting reliability. These results were obtained under almost ideal environmental conditions, with good cooling and no vibration, but if semiconductors are used at high temperatures their failure rate will be higher and not necessarily constant.

A further assumption in the above calculations is that failures are independent, that is that no failure of one component is likely to increase the chance of failure of another. The desirable goal can be partly achieved by careful design, but it generally requires some redundant components to guard against, for example a failure, or an excessive voltage on a power supply line. This kind of fault may overstress a large number of transistors and reduce their working life considerably.

### 4. Series and Parallel Systems

Two sub-systems A and B can be combined functionally in series or parallel. The series arrangement has been discussed above, and gives a combined reliability $R$ which is the product of the two sub-system reliabilities, that is, $R = R_A.R_B$. The parallel arrangement can improve reliability if the system can be arranged to deliver the correct output when either A, or B, or both are working. Thus it can only fail when *both* A and B fail. Hence if $U$ is the combined unreliability, i.e. the probability that the system will *not* work, this is the product of the separate unreliabilities $U_A$ and $U_B$, i.e.

$$U = U_A.U_B$$

We now use the rule that the probabilities for any one event must always add to unity, so that

$$U + R = 1$$
$$U_A + R_A = 1$$
$$U_B + R_B = 1$$

Therefore $U = 1-R = (1-R_A)(1-R_B)$

Thus $R = 1-(1-R_A)(1-R_B) = R_A+R_B-R_AR_B.$
$$= R_AU_B+R_BU_A+R_AR_B$$

If the two units are identical, we have

$$R = 2R_A-R_A^2$$

If the system uses three identical units in parallel

$$R = 1-(1-R_A)^3$$
$$= 3R_A-3R_A^2+R_A^3$$

More complicated systems can be analysed in a similar manner.[7]

### 5. The Reliability of Duplicate Systems

Returning to the satellite containing 1000 transistors, we may now investigate the advantages of various schemes of duplication. For a preliminary analysis, we may neglect the problem of switching to the standby channel when the main channel has failed, and the loss in service during the switching. We also assume that the system has no maintenance.

For a complete duplicate system, the combined reliability for the postulated 8000 hour period is given by $R = 2R_1-R_1^2$, where $R_1 = 0{\cdot}202$ = reliability of basic system



World Radio History

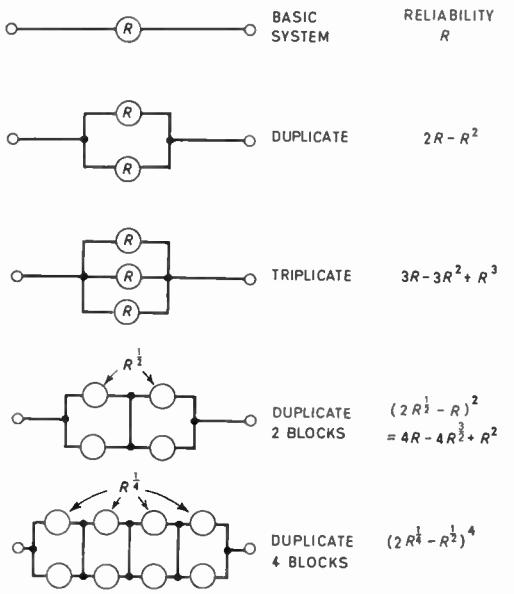



Fig. 2. Schemes of redundancy.

Thus $R = 0{\cdot}363$

This is a modest improvement in reliability for a 100% increase in cost, weight and size.

If we now split the system into two halves, and duplicate the halves separately, as shown in Fig. 2, the system reliability is enhanced.

Here $R_2 = \sqrt{R_1} = 0{\cdot}449$.

The reliability of the duplicate first half is thus

$$R' = 2R_2 - R_2^2$$
$$= 0{\cdot}697$$

The reliability of the complete system is thus $R = (R')^2 = 0{\cdot}486$, since both halves are functionally in series.

Continuing the subdivision process, we can improve the reliability by using smaller units, provided that the interconnection points cause no difficulty. Using four duplicated units, the overall reliability is 0·759, and using 8 units it rises to 0·931 for the 8000 hour period. Clearly the logical conclusion would be to arrange for each component to have its own duplicate available for use at all times, without complications such as switching. This can only be achieved under very restricted circumstances, but a computation based upon this assumption is useful as it gives an upper bound to the possible advantages of duplication. The simplest computation in this case is obtained from a consideration of failure probabilities.

In the duplicate component system, the chance of failure per component in 8000 hours is $8000 \times 2 \times 10^{-7} = 1{\cdot}6 \times 10^{-3}$.

Thus the chance that both components of a pair will fail is $(1{\cdot}6 \times 10^{-3})^2 = 2{\cdot}56 \times 10^{-6}$.

Since there are 1000 component-pairs in series, the probability of system failure is

$$U = 10^3 \times 2{\cdot}56 \times 10^{-6} = 2{\cdot}56 \times 10^{-3}$$

The reliability is thus $1 - U = 0{\cdot}9974$.

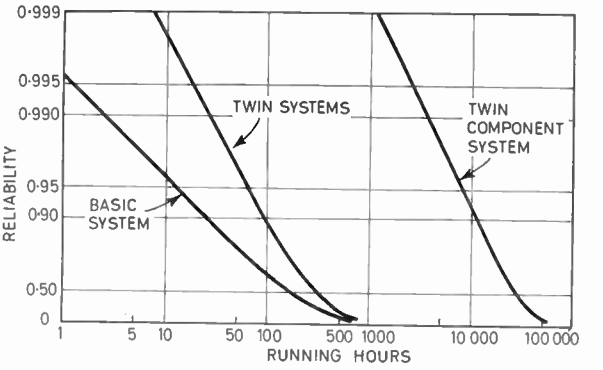
An important feature of redundant systems is shown in this calculation—the probability of failure is no longer proportional to the duration of the operating period. Here we have first-order redundancy, that is only one spare for each working component, and the failure probability is proportional to the *square* of the operating period.

Thus if the period considered is increased to 80 000 hours, the chance of failure per pair of components is $2{\cdot}56 \times 10^{-4}$, and per system is $2{\cdot}56 \times 10^{-1} = 0{\cdot}256$, an increase of 100 times.

The reliability is now $R = e^{-0{\cdot}256} = 0{\cdot}774$.

As pointed out above, it is impossible to achieve this degree of improvement in practical circuits, but the general benefit of reducing the size of the blocks in the block redundancy scheme can be obtained, and it is frequently possible to sub-divide a system to a very considerable degree before incurring very heavy penalties in the form of constraints on circuit design.[4,5,6]

The two extreme cases of duplicate systems and duplicate components are compared in Fig. 3 with a basic system resembling in its component numbers a medium sized digital computer. The slope of the reliability-time curve is twice as large for the redundant systems as it is for the basic system.




| Component numbers | Failure rate per 1000 hours |
|---|---|
| Transistors—$10^4$ | 0·027% |
| Diodes—$2{\cdot}5 \times 10^4$ | 0·0043% |
| Resistors—$6 \times 10^4$ | 0·0006% |

Fig. 3. Reliability of large system without maintenance.



World Radio History

### 6. Redundant Systems with Maintenance

The maximum benefit is obtained from a redundant system if periodic checking and repair is possible. This prevents the accumulation of single faults, which do not cause a system failure, but which involve a loss of redundancy in their part of the circuit. For systems which are accessible for maintenance the reliability is mainly a function of the ratio of time between failures to time taken to repair a fault. Thus any engineering feature which contributes to rapid fault location and repair also contributes to overall reliability. It is usual to apply the terms 'availability' or 'up-time ratio' in this case, as the probability now required is not that of the system failing to develop a fault, but of being in working order at any given time. As time increases, this tends to the value[6]

$$P_w = \frac{\text{mean time between failures}}{\text{mean time between failures} + \text{mean repair time}}$$

$$= 1 - \frac{1}{1+k} \simeq 1 - \frac{1}{k} \text{ for } k \text{ large}$$

$$\text{where } k = \frac{\text{m.t.b.f.}}{\text{mean repair time}}$$

Thus for a duplicate system in which no information is lost while changing from a faulty to a working standby system, the combined availability would be

$$P_{wD} = 1 - \left(\frac{1}{1+k}\right)^2.$$

This result depends upon using the same maintenance procedures on the duplicate and on the single systems; if there is no increase in the staff concerned it may take longer to find and repair a fault in the duplicate system and the corresponding change must be made to the value of $k$.

As an example of this, a system without repair facilities having a m.t.b.f. of 100 hours and a fault repair time of 1 hour has a reliability over a 100-hour period of only 0·368. However, if maintenance is permitted, the ultimate availability will be 0·99. A duplicate system with maintenance, allowing a repair time of 2 hours, will have an availability of 0·9996, so that the down-time in 100 hours will be reduced from 1 hour to just over 2 minutes, on average.

### 7. Comparison of Active and Passive Redundancy

The basic strategy of designing redundant systems depends upon the extent to which errors can be tolerated. If a small break in service is permissible, of duration $t$, then provided that some fault detecting system can be devised which either tests the system at intervals of not more than $t$, or can deduce in a time less than $t$ from the normal traffic that a fault has occurred, a simple standby arrangement is adequate. In this one or more identical systems are kept as standby plant, and automatically switched into circuit when a fault is detected.

If, however, only very rare and very short intervals of down time are permitted, an automatic error-correcting system, such as a triple redundancy scheme with majority voting may be used. In this any one channel may fail and the system will produce the correct output if the other two channels are working. The probability of a sustained fault, which requires a failure of two of the three channels is of the same order as that for the standby scheme. Thus, since it requires 50% more apparatus, it will be used only when the short breaks in service during changeover period cannot be tolerated. This would be the case with, for example, automatic blind landing apparatus for aircraft, or the safety circuits of a nuclear reactor, whereas in underwater telephone apparatus or parts of an electronic telephone exchange some short interruption could be tolerated, and the simpler duplicate stand-by scheme is acceptable.

As with other forms of redundancy, the effectiveness can be increased by dividing the basic system into a number of smaller blocks.

It is customary to distinguish between the above two forms of redundancy by the terms 'active' and 'passive'. Active schemes involve some change-over switching, as in the standby scheme above, whereas passive schemes have the spare channels always connected and no switching or interruption to service is required in order to permit the remaining healthy parts of the apparatus to take control when one part fails.

### 8. Passive Redundancy Schemes

The previous calculations show that applying redundancy at component level could produce a spectacular increase in reliability. Unfortunately circuit constraints usually prevent this degree of improvement, particularly with three-terminal devices such as transistors.[8,9]

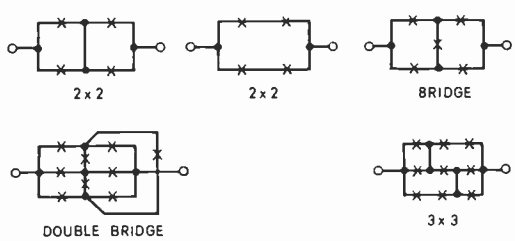



Fig. 4. Simple hammock networks.

Where only one kind of discontinuous operation is required, for example closing a switch, a duplicate switch in parallel with the first provides a complete standby, and is thus a fully effective redundancy scheme. If, however, the switch is sometimes required to be open, and is liable to an occasional fault



World Radio History

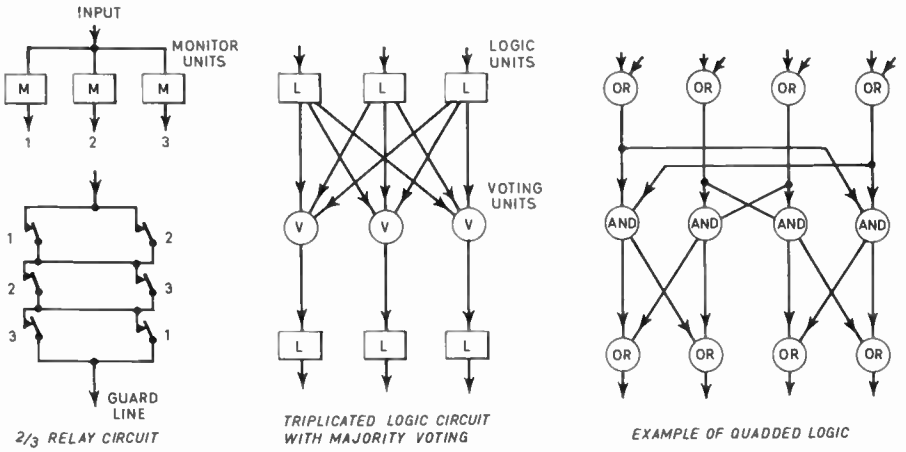



Fig. 5. Forms of redundancy.

causing it to remain closed, a parallel arrangement is less reliable, and a series connection would be better. In most cases, of course, the switch is required to be in both positions at different times, and both shunt and series redundancy is required to increase the reliability in both positions. Thus for a complete redundancy network we require at least four switches in place of one, as shown in Fig. 4.

Greater reliability can be obtained by using more switches. The theory of such 'hammock' networks has been investigated by Moore and Shannon[1] and others in relation to relay switching systems. The optimum shape, or ratio of length to breadth of such a network depends upon the ratio of the probabilities of failure in the nominally closed and the nominally open position of the switches, and the improvement over a single switch afforded by the redundancy increases with the increasing reliability of each switch.

The same hammock networks can be used for devices such as semiconductor diodes, which should ideally have zero or infinite impedance, and simple $2 \times 2$ networks with fault indicators have been used in power supply and logic circuits where extreme reliability is required. Similar circuits for transistor logic applications have been proposed by Sorensen[2] but they involve a considerable penalty in power gain, since each transistor is replaced in the hammock network by four similar transistors which require four times the base current. The fact that the input and output ports of a transistor amplifier have one common terminal also complicates circuit design. If all other components in the circuit are replicated in a similar manner, the resulting restrictions on circuit design are generally unacceptable.

The hammock network is not particularly useful for most of the components in an electric circuit which have to maintain some prescribed value of impedance. A $3 \times 3$ network, for example, consisting of 9 similar components, may in the worst case suffer an impedance change of $+36\%$ or $-26\%$ when only one of the components fails. This alteration could easily be tolerated in decoupling components, but not in resistors which determine the gain, or perhaps the bias point of an amplifier. Clearly, in most cases hammock networks would be a very uneconomical form of redundancy.

### 9. Redundant Digital Systems

An easier scheme to implement is the block redundancy arrangement, in which circuit blocks are replicated, and followed by some voting or decision circuit. In digital systems the simplest scheme involves an odd number of identical blocks, followed by a majority decision unit. In order to avoid limiting the reliability of the system to the reliability of all voting circuits in series, it is also necessary to replicate the voting circuits. Thus the simplest scheme[4] involves triple logic circuits followed by triple voting circuits as shown in Fig. 5.

Making some simplifying assumptions about component reliability, one can show that the optimum design is that in which the size of the logic block is equal to the size of the voting circuit. Since the basic system has no voting elements, an optimum triplicate scheme thus requires six times as many components as the basic system. If the ratio of reliability to cost must be optimized, rather larger logic blocks are better.

Again, in order to obtain maximum availability in a maintainable system, it is important to test the system frequently and remove single faults before they can cause a system failure.



World Radio History

A less costly system 'quadded logic' has been proposed by Tryon[3] in which the logic and voting functions are performed by the same elements. Each element is replaced by four similar elements, with twice as many inputs. By using suitable cross-connections between elements, single errors can be corrected in the following stages. This scheme is simpler in that only logical elements are needed without voting, but it is difficult to isolate individual channels for test purposes.

### 10. Analogue or Linear Systems

In analogue systems the same general pattern of replication can be used. If it is known that the three channels of a triplicate redundancy scheme are 'fail-safe', in that they cannot generate an excessive signal, but produce zero under fault conditions, it is merely necessary to choose the largest signal. If this is not so, more complicated procedures such as comparing each signal with the average of the other two are required. Many test procedures are easy to apply, for example all amplifiers can be continuously monitored by comparing an attenuated output signal with the input signal, or by measuring the level of a pilot signal outside the working frequency band, which is added to the input signal. Since most of these monitoring and detection methods are more complicated than the digital voting circuits, there will be fewer of them in an optimum design, and they will monitor larger units of the system.

### 11. Conclusion

Much of the above discussion is generally applicable to many engineering systems in addition to those which incorporate semi-conductor devices. However, the calculations based upon a failure rate which is constant with time do not apply to any component which exhibits 'wear-out' symptoms within the operating period, and in consequence a more complicated distribution of component lives (for example a Wiebull distribution) must be assumed. In most semiconductor circuits, on the other hand there is no sign of any wear-out mechanism, and an assumption of a constant failure rate gives reliability predictions which are in good agreement with observations.

The design of redundant systems is becoming of increasing interest to the user of semiconductors, despite their inherent reliability, since electronic systems are constantly growing in size and complexity, and are taking over more critical tasks where any failure may be very expensive. Some of the redundancy schemes may appear somewhat extravagant in isolation, but since the electronic apparatus frequently incurs only a small fraction of the total cost of a project (for example, in a nuclear power station or a satellite communication system) doubling or trebling this does not increase the total cost prohibitively.

In future systems assembled from micro-electronic structures, the cost penalty may be much smaller, since the cost of material forms a very small fraction of the total cost of a micro-circuit. Thus it costs very little more to make four diodes than it does to make one diode; the main penalty lies in the extra number of assemblies needed. With this cost structure, redundancy becomes a more attractive proposition, and certainly cheaper to implement than in apparatus using conventional components. As each form of micro-circuit involves certain restrictions on component values and the number of interconnections, the most effective redundancy scheme will depend upon the particular fabrication techniques available, and in future the development of redundancy schemes must be closely related to the development of micro-circuit technology.

### 12. References


1. E. F. Moore and C. E. Shannon, "Reliable circuits using less reliable relays", *J. Franklin Inst.*, **262**, pp. 191–208, pp. 281–97, September and October 1956.
2. A. A. Sorensen, "Digital circuit reliability through redundancy", *Electronics Reliab. and Micro-min.*, **1**, pp. 27–37, January–March 1962.
3. J. G. Tryon, "Quadded logic". In "Redundancy Techniques for Computing Systems", ed. by R. H. Wilcox and W. C. Mann, p. 205, (Spartan Books, Washington, 1962).
4. R. E. Lyons and W. Vanderkulk, "The use of triple-modular redundancy to improve computer reliability". *IBM J. Res. Developm.* 6, No. 2, p. 200, 1962,
5. W. G. Brown, J. Tierney and R. Wasserman, "Improvement of electronic computer reliability through the use of redundancy", *Trans. I.R.E.*, **EC-10**, 3, p. 407, 1961.
6. D. J. Creasey, "Redundancy techniques for use in an air traffic control computer", *Micro-electronics and Reliab.*, 3, No. 3, p. 175–92, 1964.
7. S. R. Calabro, "Reliability Principles and Practices", (McGraw-Hill, New York, 1962).
8. J. C. Cluley, "Low-level redundancy as a means of improving digital computer reliability", *Electronics Reliab. and Micro-min.*, **1**, pp. 203–16, July–September 1962.
9. J. C. Cluley, "A Comparison of duplicate and triplicate redundancy schemes for binary logical networks". *Micro-electronics and Reliab.*, 3, No. 1, pp. 51–59, June 1964.


*Manuscript first received by the Institution on 14th May 1965 and in final form on 28th October 1965. (Paper No. 1027).*

© The Institution of Electronic and Radio Engineers, 1966



World Radio History

U.D.C. 621.391.812.6 : 621.396.934.029.63 : 537.56

# A Discussion of Factors Affecting the Performance of Microwave Systems in Certain Plasma Channels

*By*

T. KALISZEWSKI, M.A., B.S.†

**Summary:** Principal factors affecting the performance of microwave systems in composite and transient plasma channels are reviewed. Only the macroscopic aspects of the channel characterization are considered. These include such effects as absorption, refraction, dispersion and depolarization; also nonlinear and random effects. An effort is made to assess their importance to communication with and tracking of rocket-borne systems. It is concluded that channels incorporating, for example, the re-entry plasma sheath or the rocket exhaust are inaccessible at the present time to a detailed characterization. Further research, especially of experimental nature and extending beyond the traditional study of absorption loss, is indicated.

## 1. Introduction

The reliability of the communication links between a ground terminal and a rocket in flight depends primarily on the interaction between the transmission and the propagation channel. There are at least two critical periods in each space mission during which this interaction can assume serious proportions: during the powered phase of the flight and on re-entry into the earth's atmosphere. During these periods, transient plasmas are generated which frequently cause transmission 'blackouts', thus interrupting vital tracking functions and the flow of telemetry data. Moreover, even at times of apparent transparency of the plasma component, the quality of transmission can be impaired due to the dispersive and/or turbulent character of the channel.

To reduce or to eliminate interference originating with such plasmas, it is essential that we understand their origin and properties and the nature of their interaction with the electromagnetic transmission. Without such an understanding no rational planning of new or evaluation of operational systems is possible. Ideally, this understanding would imply our ability to characterize the channel at both the microscopic and macroscopic levels. This is, however, a formidable task rarely attempted or accomplished even for a specific and fully documented mission. It is inevitable, then, that pending future development of this subject these two characterizations must be undertaken separately.

In this paper it is proposed to review the principal macroscopic factors which affect the performance of, especially, microwave systems in plasma channels. It is hoped that this review will encourage the study of plasma channel characterization by spotlighting factors other than the traditionally studied absorption loss, and by indicating directions in which the future research in this area should proceed.

† Adcom Inc., Cambridge, Massachusetts, U.S.A.

## 2. Characterization of Linear Plasma Channels

It is a customary practice when discussing plasma channels to emphasize their linear, non-linear and random characteristics separately. This practice, though quite arbitrary, serves the useful purpose of facilitating a discussion of what is otherwise a complex subject. For the same reason it is convenient also to model a plasma channel in terms of a generalized, dissipative dielectric medium with which we can associate a number of familiar parameters and effects and which is amenable to analytical manipulations. Thus, we can characterize a plasma channel by its conductivity or loss tangent, by its dielectric permittivity, propagation constant, etc. Subsequently, we can inquire about the channel's 'observables' such as the absorption, refraction or dispersion.

It is not our purpose here to delineate the criteria on which this modelling of plasma channels is based. We will note, however, the definition of a linear plasma channel as being one whose properties are unaffected by the presence of an external, electromagnetic field. Otherwise, we assume, rightly or not, that the plasma channels discussed here are susceptible to modelling and are, in fact, similar in many ways to the ionosphere.

In the interest of brevity the 'fundamentals' of plasma channels (parameters, definitions, etc.), which are dealt with at length elsewhere[1,2,3] will not be discussed and attention will be concentrated on those observable factors which are of the most immediate relevance to communication problems.

### 2.1. *Transmission Loss*

The cumulative absorption of an electromagnetic wave in plasma constitutes an important characteristic



World Radio History

of the channel which, under certain conditions, may be equated with the transmission loss. In general, however, its isolation is not a simple matter not only because of reflections from the plasma boundaries or refraction, but also because of many other factors such as antenna pattern, mismatch or depolarization contributing to the observed, overall loss. For example, the signal drop-outs observed at X-band frequencies during the retro-rocket firing on a vehicle in the *Saturn* class, are of the order of 60 decibels above the calculated free space value.[4] How much of that loss can be identified as an absorption loss is by no means clear. This question could, however, be partly resolved, by comparing the losses observed at two or more frequencies. The absorption loss $A$ is approximately inversely proportional to the square of the carrier frequency $\omega$, i.e.[5]

$$A \simeq 4{\cdot}6\times10^{-5}\frac{\nu}{\omega^2+\nu^2}\int_{\zeta_0} N(\zeta_0)\,\mathrm{d}\zeta \quad \text{decibels} \quad \ldots(1)$$

where $N(\zeta)$ is the electron density, $\nu$ is the collision frequency and the constant is evaluated in m.k.s. units; an isotropic plasma is assumed.

To isolate the flame effects it is necessary, then, to calculate scrupulously transmission loss due to all other factors leaving only the absorption, reflection and/or antenna effects to be considered. We can further aid in this process of elimination by monitoring the power of the on-board transmitters and by estimating the reflection losses. Here, again, we will depend on the information concerning the configuration and the properties of the plasma, without which such estimates are impossible. Given this information, further care must be exercised in computing the reflection losses under conditions where the linearity of the medium cannot be taken for granted and where also the far-field approximations may be questionable.

## 2.2. *Refractive Effects*

Propagation of electromagnetic waves in plasma channels is accompanied by a number of effects jointly referred to as the refraction effects. There are two observable manifestations of these effects of utmost interest to communication and especially to radio tracking of vehicles travelling in or enveloped by a plasma. Since the radio signals traversing a plasma medium are 'bent' and their transit time changed, errors will result in the determination of the range and elevation angles. In many practical situations the determination of these errors is complicated by the composite nature (plasma and atmosphere) of the radio channel and by the indeterminate character of the boundaries and distributions. Further complications arise because plasma, unlike the earth's atmosphere, is a dispersive medium.

### 2.2.1. Range errors and phase retardation

The transit time of an electromagnetic wave between two terminals in a plasma depends on the phase velocity and the actual propagation path. The latter, due to refraction, is generally greater than the geometric distance and the difference constitutes the range error. It can be shown that in a stratified medium this error will be given by

$$\Delta R = \int_{\zeta_0} \mu(\zeta)\,\mathrm{cosec}\,\theta\,\mathrm{d}\zeta - R_0 \quad \ldots\ldots(2)$$

where $\mu(\zeta)$ is the real part of a complex index of refraction $n$, $\theta$ is the local elevation angle along the path and $R_0$ is the geometric distance. In a lossless medium ($\nu = 0$), relation (2) can be written, approximately,

$$\Delta R \simeq \int_{\zeta_0} \mathrm{cosec}\,\theta\,\mathrm{d}\zeta - 1{\cdot}6\times10^{3}\frac{1}{\omega^2}\times \int_{\zeta_0} N(\zeta)\,\mathrm{cosec}\,\theta\,\mathrm{d}\zeta - R_0 \quad \ldots(3)$$

$$= \Delta R_g + \Delta R_\mu \quad \text{metres} \quad \ldots\ldots(3a)$$

Here $\Delta R_g$ represents a geometric range error and $\Delta R_\mu$ is the true refraction or velocity error. We can determine both of these, provided that we can trace the radio ray through the medium.

The phase difference introduced by plasma can be calculated likewise:

$$\Delta\phi = \frac{\omega}{c}\int_{\zeta_0}(\mu-1)\,\mathrm{d}\zeta \simeq 5{\cdot}3\times10^{-6}\left(\frac{\omega}{\omega^2+\nu^2}\right)\int_{\zeta_0} N(\zeta)\,\mathrm{d}\zeta \text{ radians} \quad \ldots(4)$$

In contrast to the frequency dependence of absorption, the phase change introduced by plasma is inversely proportional to the carrier frequency.

### 2.2.2. Elevation angle errors

An elevation angle error is, by definition, equal to the difference between the apparent and true elevation. In general, its determination requires that the ray tracing techniques be used unless the medium shows certain regularity which makes it then accessible to purely analytical treatment. For example, it can be shown that in a spherically stratified medium, the total bending angle $\beta$, is given by

$$\beta = -\int_{\mu_1}^{\mu_2}\frac{\mathrm{d}\mu}{\mu}\cot\theta \quad \ldots\ldots(5)$$



World Radio History

where $\mu_1$, $\mu_2$ correspond to the index of refraction along the path.

Knowing $\beta$ and also the apparent elevation angles at the terminal points, the angle error can be determined,† i.e.

$$\varepsilon = \beta - \tan^{-1} \frac{\frac{\mu_1}{\mu_2} - \cos\beta - \sin\beta \tan\theta_1}{\sin\beta - \cos\beta \tan\theta_1 + \frac{\mu_1}{\mu_2}\tan\theta_2} \quad \ldots(6)$$

Incidentally, a zero bending angle does not imply zero angle error. For example, in the frequently employed model of a homogeneous plasma sheath, bounded by a vacuum on the incident and transmitted side, the bending angle $\beta$ is zero but the angular elevation error can be shown to be,

$$\varepsilon = \phi_1 - \tan^{-1} H/X = f(N) \quad \ldots(7)$$

where $X$, $H$ are the total horizontal and vertical separation of the terminals, $\phi_1$ is the apparent (incident) angle of elevation and the error $\varepsilon$ is a function of sheath electron density.

### 2.3. *Dispersion Effects*

In discussing the range error in the preceding sub-section we have ignored one important property of the plasma medium, namely its dispersive character. For this reason it is more appropriate, when discussing the transit time or range errors, to use the notions of the group velocity and group index of refraction; these are related to the phase velocity and the non-dispersive index of refraction as follows[6]:

$$\mu' = \frac{c}{v_g} = \mu + \omega \frac{d\mu}{d\omega} \quad \ldots(8)$$

$$v_g = \frac{1}{\frac{1}{v_p} + \frac{\omega}{c}\frac{d\mu}{d\omega}} \quad \ldots(9)$$

Consequently, the range error of (3) must include also a 'dispersion error', i.e.

$$\Delta R' = \int_{\zeta_0} \mu'(\zeta) \operatorname{cosec}\theta \, d\zeta - R_0$$

$$= \Delta R + \omega \int_{\zeta_0} \frac{d\mu}{d\omega} \operatorname{cosec}\theta \, d\zeta \quad \ldots(10)$$

The dispersion effect in a plasma channel also constitutes one of the factors limiting its effective frequency bandwidth. Staras[7] has pointed out in this connection that unless the differential group delay $\Delta\tau$ is less than $(\Delta f)^{-1}$, serious distortions must be expected; here $\Delta f$ is of the order of the r.f. bandwidth and

$$\Delta\tau = \left(\frac{\Delta f}{f^3}\right) \frac{1}{4\pi^2 c\varepsilon_0 m} \int_{\zeta_0} N(\zeta) \, d\zeta \quad \ldots(11)$$

Another approach to the problem of dispersion and its effects on pulse distortion is exemplified by the work of Marquedant *et al.*[8]; their results will be summarized later.

### 2.4. *Polarization Effects*

Another consideration of great importance to communication in plasma channels is the polarization of the transmitted signal. In space applications this consideration is, of course, especially important even apart from the plasma-wave interaction on account of the vehicle's motion. However, we are interested here only in the effects originating in the channel. There are at least two physical reasons why a signal transmitted in a plasma channel may experience depolarization. One is the magnetic field, natural or artificially imposed on the channel as, for instance, when 'magnetic windows' are created to facilitate communication on re-entry.[9] Inhomogeneities, too, can cause depolarization; normally, waves scattered from inhomogeneities specularly preserve their polarization. One must reckon, however, with a diffuse component of scattering which, as a rule, will be depolarized.

The effect of the Earth's magnetic field at microwave frequencies can, for all practical purposes, be discounted. Thus, unless an external field is present the medium can be assumed to be isotropic and effects such as the magneto-ionic splitting, Faraday effects, etc., can be ignored. Should, however, the medium be purposely rendered anisotropic, an appropriate analysis will have to be invoked. The general question of propagation in magneto-active plasma channels, and especially the polarization effects, are adequately discussed elsewhere.[1,2,3]

## 3. Some Non-linear Effects in Plasma

The assumption of linearity for practical plasma is a convenience which may be occasionally correct. To ascertain the validity of this assumption, it is necessary to examine both the properties of the plasma and the strength of externally generated electromagnetic fields. The appearance of non-linear effects can have serious consequences for the reliability and quality of radio transmission. In certain extreme cases, the utilization of the link may become impossible unless measures are taken to suppress these effects below the limit of observability. There are several ways in which this can be done, most simply by reducing the transmitted power. A criterion useful in assessing the observability of non-linear effects is the so-called 'plasma field',[10] defined as

† This expression is attributed to K. A. Norton of the National Bureau of Standards, Boulder, Colo.



World Radio History

follows:

$$E \ll E_p = 4{\cdot}2 \times 10^{-10}\sqrt{GT(\omega^2+\nu^2)} \quad \text{volt/cm} \qquad \ldots\ldots(12)$$

where $T$ = electron temperature in the absence of external field,

$G$ = fractional energy loss factor,

$\omega$, $\nu$ are signal and collision frequencies, respectively.

An incident signal field much less than $E_p$ is referred to as a 'weak' field and its effects on plasma can be ignored. On the other hand, fields such that $E_i > E_p$ will, as a rule, give rise to non-linear effects of which there are several manifestations.

It is well known that 'strong' waves propagating in plasma can modify its properties, such as, for instance, the index of refraction. This modification can be effected in several ways, either through the change in collision frequency or in the rate coefficients (recombination, etc.). The resulting changes in absorption can be quite drastic and so can be other manifestations of non-linear effects. We are familiar with the self- and cross-modulation in the ionosphere and there is good reason to believe that similar effects can be expected in, for instance, the re-entry plasma.[11] It is rather common practice to use collinear antennas which simultaneously illuminate a common volume of plasma. In such applications the cross-modulation effect can be effectively reduced by separating the radiating elements or, as we have already observed, by reducing the radiated powers. Further minimization of the non-linear effects can be achieved by taking recognition of the finite electron energy relaxation time. This constant determines the lowest permissible pulse width below which serious distortion must be expected. For practical purposes, this constant is approximately equal to

$$\tau = \frac{1}{G\nu_0} \quad \text{seconds} \qquad \ldots\ldots(13)$$

For example, in a typical re-entry plasma $G \simeq 1{\cdot}2 \times 10^{-3}$ and $\nu_0 = 10^9$ seconds$^{-1}$, and the relaxation constant $\tau$ will be of the order of one microsecond.

Further complications must be expected in magneto-active plasmas, such as would be created through deliberate imposition of magnetic field.[9,12]

## 4. Antenna Performance in Plasma

Experience shows that the performance of the antennas used in applications such as re-entry communication cannot be considered apart from the plasma into which they radiate and in which they are embedded. The usual criteria of antenna performance such as the impedance, gain and bandwidth must be re-examined inasmuch as the proximity of a dissipative and dispersive medium may lead to their drastic modification. Recent studies of this subject demonstrate the real possibilities of antenna breakdown, pattern redistribution and impedance mismatch in environments comparable to that encountered in aerospace applications.

A complete assessment of all the factors contributing to the degradation of antenna performance in plasma is not quite possible at this time. Certain available results point, however, to simple measures which can be taken to minimize such degradation. These include careful selection of the carrier frequencies, average and peak powers and the design of suitable antennas.

### 4.1. *Admittance of a Slot Antenna*

Because of aerodynamic considerations a flush-mounted array of slots is a logical choice for the radiating structures in many aerospace applications. It is undoubtedly for this reason that the performance of this type of antenna is receiving increased attention. For obvious reasons, the analytical studies are restricted to structures and plasmas which are stationary, isotropic and most often, also homogeneous. An example of such studies is a recent work by Galejs[13] who considers a rectangular, waveguide backed slot covered by a thin homogeneous plasma layer. Other examples are annular slots, either waveguide or cavity backed, covered by layers having also density stratifications.[14]

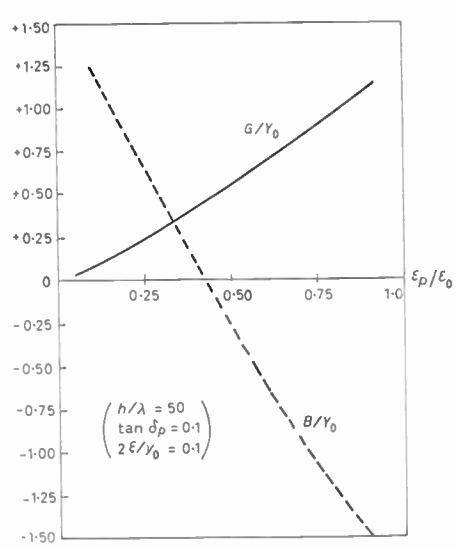



Fig. 1. Normalized conductance and susceptance of a waveguide radiating into plasma layer. (Adapted from Ref. 13.)



World Radio History

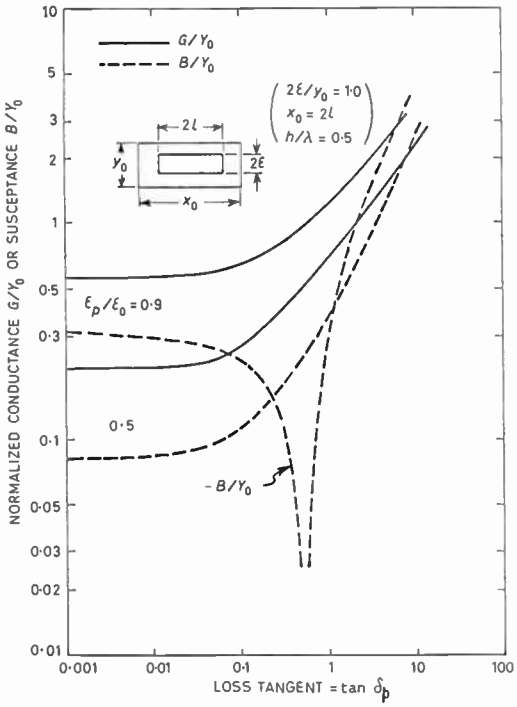



Fig. 2. Slot admittance with increasing plasma loss.[13]

For a rectangular slot, Galejs cites the following results:

(1) For a particular set of parameters (tan $\delta$, $\varepsilon_p$) the admittance of the slot shows little dependence on the thickness of the plasma layer, except in the region of $h/\lambda < 0{\cdot}5$ where the conductance $G$ increases slightly.

(2) Increase in the plasma density of the layer causes the conductance to decrease and make the susceptance more inductive (Fig. 1).

(3) Increase in the collision frequency (i.e. loss tangent) causes the conductance to increase and also affects the susceptance (Fig. 2).

Here the increase or decrease in conductance or susceptance is meant to be relative to their free-space values.

The effect of plasma stratification on the admittance of an annular slot is shown in Fig. 3. The important feature of this study by Galejs[14] is the behaviour of the radiation conductance $G_r$ in the vicinity of the plasma frequencies corresponding to the densities of the assumed two-layer model of a re-entry sheath. It is obvious that for the minimum mismatch operation above the highest plasma frequency is desirable.

### 4.2. *Radiation Pattern Distortion*

A decrease of the radiation conductance in the vicinity of the plasma resonance implies a corre-

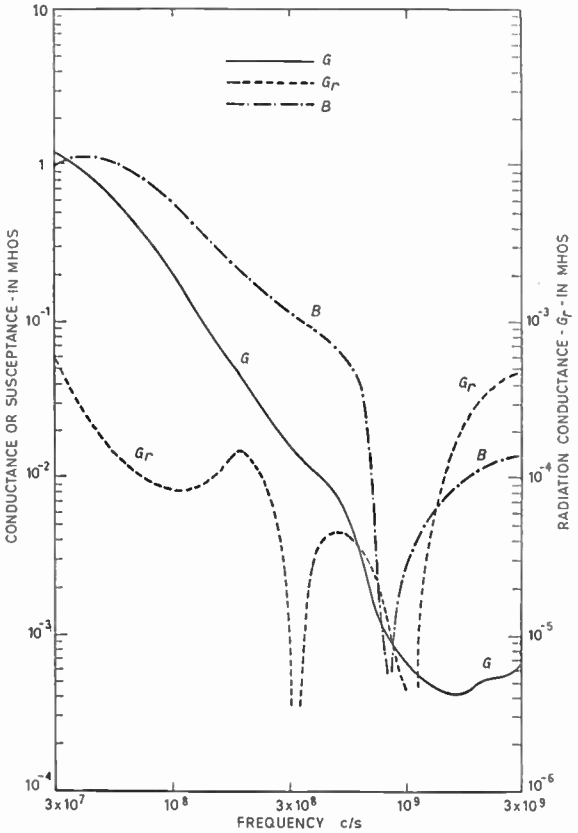



Fig. 3. Slot admittance for a two-layer approximation of a re-entry plasma sheath.[14]

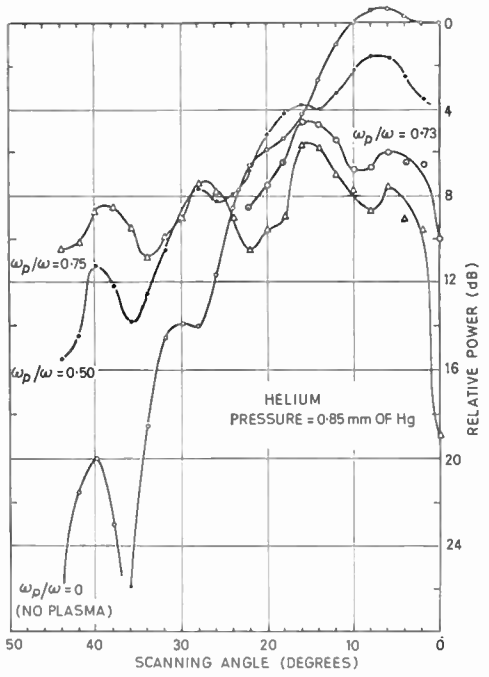



Fig. 4. Effects of plasma on radiation pattern, at X-band.[15]



World Radio History

sponding decrease in antenna efficiency. Another effect which leads to similar consequences is the distortion of the antenna radiation pattern, or redistribution of its radiated power. This effect was studied under laboratory conditions by Bachynski *et al.*[15] who have found at both X and Q-band a tendency to lower the directivity and to increase antenna side-lobe level. An example of their finding at X-band is shown in Fig. 4. Unfortunately, the plasma density available in that experiment was not high enough to show what happens to the pattern close to and at the plasma resonance. The trend, however, is clear: an originally highly directive pattern flattens out and develops additional side-lobes. It is obvious that under those conditions the frequently observed discrepancies between the calculated and measured values of the signal strength can be attributed to either the plasma absorption or to the effect of pattern distortion.

### 4.3. *Effect of Ambient Ionization on Antenna Breakdown*

An antenna operating in plasma at a 'safe' frequency can initiate a breakdown condition even at power levels which would be inadequate for that purpose in a neutral gas. The effect of the ambient ionization is, as shown by Whitmer and MacDonald,[16] to lower drastically the required breakdown field as the ionization approaches the critical density. Thus, the frequently repeated conclusion that it is necessary to transmit at a frequency above the plasma frequency may have to be modified to read 'well above' if the consequences of a breakdown are to be avoided. How much above will depend, of course, on the ambient electron density and on the breakdown field in a neutral gas at the given frequency. For air MacDonald[17] has computed a number of curves for both c.w. and pulsed fields as a function of pulse width, frequency and pressure; these are summarized in Figs. 5 and 6.

A special consideration attending the use of pulsed transmission merits a further comment. It is evident from Fig. 6 that the value of the breakdown field depends greatly on the pulse width and this dependence is likely to be even more important in a pre-ionized medium. Hence the selection of the pulse width and of the peak power for a given application must reflect this as well as the frequency limitation of the medium.

### 4.4. *Some Noise Considerations*

Performance of a communication system, its reliability and efficiency depends to a large extent on the total noise level at the input to the antenna terminals. With the advent of low-noise receivers and amplifiers, the principal sources of this noise at microwave frequencies are the atmosphere, the earth and other discrete celestial sources, hot surfaces in the vicinity of antenna (rocket nozzles, etc.) and plasma. The relative importance of these sources in contributing to the total noise level depends, of course, on application. For example, on re-entry, the principal source of noise at frequencies below the plasma frequencies are the hot, ablating surfaces of the vehicle; at frequencies above the plasma frequencies the contributions from sources such as the atmosphere or earth predominate. Finally, close to the plasma frequency, noise from plasma exceeds all other contributions. This is shown graphically in Fig. 7, where the effective temperature of an antenna in a typical re-entry environment ($N = 4\times10^{18}$ electrons/m$^3$; $\nu = 1{\cdot}12\times10^9$ s$^{-1}$; $T_p = 5000$°K; $T_v = 500$°K; $T_{ex} = 300$°K; $d/\lambda_p = 2{\cdot}5$) is shown.[18]

## 5. Transmission of Digital Signals

In Section 2.3 we have discussed the effects of dispersion in plasma channels on the effective bandwidth, range and elevation errors. A different approach to the same problem is exemplified by the work of Marquedant *et al.*[8] who have attempted to assess the effect of dispersion on the transmission of digital signals. Since the dispersion is likely to result in the distortion of individual pulses an intersymbol interference, unless controlled or minimized by the choice of a suitable modulation scheme, must be expected. Hence, the overall effect of dispersion on digital communication will be to reduce the latter's reliability.

A classical approach to the transmission problem is to define a suitable transfer function for the channel. Such a function was derived in the study cited here for a narrow, half-wavelength rectangular radiation slot covered by a lossless plasma sheath of thickness $L$. On the assumption of transparency ($\omega > 2\omega_p$) this expression was shown to be

$$G(\omega) \simeq \left(1 - \frac{\delta_0}{2}\right) \exp\left[\mathrm{j}\left(\frac{\omega}{\omega_0} - 1\right)^2 \cdot \frac{\omega_p^2}{2\omega_0 c} L\right] \quad ...(14)$$

where

$$\delta_0 = \left(\frac{1}{\sqrt{\varepsilon}} - 1\right) \ll 1_{\omega=\omega_0}$$

and $\varepsilon$ is the real part of the plasma's equivalent dielectric constant and $\omega_0$ is the carrier angular frequency.

Subsequent computations of the radiated field, excited by a rectangular r.f. pulse of width $T$, disclose distortions which, in general, depend on the plasma density, transmission path length, frequency and width of the pulse. A parameter found useful in conveying this dependence, is the so-called Elliott's number, $a$,



World Radio History

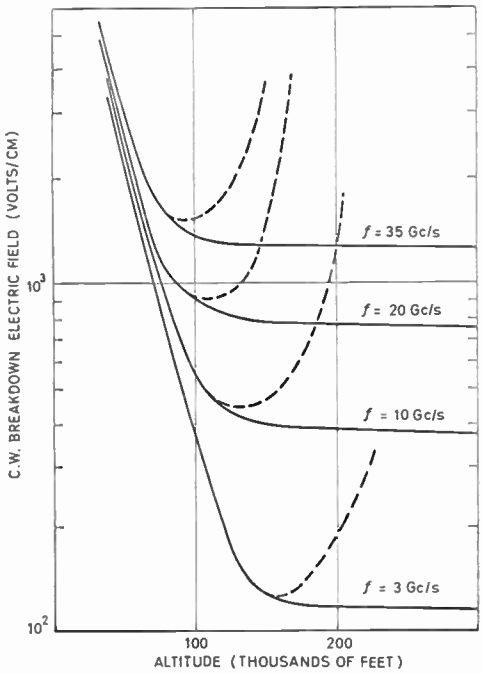



Fig. 5. C.w. breakdown threshold electric field in air. Solid line—diffusion length very large. Dashed line—diffusion length equal to half-wavelength of the incident wave.[17]

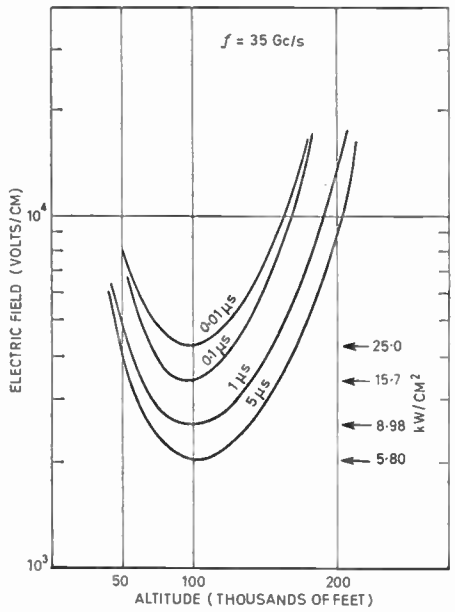



Fig. 6. Peak electric field for which over 90% of pulse is transmitted at 35 Gc/s.[17]

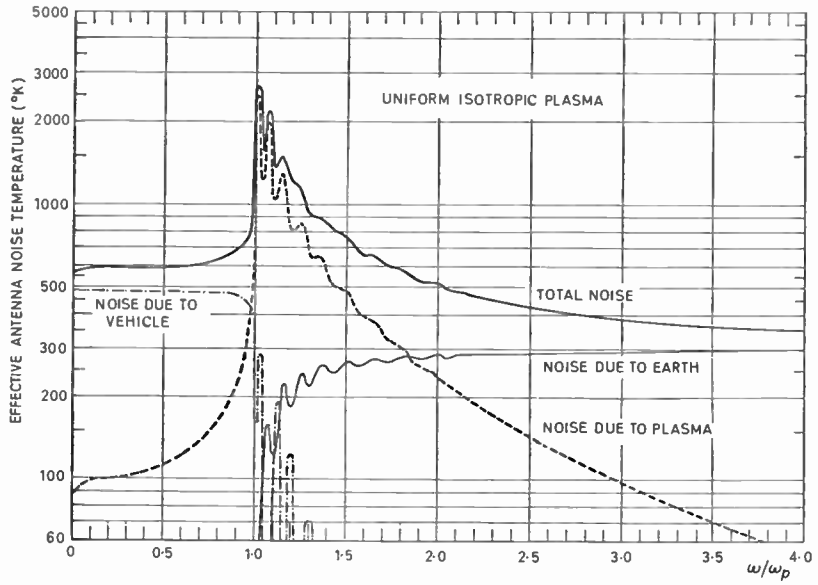



Fig. 7. Antenna noise temperature in plasma environment.[18]

$$a = \frac{8\sqrt{\pi}\delta_0}{T\omega_p}\sqrt{L/\lambda_0}\left(1+\frac{\delta_0}{2}\right) \qquad \text{.....(15)}$$

A large value of $a$ ($a \simeq 1$) indicates serious pulse distortion. An example, cited by Marquedant *et al.*[8] indicates this to be the case in a typical re-entry plasma sheath, where $L = 0{\cdot}25$ metres and $T = 10^{-9}$ seconds. An even more serious distortion can be expected at frequencies close to the plasma frequency.

In a related study, Marquedant *et al.* compute the error probabilities for a number of modulation



World Radio History

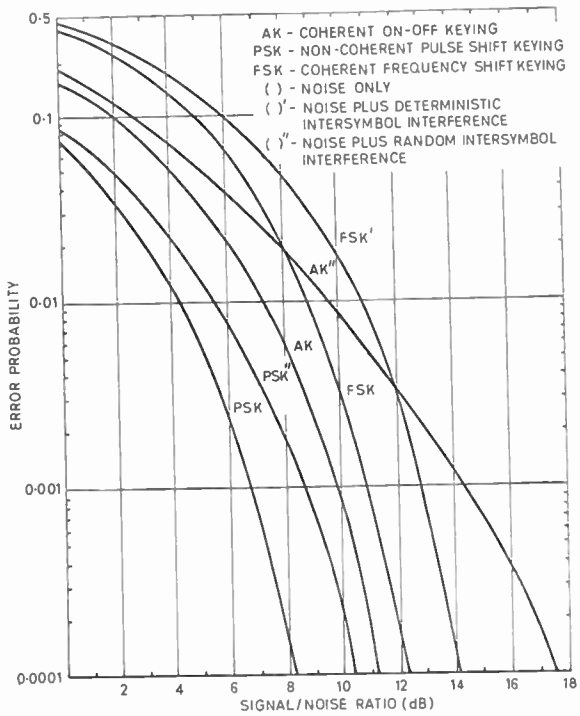



Fig. 8. Error probability vs. signal-to-noise ratio.[8]

schemes used in a plasma channel which is subject to perturbations by a Gaussian noise and also by the maximum dispersive distortion ($a = 1$). A summary of these conclusions is reproduced in Fig. 8.

## 6. Random Effects

All of the preceding discussion of plasma channels rests on the assumption that the channels are essentially static, i.e. time invariant or, at the most, varying in a predictable manner. This assumption, like so many others, constitutes a simplification dictated by the need for tractability. In reality, both natural and man-generated channels are random in that the parameters which describe them exhibit random fluctuations about their average values. The existence of these fluctuations together with spatial inhomogeneities in plasma gives rise to a number of effects which still further limit the performance of communication systems. Thus, one must reckon with fluctuations in both the phase and amplitude of a transmitted signal, with multipath and Doppler effects; also with random errors in the determination of parameters such as range and elevation angle.

To characterize a randomly inhomogeneous plasma channel one would like to know, among other things, the frequency spectrum of the fluctuations, their correlation (or lack of it) in space, scale of inhomogeneities, etc. This type of information can rarely, if at all, be obtained from consideration of first principles, since both the origin and the mechanisms responsible for randomization of plasma are obscure or, at the most, controversial. However, a good deal of information can be gathered from empirical data and from their confrontation with simple models. Examples of this can be found in the literature on ionospheric and tropospheric radio communication; there is a good reason to believe that this experience is both relevant and mandatory to the characterization of plasma channels discussed here.

### 6.1. *Phase Fluctuations*

Phase fluctuations caused by the random inhomogeneities in plasma channels are of special interest to tracking systems which make use of the principle of phase interference. The mean square value of the phase difference in an interferometer is related to the mean square of phase fluctuation at either of the terminals as follows,

$$\overline{\Delta(\phi_1-\phi_2)^2} = 2(1-\rho)\overline{\Delta\phi_i^2} \qquad \text{.....(16)}$$

Here, $\rho$ is a correlation function characterizing the dependence of the fluctuations of phases $\phi_1$ and $\phi_2$. As it can be readily shown, eqn. (16) is related to the mean square of angular fluctuations, i.e.

$$\overline{\Delta\theta^2} = \left(\frac{\lambda}{2\pi d}\right)^2 2(1-\rho)\overline{\Delta\phi_i^2} \qquad \text{.....(17)}$$

where $d$ is the length of the baseline.

The evaluation of eqn. (17) is predicated on our knowledge of the exact value, or functional dependence of $\rho$ and $\overline{\Delta\phi_i^2}$. For a lossless plasma ($\nu = 0$), Booker[19] relates the latter quantity to the variation in the elevation density along the path,

$$\overline{\Delta\phi_i^2} = 2\pi^2 r_e \lambda^2 \int \xi_0 \overline{\Delta N^2}\, dz \qquad \text{.....(18)}$$

where $r_e$ is the classical radius of electron ($r_e = 2{\cdot}8 \times 10^{-15}$ metre), and $\xi_0$ is referred to as the scale of inhomogeneities and is defined, simply, as a distance for which the correlation between the density fluctuations is equal to ($e^{-1}$). For $\overline{\Delta\phi_i^2} < 1$ this parameter can also be identified with the correlation distance, to be used in the computation of $\rho$ in eqn. (17).

### 6.2. *Amplitude Fluctuations*

A signal transmitted through a random plasma channel will also show fluctuations of its amplitude, even if the medium is lossless. It can be shown[19] that in the case of so-called 'weak' scattering (no multiple reflection, strong specular component) the mean square fractional deviation of amplitude is equal to the mean square value of the phase variation, i.e.



World Radio History

$$\left(\frac{\overline{\Delta A}}{A}\right)^2 = \overline{\Delta\phi_i^2} < 1 \qquad \text{.....(19)}$$

Implied also in the above relation are two additional assumptions: firstly, the source of illumination is supposed to be removed far from the inhomogeneities, and secondly, the observation of fluctuations takes place at distances greater than the Fresnel zone distance $z_F$ where

$$z_F = \frac{(2\pi\xi_0)^2}{\lambda} \qquad \text{.....(20)}$$

The assumption of the far zone and thus of a plane wave illumination cannot be maintained in cases where the source is close to or embedded in plasma. The required modifications of (19) for cases in which the above assumptions do not obtain, can be effected[20] by multiplying it with $(z_{\text{eff}}/z_F)^2$ where

$$z_{\text{eff}} = \frac{zz'}{z+z'} \qquad \text{.....(21)}$$

and $z, z'$ are the distances from the inhomogeneities to the source and observer, respectively, and $z < z_F$.

An important conclusion to be drawn from this modified expression is that no or only very small amplitude fluctuations can be expected in situations in which either $z$ and $z'$ is small.

### 6.3. *Correlation Bandwidth*

One of the features characterizing a wave transmitted through (or scattered by) the randomly inhomogeneous medium is its angular spectrum. It can be shown[21] that if the scale of inhomogeneities $\xi_0$ is large compared with the wavelength, this scattering will be confined to a conical beam having a width of

$$\theta_0 \simeq \frac{\lambda}{2\pi\xi_0} \qquad \text{.....(22)}$$

and whose axial direction coincides with the direction of incidence.

Since this beam width is a function of frequency, there must be a finite bandwidth over which the signals confined by the cone will still remain correlated. We can determine[5] this bandwidth by comparing the phases of the outermost signals having frequencies $f_1$ and $f_2$ and restricting their difference at the observation point to one radian, i.e.

$$\Delta\phi = \frac{2\pi}{\lambda_2}L_2 - \frac{2\pi}{\lambda_1}L_1 \leq 1 \qquad \text{.....(23)}$$

where

$$L_j \simeq z\left(1 + \frac{\theta_j^2}{2!}\right) \qquad \text{.....(24a)}$$

$$j = 1, 2 \qquad \text{.....(24b)}$$

and $z$ is the (axial) distance of the irregularities from the observer.

Now, since

$$\theta_j = \frac{\lambda_j}{2\pi\xi_0}$$

and

$$\lambda_2 = \lambda_1 + \Delta\lambda \qquad \text{.....(25)}$$

we can also write,

$$\Delta\lambda \simeq \frac{4\pi\xi_0^2}{z} \qquad \text{.....(26)}$$

or also

$$\Delta f \simeq \frac{4\pi\xi_0^2}{zc} f_1^2 \qquad \text{.....(27)}$$

Application of (27) to a large thrust rocket exhaust plasma channel at 400 Mc/s yields correlation bandwidths of the order of 10 Mc/s. Crucial to this estimate is, of course, the value of the scale size $\xi_0$ which was assumed to be equal to the transverse dimension of the plume at 150 000 ft.[22]

### 6.4. *Analytical Characterization of the Transmitted Signals*

For signal processing purposes it is desirable to represent the signal transmitted through random inhomogeneities in a reasonably simple analytical form. Such representations can rarely be derived from a consideration of the first principles (turbulence, etc.). However, useful information can be obtained from simple channel models such as, for instance, drifting irregular diffraction screen[23] or cloud of randomly disposed and agitated irregularities (frequently referred to as 'blobs').[24] Furthermore, an examination of a particular signal can yield clues as to its statistical properties and assist in the verification and/or construction of a model which will approximate the properties of a real channel.

The statistical character of the signals transmitted through random plasma channels determines the kind of information which must be searched for. Thus, one is interested in knowing the amplitude probability density function of the signal, its autocorrelation function and (frequency) power density spectrum. Frequently, information about the angular power density spectrum and the spatial autocorrelation function is also desirable since, as we have seen, it plays an important role in the assessment of the channel's bandwidth and other effects as well.

As an illustration of the results which can be obtained from such modelling, let us consider transmission of a monochromatic signal from a source (say, an artificial satellite of the earth) above the ionosphere. In addition to a large specular component, this signal will exhibit also fluctuations in both its phase and amplitude, a testimony to the fact that the ionosphere is a random channel. An analysis shows[24] that the amplitude probability density function of this signal is given by



World Radio History

$$W(A) = \frac{2A}{\sum_s \bar{E}_s^2} \exp\left(-\frac{E_0^2 + A^2}{\sum \bar{E}_s^2}\right) I_0\left(\frac{2E_0 A}{\sum \bar{E}_s^2}\right) \quad (28)$$

where $E_0$ = amplitude of the specular component,

$\bar{E}_s^2$ = mean square of the random amplitudes,

$A$ = amplitude of the combined signal,

$I_0$ is a zero-order Bessel function of imaginary argument.

For $E_0 = 0$, eqn. (28) reduces to the familiar form of the Rayleigh distribution.

To obtain the amplitude autocorrelation function of the signal it is necessary to know its frequency power spectrum. For example, if we assume the drift velocity of the 'blobs' ensemble to be zero and the distribution of the random component of velocity to be Gaussian, the autocorrelation $\rho(\tau)$ for

$$E_0 > (\sum \bar{E}_s^2)^{\frac{1}{2}}$$

is given approximately by

$$\rho(\tau) = \exp\left(-\frac{8\pi^2\tau^2}{\lambda^2} v_0^2\right) \quad \text{.....(29)}$$

where

$$v_0^2 = \overline{v_s^2}$$

Similar expressions can, of course, be obtained for a much more general case when both the drift and random velocities are included. Similarly, a spatial amplitude correlation function can be obtained when the angular power spectrum of the transmitted wave is known. Suppose again that the spectrum is Gaussian, i.e.

$$W(\theta) = \frac{W_0}{\sqrt{2\pi\theta_0}} \exp\left(-\frac{\theta^2}{2\theta_0^2}\right) \quad \text{.....(30)}$$

where

$$\theta_0^2 = \overline{\theta^2}$$

and

$$W_0 = \int_0^{\pi} W(\theta)\, d\theta$$

With eqn. (30) and on the assumption that the scatter amplitude components predominate ($E_0 < (\bar{E}^2)^{\frac{1}{2}}$) the spatial amplitude correlation function is given by

$$\rho(\xi) = \exp\left(-\frac{4\pi^2\xi^2\theta_0^2}{\lambda^2}\right) \quad \text{.....(31)}$$

As we have pointed out earlier, $\rho(\xi_0) = e^{-1}$ defines the scale of inhomogeneities $\xi_0$,

$$\xi_0 = \frac{\lambda}{2\pi\theta_0} \quad \text{.....(32)}$$

The above relations have been repeatedly confronted with the experimental data and, within the restrictions used in their derivations, found remarkably consistent and accurate. We should not construe, however, this observation to mean that similarly simple relations will be valid for other plasma channels. The chances are that they will not, but the approach and the tools used in this example are applicable no matter what the properties of an individual channel may be. (See, for example, Ref. 25.)

## 7. Conclusion

In this review an endeavour has been made to identify and to interpret some of the factors which affect microwave communication in plasma channels normally encountered in aerospace applications. No claim is being made as to the completeness of this review; however, enough details have been included to illuminate the intrinsic complexity of such channels and to indicate the scope of the required characterization. It is hoped that the impression of the telecommunication problems emerging from this review will have a stimulating effect on the future course of their investigations. It is quite clear that despite vigorous research a completely analytical characterization of real plasma channel is not possible at the present time. There are a number of parameters and effects which can only be ascertained empirically. This conclusion leads us to believe that an experimental programme, conceived along the experience charted here, merits more attention than it has received thus far. Needless to say, many of the present-day studies concerned with the attenuation by rocket exhaust, for instance, could profit from including in their programmes effects other than absorption which are of paramount importance to communication. There can be little doubt that such experiments would contribute to the solution of specific communication problems and also promote better understanding of plasma channels, in general. There are impressive precedents for this approach amply documented in the annals of ionospheric and tropospheric radio communication.

## 8. References


1. J. A. Ratcliffe, "The Magneto-Ionic Theory and its Applications to the Ionosphere". (Cambridge University Press, 1959.)
2. V. L. Ginzburg, "The Propagation of Electromagnetic Waves in Plasmas". (Pergamon Press, Oxford, 1964.)
3. K. G. Budden, "Radio Waves in the Ionosphere". (Cambridge University Press, 1961.)
4. Olen Ely, R. W. Hockenberger, P. Howell and G. Boone, "Radio Frequency Evaluation of SA-5 Vehicle", NASA Tech. Memo. X-53073, June 1964.
5. R. S. Lawrence, C. G. Little and H. J. A. Chivers, "A survey of ionospheric effects upon earth-space radio propagation", *Proc. Inst. Elect. Electronics Engrs*, 52, p. 4, 1964.
6. J. A. Stratton, "Electromagnetic Theory", Chapter IX. (McGraw-Hill, New York, 1941.)
7. H. Staras, "The propagation of wideband signals through the ionosphere", *Proc. Inst. Radio Engrs*, 49, p. 1211, 1961.
8. R. Marquedant, C. M. Knop and H. Hodara, "Performance of digital signals through plasmas", *Trans. I.E.E.E. on Communication Systems*, CS-12, p. 74, 1964.




World Radio History

9. M. P. Bachynski, "Electromagnetic wave penetration of re-entry plasma sheaths", *J. Res. Nat. Bur. Stand.* (*U.S.A.*), 69D, No. 2, p. 147, February 1965.

10. V. L. Ginzburg and A. V. Gurevich, "Nonlinear phenomena in a plasma located in an alternating electromagnetic field", (English translation) in *Soviet Physics-Uspekhi* (*U.S.A.*), 3, No. 2, pp. 175–194, September–October 1960.

11. R. J. Papa, "The nonlinear interaction of an electromagnetic wave with a time-dependent plasma medium", *Canadian J. Physics*, 43, p. 38, 1965.

12. R. J. Papa, "Radio Frequency Propagation Through an Inhomogeneous, Magnetoactive, Nonlinear Plasma Medium", Air Force Cambridge Res. Lab. (Bedford, Mass.), Research Paper No. 49, September 1964.

13. J. Galejs, "Slot antenna impedance for plasma layers", *Trans. I.E.E.E. on Antennas and Propagation*, AP-12, p. 738, 1964.

14. J. Galejs, "Admittance of annular slot antennas radiating into a plasma layer", *J. Res. Nat. Bur. Stand.* (*U.S.A.*), 68D, No. 3, pp. 317–323, March 1964.

15. M. P. Bachynski, G. G. Cloutier and B. W. Gibbs, "Millimeter Wave Antennas in Ionized Media", RCA Victor (Montreal, Canada) Res. Rep. No. 7-801-21, October 1962.

16. R. F. Whitmer and A. D. MacDonald, "Radio Frequency Breakdown Conditions in the Presence of Plasma Sheath", Sylvania (Mountain View, Cal.) Report MPL-B3, May 1959.

17. A. D. MacDonald, "High frequency breakdown in air at high altitudes", *Proc. I.R.E.*, 47, p. 436, 1959.

18. M. P. Bachynski, "Antenna noise temperature in plasma environment", *Proc. I.R.E.*, 49, p. 1846, 1961.

19. H. G. Booker, "The use of radio stars to study irregular refraction of radio waves in the ionosphere", *Proc. I.R.E.*, 46, p. 298, 1958.

20. A. Hewish, "The diffraction of radio waves in passing through a phase changing ionosphere", *Proc. Roy. Soc.*, 209, p. 81, 1951.

21. H. G. Booker and W. E. Gordon, "Theory of radio scattering in the troposphere", *Proc. I.R.E.*, 38, p. 401, 1950.

22. T. Kaliszewski, "Disturbances in the radio frequency links caused by a rocket exhaust". Paper in preparation, 1965.

23. J. A. Ratcliffe, "Diffraction from ionosphere and the fading of radio waves", *Nature*, 162, 9th July 1948.

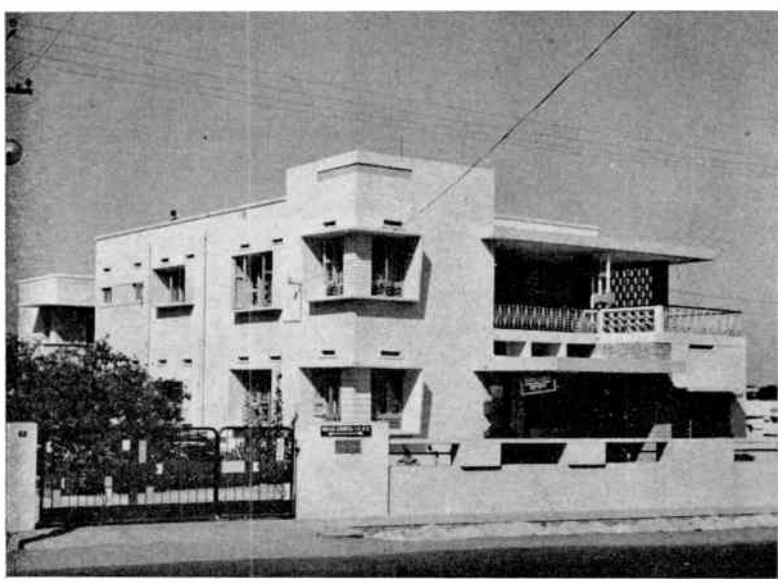
24. Y. L. Alpert, "Radio Wave Propagation and the Ionosphere" pp. 31–81 (Consultants Bureau, New York, 1963.)

25. A. A. Geiger, "Analysis of the effects of rocket exhaust on FSK telemetry", 6th Solid Propellant Rocket Conf., Washington, D.C., AIAA paper No. 65-184, February 1965.

*Manuscript first received by the Institution on 28th January 1965 and in final form on 9th September 1965.* (*Paper No. 1028.*)

© The Institution of Electronic and Radio Engineers, 1966

## Indian Division of the Institution



The Institution's new office in Bangalore is situated at 7 Nandidurg Road (Telephone: Bangalore 29640). Mrs. Anand, the Administrative Secretary at this address, will deal with membership subscriptions and applications.



World Radio History

# Radio Engineering Overseas . . .

*The following abstracts are taken from Commonwealth, European and Asian journals received by the Institution's Library. Abstracts of papers published in American journals are not included because they are available in many other publications. Members who wish to consult any of the papers quoted should apply to the Librarian, giving full bibliographical details, i.e. title, author, journal and date, of the paper required. All papers are in the language of the country of origin of the journal unless otherwise stated. Translations cannot be supplied.*

## ERROR DETECTION

The early detection of errors in data, for example during input and transmission, is possible when the data are provided with check characters which form an integral part of the data. These check characters should be formed from the original data in such a way that as many errors as possible are detectable. The conditions under which all commonly occurring types of errors, such as single-character errors and 'interchange errors', can be detected with certainty are given in a Dutch paper and it is shown that the residual-error probability, which is inevitable in any checking method, can in this way be made very low (e.g. about $10^{-5}$: with one check character and less than $10^{-7}$ with two). The use of suitable principles for the formation of the check characters allows this process to be carried out very simply in practice by an electronic system. Data verification units working on this principle are described, which serve not only to determine the check character but also to carry out the actual checking of the data, by means of a zero check. This gives optimum checking with one or two check characters using very simple equipment.

'Data checking during input and transmission by means of one or two check characters', G. Renelt and J. Schröder, *Philips Technical Review*, 26, No. 4/5/6, pp. 156–63, 1965. (In English.)

## RANGE GATED M.T.I.

Studies are made in a Swedish paper of a coherent moving target indication system with range gating, two slightly different alternative systems having been studied. System 1 has no device for pulse lengthening and Doppler filtering is introduced immediately after the range gating. System 2 has pulse lengthening after the range gating and the Doppler filtering takes place after the pulse lengthening.

The studies revealed that in System 1 the signal/noise ratio in one of the range channels is determined principally by the limitation of bandwidth immediately prior to the non-linear detection. The influence of a large bandwidth prior to the range gating is of little significance. The optimal length of gate is approximately 0·9 of the pulse length $t_p$, when the pulses are regarded as rectangular.

System 2 gives a signal/noise ratio which is dependent on the frequency of the Doppler shift. The mean signal/noise ratio in the range channel is at most as large as the signal/noise ratio before the lengthening of the pulse.

'Detectability of signal in noise when using range gated m.t.i. radar', A. Törby, *Ericsson Technics*, 21, No. 2, 1965 pp. 241–68. (In English.)

## NETWORK ANALYSIS BY COMPUTER

In an Australian paper methods of analysing passive and active linear networks are assessed for application in automatic computation. A versatile digital computer program is described, in which the elements of a three-terminal two-port network are specified by element value and terminating nodes, and nodal elimination techniques are employed to compute the overall performance. Applications to networks with up to 20 independent nodes are illustrated, together with some discussion on computational error.

'Digital computer analysis of three-terminal two-port networks', J. V. Fall, *Proceedings of the Institution of Radio and Electronics Engineers Australia*, 26, No. 5, pp. 167–73, May 1965.

## PSYCHOPHYSIOLOGICAL CONSIDERATIONS IN STEREOPHONY

The author of a Czech paper describes the results of psychophysiological and physical tests made to determine the influence of the microphone position and the distance of the sound source upon true localization, as well as the localization dependance upon the parameters in pseudo-stereophony. The tests enabled a check to be made on the acoustical pressure course of microphones in relation to sound location and a comparison of the qualitative effects in pseudo-stereophony with those in three-channel stereophonic systems. The optimal recording conditions in three-channel stereophonic systems were verified by disturbing the recorded word with noise and spoken word. Conclusions of the psychophysiological evaluation were drawn from 89 000 tests.

'The determination of acoustical conditions for sound recording in three channel stereophonic systems', F. Kolmer, *Slaboproudý Obzor*, 26, No. 6, pp. 321–30, June 1965.

## ANALOGUE-DIGITAL CONVERSION

Following a brief description of known principles for analogue-digital conversion, a new principle, the so-called division method, is outlined in a German paper. It is shown that the successive division of an analogue quantity by a natural number and subsequent subtraction of the remainder produced in each division, yields an analogue-digital conversion. The theoretical outline is complemented by a brief discussion of the practical realization of such converters, and some experimental results are given.

'A new method of analogue-digital conversion', P. Leuthold, *Archiv der Elektrischen Übertragung*, 19, No. 9, pp. 453–58, September 1965.



World Radio History