THE RADIO AND ELECTRONIC ENGINEER

The Journal of the Institution of Electronic and Radio Engineers

Economics and the Engineer

THE views of engineers on economic policy are often discounted on the grounds that scientific and technical training has no relevance to the understanding of economics. Such misconception may be a contributory reason for the small number of engineers in Parliament and, compared with other professions, the relative minority of engineers who are active company Directors. This in spite of the indefatigable endeavour of engineers to reduce costs in research, design, and production. Indeed, the economics of production and the control of costs were the basis of the Institution's recent conference on 'The Integration of Design and Production in the Electronics Industry'.

Production cost is an integral part of national economics. Failure to recognize this fact creates serious unbalance of the entire economy of any industrial nation.

Just as the philosophy of integration can be applied in circuit design and between design and production, so can it operate between the member firms of an industry and, on a broader scale, between the many industries on which the well-being of a country depends. An engineer does not, and cannot, work in isolation. The internal organization of his industry and the way his industry fits into the overall economy of the country are of paramount importance to him. It is for these reasons that the recent announcement of a proposed merger between two large British companies in the electronics industry represents a development of concern to every engineer, affecting as it does, the economic prospects of his profession and industry.

The computer field is an outstanding example of the need for rationalization of production because it is a highly capitalized sector of the electronics industry. Indeed, the National Institute of Economic and Social Research, and the Economics Development Committee for Electronics, have already pointed out that in a rapidly progressing scientifically based industry such as that of electronic capital goods, competition mainly takes the form of technical innovation and technical service to customers. Any firm which wishes to survive must be capable at least of matching within a short time any major innovation made by its more advanced competitors. To do this, it must have a certain research and development capacity, and if it falls below this defensive threshold of R & D it will fail.

Similar industrial mergers in the corresponding sectors of the French electronics industry are also being fostered so as to create industrial units which can work within the economic limits imposed by the buyer. The incentive of competition within highly capitalized, science based industries, has proved insufficient to ensure profitable consumer progress in both Britain and France. Notwithstanding the capacity of the larger consumer market in America, there are indications that similar pressures are operating there—hence American interest in the European market.

Rationalization for good economic reasons does not necessarily mean the end of the inventive skill of the smaller company from which the original radio industry sprung. The ability to assess and successfully use the craft and skills of emerging industries within the larger production units calls increasingly for the 'practical' economics of the engineer as distinct from the planning of theorists.

G. D. C.

INSTITUTION NOTICES

Special General Meeting

A Special General Meeting of Corporate Members is being convened at 9 Bedford Square, London, W.C.1, in the near future for the purpose of amending certain of the Institution's Bye-Laws relating to the designation and requirements for membership.

A formal communication will be sent shortly to all corporate members, giving statutory notice of the Special General Meeting and setting out the precise wording of the amendments. Briefly, however, the changes will mean that with effect from a date to be stated, those who are at present 'Members' will be known as 'Fellows', and 'Associate Members' as 'Members'. In addition the requirements for entry to the class of senior technicians possessing educational qualifications of a standard not less than that of Higher National Certificate or the City and Guilds Full Technological Certificate, or such other similar qualifications as the Council may prescribe.

Third U.K.A.C. Control Convention

The third U.K.A.C. Control Convention, on 'Advances in Control of Systems', will be held at the University of Leicester from 2nd to 4th April 1968. It is being organized by the Society of Instrument Technology in conjunction with the other Engineering Institutions and Societies of the United Kingdom Automation Council. The I.E.R.E. has been represented on the Organizing Committee by Mr. W. E. Willison (*Associate Member*).

The theme of this Convention is primarily concerned with advances in the control of commercial and industrial systems. The subject has been divided into the following categories:

Systems Theory

Design methodology; statistics; information theory; queueing; games and decision theory; theory of graphs and networks; mathematical programming; computer and symbolic programming; and other appropriate subjects.

Applications

Industrial Systems—manufacturing systems (such as production control, network planning, stock control, distribution and routing scheduling, plant loading, part programming and numerical control, design automation, etc.); process systems (including real-time control of plant in petroleum, chemical, steel, food, electric power and other appropriate industries).

Commercial Systems—economics, industrial dynamics, banking, insurance, data collection and distribution, reservations, etc., with particular stress on real-time systems. Also organization modelling and planning, resource allocation and forecasting.

Contributions of between 2,000 and 4,000 words are invited, and synopses of approximately 300 words should be submitted to the Convention Secretariat, The Society of Instrument Technology, 20 Peel Street, London, W.8 by 15th September 1967. Full texts of contributions should be submitted by 15th November.

Registration forms and further information for those wishing to attend the Convention will be available a little later in the year from I.E.R.E., 9 Bedford Square, London, W.C.1.

I.F.A.C. Symposium on Automatic Control

Advice has been received that the International Symposium on Computer Control of Natural Resources and Public Utilities, which is being organized by the Israel Committee for Automatic Control, under the auspices of the International Federation for Automatic Control, will take place, as planned, at Technion City, Haifa, from 11th to 14th September 1967.

The subjects to be discussed are: computer control of water supply, road traffic, oil refineries and power stations; automatically controlled telephone systems; and theoretical aspects of automatic control.

Registration forms and further details may be obtained on application to the Honorary Secretary, United Kingdom Automation Council, c/o Institution of Electrical Engineers, Savoy Place, London, W.C.2. Members wishing to attend the Symposium are advised to register without delay.

Proceedings of Thin Films Conference

The final Proceedings of the Conference on 'Applications of Thin Films in Electronic Engineering', which was held at Imperial College, London, in July 1966, is now published. This revised edition contains all papers presented at the Conference, including some which were not preprinted, and also Reports of Discussions, Corrigenda and List of Delegates.

Those persons who received the first edition, either at the time of the Conference, or afterwards, will be sent, free of charge, the Supplement Volume which contains all the additional matter.

Copies of the complete, Revised Edition may now be purchased from the Institution's Publications Department, 8–9 Bedford Square, London, W.C.1, price £6.

Index to Volume 33

This is the first issue of Volume 34 of *The Radio and Electronic Engineer*. An Index to Volume 33, which covered the period January–June 1967, will be included with the August issue.

Amplitude Limiting Applied to a Sensitive Correlation Detector

By

M. J. YERBURY, B.Sc.(Eng.)

(Graduate)†

Summary: It is shown that the stability of a sensitive correlation detector can be improved by limiting the amplitudes of the voltages in either one or both input channels. The loss of output signal/noise ratio compared to an ideal multiplier-correlator is determined as a function of limiting level for both cases.

The output signal/noise ratio of an unstable multiplier-correlator can be considerably less than the theoretical maximum. Thus, depending upon the source of the instability, it is possible to increase the output signal/noise ratio until an optimum value is reached by the technique of amplitude limiting. Further advantages of this technique are discussed in the paper.

List of Symbols

I_i	scale factor of the limiter, $i = 1, 2$
$\overline{R(\sigma)}$	average output of a square law device after synchronous rectification
T	integration time
$V_i(t)$	output voltage of channel i
$w_i(f)$	spectral density of background noise in the output of channel i
x	an arbitrary variable
σ	r.m.s. value of the noise input to a limiter
σ_i	r.m.s. value of the limiter in channel i
σ_T	r.m.s. noise output of a correlator
τ	time-delay between the two input voltages

$\psi_{,ii}(x)$	auto-correlation function of the background noise in the output of channel i
$\psi_{Lii}(x)$	auto-correlation function of the noise output of a limiter in channel i
$\psi_{12}(\tau)$	cross-correlation function of the undistorted outputs of channels 1 and 2
$\psi_{12}(0)$	value of $\psi_{12}(\tau)$ at $\tau = 0$
$\psi_{ii}(0)$	value of $\psi_{ii}(x)$ at $x = 0$
$\psi_{L1,L2}(\tau)$	cross-correlation function of the outputs of limiters in channels 1 and 2
$\psi_{L1,2}(\tau)$	cross-correlation function of the output of a limiter in channel 1 and the undistorted output of channel 2

1. Introduction

The technique of cross-correlation is widely used to detect small signals obscured by noise, for example in the fields of radio and optical astronomy, acoustics and geophysics. Such varied fields of investigation demand correspondingly diverse methods of processing the incoming data. Thus, when electronic correlation systems are used, their specifications must depend greatly on the particular requirements of each application and, as a result, it is difficult to generalize in any discussion on the relative merits of correlation systems.

The work described here was initiated by a need to improve on the stability of an analogue correlator which forms part of the optical stellar interferometer[1] at Narrabri Observatory, New South Wales.

---

† Cornell-Sydney University Astronomy Centre, Narrabri, New South Wales.

A simplified block diagram of the correlation system is shown in Fig. 1. Due to the extremely small signal/noise ratio encountered ($\sim -60\,\mathrm{dB}$), it is usually necessary to integrate for several hours before the signal can be estimated to within an accuracy of $\pm 5\%$. Therefore an important parameter in what follows is the signal threshold or, what amounts to the same thing, the output signal/noise ratio in a given time.

2. Correlator Instability

Instability in this context exists when with two uncorrelated inputs, samples of the correlator output at any chosen time have either a non-Gaussian probability distribution or a Gaussian distribution with a non-zero mean.

Instability of the first type is characterized by short term drifts in the correlator output voltage with an equal probability of having positive or negative directions. The second type of instability is a steady,

Fig. 1. Block diagram of the correlator.

unidirectional drift in the correlator output voltage which either detracts from or enhances the signal output. Ideally, this drift should be made negligibly small compared to the smallest expected value of the integrated correlation. If it cannot be made small, a simple method of dealing with this problem would be to measure the drift accurately and use this result as a zero error correction on subsequent signal estimates. This has two major disadvantages. The first is that the drift may not remain constant over a long period of time or even over the relatively short period of a few hours corresponding to a single integration. The second disadvantage arises from the fact that any one estimate of the drift requires at least the same integration time as that used in estimating the signal and preferably several times this period. This is because the uncertainty in the drift measurement contributes to the overall uncertainty in the signal measurements and thus degenerates the output signal/noise ratio. Also if any of the time required by the instrument in estimating its own zero error could be used in measuring the signal, this would constitute a further loss of signal/noise ratio.

The correlator described here suffers from drift due to amplitude modulation in the phase-reversing process and subsequent detection in the non-linearity of the analogue multiplier. Another source of drift

is the presence of unidirectional, non-Gaussian pulses at the input, whose amplitude considerably exceeds the r.m.s. noise background. It was felt that the effectiveness of both sources of instability could be reduced by the use of amplitude limiting.

3. Amplitude Limiting

It is possible to apply amplitude limiting in either one channel or two channels of the correlator and some work has already been published[2-7] concerning the effect on the output signal/noise ratio for certain types of correlator, of two channel infinite clipping. This work does not take into account the fact that infinite clipping cannot be reproduced perfectly in a practical device. Among other things, the precision with which this process is effected depends greatly on the spectral density of the input to the device. For example, in the correlator described here the advantages often associated with infinite clipping, namely comparison of signs and ease of computation, do not apply. According to the work of Rice,[8] the expected number of zero crossings of the voltage in one channel of the correlator is about 150×10^6 in one second. At the present state of development of electronic components and techniques it is unlikely that an accurate zero-crossing sensor capable of dealing with such an input could be constructed. Further, processing binary information arriving at this rate by a digital computer does not, at this time, seem to be a practical proposition. In such cases it is desirable to examine the effect of limiting at an arbitrary level on the correlator stability and output signal/noise ratio.

A non-linear operation such as amplitude limiting might be expected to cause a loss of signal/noise ratio, since some details of the nature of the signal are conveyed by the amplitude fluctuations in each channel. However, the determination of the actual loss in signal/noise ratio is complicated by the fact that the effectiveness of the noise background is also reduced by the limiting. In the case discussed here, drift can be considered to cause a loss of output signal/noise ratio. Thus a correlator possessing instability may have a poor output signal/noise ratio when compared with the ideal multiplier-correlator. If amplitude limiting is applied to a correlator so as to reduce the drift there may be an overall improvement in output signal/noise ratio even when the inherent loss due to the limiting process is taken into account.

The mathematical theory which follows examines the effect on the output signal/noise ratio of a multiplier-correlator when amplitude limiting occurs in one or both input channels. The last section of the theory is devoted to an assessment of the effect of limiting on the drift caused by amplitude modulation of the input voltages.

4. The Multiplier-Correlator

4.1. Signal Output

The process carried out by the correlator is

$$\psi_{12}(\tau) = \lim_{T \to \infty} \frac{1}{T} \int_0^T V_1(t) . V_2(t+\tau) . dt$$

$$= \overline{V_1(t) . V_2(t+\tau)} \qquad \ldots\ldots(1)$$

where

$$V_1(t) = V_{s1}(t) + V_{n1}(t) \qquad \ldots\ldots(2)$$

$$V_2(t+\tau) = V_{s2}(t+\tau) + V_{n2}(t+\tau) \qquad \ldots\ldots(3)$$

and where the bar denotes an ensemble average taken over the random components of $V_1(t)$ and $V_2(t+\tau)$. This is equal to the time average in the limit $T \to \infty$ if $V_1(t)$ and $V_2(t+\tau)$ are stationary. In practice, the correlator integrates for a finite time T and, due to the presence of a large amount of noise, the result of the integration is likely to be in error by an amount which can be estimated.

The signal output of the multiplier-correlator will be defined as $T . \psi_{12}(\tau)$.

4.2. Noise Output

The r.m.s. value of the fluctuations in the integral

$$I = \int_0^T V_1(t) . V_2(t+\tau) . dt$$

is required; the mean square value of this is written

$$\overline{I^2} = \int_0^T dt_1 \int_0^T dt_2 \overline{V_1(t_1) . V_1(t_2) . V_2(t_1+\tau) . V_2(t_2+\tau)} \qquad \ldots\ldots(4)$$

In this work the input signal/noise ratio is taken to be very small (Sect. 8) so that the contribution of the correlated components of $V_1(t)$ and $V_2(t+\tau)$ to eqn. (4) is negligible. Accordingly it is permissible to write

$$\overline{I^2} = \int_0^T dt_1 \int_0^T dt_2 \overline{V_1(t_1) . V_1(t_2)} . \overline{V_2(t_1+\tau) . V_2(t_2+\tau)}$$

$$= \int_0^T dt_1 \int_0^T dt_2 \psi_{11}(t_2-t_1) . \psi_{22}(t_2-t_1)$$

which reduces to

$$\overline{I^2} = 2 \int_0^T (T-x) . \psi_{11}(x) . \psi_{22}(x) \, dx \qquad \ldots\ldots(5)$$

This is a general result which can be used to find the r.m.s. deviation in the integral I as a function of the time of integration T.

Representing the mean square deviation of the integral by σ_T^2 gives

$$\sigma_T^2 = \overline{(I - \bar{I})^2}$$

However, $\bar{I}$ is negligibly small since the background noise voltages have zero mean value, giving

$$\sigma_T^2 = \overline{I^2} \qquad \ldots\ldots(6)$$

The Wiener-Kintchine theorem states that

$$\psi_{ii}(x) = \int_0^\infty w_i(f) . \cos 2\pi f x . df \qquad \ldots\ldots(7)$$

so that substituting eqn. (7) into eqn. (5) with $i = 1, 2$ and integrating with respect to x gives

$$\sigma_T^2 = \tfrac{1}{2} \int_0^\infty w_1(f_1) . df_1 \int_0^\infty w_2(f_2) . df_2 \times$$

$$\times \left[\frac{\sin^2 \pi(f_1+f_2)T}{\pi^2(f_1+f_2)^2} + \frac{\sin^2 \pi(f_1-f_2)T}{\pi^2(f_1-f_2)^2} \right] \ldots\ldots(8)$$

This equation can be considerably simplified by noting that the first term in the square brackets is negligibly small compared with the second term when the lowest frequencies passed by the amplifiers exceed a few kilohertz and where T is much greater than the period of the lowest frequency. Further, the values of f_2 for which the second term in the brackets differs significantly from zero are grouped closely around f_1 so that as far as the first integration with respect to f_2 is concerned, $w_2(f_2)$ can be assumed to be independent of f_2 and a function of f_1 only. Equation (8) becomes,

$$\sigma_T^2 = \tfrac{1}{2} \int_0^\infty w_1(f_1) . w_2(f_1) \, df_1 \times$$

$$\times \frac{T}{\pi} \int_{-\infty}^\infty \frac{\sin^2 \pi(f_1-f_2)T}{\pi^2(f_1-f_2)^2 T^2} . \pi T \, df_2$$

$$= \frac{T}{2} \int_0^\infty w_1(f) . w_2(f) . df \frac{1}{\pi} \int_{-\infty}^\infty \frac{\sin^2 X}{X^2} . dX$$

Thus,

$$\sigma_T^2 = \frac{T}{2} \int_0^\infty w_1(f) . w_2(f) . df \qquad \ldots\ldots(9)$$

Note that the r.m.s. noise output is proportional to the square root of the time of integration T.

4.3. The Signal/Noise Ratio

The signal/noise ratio at a time T will be defined for the multiplier-correlator as

$$\left(\frac{S}{N}\right)_0 = T \frac{\psi_{12}(\tau)}{\sigma_T} \qquad \ldots\ldots(10)$$

a similar definition will apply to other systems where $\psi_{12}(\tau)$ will be replaced by the particular cross-correlation function and σ_T will be used to represent the r.m.s. noise output at the time T for that system.

Thus for the multiplier-correlator

$$\left(\frac{S}{N}\right)_0 = \sqrt{2T} \cdot \frac{\psi_{12}(\tau)}{\left[\int_0^\infty w_1(f) \cdot w_2(f) \cdot df\right]^{\frac{1}{2}}} \quad \text{......(11)}$$

The signal threshold is reached at a time T which makes the right-hand side of eqn. (11) equal to unity.

5. The Amplitude Limited Correlator

5.1. *Mathematical Representations*

Several methods are available for dealing with non-linear operations on Gaussian noise. A general technique has been discovered by Price[9] which can be used to obtain all the cross-correlation functions used in this paper so that detailed analyses of particular cases will not be presented here. Figure 2 shows three possible mathematical representations for a limiter. Of these, the abrupt clipper of Fig. 2(a) leads to an enclosed form for the correlation function only in the special case of infinite clipping whereas the limiter of Fig. 2(b) does not yield a definite correlation function.[10]

The representation of Fig. 2(c), due to Baum[11] and termed 'smooth limiting' is useful, in that it leads to a definite and relatively simple correlation function. This representation is

$$I = I_i \sqrt{\sigma_i} \operatorname{erf} \frac{V}{\sqrt{2\sigma_i}} \quad \text{......(13)}$$

a clear understanding of which may be had when it is realized that σ_i is the value of $|V|$ at which the slope of the integrand has its maximum rate of change. It can be thought of as the level at which distortion is becoming significant and is known as the r.m.s. value of the limiter. It is also worth noting that for the region close to the origin where distortion is negligible the slope of the function is $\sqrt{2/\pi} \cdot I_i$.

If the representation is used to obtain the cross-correlation function of the outputs of two limiters with inputs $V_1(t)$ and $V_2(t+\tau)$ respectively, this becomes

$$\psi_{L1,L2}(\tau) = \frac{2}{\pi} \sigma_1 \sigma_2 I_1 I_2 \times$$

$$\times \arcsin\left\{\frac{\psi_{12}(\tau)}{\sqrt{[\psi_{12}(0) + \psi_{11}(0)][\psi_{12}(0) + \psi_{22}(0)]}} \times \right.$$

$$\left. \times \frac{1}{\sqrt{(1+\alpha')(1+\beta')}}\right\} \quad \text{......(14)}$$

where

$$\alpha' = \frac{\sigma_1^2}{\psi_{12}(0) + \psi_{11}(0)} \quad \text{......(15)}$$

$$\beta' = \frac{\sigma_2^2}{\psi_{12}(0) + \psi_{22}(0)} \quad \text{......(16)}$$

The quantities α' and β' relate the input noise levels to the so-called r.m.s. values of the limiters and will be referred to as the normalized r.m.s. limiting levels. If the r.m.s. noise inputs are made much greater than the corresponding r.m.s. values of the limiters so that both α' and β' tend to zero, the expression (14) reduces to that of the infinite clipper, apart from the constant term $\sigma_1 \sigma_2 I_1 I_2$.
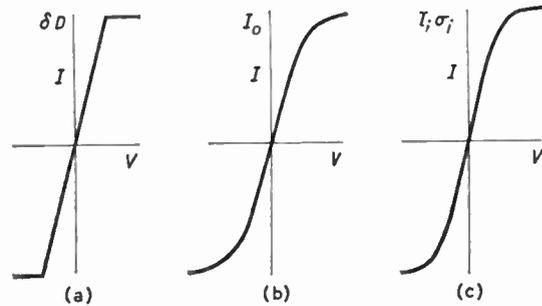
Fig. 2. Three different forms of amplitude limiter.

(a) $I = \frac{\delta}{j\pi} \int_a \exp(jVu) \sin Du \cdot \frac{du}{u^2} - \delta D.$

(b) $I = I_0 \tanh \frac{V}{C}.$

(c) $I = I_i \sqrt{\sigma_i} \operatorname{erf} \frac{V}{\sqrt{2\sigma_i}}$ (after Baum[11]).

In this latter case, it is seen that the argument of the arcsine is the ratio of two quantities, each of which is ideally proportional to the product of the gains of any amplifiers preceding the limiters, so that the measured value of $\psi_{L1,L2}(\tau)$ should be unaffected by frequency independent gain variations in these amplifiers. This is discussed further in a later section.

5.2. *Output Signal*

If it is assumed that the input signal/noise ratio is very small, then from eqn. (14)

$$\psi_{L1,L2}(\tau) \simeq \psi_{12}(\tau) \cdot \frac{2}{\pi} I_1 I_2 \left[\frac{\alpha'\beta'}{(1+\alpha')(1+\beta')}\right]^{\frac{1}{2}} \quad \text{......(17)}$$

and the signal output of the correlator with limiting on both input channels is given by $T \cdot \psi_{L1,L2}(\tau)$. If $\alpha' = \beta' = \alpha$, for example by a suitable choice of σ_1, σ_2 or of the gains of the two input channels, then

$$\psi_{L1,L2}(\tau) \simeq \psi_{12}(\tau) \cdot \frac{2}{\pi} I_1 I_2 \cdot \frac{\alpha}{1+\alpha} \quad \text{......(18)}$$

The smooth limiting representation enables the cross-correlation function to be studied as a function of the limiting level, including the hypothetical case

of infinite clipping. Further, the shape of the limiting function is probably closer to that of two semiconductor diodes connected as a limiter than the infinite clipping function which hitherto was the only one amenable to calculation.

5.3. Output Noise

The noise output of the correlator with amplitude limiting on both channels is found by using eqn. (5). In this case

$$\overline{I^2} = 2 \int_0^T (T-x) . \psi_{L11}(x) . \psi_{L22}(x) \, dx \quad ......(19)$$

where

$$\psi_{L11}(x) = \sigma_1^2 \frac{2}{\pi} I_1^2 \arcsin \left[\frac{\psi_{11}(x)}{\psi_{11}(0)} . \frac{1}{(1+\alpha)} \right] \quad ......(20)$$

$$\psi_{L22}(x) = \sigma_2^2 \frac{2}{\pi} I_2^2 \arcsin \left[\frac{\psi_{22}(x)}{\psi_{22}(0)} . \frac{1}{(1+\alpha)} \right] \quad ......(21)$$

and where, as before, the contribution of the small correlated component to the noise output has been neglected. It is now necessary to substitute eqns. (20) and (21) into eqn. (19) and solve the resulting integral.

The method adopted is an extension of that used for the multiplier-correlator. Initially the arcsines are expanded into power series and the terms resulting from the product of the series are grouped according to their order. Thus with

$$\rho_1 = \frac{\psi_{11}(x)}{\psi_{11}(0)} . \frac{1}{1+\alpha} \quad ......(22)$$

and

$$\rho_2 = \frac{\psi_{22}(x)}{\psi_{22}(0)} . \frac{1}{1+\alpha} \quad ......(23)$$

the right-hand side of eqn. (19) becomes

$$\overline{I^2} = \frac{8}{\pi^2} . (\sigma_1 \sigma_2 I_1 I_2)^2 \left\{ \int_0^T (T-x)\rho_1\rho_2 \, dx + \right.$$

$$+ \int_0^T (T-x) . \frac{(\rho_1\rho_2^3 + \rho_2\rho_1^3) \, dx}{6} +$$

$$\left. + \int_0^T (T-x) \left[\frac{\rho_1^3\rho_2^3}{36} + \frac{3(\rho_1\rho_2^5 + \rho_2\rho_1^5)}{40} \right] dx + ... \right\} \quad ......(24)$$

Representing these integrals by $J_1, J_2 ... J_n$, etc., gives

$$\overline{I^2} = \frac{8}{\pi^2} . (\sigma_1 \sigma_2 I_1 I_2)^2 (J_1 + J_2 + ... J_n + ...) \quad ......(25)$$

A method for solving the integrals J_n is given in the Appendix, where it is found necessary to assume particular forms for the spectral densities $w_1(f)$ and $w_2(f)$. In this case they are taken to be rectangular, extending over a range of frequencies from f_A to f_B with $f_B \gg f_A$. In practice the two amplifiers will not be identical and $f_B - f_A$ will be their mutual bandwidth.

Finally, from eqns. (24), (25) and (57) it is clear that the expansion of $(\arcsin \phi)^2$ can be used as a generating function, the relevant coefficients in the resulting summation being obtained by putting $\phi = 1/(1+\alpha)$. Hence,

$$\sum_{n=1}^{\infty} J_n = \frac{T}{4} \int_{f_A}^{f_B} \frac{w_1 . w_2 \, df}{\psi_{11}(0)\psi_{22}(0)} \sum_{n=1}^{\infty} \frac{[(n-1)!]^2}{n(2n-1)!} \times$$

$$\times \frac{2^{2n-2}}{(1+\alpha)^{2n}} . \frac{(2n)!}{(n!)^2} . \left(\frac{1}{2} \right)^{2n-1}$$

$$= \frac{T}{4} \int_{f_A}^{f_B} \frac{w_1 . w_2 \, df}{\psi_{11}(0) . \psi_{22}(0)} \sum_{n=1}^{\infty} \frac{1}{n^2} . \left(\frac{1}{1+\alpha} \right)^{2n} \quad ......(26)$$

Substituting eqn. (26) into eqn. (25) gives the mean square noise at the output of the dual channel amplitude limited correlator as

$$\sigma_T^2 = \overline{I^2} = \alpha^2 (I_1 I_2)^2 \frac{2T}{\pi^2} \int_{f_A}^{f_B} w_1 . w_2 \, df \sum_{n=1}^{\infty} \frac{1}{n^2} . \left(\frac{1}{1+\alpha} \right)^{2n} \quad ......(27)$$

where w_1, w_2 are the spectral densities of the input noise voltages to the respective limiters, assumed constant over the frequency range f_A to f_B and zero outside this range.

5.4. Signal/Noise Ratio

As before the signal/noise ratio is defined by

$$\left(\frac{S}{N} \right)_1 = T . \psi_{L1,L2}(\tau) \quad ......(28)$$

which from eqns. (18) and (30) becomes

$$\left(\frac{S}{N} \right)_1 = \sqrt{2T} \frac{\psi_{12}(\tau)}{\left[\int_0^{\infty} w_1 . w_2 \, df \sum_{n=1}^{\infty} \frac{1}{n^2} . \frac{1}{(1+\alpha)^{2n-2}} \right]^{\frac{1}{4}}} \quad ......(29)$$

Equation (29) thus gives the signal/noise ratio at a time T for the correlator with amplitude limiting on both input channels as a function of the limiting level through α. The two extreme cases $\alpha \to 0$ and $\alpha \to \infty$ may be examined, corresponding to infinite clipping and zero clipping respectively. In the former

the infinite summation becomes

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$$

so that

$$\left(\frac{S}{N}\right)_1 \bigg|_{\alpha \to 0} = \sqrt{T} \cdot \frac{2\sqrt{3}}{\pi} \cdot \frac{\psi_{12}(\tau)}{\left[\int_0^{\infty} w_1 w_2 \, df\right]^{\frac{1}{2}}} \quad \ldots\ldots(30)$$

in the latter case the infinite summation can be written

$$\sum_{n=1}^{\infty} \frac{1}{n^2} \cdot \frac{1}{(1+\alpha)^{2n-2}} = 1 + \sum_{n=1}^{\infty} \frac{1}{(n+1)^2} \cdot \frac{1}{(1+\alpha)^{2n}}$$

which reduces to unity as $\alpha \to \infty$, giving

$$\left(\frac{S}{N}\right)_1 \bigg|_{\alpha \to \infty} = \left(\frac{S}{N}\right)_0 = \sqrt{2T} \frac{\psi_{12}(\tau)}{\left[\int_0^{\infty} w_1 w_2 \, df\right]^{\frac{1}{2}}} \quad \ldots\ldots(31)$$

5.5. Comparison with Multiplier-Correlator

A useful measure for comparison is the loss of signal to noise ratio at a given time T which occurs when limiting is applied to the multiplier-correlator. A factor F is therefore introduced defined by

$$F = 10 \log_{10} \frac{\left(\frac{S}{N}\right)_1}{\left(\frac{S}{N}\right)_0} \, \text{dB} \quad \ldots\ldots(32)$$

which gives the loss of signal/noise ratio compared with the multiplier-correlator.

Hence for the amplitude-limited correlator

$$F = -5 \log_{10} \left[1 + \sum_{n=1}^{\infty} \frac{1}{(n+1)^2} \cdot \frac{1}{(1+\alpha)^{2n}}\right] \, \text{dB} \quad \ldots\ldots(33)$$

which in the case of infinite clipping reaches its maximum value

$$F_{max} = -5 \log_{10} \frac{\pi^2}{6} \, \text{dB} = -1 \cdot 081 \, \text{dB} \quad \ldots\ldots(34)$$

This represents a surprisingly small loss when it is considered that all the signal information carried by the amplitudes of the input waves is rejected. Inspection of Fig. 3, a plot of eqn. (33) shows that the loss falls quite rapidly with increase in the level at which limiting becomes significant.

6. Amplitude Limiting in One Channel

Due to the extreme sensitivity requirements of the correlator discussed here it was found necessary to include phase switching on both input channels. In less exacting applications single-channel phase switching may be adequate. Also in the particular case of the stellar interferometer, the correlator is driven by the outputs of two photomultiplier tubes where relatively large undirectional pulses are evident, making double-channel limiting desirable. In many cases it may well be advantageous to use single-channel limiting while double-channel limiting is unnecessary. Further, in any correlator with a non-linear device in one channel and inputs consisting of a correlated signal and uncorrelated output noise (not necessarily Gaussian) the signal output will be linearly proportional to the cross-correlation function of the input voltages. This follows from Bussgang's work on the cross-correlation of the input and output of a non-linear device.[12]

6.1. Output Signal

In this case a limiter is put in channel 1 of the correlator and channel 2 is a linear amplifier. Price's theorem[9] yields for the correlation function

$$\psi_{L1,2}(\tau) = \sqrt{\frac{2}{\pi}} \cdot I_1 \cdot \frac{\psi_{12}(\tau)}{\sqrt{1 + \frac{\psi_{11}(0) + \psi_{12}(0)}{\sigma_1^2}}} \quad \ldots\ldots(35)$$
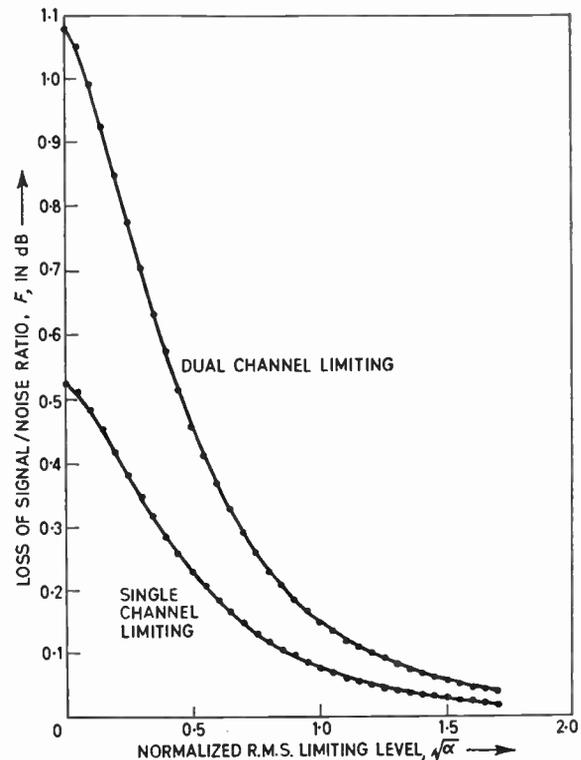
Fig. 3. Loss of signal/noise ratio as a function of limiting level.

which, from eqn. (15) may be written

$$\psi_{L1,2}(\tau) = \sqrt{\frac{2}{\pi}} \cdot \psi_{12}(\tau) \cdot I_1 \cdot \sqrt{\frac{\alpha'}{1+\alpha'}} \qquad \ldots\ldots(36)$$

If σ_1 tends to zero, keeping $\sigma_1 I_1 = I_0$, a constant, the correlation function for the case of infinite clipping in one channel of the correlator is obtained in the form

$$\psi_{L1,2}(\tau) = \sqrt{\frac{2}{\pi}} \cdot \frac{I_0 \psi_{12}(\tau)}{\sqrt{\psi_{11}(0)+\psi_{12}(0)}} \qquad \ldots\ldots(37)$$

Equation (37) shows that the measured correlation function should be independent of uniform gain variations in one channel of the correlator excluding any changes in the limiter output ($\pm I_0$). Comments on this point appear later.

For analogue correlators operating with relatively low input frequencies this system presents interesting possibilities in that it is merely required to invert the phase of one channel when the voltage in the other channel is negative and to return the phase of the former to normal when the latter becomes positive. Such a device will be termed a cross-phase correlator.

6.2. Output Noise

In order to calculate the r.m.s. noise output σ_T in this case, eqn. (5) is used in the form

$$\sigma_T^2 = 2 \int_0^T (T-x) \cdot \psi_{L11}(x) \cdot \psi_{22}(x) \cdot dx \qquad \ldots\ldots(38)$$

where $\psi_{L11}(x)$ is the autocorrelation function of the output of the limited channel and is given by eqn. (20). As before the presence of the small correlated component has been neglected.

Equation (38) is reduced to the form shown below by arguments similar to those used in Sect. 5.3, thus

$$\sigma_T^2 = \alpha' I_1^2 \frac{4T}{\pi} \int_0^\infty w_1 \cdot w_2 \cdot df \sum_{n=1}^\infty \frac{1}{[4(1+\alpha')]^{2n-1}} \times$$

$$\times \frac{1}{n} \cdot \frac{[(2n-2)!]^2}{[(n-1)!]^4} \qquad \ldots\ldots(39)$$

where α' is given by eqn. (15).

6.3. Signal/Noise Ratio

Equations (36) and (39) give as the signal/noise ratio for the single limited correlator

$$\left(\frac{S}{N}\right)_2 = \sqrt{\frac{T}{2(1+\alpha')}} \times$$

$$\times \frac{\psi_{12}(\tau)}{\left\{\int_0^\infty w_1 \cdot w_2 \cdot df \sum_{n=1}^\infty \frac{1}{[4(1+\alpha')]^{2n-1}} \cdot \frac{1}{n} \frac{[(2n-2)!]^2}{[(n-1)!]^4}\right\}^{\frac{1}{2}}} \qquad \ldots\ldots(40)$$

It is worth while considering the special case of the cross-phase correlator which arises when

$$\alpha' \to 0$$

The summation in eqn. (40) becomes

$$\sum_{n=1}^\infty \frac{1}{4^{2n-1}} \cdot \frac{1}{n} \cdot \frac{[(2n-2)!]^2}{[(n-1)!]^4} = \frac{1}{\pi}$$

so that

$$\left(\frac{S}{N}\right)\Big|_{\alpha \to 0} = \sqrt{\frac{\pi T}{2}} \cdot \frac{\psi_{12}(\tau)}{\left[\int_0^\infty w_1 \cdot w_2 \cdot df\right]^{\frac{1}{2}}} \qquad \ldots\ldots(41)$$

6.4. Comparison with the Multiplier-Correlator

The factor F (Sect. 5.5) giving the loss of signal/noise ratio compared with the multiplier-correlator is

$$F = 10 \log_{10} \frac{\left(\dfrac{S}{N}\right)_2}{\left(\dfrac{S}{N}\right)_0}$$

Thus,

$$F = -5 \log_{10} \left\{ 4(1+\alpha') \sum_{n=1}^\infty \frac{1}{[4(1+\alpha')]^{2n-1}} \times \right.$$

$$\left. \frac{1}{n} \frac{[(2n-2)!]^2}{[(n-1)!]^4} \right\} \text{ dB} \qquad \ldots\ldots(42)$$

This has its maximum value in the case of the cross-phase correlator of

$$F_{max} = -5 \log_{10} \frac{4}{\pi} \text{ dB} = -0.525 \text{ dB} \qquad \ldots\ldots(43)$$

This loss of 11·4% is about half that occurring in the polarity coincidence correlator. For the general case of smooth limiting in one channel the loss given by eqn. (42) is plotted in Fig. 3 as a function of the normalized r.m.s. limiting level

$$\sqrt{\alpha'} = \frac{1}{\sqrt{\psi_{12}(0)+\psi_{11}(0)}}$$

and is seen to be almost exactly half the loss incurred by dual channel limiting over the entire range of limiting levels.

7. Square-law Detection of the Output of a Limiter

The inevitable consequence of using a practical limiting device is that some amplitude-modulated

noise will pass through it and after detection in the multiplier non-linearity will lead to drift at the correlator output.

The question as to how effective a limiter will be in reducing the drift due to this source can be simply resolved if the autocorrelation function of the output of the limiter is known and if the multiplier non-linearity is assumed to obey a square-law. The function of the square-law device and integrator is to find a quantity whose rate of change is proportional to the mean square output of the limiter.

The phase reversing process is assumed to produce square-wave amplitude modulation of the noise input to the limiter, thus the r.m.s. value σ, of the input is also in general square-wave modulated and the output of the phase-sensitive rectifier will be a direct current related to the change in σ over a phase reversal.

The autocorrelation function of the output of the limiter is

$$\psi_{Lii}(x) = \frac{2}{\pi}(\sigma_i I_i)^2 \arcsin\left[\frac{\psi_{ii}(x)}{\psi_{ii}(0)} \cdot \frac{1}{1+\alpha}\right] \quad \ldots\ldots(44)$$

where $\alpha = \left(\frac{\sigma_i}{\sigma}\right)^2$.

The average value of the output of the square law device is therefore

$$\overline{R(\sigma)} = k\psi_{Lii}(0)$$
$$= \frac{2k}{\pi}(\sigma_i I_i)^2 \arcsin\left(\frac{1}{1+\alpha}\right) \quad \ldots\ldots(45)$$

where k is a constant of the square-law device.

If the modulation depth is small, then

$$\Delta\overline{R(\sigma)} \simeq \Delta\sigma \cdot \frac{\partial\overline{R(\sigma)}}{\partial\sigma}$$

and hence

$$\Delta\overline{R(\sigma)} \simeq \Delta\sigma \cdot \frac{4k}{\pi}I_i^2\sigma_i \cdot \frac{\alpha}{(1+\alpha)\sqrt{2+\alpha}} \quad \ldots\ldots(46)$$

The drift in the correlator output voltage due to amplitude modulation is proportional to $\overline{R(\sigma)}$. This can be reduced in several ways, one of which is to insert some extra gain between the modulating source and the detector. If, for the sake of simplicity, this extra gain is assumed to have a flat frequency response within the band, of amplitude $A \gg \sigma_i/\sigma$, then it can be shown that the rejection of drift defined by

$$\text{rejection} = \frac{\text{drift without limiting}}{\text{drift with limiting}} - 1 \ldots\ldots(47)$$

becomes

$$\text{rejection} \simeq \sqrt{2}A \cdot \left(\frac{\sigma}{\sigma_i}\right)^3 \quad \ldots\ldots(48)$$

When interpreting eqn. (48) the r.m.s. value of the input is taken to be fixed. In the case where two semiconductor diodes are connected to form a limiter, the minimum value of σ_i is restricted by the inherent saturation voltage of the diodes so that normally the rejection will be improved by inserting more gain between the modulator and detector.

8. Conclusions

If a signal and a noise voltage each have a spectral density which extends more or less uniformly over a very wide frequency band, then the design of a correlation detector should be aimed at accepting two samples of the signal in the uncorrelated noise background over as wide a bandwidth as possible and at maintaining a stable zero datum for as long a period of time as possible.

Broadly speaking, the signal/noise ratio at the correlator output will be proportional to the square root of the product of the correlator bandwidth and time of integration so that, in theory at least, it would be possible to detect the presence of any signal, no matter how small, by integrating long enough in the presence of a signal. In practice the limit to the correlator sensitivity is set by instability in the output zero datum so that it is desirable to increase the bandwidth as far as it is practicable.

It has been shown that, for a wideband correlation detector, instability, if it is caused by amplitude modulation of the input voltages and subsequent detection in multiplier non-linearity, can be reduced by amplitude limiting. There is a loss of signal/noise ratio at the output of the correlator due to limiting, which is greater in the dual channel limiting case, than with single channel limiting. However, there may be an overall gain in signal/noise ratio if the drift without limiting is such that it must be measured at the expense of time reserved for signal detection, or if the uncertainty in the drift contributes significantly to that in the signal. Theoretically, the drift from the above-mentioned source can be reduced to zero by infinite clipping, causing a loss of about 1·08 dB of signal/noise ratio with dual-channel clipping and about 0·525 dB loss with single-channel clipping. The loss falls off quite rapidly (Fig. 3) as the degree of limiting is reduced but an increasing amount of instability is to be expected as a result.

The method adopted here in calculating the output signal/noise ratio can also be applied to other examples provided that the correlation functions of the noise background voltages after passage through a non-linear device can be put into closed forms.

The signal output of the two-channel-limited correlator is not linearly proportional to that of the multiplier-correlator. The maximum departure from linearity can occur when both channels are infinitely

clipped, although even in this case the signal output is linear to within $\pm 0.1\%$ if the input signal/noise ratio does not exceed about $0.25 : 1$.

In using the smooth limiting function of Baum,[11] the effect of departure from the hypothetical case of infinite clipping can be examined. This is particularly advantageous in analogue correlators with high-frequency inputs. An application to low-frequency correlators arises from consideration of single-channel limiting where it is merely necessary to invert the phase of the voltage in one channel when the voltage in the other channel is negative, returning the former to normal when the latter is positive. This has been referred to here as a cross-phase correlator and has the advantage that an analogue multiplier is not required.

A further advantage appears when one or both input channels are infinitely clipped. The correlation function in these cases is normalized to the mean square noise in the clipped channel or channels respectively. Thus any relatively slow gain variations which occur in the clipped channel or channels and which do not alter the shapes of the frequency responses, will not affect the measured value of the output signal.

For the smooth limiting examples in which the correlation functions are not completely normalized to the noise backgrounds, gain variations affect the signal output through α' and β'. These can be made very small and in fact the method adopted in calculating the rejection of modulation can also be applied to estimating the immunity of the measured signal output against gain variations.

The approximation made in the Appendix to simplify eqn. (55) causes the calculated noise outputs of both types of correlator to be slightly higher than the true values. The error is a maximum for dual-channel infinite clipping, where it can be shown to lie between $+2\%$ and $+3\%$. This error falls rapidly as the limiting level is increased. For the cross-phase correlator the errors are roughly halved. It follows that the estimated losses of signal/noise ratio are proportionately high. McNeil[13] has shown independently by numerical integration that the losses obtained for dual channel and single channel infinite clipping are 0.97 dB and 0.46 dB respectively. These results are respectively 2.7% and 1.5% lower than those derived here, agreeing closely with the above estimates of the errors.

9. Acknowledgments

10. References

1. R. Hanbury Brown, 'The stellar interferometer at Narrabri observatory', *Sky and Telescope*, **28**, No. 2, August 1964.

2. J. J. Faran, Jr., and R. Hills, Jr., 'Correlators for Signal Reception', Harvard University Acoustics Research Laboratory, Tech. Memos, Nos. 27 and 28, September and November 1952.

3. E. J. Riisnaes, 'The Effect of Clipping and Sampling on the Performance of some Signal Detection Systems', Admiralty Research Laboratory, A.R.L./N30/L, December 1958.

4. J. B. Thomas and T. R. Williams, 'On the detection of signals in non-stationary noise by product arrays', *J. Acoust. Soc. Amer.*, **31**, p. 453, 1959.

5. S. Weinreb, 'A Digital Spectral Analysis Technique and its Application to Radio Astronomy', Massachusetts Institute of Technology, Tech. Report, No. 412, August 1963.

6. H. S. Heaps, 'Maximum error caused by using completely clipped functions in the computation of Fourier transforms and correlation functions', *Quart. Appl. Maths*, **19**, p. 321, 1962.

7. J. H. Laning and R. H. Battin, 'Random Processes in Automatic Control', (McGraw-Hill, New York, 1956).

8. S. O. Rice, 'Mathematical analysis of random noise', I–IV, *Bell Syst. Tech. J.*, **23**, pp. 282–332, July 1944 and **24**, pp. 47–156, January 1945.

9. R. Price, 'A useful theorem for non-linear devices having Gaussian inputs', *Trans. Inst. Radio Engrs on Information Theory*, IT-4, No. 2, pp. 69–72, June 1958.

10. J. H. Van Vleck, 'The Spectrum of Clipped Noise', Report No. 51, Radio Research Laboratory, Harvard University, 21st July 1943.

11. R. F. Baum, 'The correlation function of smoothly limited Gaussian noise', *Trans. I.R.E.*, IT-3, No. 3, pp. 193–7, September 1957.

12. J. J. Bussgang, 'Cross-Correlation Functions of Amplitude Distorted Gaussian Signals', Res. Lab. of Electronics, M.I.T., Cambridge, Mass., Tech. Rep. 216, Sect. 3; 26th March 1952.

13. D. R. McNeil (private communication).

11. Appendix

The evaluation of I_n necessitates solving integrations of the form

$$I_n = [\psi_{11}(0)]^{-l}[\psi_{12}(0)]^{-m}(1+\alpha)^{-2n} \int_0^T (T-x) \times$$

$$\times \left[\tfrac{1}{2} \int_{-\infty}^{\infty} w_1(f)\cos 2\pi fx \, . df \right]^l \times$$

$$\times \left[\tfrac{1}{2} \int_{-\infty}^{\infty} w_2(f)\cos 2\pi fx \, . df \right]^m dx$$

$$\ldots\ldots(49)$$

where $l+m = 2n$ and l, m are odd integers. These integrations arise when the substitutions

$$\rho_1 = [\psi_{11}(0)(1+\alpha)]^{-1} \tfrac{1}{2} \int_{-\infty}^{\infty} w_1(f)\cos 2\pi f x \,.\, \mathrm{d}f$$

$$\ldots\ldots(50)$$

and

$$\rho_2 = [\psi_{22}(0)(1+\alpha)]^{-1} \tfrac{1}{2} \int_{-\infty}^{\infty} w_2(f)\cos 2\pi f x \,.\, \mathrm{d}f$$

$$\ldots\ldots(51)$$

are made in eqn. (24); note that $w_1(f) = w_1(-f)$, $w_2(f) = w_2(-f)$. Equation (49) becomes

$$I_n = [\psi_{11}(0)]^{-l}[\psi_{22}(0)]^{-m}(1+\alpha)^{-2n} \times$$

$$\times (\tfrac{1}{2})^{2n} \prod_{i=1,\,j=l+1}^{i=l,\,j=2n} \int_{-\infty}^{\infty} w_1(f_i)\,.\,\mathrm{d}f_i \times$$

$$\times \int_{-\infty}^{\infty} w_2(f_j)\,.\,\mathrm{d}f_j \int_{0}^{T} (T-x)\cos 2\pi f_i x\,.\,\cos 2\pi f_j x\,.\,\mathrm{d}x$$

$$\ldots\ldots(52)$$

Expanding the products of the $2n$ cosines on the right-hand side of this equation and integrating with respect to x as in Section 4.2 gives for $n > 1$,

$$I_n = [\psi_{11}(0)]^{-l}[\psi_{22}(0)]^{-m}(1+\alpha)^{-2n} \times$$

$$\times (\tfrac{1}{2})^{2n} \prod_{i=2,\,j=l+1}^{i=l,\,j=2n} \int_{-\infty}^{\infty} w_1(f_i)\,.\,\mathrm{d}f_i \times$$

$$\times \int_{-\infty}^{\infty} w_2(f_j)\,.\,\mathrm{d}f_j \int_{-\infty}^{\infty} w_1(f_1)\,.\,\mathrm{d}f_1 \times$$

$$\times (\tfrac{1}{2})^{2n} T^2 \sum_{r=1}^{2^{2n-1}} \left(\frac{\sin X_r}{X_r}\right)^2$$

$$\ldots\ldots(53)$$

where

$$\left.\begin{array}{ll} X_1 & = \pi(f_1+f_2+\ldots+f_{2n})T \\ X_2 & = \pi(f_1+f_2+\ldots-f_{2n})T \\ \cdot & \\ \cdot & \\ X_{2^{2n-1}} & = \pi(f_1-f_2-\ldots-f_{2n})T \end{array}\right\} \ldots\ldots(54)$$

It is now possible to integrate with respect to f_1 when it is noticed that the contribution of the $(\sin X_r/X_r)^2$ terms may be neglected except where X_r lies in the region of zero. One further requirement is that $w_1(f_1)$ is a relatively slowly varying function of f_1 so that $w_1(f_1)$ can be considered constant over that range of f_1 corresponding to $X_r \simeq 0$. The integration is performed using the fact that

$$\int_{-\infty}^{\infty} \left(\frac{\sin X}{X}\right)^2 \,.\, \mathrm{d}X = \pi$$

giving, after some rearrangement

$$I_n = [\psi_{11}(0)]^{-1} \,.\, (1+\alpha)^{-2n} \frac{T}{4} \int_{-\infty}^{\infty} w_1(f_2)\,\mathrm{d}f_2 \times$$

$$\times \prod_{i=3,\,j=l+1}^{i=l,\,j=2n} \int_{-\infty}^{\infty} \eta_1(f_i)\,\mathrm{d}f_i \times$$

$$\times \int_{-\infty}^{\infty} \eta_2(f_j)\,.\,\eta_1(f_2+f_3+\ldots+f_{2n})\,\mathrm{d}f_j$$

$$\ldots\ldots(55)$$

where

$$\eta_k(f) = \frac{w_k(f)}{2\psi_{kk}(0)} = \frac{w_k(f)}{\int_{-\infty}^{\infty} w_k(f)\,.\,\mathrm{d}f} \qquad k = 1, 2$$

$$\ldots\ldots(56)$$

Consider the case where $w_k(f)$ is a uniform spectral density is defined in Section 5.3. When used in eqn. (55) $w_k(f)$ takes the value w_k for frequencies extending from $-f_B$ to $+f_B$ but is zero outside this range and in a small region $-f_A$ to $+f_A$, where $f_A \ll f_B$. It is possible to evaluate these integrals, in principle, for any finite value of n but the calculations become very tedious for $n > 3$. However, the same result is obtained in each case by treating f_2 as the sum of $(2n-1)$ independent random variables each with zero mean and with probability density function given by the appropriate value of $\eta_k(f)$. In this way the problem is reduced to finding the probability that f_2 lies between the limits defined by $w_1(f_2)$. For $n \gg 1$ this probability becomes $\sqrt{3/\pi n}$.

An approximate solution can now be found since the desired probability is closely related to the probability that out of the $(2n-1)$ variables, exactly n are positive or exactly n are negative. This probability is given by

$$\frac{(2n)!}{(n!)^2} \,.\, (\tfrac{1}{2})^{2n-1}$$

Using this approximation, eqn. (55) reduces to

$$I_n \simeq \frac{T}{4} \frac{\int_{f_A}^{f_B} w_1 w_2 \,\mathrm{d}f}{\psi_{11}(0)\psi_{22}(0)} \cdot \frac{(2n)!}{(n!)^2} \cdot (\tfrac{1}{2})^{(2n-1)} \left(\frac{1}{1+\alpha}\right)^{2n}$$

$$\ldots\ldots(57)$$

The effect of the approximation on the accuracy of the result is discussed in Section 8.

Space Research at University College London

Scientific space research in this country began in 1953 when plans were prepared for the development of a small sounding rocket to carry into the upper atmosphere instruments built by University groups. Prominent amongst these was the group in the Department of Physics at University College London led by Professor R. L. F. Boyd, under the general direction of Professor Sir Harrie Massey, F.R.S. This group played a pioneering role in the British space research programme and was responsible for two-thirds of the experiments in *Ariel I*, the first Anglo-American satellite. The group has since had experiments in two N.A.S.A. *Explorer* satellites, and future experiments will be carried in no less than eight satellites and many more high altitude rockets over the next four years.

The potential effectiveness of the University College group has now been greatly increased by the establishment of an outstation at Holmbury St. Mary, near Dorking, some 30 miles south of London. The purchase of a country mansion, Holmbury House, was made possible by a donation of £65 000 by Mullard Ltd. in 1965 and the new laboratory for the space research group came into operation in October last year. It is to be known as the Mullard Space Science Laboratory, and a formal opening ceremony was performed on 3rd May 1967 by Dr. F. E. Jones, F.R.S., C.Eng., F.I.E.E., M.I.E.R.E., managing director of Mullard, who recalled the origins of the British space research programme with which he had been associated in his capacity of Deputy Director of the Royal Aircraft Establishment. Dr. Jones paid particular tribute to the world-wide reputation which the space research group of the already famous Physics Department of UCL had achieved. His company had believed that they could fill a gap by helping finance certain university research activities which could not properly be financed from other sources.

The six laboratories at Holmbury House provide facilities for 23 scientists and seven research students. Other staff include 26 laboratory and workshop technicians and 11 administrative workers. Current annual expenditure is supported largely by a grant of about £150 000 per annum from the Science Research Council.

Practical work in the Space Research Group is largely devoted to originating new measurement methods, including detection devices, and to calibration of the experiments before flight. A great deal of development work and the production of mechanical, optical and electronic hardware is put out to contract at Government establishments and industry to relieve the load on the Group's own support services. The really scientific, as distinct from technological, work is in the interpretation and analysis of the data obtained. Data analysis will be eased by the installation at the Laboratory of an IBM 1130 computer which will operate on-line via a data link with a large IBM 360 computer at University College. The data link is expected to be in operation next year.

The main investigations carried out by the Laboratory are in the fields of ultra-violet and x-ray astronomy both of the sun and of other celestial objects, and in the study of the ionosphere, the magnetosphere and the interplanetary medium. While the primary aim of the experiments is to increase the fundamental knowledge of the world around us, the findings are expected to have a bearing on such widely different spheres as satellite communications and space travel.

Following the success of the experiments in *Ariel I*, the National Aeronautics and Space Administration in the U.S.A. has provided facilities for satellite experiments comparable with those given to American universities: UCL experiments were carried on *Explorer XX*, launched in 1964, and DME-A (*Explorer XXXI*) launched in November 1965. The UCL experiments are still operating and supplying information on electron temperatures and ion mass spectra which is the main data source for the Laboratory today. Sounding rocket flights continue to be provided through the Ministry of Technology and Woomera Rocket Range in Australia, and it is certain that the Mullard Space Science Laboratory will be one of the largest users of the satellite and rocket facilities now being built up by the European Space Research Organization (ESRO).

The future programme involves experiments on eight satellites and over thirty sounding rockets. Launchings will be under the auspices of the domestic U.K. programmes, the ESRO programme and the U.S.–U.K. co-operative programme. The investigations will be a logical extension of the work already completed and will include new work on ultra-violet and x-ray astronomy both of the sun and other celestial objects. The u.v. and x-ray studies are part of what is called the 'New Astronomy' which now accounts for over 50% of the current expenditure. The ionospheric studies, however, still have a great number of individual-experiments. Britain's long interest in the ionosphere has placed her in a commanding position in this and associated areas of study. In fact, ionospheric studies have been dubbed 'a British institution'. Being so well established in this field, it is probable that future research at UCL will continue on logical extensions of the current work.

The time scale of space research is necessarily long. It has taken a decade to develop and perfect what may be called first generation experiments. The techniques of measurement in a very difficult environment and interpretation of results have now been learnt and are being applied to current work. Even so, quick or dramatic results cannot be expected. As experiments become more and more complex and equipment more refined the time scale from inception of an experiment to its successful analysis of data may still extend over a period of up to seven years.

The next two years' work is already clearly defined and planned but, working on a time scale of seven years, future experiments are already in the planning stage. The new laboratories, the acquisition of expensive capital equipment, and the promise of continuing finance reflects at UCL the new order of stability now appearing in the British space research programme.

The Radio and Electronic Component Show, 1967

The twentieth Radio and Electronic Component Show, sponsored by the Radio and Electronic Component Manufacturers' Federation, was held at Olympia, London, from 23rd to 26th May. With over 300 exhibitors, occupying a floor space of 100 000 ft², it was the largest component exhibition yet held in this country. Space does not allow a comprehensive review of the exhibition, but among the many interesting exhibits were the following.

Digital Coded Typewriter (*National Physical Laboratory*). While there is an increasing need for typewriters producing an electrically-coded output, conventional methods of achieving this tend to be costly, particularly when conversion of an existing typewriter is considered.

Recent work at N.P.L. has been directed towards a cheap and simple method of conversion. The basis of this method is that each typewriter hammer has coded markers on its side. These markers are sensed by a photo-electric reading head as the hammer strikes the platen.

The reading head consists of a number of semiconductor photo-diodes mounted in a transparent block, behind which there is a lamp. This assembly is mounted near the point where the typing hammers enter the guide immediately prior to striking the platen. When a hammer moves near the head, light from the lamp is reflected from the coding markers and picked up by the corresponding photo-diodes. A pair of photo-diodes mounted on the opposite side of the guide registers the presence of a hammer travelling in the correct direction, i.e. towards the platen.

The method has a number of advantages over conventional procedures: (a) most typewriters can be converted without mechanical alteration; (b) any code can be generated; (c) coding is very easily changed or corrected; (d) assembly is not critical.

Flat Strip Cable Connectors (*Signals Research and Development Establishment*). The flat cable, which can at present carry up to a maximum of 19 circuits, has the advantage over traditional cable of greatly increased heat transfer characteristics and is robust, flexible, compact and light. Its copper conductors are visible in the plastic sheath.

The initial development of the basic connector developed by S.R.D.E. was to meet military requirements, but it offers vast possibilities in the civil field; for instance, in civil aviation, data processing, telephone and communications equipment and possibly even domestic wiring.

The connectors, which 'plug into' the cable, can be positioned at any point. The S.R.D.E. connector is intended to provide a base to which can be bolted control boxes and service outlets. Miniature plugs and sockets being developed will allow for flexible connection.

High-power Travelling Wave Tubes (*Standard Telephones Cables Ltd.*). The Valve Division of S.T.C. has introduced a high-power t.w.t. for use in Bands IV and V television transmitters and transposers, f.m. sound transmitters and link amplifiers.

Designated the W45B/5E, the tube operates in the 470 to 960 MHz frequency band and provides 200 W for transmitter service, or 50 W for common-sound-and-vision transposer service. A saturation output of 500 W can be obtained in pulsed service at duty ratios up to 10%. The tube is operated in permanent magnet mounts which have a lower external magnetic field and permit easy replacement of tubes under field conditions. A feature of the W45B/5E is that all power supplies, including the heater supply, can be switched on simultaneously. Thus it can be employed in remote-controlled equipment.

Precision Voltage Reference Source (*Mullard Ltd.*). What is believed to be the first available voltage reference 'diode' capable of maintaining a voltage within limits only previously obtained by using a standard cell was shown as a working exhibit.

Development type 84BZY, the new diode is available in three versions, having temperature coefficients of 0·001% per degC, 0·002% per degC and 0·005% per degC respectively. All are intended for general industrial applications and for use in laboratory instruments. The 'diode' comprises a resistor, transistor and Zener diode fabricated using integrated circuit techniques. In this circuit the temperature coefficient of the voltage between the base and emitter of the transistor is backed-off against the temperature coefficient of the Zener diode. A value of resistance (R) is chosen so that (I_R) is several times the base current (I_B) of the transistor, ensuring that the current in the Zener diode is relatively unaffected by the total device current.

Thermoelectric Generators (*The Plessey Co. Ltd.*). Semiconductor thermoelectric modules made from iron disilicide were shown by the Chemical and Metallurgical Division of the Plessey Co. The material is hot pressed into a monolithic ceramic mass capable of operating in direct flame at up to 900°C. There are only two external connections, both on the cold side.

Smallest is the TEG1 flame failure detector, which produces 25 mW at 200 mV for 500°C between hot and cold junctions—five times the power and ten times the voltage of the metal to metal thermocouples normally used for this purpose. Larger modules give up to 400 mW at 600 mV.

Flat-faced Display C.R.T. (*Ferranti Ltd.*). The type 24/48L3MS2, all-glass, flat-faced c.r.t. has been specially developed for use in large area direct-view display systems. It has a nominal faceplate diameter of 600 mm (24 in), while the height of the centre of the face above the edge (the chord height) is only 9 mm (0·35 in). With this degree of flatness, the radius of curvature is five times greater than the diameter of the face. The tube shown had an L3 phosphor which is designed for fine symbol presentation and is capable of providing flicker-free images down to 5 Hz depending on viewing conditions.

Mechanisms in Gunn Effect Microwave Oscillators

By

J. E. CARROLL,
Ph.D.†

Summary: Four microwave mechanisms found in Gunn diodes are discussed. Firstly, growing space-charge waves give a frequency dependent negative resistance which can be used to obtain amplification as well as to make oscillators. When any region or domain of high electric field is formed in a Gunn diode, three other microwave mechanisms are found: a current or voltage generator is created by the collapse of the domain if it drifts out of the diode at the anode contact; a negative resistance is sustained by the presence of the domain; finally, d.c. can be converted into r.f. by parametric mechanisms. This last effect is caused by the modulation of the circuit reactance due to the insertion and removal of the capacitance associated with the high-field domain. An equivalent circuit is discussed and is used to explain how these effects contribute to the various modes of oscillation exhibited by Gunn diodes.

List of Principal Symbols

C	circuit capacitance
C_D	domain capacitance
C_0	sample capacitance at low electric fields
D	diffusion coefficient $\sim 200 \text{ cm}^2/\text{s}$ at E_c for GaAs
e	electronic charge, $1 \cdot 6 \times 10^{-19}$ coloumbs
E_c	critical electric field for domain formation $\sim 3\text{kV/cm}$ for GaAs
I_c	critical current for domain formation
I_D	current shunting domain
$I_a(t)$	current generator from domain discharging
I_0, I_1	d.c. and r.f. components of current respectively
$k =$	v_s/v_c, ($1/k$ is peak/valley ratio)
l	specimen length
l_D	domain length
n	donor density $\sim 1 \cdot 1 \times 10^{15}$ donors/cm³ for 1 ohm-cm GaAs
L	circuit inductance
Q	charge (see Appendix)
Q_0	quality factor for low field sample
Q_D	quality factor for domain
R_0	sample resistance at low electric fields
$-R_D$	negative domain resistance
R_L, R	circuit resistance
v_0, v_1	electron velocity; d.c. and r.f. components

v_c	critical velocity for domain formation
v_s	saturated electron velocity at high electric fields
V_0, V	applied d.c. voltage across diode
V_c	critical voltage for domain formation
$V_a(t)$	voltage generator [series equivalent to $I_a(t)$]
ε	permittivity of GaAs ($\sim 1 \cdot 1 \times 10^{-10}$ F/m)
σ	conductivity
σ_0	conductivity at low electric fields
σ_1	differential or a.c. conductivity
ρ, ρ_0	donor electric charge density $= ne$
ρ_1	r.f. electric charge density
μ, μ_0	mobility, at low electric fields $\sim 7000 \text{ cm}^2/\text{Vs}$
μ_1	differential or a.c. mobility
ω	angular frequency
ω_c	dielectric relaxation frequency $\sim 0 \cdot 9 \times 10^{12} \text{ s}^{-1}$ for 1 ohm-cm GaAs

1. Introduction

The Gunn effect semiconductor oscillator[1] generates microwave power using a small specimen of gallium arsenide to which two ohmic contacts have been made. It is made to oscillate simply by applying a voltage across the resulting diode.† These oscillations are essentially caused by a.c. negative resistivity in the semiconductor GaAs which, at present, appears to be one of the most promising materials with this property.

† Services Electronics Research Laboratory, Microwave Division, Harlow, Essex.

† The term 'diode' is used here in its sense of two electrodes. No rectifying action is implied.

At first sight the device is a very simple one because there are no p–n junctions. This simplicity is deceptive and in fact the mechanism of operation is quite complex, with a number of different modes of oscillation.[2,3] The purpose of this paper is to indicate how the important modes can be explained in terms of competing microwave mechanisms. Since most of the effects have been previously recorded, the paper is to some extent a review which tries to present a collated picture of the microwave mechanisms. Simple physical arguments have been used wherever possible. Mathematical work that is relevant but not found in the references has been placed in the Appendix.

The paper opens with a review of the necessary physics and then examines an approximate but helpful equivalent circuit for the Gunn diode to show how the relevant microwave mechanisms contribute to the various modes of oscillation. The paper concludes with speculations on the future of these devices.

2. Review of Gunn Effect Physics

In 1963, Gunn discovered microwave oscillations in the current flowing through samples of n-type GaAs, and InP, to which sufficient voltage had been applied. His measurements, with a capacitive probe, of the voltage along the sample revealed that there was a region, or domain, of very high electric field. This formed near the cathode, drifted through the specimen at about 10^7 cm/s, and left the specimen at the anode. Whenever the domain left the specimen, a new domain was nucleated at the cathode and simultaneously a pulse of current was generated in the external circuit. This gave an oscillator whose fundamental frequency was determined by the length of the semiconductor specimen. Typically, the oscillation frequency was 1 GHz for 100 µm of specimen length.
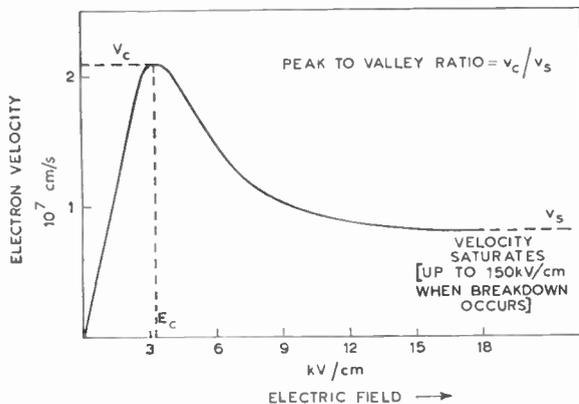
The effect has its roots in the band structure of semiconductors, such as GaAs, where at high enough electric fields (above 3000 V/cm for GaAs) conduction electrons transfer to higher energy states[4,5] which give a lower mobility to the electrons. Calculations[6] of the electrons' drift velocity in a known uniform electric field show that there is a region of negative differential mobility caused by this transfer of the electrons to lower mobility states (Fig. 1). Because of the relation ($\sigma = \rho_0\mu$) between mobility and conductivity, material with these properties is said to have a negative differential conductivity although a negative conductance is not always observed between two ohmic contacts to a specimen. The negative conductivity can lead to the spontaneous formation of a high electric field domain[7] in the specimen and this leads to a complex behaviour of the diode.

To appreciate how this high field domain forms, suppose that there is a sudden drop in the electrons' velocity at fields above the critical value E_c. Suppose also that there is a region in the semiconductor with a higher field than the rest of the specimen. Immediately† this region reaches E_c, the electrons in it will move slower than those outside. Electrons on the cathode side will accumulate at the rear of the region while those on the anode side will run away leaving a depletion region. The result is a dipole layer of charge (Fig. 2) which maintains a region of high electric field[7-12] and which drifts through the specimen at about 10^7 cm/s. This is the classical Gunn domain which usually is formed near the cathode contact whenever the voltage across a diode rises above the critical value V_c (~ 0.3 V/µm in GaAs).

Other types of high field domain can exist[11] and the accumulation domain deserves special mention because of its practical effects. The voltage across a diode can rise sufficiently fast so that the field across the whole specimen exceeds the critical value before the dipole domain has had time to form. Charge accumulates at the cathode contact and depletes in the anode contact. An accumulation of charge subsequently detaches itself from the cathode (Fig. 3) and moves through the specimen. In this type of domain the charge accumulation layer is detached from the corresponding depletion layer, which now lies in the anode thus still preserving the overall electrical neutrality of the semiconductor. In a practical (n⁺)n anode contact,‡ the ratio of the contact donor

Fig. 1. Velocity-field characteristic (after Butcher and Fawcett[6]).

---

† In fact the electrons take a finite time to transfer to the lower mobility states. This is usually assumed to be too short a time to matter: 10^{-13} to 10^{-14} seconds.

‡ The technology of ohmic contacts is not discussed in this paper. However, useful anode and cathode contacts can be made by an epitaxial growth of highly doped (n⁺) material on top of the n-type Gunn layer, and vice versa. Evaporated metal contacts can easily be made to the n⁺ layers.

Fig. 2. Dipole domain.

put in the form $nl > 3 \times 10^{11}$ cm^{-2} where the donor density n has been related to the negative a.c. conductivity σ_1 through appropriate assumptions; the numerical value for the condition varying slightly from reference to reference depending on the assumptions. The condition for domains to form needs a more detailed consideration of the growth of space charge waves in a semiconductor.[14,15] It is found that diffusion causes damping of the growth rate for such waves and that spatial disturbance cannot grow in time unless $\sigma_1 l^2 > 10^{-8}$ mho cm (see Appendix). This condition has been given previously[14] in the form $nl^2 > 10^9$ cm^{-1} where σ_1/σ_0 was taken to be as low as 10^{-2}, a value that could be appropriate for fields just above threshold where the a.c. negative mobility is low. This relation gives a necessary criterion for domain formation. It is therefore apparent that suitable choice of donor density and length can yield specimens in which domains do not form but there is still significant growth of r.f. space charge waves. This material is stable at low frequencies with the d.c. increasing monotonically with the applied voltage. The growing r.f. space-charge waves that can occur radically modify the sample impedance so that an r.f. negative resistance can be formed. This is discussed in the next section.

Although avalanche breakdown can occur at very high electric fields[16] or even at low fields with unsuitable contacts to the GaAs,[17] its discussion lies outside the scope of this review.

density to the donor density in the bulk of the specimen may be as high as 10^3 instead of about 5 as shown in Fig. 3. As a consequence, in this figure, the penetration of the electric field into the anode contact is greatly exaggerated. It is found that any inhomogeneities in the donor density of the bulk of the specimen readily nucleate a depletion region ahead of the accumulation layer so that this type of domain rapidly turns into a standard dipole domain.[8,9] Consequently if the properties of the accumulation domain are to be used, they must be used inside a time interval which is too short for dipole domains to form. Such a mode of operation is possible[13] and is discussed later.

The speed with which charge reverts to its equilibrium value inside a conductor is largely governed by the dielectric relaxation time ε/σ; typically 10^{-12} s in 1 mho cm^{-1} GaAs. Similarly in material with a negative differential conductivity $-\sigma_1$, charge builds up, as it travels along, with a time constant ε/σ_1. These relaxation times in general limit the rate of growth or decay of charge accumulation inside the diode. Thus if the negative conductivity $-\sigma_1$ is too low in value and the specimen length l is too short then no useful amount of charge build up can occur in the time taken for the charge to travel through the specimen. The diode is then completely stable.[8,14] For products $\sigma_1 l > 10^{-4}$ mhos oscillations can usually be observed though domains need not always form. This condition for oscillation is more usually

Fig. 3. Accumulation domain.

3. The Amplifying Mode

It was previously stated that it was possible to obtain specimens of material in which domains are inhibited from forming but changes in charge density grow as they drift through the specimen. A negative resistance can then be formed which can be used for amplification[18-20] or oscillation.

The following argument indicates how the negative resistance arises. It must first be noted that significant r.f. changes of charge density are not possible inside the highly conducting cathode contact because of the correspondingly high dielectric relaxation frequency. It then follows from Gauss's theorem that the r.f. electric field at the cathode contact is the same as it is right inside the contact, namely zero. However inside the bulk of the semiconductor specimen the r.f. charge density, ρ_1, can change significantly. At the cathode contact ρ_1 must be in phase with the total r.f. current flowing through the specimen since $I_1 = \rho_1 v_0 + v_1 \rho_0$, neglecting diffusion, but $v_1 = \mu E_1 = 0$ at the cathode, hence $I_1 = \rho_1 v_0$. However, as ρ_1 drifts through the semiconductor it grows with a time constant given by the negative dielectric relaxation time so that

$$\rho_1 = (I_1/v_0) \exp[j\omega(t - z/v_0)] \exp[\sigma_1 z/\varepsilon v_0]$$

For small rates of growth, it can be seen from Poisson's equation $(\partial^2 V_1/\partial z^2 = \rho_1/\varepsilon)$ that the r.f. voltage is approximately in antiphase with the charge density. Now the exponential growth ensures that most of the voltage is dropped at the anode end of the specimen. Thus, provided that V_1 is in antiphase to I_1 at the anode of the diode a negative resistance will result. This clearly is the case if ρ_1 is in phase with the applied current at the anode as well as at the cathode. Thus when the specimen length l is an integral number of space charge wavelengths long $(\omega l/v_0 = 2m\pi)$ negative resistance should be observed. In other words, negative resistance occurs at frequencies around the transit-time frequency (and its harmonics) in spite of the absence of a proper domain.

By using a simple linear analysis[18] the following useful results can be obtained. At the transit-time frequency the sample appears to behave approximately as if the low field capacitance C_0 were shunted by a negative resistance $-\frac{1}{2}R_0(\sigma_1/\sigma_0)$ where R_0 is the low field ohmic resistance and $-\sigma_1$ is the a.c. negative conductivity, suitably averaged over the diode. More rigorous answers need computation.[19,20]

Careful selection of the semiconductor material is of paramount importance for the successful application of these ideas. Domains have to be inhibited yet the sample must not be completely stable. Most materials exhibit negative resistance at applied voltages just below those required for domain formation but in this case it is difficult to prevent domains from forming thus causing severe non-linearities or oscillation. The amplifier's efficiency and noise figure are both poor but it is too early to say what the fundamental limitations are.

4. Gunn Diode Equivalent Circuit

In this and subsequent sections consideration is given only to samples in which domains form. In this case it is possible to explain qualitatively most of the features of a Gunn effect oscillator by referring to an equivalent circuit based on Fig. 5. The domain is essentially considered as an extra circuit element in series with the low electric field properties of the diode, namely its ohmic resistance R_0 and shunt capacitance C_0. The domain, which is switched in and out of the circuit at appropriate times, has a capacitance C_D, a resistance $-R_D$ and a current generator associated with it. This section indicates how the equivalent circuit can be justified, firstly by considering a dipole domain and then generalizing to the accumulation domain.

Associated with a dipole domain (Fig. 2) there is a voltage V_D in excess of the value calculated from the low field properties of the sample. There is also a domain charge q_D with a positive and negative layer as in a capacitor. This is the basis of the domain capacitance C_D, whose value can be roughly estimated from the charge separation in the dipole layer. Thus if a domain fills the whole specimen, C_D will be of the order of C_0 whereas if the dipole domain is very short C_D will be many times the low field capacitance C_0. Thus in general C_D varies with the domain voltage. The upper value to the domain capacitance is limited by the diffusion of the electrons which prevents the depletion and accumulation layers from approaching closer than the order of $0.1\sqrt{\sigma}$ µm in GaAs of conductivity σ (mho cm^{-1}).

Current flows through the specimen to charge up the domain but even if the domain charge is constant a current still flows because of the drift of the domain towards the anode. Consequently, in the equivalent circuit there must be, shunting the domain capacitance, a current whose value is determined by the drift velocity of the domain. This drift velocity has to be calculated from the equations for charge motion.[21-23] One of the simplest calculations[21] assumes that the diffusion constant for the semiconductor is independent of the electric field. The results can be summarized using the velocity-field characteristic and an 'equal areas rule' (Fig. 4). This rule relates the domain velocity, the highest field (E_{high}) in the domain, and the field (E_{low}) outside the domain. If the domain's charge (or voltage) increases, then by Gauss's theorem ($E_{high} - E_{low}$) must increase but consequently, from the equal areas rule, the domain velocity decreases. Thus the current shunting the domain decreases with increasing domain volts. This
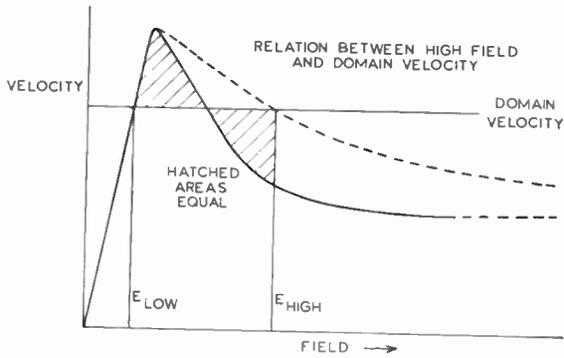
Fig. 4. Equal areas rule.
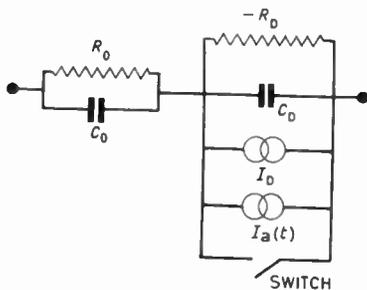
(a) Parallel form.
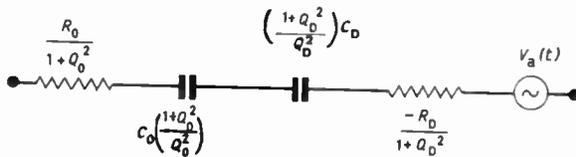
(b) Equivalent series form at frequency ω.

Fig. 5. Equivalent circuit for Gunn diode.

effect is represented as a fixed shunting current I_D together with a voltage dependent negative resistance $-R_D$ in parallel with the domain capacitance. I_D in general is equal to I_c, the critical current at which domains are initiated. The negative resistance, $-R_D$, is essentially a direct consequence of the negative differential resistivity of the semiconducting material in the high-field domain. Its value is therefore inversely proportional to the average negative differential mobility $-\mu_1$ in the high-field region and also proportional to the length of this region. Thus $-R_D \simeq R_0(\mu_0/\mu_1)(C_0/C_D)$, where C_0/C_D has been used as an estimate of the domain's length in comparison with the sample's length (also see Appendix). At very high domain voltages, E_{high} becomes so large that the domain velocity remains close to the satu-

rated value v_s. There is then virtually no change in the current, shunting the domain, with domain volts. In this case one may approximate by putting $-R_D$ infinite with $I_D = I_c(v_s/v_c)$ which is the value appropriate to the domain moving with a constant velocity v_s.

If the domain is allowed to drift into the anode then it is forced to be discharged. This effect is approximated to here by a discharging current generator $I_a(t)$ across the domain capacitance. This generator is only present if the domain reaches the anode and its magnitude is proportional to the charge on the domain when it arrives at the anode, for clearly if there is no charge to be discharged then $I_a = 0$. The spectrum of $I_a(t)$ depends on the domain size but it is regarded here as essentially a pulse with a pulse width that is short compared to the domain's transit time.

The switch in the equivalent circuit shorts out the domain circuit when the domain is absent so that the diode then behaves with its low electric field properties: a resistance R_0 shunted by a capacitance C_0. The switch must open when the voltage across R_0 exceeds the critical value V_c required for domain formation. It only closes again if the domain disappears either by the domain being discharged at the anode or by the r.f. current draining out the domain's charge.

Although the arguments for the validity of the equivalent circuit have been presented from the basis of a steady state dipole domain[24, 21, 10] the results are still qualitatively valid for the true dynamic state and the accumulation domain. A dynamic form[22] of the equal areas rule is needed to show this. The arguments for the accumulation domain are qualitatively the same as those for the dipole domain and the detailed justification is given by a mathematical treatment in the Appendix.

It is again emphasized that in general the accumulation domain is only present for a short period after its formation, otherwise inhomogeneities (as small as one part in a thousand), in the material, turn the domain into a dipole domain by nucleating a depletion region ahead of the charge accumulation layer. Consequently one is only interested in the accumulation domain's properties just after it has formed. From the separation of the accumulation and depletion of charge it can be seen that the accumulation domain's capacitance is the order of C_0, the low field capacitance. It is not always exactly this value because the domain is calculated as being in series with the low field properties of the sample. This calculation of the properties of the accumulation domain, which virtually fills the whole specimen, as an element in series with the low field specimen properties may seem to be artificial. However the mathematics in the Appendix justifies this approach and moreover it preserves the

generality of the one equivalent circuit. Because of this method of calculating the properties of the accumulation domain it follows that $R_0 - R_D$ must be the conventional a.c. resistance $(\mu_0/\mu_1)R_0$. It can then be seen from Fig. 1 that $-R_D$ has a minimum negative value of about $-4R_0$ (see also Appendix).

It must be clearly stated that the equivalent circuit is not being claimed as a method of calculating the detailed performance of a Gunn diode without the use of a computer, for the circuit parameters can only properly be estimated by recourse to computation. It is however claimed that it is helpful for interpreting experimental data, explaining device performance and designing approximately microwave circuits to ensure operation in a required mode.

5. Domain Modes of Oscillation

Three mechanisms of microwave generation will now be discussed using the equivalent circuit of Fig. 5 for guidance. Firstly, there is the current generator $I_a(t)$. This is only present if the domain is allowed to transit through the specimen and leave at the anode. It gives rise to modes which are commonly referred to as transit time modes because the oscillation period is to some extent tied to the transit-time. Secondly, there is a negative resistance associated with the domain's presence and this can give rise to modes in which the domain does not have to travel the full length of the specimen because it is 'quenched' by the r.f. current draining out the domain charge. The power for oscillation is derived from the negative resistance. These modes are commonly called quenched domain modes. However when the dipole domain is an accumulation domain the quenched domain mode has been referred to as the 'limited space-charge accumulation' (l.s.a.) mode.[13] This is discussed in further detail later. Finally, the domain capacitance can cause modulation of the circuit reactance and so lead to parametric conversion of d.c. into r.f.[25]

6. Transit-time Modes

From the series form for the equivalent circuit it can be seen that the total series specimen resistance R_s is given by $R_s = R_0/(1 + Q_0{}^2) - R_D/(1 + Q_D{}^2)$. Thus it can be seen that even if the domain negative resistance, $-R_D$, is greater than the low field specimen resistance R_0, the domain susceptance can reduce its effect so that the overall series resistance is positive at high enough frequencies. If R_s is positive at the domain transit-time frequency then negative resistance oscillations will not occur at this frequency. However the current generator $I_a(t)$ can generate power provided the domain completes a transit across the specimen. This leads to the transit-time modes of which two forms should be distinguished: resistive load oscillations and resonant load oscillations.

6.1. Resistive Load

With a resistive r.f. load across the Gunn diode, oscillations can be maintained by having a constant voltage, across the diode, at some value above the threshold V_c. As soon as one domain leaves at the anode a new domain must be formed in the specimen in order to absorb the diode voltage that is in excess of V_c. The effect of the current generator $I_a(t)$ is therefore to cancel the domain voltage sufficiently to allow the low electric field in the specimen to reach E_c and initiate a new domain. The current generator then behaves as a pulse generator with the pulse repetition frequency timed by the domain's transit time through the specimen. This is the classic transit time mode first discovered by Gunn and a characteristic waveform is shown in Fig. 6(a). Voltage tuning is to some extent possible because the domain velocity depends on the domain voltage which in turn depends on the applied voltage.[10] It is not a good oscillator mode because the output is rich in harmonics and not well stabilized in frequency.

6.2. Resonant Load

If a parallel resonant LCR circuit is placed across the Gunn diode, it will act as a frequency filter for the transit time mode. The domain capacitance will lower the effective resonant frequency of the external circuit which should be selected close to the transit time frequency. The circuit then provides a low impedance at the harmonics, thus forcing a nearly sinusoidal voltage across the diode. However at the fundamental frequency, the r.f. voltage can be made large by having

(a) Resistive transit-time.

(b) Resonant transit-time.
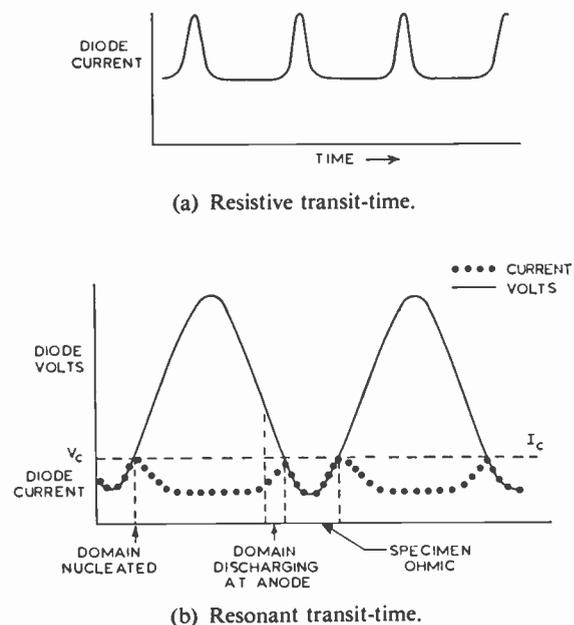
Fig. 6. Current generator modes.

a high enough circuit impedance. The total voltage across the diode falls below V_c for part of the oscillation period (Fig. 6(b)). Consequently, if a domain leaves the specimen during this part of the cycle no new domain is formed in the specimen since the voltage is below the critical value for domain formation.[26] The new domain is formed when the total diode voltage once more exceeds V_c. Thus the external circuit can control the frequency by altering the proportion of the oscillation period taken by a transit of the domain, even though the transit time need not alter. The variation of the transit time with domain voltage can supply yet a further tuning mechanism. Approximately an octave of tuning with efficiencies around 5% are theoretically claimed for this mode,[27] in fair agreement with experimental results.

For optimum efficiency the domain must not give a pulse of current with too high a harmonic content. In other words the domain must not be too short compared to the specimen length. The relative domain size is controlled by choosing the donor density (n) and length (l) product of the material to lie in a suitable range. Computations[27] indicate that the optimum choice for nl is between 1 to 2×10^{12} cm^{-2}. The load resistor R, in the parallel LCR load across the diode needs to be about 30 R_0 with the circuit Q higher than 10.

A series load circuit is also worthwhile examining. From the series equivalent circuit (Fig. 5(b)) it appears that the diode behaves as a voltage generator with a series internal resistance and capacitance. The internal resistance can sometimes be very low because the domain negative resistance helps to cancel the low field diode resistance (if this internal resistance is negative then a new mode can operate as discussed later). Thus the required optimum load for maximum power transfer would be an inductance, conjugate in impedance to the averaged domain capacitance, together with a low series resistance. In practice, the efficiency of this mode has been increased by adding a series inductance[28] to the external tuned circuit. The inductance has an impedance of several times R_0. It is suggested here that the increase in efficiency is possible, firstly, because the diode voltage is no longer constrained to be approximately sinusoidal. This allows the domain to arrive at the anode with a bigger charge and voltage although this means that the diode voltage has to drop more rapidly than previously as the domain is discharged. The arrival at the anode of more charge in the domain means that $I_a(t)$ has a bigger value, or equally $V_a(t)$ is greater and so more power is given out. Secondly, because a large inductance gives a high impedance at the harmonics of oscillation, the r.f. power can be enhanced by the parametric conversion of d.c. into r.f. power. This novel effect is discussed later.

The LCR circuit can also be chosen to have an effective resonance near to some harmonic of the transit-time frequency, thus turning the Gunn diode into a harmonic generator. However frequencies higher than the transit-time value can be more efficiently obtained by other modes which are treated later.

As previously stated the domain velocity alters with the domain voltage drop and hence with the applied bias, thus giving a voltage tuning mechanism for the diode. It has not in practice been found to be very effective. Another mechanism for changing the transit time has been established.[29] This uses a slowly increasing donor-density at the anode contact. A highly-charged domain takes longer to discharge at this graded anode and travels further than a lightly-charged domain. The frequency then decreases as the applied voltage, and hence the domain charge, increases. Tapering the dimensions of the sample has a similar effect to increasing the conductivity since it lowers the local current density. This has given an octave of tuning.[30] Whether the power output can be made to be sufficiently uniform for a practical device is not yet known.

7. Negative Resistance Modes

7.1. *Self-modulated Transit-time modes*

Provided that the domain negative resistance is greater than the low field sample resistance, then R_s can be negative at low frequencies, where the domain capacitance has a high impedance; yet at higher frequencies, where the domain capacitance has to some extent shorted out the negative resistance, R_s can be positive. Thus it is possible to have transit time oscillations occurring with R_s positive at this oscillation frequency, but simultaneously the domain negative resistance can excite low frequency oscillations, provided that there is a suitable r.f. circuit surrounding the diode. The low frequency oscillations then modulate the transit-time oscillations. Indeed this effect can produce unwanted modulation of the Gunn diode if there are parasitic resonant circuits associated with the power supply. Suitable damping to the unwanted resonances can naturally cure the trouble but driving the Gunn diode at a higher bias voltage will also cure the parasitic oscillations since R_D increases and consequently $R_D/(1 + Q_D^2)$ decreases with increasing bias. Then R_s becomes positive at the parasitic resonance frequency.

One might also be able to make a negative resistance r.f. amplifier, using the low frequency negative resistance, combined with a local oscillator, using the transit time mode, and rely on the inherent nonlinearities to produce the difference frequency as the i.f. A similar scheme has previously worked,[31]

though the negative resistance was caused by growing r.f. space charge waves at the lower frequency.

7.2. Quenched Domain and L.S.A. Modes

When the total series resistance R_s is negative at frequencies above the transit time, oscillations can be observed at these higher frequencies.[3,13,32] With a LCR resonant circuit, the r.f. voltage and current can build up at the effective resonant frequency so that the r.f. current and d.c. current through the diode are comparable. After a domain has formed, the total current through the diode can fall to such a low value that the domain is drained of its charge before it has reached the anode. As a consequence the domain disappears from the specimen which then reverts to its ohmic behaviour until the current and voltage rise again to the level required to initiate a domain. This mode is referred to either as the quenched domain mode or, as Gunn[33] refers to it, an under-voltage resonant mode. The power must now be derived from the domain's negative resistance. Figure 7 shows the qualitative relation between the specimen current and voltage. Figure 7(a) indicates a better waveform to the current than for Fig. 6(b) and hence more power output. The domain reactance causes the phase difference between the current minimum and the voltage maximum.

The accumulation domain, with its low capacitance shunting the domain negative resistance, readily yields an overall negative resistance at high frequencies. This can lead to a quenched accumulation domain mode at frequencies well above the transit-time frequency. The frequency can be so high that charge only has time to accumulate near the cathode contact and any accumulations or depletions of charge elsewhere in the specimen are limited to negligible proportions. Then most of the specimen has a more or less uniform high electric field which is operating in the negative a.c. mobility region of the velocity-field characteristic. The limitation on any accumulations of charge, except near the cathode, has caused this mode to be called the l.s.a. mode.[13,27,33] The prevention of unwanted accumulations of charge due to any non-uniformities in the specimen is most easily accomplished if the electric field only spends a short time in the steep region of negative a.c. mobility which leads to spontaneous growth of charge disturbances. Large r.f. voltages help to swing the field rapidly through this rapid growth region and consequently applied bias voltages well above threshold are required to make this mode work well. Figure 7(b) gives the qualitative relation between current and voltage indicating essentially a pure negative resistance with the domain reactance almost negligible.

Since this mode essentially makes the diode behave as a uniform negative resistance a very simple model
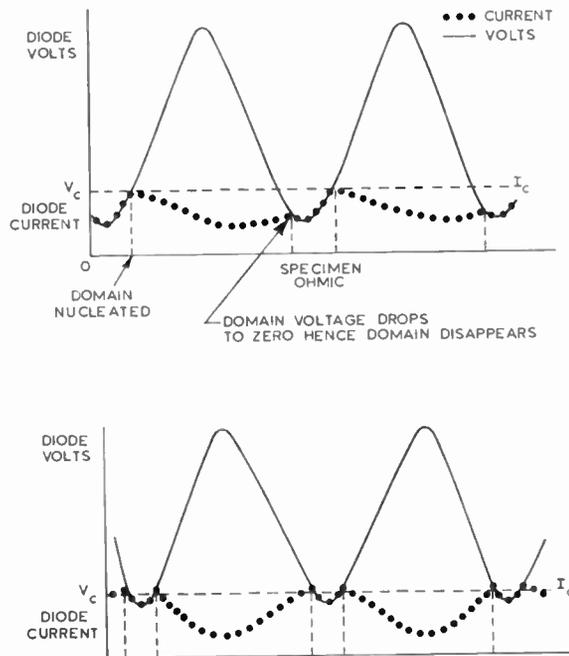
Fig. 7. Negative resistance modes (quenched domain modes or l.s.a. mode).

can be formed to estimate the limiting efficiency The voltage is assumed to swing from V_c to $V_c + 2V_1$ and the device is assumed to behave like a constant negative resistance over this range with the current decreasing from its peak value I_c to the value, appropriate to the saturated electron velocity, I_s. Thus the r.f. current has a peak to peak value of $I_c - I_s$. Provided that V_1 is considerably greater than V_c one can write down the efficiency as $n = \frac{1}{2}[(I_c - I_s)/(I_c - I_s)]$ 100% giving a value of 16·6% for a 2.1 peak to valley ratio $I_c : I_s$. From Fig. 1 it can be seen that the electron velocity does not really saturate until about 20 kV/cm. Over the range of fields below this value one can obtain an average negative mobility of about $-\mu_0/10$, so that the average negative resistance $-R_s$ for the whole device is the order of $-10 R_0$. Thus the maximum loading (minimum value of R) for a parallel LCR circuit that will maintain oscillations is expected to be the order of $R = 10 R_0$. Detailed calculations depend on the velocity field characteristic, but one set of computations[33] confirms the results of this simple approach.

In terms of the circuit of Fig. 5(b), the average negative domain resistance $-R_D$, although greater in magnitude than R_0, is still such that, combined with the low domain capacitance C_D, Q_D is small. Then $R_s \simeq -(R_D - R_0)$ which as stated previously is just the resistance calculated from the differential mobility and the whole sample length.

All quenched domain modes have two useful properties. Firstly, the efficiency need not be affected by the sample length (see next section for dipole domains). Secondly, the length is not controlled by the electron transit-time. In the l.s.a. mode the length must be short compared to the wavelength of an electromagnetic wave propagating through the specimen.[33] This limits the length to the order of $10^9/f$ cm instead of $10^7/f$ cm for a transit-time mode. Thus the power input (at least on a pulsed basis) can potentially be orders of magnitude greater, for a given impedance, than in a transit time mode. Moreover in the l.s.a. mode the applied voltage can be greater than that applied to a dipole domain mode before high electric field breakdown occurs. As a consequence of all these properties, the l.s.a. mode may be able to give powers in the kilowatt range at 10 GHz.

It has been suggested that multiple dipole domain formation[9] is an important source of noise in dipole domain modes.[36] If this theory is well founded then the l.s.a. mode may give quieter oscillations than other modes.

8. Parametric Energy Conversion

It can be seen most easily from the series equivalent circuit that when R_D has a high impedance compared to $1/\omega C_D$ then the presence of the domain can seriously modify the reactance of the diode circuit. R_D can be large enough for this effect at high enough domain voltages (V_D above about 10 V for dipole domains in 1 ohm cm GaAs but for accumulation domains V_D probably needs to be greater than 5 V_c). The domain field must not of course be so great (around 150 kV/cm) that impact ionization occurs with the consequent breakdown of the domain and associated theory. Thus at high enough applied voltages the presence and absence of a domain can modulate the circuit reactance. This can be expected to lead to parametric effects. At present the only effect that has been observed is essentially the action obtained in a degenerate parametric amplifier, namely, if a reactance is modulated at a frequency $2f_0$ then negative resistance can be seen at f_0. However this can be extremely important in enhancing the power output from a Gunn diode working in a dipole domain mode.

An experiment[29] has demonstrated that when a Gunn oscillator is oscillating at f_0 then pumping power into the device at $2f_0$ can enhance the power output at f_0 and moreover reduce the d.c. power required. In the experiment the down conversion efficiency was approximately 50%, the maximum allowed if the Manley and Rowe relations are applicable. As might be expected if the input power was at $2f_0 + \delta$ then both f_0 and $f_0 + \delta$ could be seen on a spectrum analyser over a range of δ of tens of megahertz depending on the circuit. This experiment suggests that if second har-

monic power is trapped in the diode by high external circuit impedances at the second harmonic then the power output could be enhanced by parametric action. An experiment on changing the second harmonic circuit impedances confirmed that there was a mode for the test diode in which changing the second harmonic circuit impedances drastically altered the power output. Moreover direct measurement of the second harmonic circuit impedance by a slotted line showed a high inductance at the optimum power output condition. A simple theory has been developed for this mode of oscillation[22] and it can be shown formally that the negative resistance calculated directly can also be obtained by calculating the effective capacity variation at the second harmonic of the oscillation frequency, and then applying standard parametric oscillator theory. Thus the statement that the action is essentially parametric in nature seems justifiable.
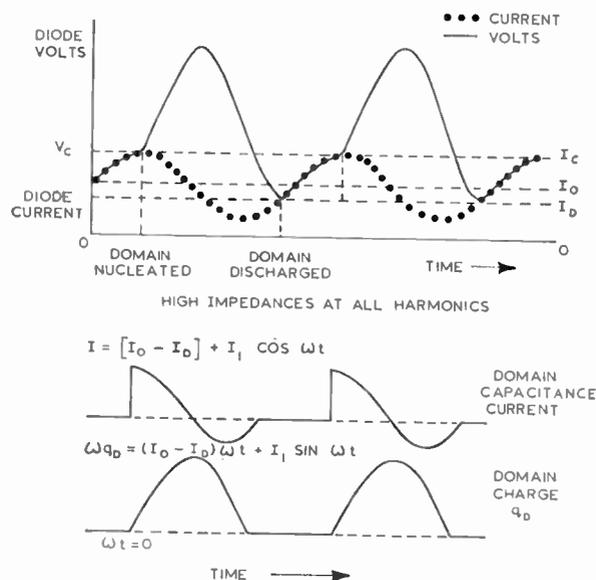
Fig. 8. Idealized parametric mode.

Figure 8 shows an idealized form for a quenched domain mode in which it is only the parametric negative resistance that maintains the oscillations. I_D is assumed constant, with R_D infinite, and $I_a(t)$ is zero since the domain is quenched before it reaches the anode. Thus neither the domain's generator nor the domain's negative resistance can contribute to the output power. An analysis is given, for a fixed domain capacitance C_D, in the Appendix. Figure 9 shows how the frequency for optimum power output varies with applied bias. It also shows the optimum value of circuit inductance L and the series load resistance R_L required across the diode. It must however be remembered that the domain's form in general varies
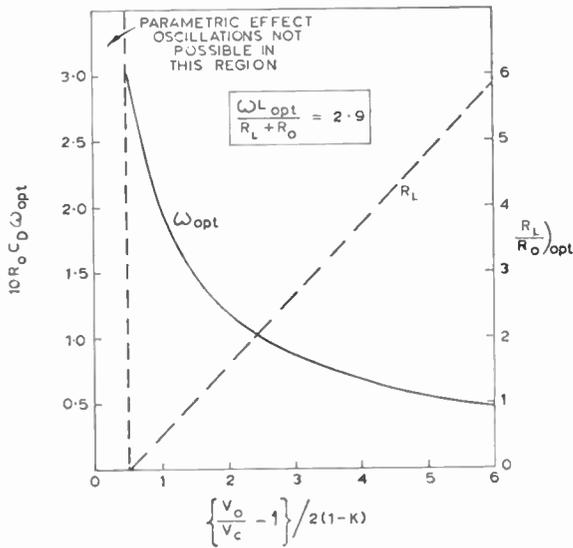
Fig. 9. Variation of frequency for optimum power output with applied bias.

$$\frac{1}{k} = \frac{v_o}{v_s} \text{ (peak/valley ratio)}$$

with time or the applied voltage so that the effective value of C_D may vary with applied bias. The case of the dipole domain is referenced.[25]

The source of r.f. power can also be qualitatively seen from the model described by the waveforms in Fig. 8. The battery current assists the r.f. current in driving the charge into the domain capacitance at a lower r.f. voltage than that at which it leaves. This is the basis of all parametric effects. One should also note that there is no charge on the domain's capacitance when it forms or disappears so that no energy is lost or gained on this account, the energy coming from the charge stored by the steady current.

This effect, used as above, can allow over an octave of tuning by varying the external circuit and with applied voltages above three times the threshold value. It also allows a small amount ($\sim 1\%$) of voltage tuning with a fixed circuit though this depends on the precise nature of the circuit and needs more experimental work to establish its usefulness. If the circuit is continually readjusted for optimum power as the applied voltage is increased then the frequency decreases. This may explain some experimental results.[30] The reasons for this decrease are given in the Appendix.

The efficiency of the effect increases with applied voltage and tends to a limit which depends on the peak/valley ratio ($1/k$) for the semiconductor material. The theoretical limiting efficiency is largest for the greatest peak/valley ratio: thus the efficiency is approximately 8%, $1/k = 1.5$; 15%, $1/k = 2$; 26%, $1/k = 3$. Thus

as for the l.s.a. mode the quality of the material plays an important part in deciding how effective the mode can be.

The effect should also be able to enhance the power output in the resonant transit time mode, and it is interesting to note that in a practical oscillator working with dipole domains, where high efficiencies[25] (up to 19%) have been observed, a high impedance at the harmonics was in effect inserted by using a highly inductive load. The parametric mode can account theoretically for such high efficiencies.

9. Conclusions

Four mechanisms of generating power from Gunn diodes have been discussed: firstly, growing r.f. space-charge waves; secondly, a voltage generator caused by the domain discharging at the anode; thirdly, a negative resistance associated with the negative resistivity of the material in the domain, and finally parametric effects caused by the modulation of the circuit reactance due to the periodic introduction of the domain. An equivalent circuit has been given which helps to account for the varied performance and wide tuning ranges that have been observed from Gunn diodes.

The required circuit for optimum power transfer always needs to be inductive at the oscillation frequency to cancel the capacitative effects of the domain. In the l.s.a. mode the capacitative effects can be very small. It appears that higher efficiencies can be obtained for dipole domain modes by having a high impedance at the harmonics rather than the low harmonic impedance that is in general given by the parallel LCR circuit. This improvement in efficiency is attributed in this paper to parametric conversion of d.c. into r.f. as described in the text. It may be possible to improve the efficiency of the l.s.a. mode by having high impedance at the harmonics and applying higher voltages. The parametrically derived negative resistance could enhance the negative resistance due to the negative a.c. mobility. This could be a very powerful mode with efficiencies over 20% for good material.

Although large mechanical tuning ranges have been observed, similar voltage tuning with a fixed circuit has not yet been achieved. It may be that it is only a question of circuit design for the frequency has been observed to decrease with applied voltage[30] (also predicted in the parametric mode) but the circuit values have to change appropriately with frequency. The transit-time mode may also provide the answer for electronic tuning, with perhaps having the donor density near the anode modified as discussed in the text. A third electrode might be used to control the current flow, as in a field effect transistor, and hence

control the domain formation and alter the domain capacitance, transit time, or both.

Other uses for the Gunn effect may be found. The parametric effects discussed here might be used to create a parametric amplifier, though the current effectively shunting the domain due to its movement would prevent the lowest possible noise figures. Applications to logic circuits have also been discussed elsewhere.[29]

Unfortunately, at present, theoretical predictions are more exciting than practical reality. Pulsed efficiencies are typically less than 8% while c.w. efficiencies are usually less than 3%. Occasionally these figures are exceeded but far too often the results are poor with a tremendous variability between slices of nominally similar GaAs. Poor practical efficiencies could be accounted for by a low peak/valley ratio for the semiconductor material. Unfortunately it is not yet known what factors control the effective value of this ratio in a diode. It is also generally found that the efficiency is sensitive to the nature of the cathode contact and how it is made. Thus, as always, technology provides the key to the future exploitation of the theory.

Output frequencies have mostly been between 1 and 20 GHz for these oscillators, with useful powers at much higher frequencies, especially in the l.s.a. mode. The difficulties of dealing with power densities around 10^7 W/cm^3 into the diodes has at present limited the output power to around 100 mW c.w. or the order of $100/f^2$ W peak pulse (f in GHz). These laboratory results will be considerably lowered for field tests in the near future but should be increased by at least an order in the foreseeable future.

10. Acknowledgments

11. References

1. J. B. Gunn, 'Instabilities of current and potential distribution in GaAs and InP', Plasma Effects in Solids, 2, 7th Int. Conf. Semiconductor Physics, Paris 1964, (Academic Press, Dunod 1965).

2. J. B. Gunn, 'Microwaves à la mode', 6th Int. Conf. on Microwave and Optical Generation, Cambridge 1966, p. 308 (I.E.E., London, 1967).

3. J. B. Gunn, 'Effect of domain and circuit properties on oscillations in GaAs', IBM J. Res. Devel. 10, pp. 310–20, July 1966.

4. B. K. Ridley and T. B. Watkins, 'The possibility of negative resistance effects in semiconductors', Proc. Phys. Soc. (London), 73, pp, 293–304, August 1961.

5. C. Hilsum, 'Transferred electron amplifiers and oscillators', Proc. Inst. Radio Engrs, 50, pp. 185–9, February 1962.

6. P. N. Butcher and W. Fawcett, 'Calculation of the velocity-field characteristic for gallium arsenide', Phys. Letters, 21, pp. 489–90, June 1966.

7. B. K. Ridley, 'Specific negative resistance in solids', Proc. Phys. Soc. (London), 82, pp. 954–66, December 1963.

8. D. E. McCumber and A. G. Chynoweth, 'Negative conductance amplification and Gunn instabilities in two valley semiconductors', Trans. Inst. Elect. Electronics Engrs on Electron Devices, ED-13, pp. 4–21, January 1966.

9. H. Kroemer, 'Non-linear space charge domain dynamics in a semiconductor with negative differential mobility', Trans. I.E.E.E., ED-13, pp. 27–40, January 1966.

10. P. N. Butcher and W. Fawcett, 'Stable domain propagation in the Gunn effect'. To be published: see also earlier version 'A simple analysis of stable domain propagation in the Gunn effect', Brit. J. Appl. Phys., 17, pp. 841–50, July 1966.

11. J. A. Copeland, 'Stable space charge layers in two valley semiconductors', J. Appl. Phys., 37, pp. 3602–9, August 1966.

12. J. B. Gunn, 'Properties of a free steadily travelling electrical domain in GaAs', IBM J. Res. Devel., 10, pp. 300–9, July 1966.

13. J. A. Copeland, 'A new mode of operation for bulk negative resistance oscillators', Proc. I.E.E.E., 54, p. 1479, October 1966.

14. B. K. Ridley, 'The inhibition of negative resistance dipole waves and domains in GaAs', Trans. I.E.E.E., ED-13, pp. 41–3, January 1966.

15. A. Susaki and T. Takagi, 'Conditions for amplification, oscillation in GaAs bulk semiconductors', Proc. I.E.E.E., 54, pp. 2027–8, December 1966.

16. J. S. Heeks, 'Some properties of the moving high field domain in Gunn effect', Trans. I.E.E.E., ED-13, pp. 68–79, January 1966.

17. J. A. Copeland, 'Switching and low field breakdown in n GaAs bulk diodes', Appl. Phys. Letters, 9, pp. 140–1, August 1966.

18. R. W. H. Engelmann and C. F. Quate, 'Linear or small signal theory for the Gunn effect', Trans. I.E.E.E., ED-13, pp. 44–52, January 1966.

19. A. L. McWhorter and A. G. Foyt, 'Bulk GaAs negative conductance amplifiers', Appl. Phys. Letters, 9, pp. 300–2, October 1966.

20. P. N. Robson and S. Mahrous, 'Small signal impedance of stable transferred-electron diodes', Electronics Letters, 2, pp. 107–8, March 1966.

21. P. N. Butcher, 'Theory of stable domain propagation in Gunn effect', Physics Letters, 19, pp. 546–7, December 1965.

22. J. E. Carroll, 'Non-uniform motion of high-field domains in the Gunn effect', Electronics Letters, 2, pp. 194–5, June 1966.

23. J. W. Allen, W. Shockley and G. L. Pearson, 'Gunn domain dynamics', J. Appl. Phys., 37, pp. 3191–4, July 1966.

24. G. S. Hobson, 'Small signal admittance of a Gunn effect device', *Electronics Letters*, 2, pp. 207–8, June 1966.

25. J. E. Carroll, 'A Gunn diode self-pumped parametric oscillator', 6th Int. Conf. on Microwave and Optical Generation and Amplification, Cambridge 1966, pp. 309–13 (I.E.E., London, 1967).

26. P. N. Robson and S. M. Mahrous, 'Some aspects of Gunn effect oscillators', *The Radio and Electronic Engineer*, 30, pp. 345–5, December 1965.

27. J. A. Copeland, 'Theoretical study of a Gunn diode in a resonant circuit', *Trans. I.E.E.E.*, ED-14, p. 55–8, February 1967.

28. D. G. Dow, C. H. Mosher and A. B. Vane, 'High pulsed power using the Gunn effect', 6th Int. Conf. on Microwave and Optical Generation and Amplification, Cambridge 1966, p. 273 (I.E.E., London, 1967).

29. C. P. Sandbanks, 'Synthesis of complex electronic functions using solid state bulk effects'. To be published in *Solid State Electronics*.

30. M. Shoji, 'A voltage tuneable Gunn oscillator', *Proc. I.E.E.E.*, 55, pp. 130–1, January 1967. (Letters.)

31. B. W. Hakki, 'GaAs post threshold microwave amplifier, mixer, and oscillator', *Proc. I.E.E.E.*, 54, pp. 299–300, February 1966.

32. J. E. Carroll, 'Oscillations covering 4 to 31 Gc/s from a single Gunn diode', *Electronics Letters*, 2, p. 141, April 1966.

33. J. A. Copeland, 'L.S.A. oscillator diode theory', *Trans. I.E.E.E. on Electron Devices*. (To be published.)

34. J. E. Carroll, 'Resonant circuit operation of Gunn diodes: a self-pumped parametric oscillator', *Electronics Letters*, 2, pp. 215–6, June 1966.

35. T. E. Hasty, P. A. Cunningham and W. R. Wisseman, 'Microwave oscillations in epitaxial layers of GaAs', *Trans. I.E.E.E.*, ED-13, pp. 114–6, January 1966.

36. G. S. Hobson, 'Source of f.m. noise in cavity controlled Gunn effect oscillators', *Electronics Letters*, 3, p. 63, February 1967.

12. Appendix

12.1. *Current Equation*

$$I = \rho v(E) - D\frac{\partial \rho}{\partial z} + \varepsilon\frac{\partial E}{\partial t}$$

relates the total current I with the conduction, diffusion and displacement current. For simplicity the diffusion coefficient D is assumed to be independent of the electric field. In normalized form for a specimen of length l:

$$J = pU(F) - d\frac{\partial p}{\partial X} + \frac{\partial F}{\partial T} \qquad \text{......(1)}$$

where

$$F = E/E_c, \qquad J = I/I_c, \qquad p = \rho/\rho_0,$$
$$T = \omega_c t, \qquad \omega_c = \rho_0\mu_0/\varepsilon = \sigma_0/\varepsilon$$
$$X = z/l, \qquad d = D/v_c l, \qquad U = v/v_c$$

12.2. *Space Charge Waves*

Gauss's theorem yields $\partial F/\partial X = ap$, $a = \omega_c l/v_c$ hence from eqn. (1), for small disturbances varying as $\exp(AT - BX)$,

$$BJ_1 = (-\sigma_1/\sigma_0)p_0 ap_1 + Bu_0 p_1 - dB^2 p_1 + Aap_1$$

where

$$p = p_0 + p_1, \qquad U = u_0 + (-\sigma_1/\sigma_0)F_1, \qquad J = J_0 + J_1$$

A negative differential conductivity $-\sigma_1$ has been assumed and squares of small quantities p_1 etc. have been neglected. Hence for time variations $\exp j\omega t$ one obtains space variations

$$\exp(-j\omega z/v_0)\exp(\sigma_1 z/\varepsilon v_0)$$

assuming d is negligible. Diffusion lowers the rate of growth of the wave. In particular a spatial disturbance only grows in time if the diffusion term is small enough. One can Fourier analyse any spatial disturbance in the specimen (e.g. a change in the charge density distributed over the length l) into waves with a wavelength $2l/m$ (m integer) hence $B = jm\pi$. The wave will only grow in time if the real part of the corresponding value for A is positive i.e. if

$$\sigma_1 p_0 a/\sigma_0 > d\pi^2 m^2 \qquad p_0 \simeq 1$$

and

$$D \simeq 100 \text{ cm}^2/\text{s}$$

so that

$$\sigma_1 l^2 > 10^{-8} \text{ mho cm}^2$$

is approximately required for temporal growth of any spatial disturbance (e.g. domains) in negative resistivity material.

12.3. *Dipole Domain*

If d tends to zero, the charge accumulation layer is very thin compared to the specimen length. The layer travels with a velocity u which can be estimated from the dynamic equal areas rule[20] (Fig. 10) relating u, the high field F_H and the low field F_L on either side of the accumulation layer. A straight line approximation to the electron velocity field relation, $U(F)$, is used in Fig. 10. This yields

$$u = 1 + \tfrac{1}{3}(1 - F_L) - \tfrac{1}{6}(F_H - F_L) - \tfrac{2}{3}(F_L - 1)^2/(F_H - F_L)$$
$$F_L < 1 \quad F_H < 3$$
$$= \tfrac{1}{3} + \tfrac{1}{6}[(F_L + 1)(5 - 3F_L)/(F_H - F_L)],$$
$$F_L < 1 \quad F_H > 3$$

Either $(F_H - F_L)$ is large or $(F_L - 1)^2$ is small in general so that the approximation is made that

$$u = 1 + \tfrac{1}{3}(1 - F_L) - \tfrac{1}{6}(F_H - F_L), \qquad u > \tfrac{1}{3} \qquad \text{......(2)}$$
$$= \tfrac{1}{3} \qquad\qquad\qquad\qquad \text{otherwise}$$
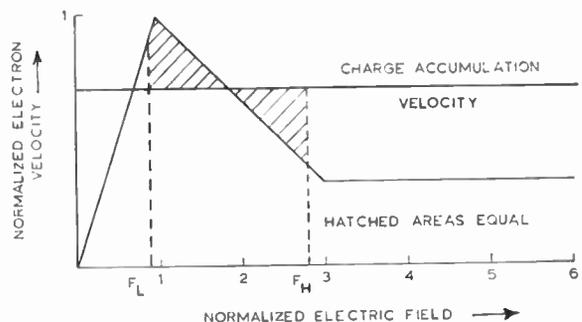
Fig. 10. Dynamic equal areas rule.

From Gauss's theorem

$$F_H - F_L = Q$$

where

$$Q = q\omega_c/\rho_0 v_c \qquad \ldots\ldots(3)$$

and q is charge/unit area in accumulation layer.

The dipole domain voltage is given by

$$V_D/V_c = \phi = \tfrac{1}{2}(F_H - F_L)l_D$$

where l_D is the effective length of the domain. If diffusion is neglected then $l_D = q/\rho_0 l$ thus

$$\phi = A^2 Q^2 \qquad \ldots\ldots(4)$$

where

$$A^2 = \tfrac{1}{2}f(v_c/\omega_c l)$$

with

$$= 1 \quad \text{if} \quad d = 0$$

If diffusion is not zero a suitable value must be inserted for f.

The rate of increase of charge in the layer is given by the rate of flow into the layer hence

$$\frac{dQ}{dT} = F_L - u,$$

substituting from eqns. (2) and (3) yields

$$F_L = 1 - \tfrac{1}{8}Q + \tfrac{3}{4}\frac{dQ}{dT} \qquad u > \tfrac{1}{3}$$
$$\text{otherwise}$$
$$= \tfrac{1}{3} + \frac{dQ}{dT}$$

But total current J is given by $F_L + dF_L/dT$ hence

$$J = 1 - \tfrac{1}{8}Q + \tfrac{3}{4}\frac{d^2Q}{dT^2} + \tfrac{5}{8}\frac{dQ}{dT} \qquad u > \tfrac{1}{3}$$
$$\text{otherwise}$$
$$= \tfrac{1}{3} + \frac{d^2Q}{dT^2} + \frac{dQ}{dT}$$

In general $(\omega/\omega_c)^2$ is small, and from eqn. (4)

$$J = 1 - \tfrac{1}{4}\phi(C/C_0) + \tfrac{5}{8}(C/C_0)\frac{d\phi}{dT} \qquad u > \tfrac{1}{3}$$
$$\text{otherwise} \qquad \ldots\ldots(5)$$
$$= \tfrac{1}{3} + (C/C_0)\frac{d\phi}{dT}$$

where $C/C_0 = 1/2A\sqrt{\phi}$.

Equation (5) shows that the domain behaves as an extra element in series with the low field properties of the sample. The element is a combination of capacitance shunted by a variable negative resistance just as outlined in the main text: the capacitance varies from $\tfrac{1}{8}C$ to C as the domain velocity alters while the negative resistance varies from $-R_D = -4R_0(C_0/C)$ to infinity. This is the mathematical basis of the statements in the text.

12.4. *Accumulation Domain*

The rate of flow of current into the accumulation layer in this form of domain is given by

$$\frac{dQ}{dT} = U(F_L) - U(F_H) = F_L - \tfrac{4}{3} + \tfrac{1}{3}F_H \quad F_H < 3 \quad F_L < 1$$
$$= F_L - \tfrac{1}{3} \qquad F_H > 3 \quad F_L < 1$$

From eqn. (3)

$$F_L = 1 - \tfrac{1}{4}Q + \tfrac{3}{4}\frac{dQ}{dT} \qquad F_H < 3$$
$$= \tfrac{1}{3} + \frac{dQ}{dT} \qquad F_H > 3$$

But in the accumulation domain where the charge layer has not had time to move any significant distance the domain voltage $\phi \simeq Q$ so that exactly as for the dipole domain

$$J = 1 - \tfrac{1}{4}\phi(C/C_0) + \tfrac{1}{2}(C/C_0)\frac{d\phi}{dT} \qquad F_H < 3$$
$$\qquad\qquad \ldots\ldots(6)$$
$$= \tfrac{1}{3} + (C/C_0)\frac{d\phi}{dT} \qquad F_H > 3$$

where $C \simeq C_0$, and $(\omega/\omega_c)^2$ has been assumed much less than unity. Equation (6) shows similar features to eqn. (5) so that the accumulation domain can be considered as an extra circuit element in series with the low field properties of the sample. The domain's capacitance varies from $C_0/2$ to C_0 approximately while $-R_D$ varies from $-4R_0$ to infinity.

12.5. *Idealized Parametric Mode*

To demonstrate the principle, C_D is assumed constant, with R_D infinite and $I_D = kI_c$, $(k = v_s/v_c)$. $\omega C_0 R_0$ is assumed small so that C_0 is neglected. Infinite impedances at all harmonics are assumed so that

$$I = (I_c - I_1) + I_1 \cos \omega t \qquad \ldots\ldots(7)$$

The domain forms at $\omega t = 0$ when $I = I_c$. The current flowing into the domain is $dq/dt = I - I_D$ where q is the charge on the domain. The domain voltage

$$V_D = q/C_D$$

Hence

$$\omega C_D R_0(V_D/V_c) = (1 - a - k)\omega t + a \sin \omega t \qquad \ldots\ldots(8)$$
$$a = I_1/I_c, \qquad V_c = I_c R_0$$

The domain disappears when $\omega t = T_0$.

Maximum charge is transferred from battery to circuit when $T_0 = \pi - 1 \cdot 35$ when the domain is only just extinguished. The d.c. domain voltage

$$V_{D0} = 1/2\pi \int V_D \, d(\omega t)$$

The load resistance, R_L, required to maintain the oscillations is given by

$$-I_1(R_0 + R_L) = 1/\pi \int V_D \cos \omega t \, d(\omega t)$$

But using
$$(I_0 - I_D)/I_1 = (1 - a - k)/a = -\sin T_0/T_0$$
one can show that
$$V_{D0}(I_0 - I_D) = \tfrac{1}{2} I_1^2 (R_L + R_0) \qquad \ldots\ldots(9)$$
This yields the required load resistance while the series inductance L is found from
$$I_1 \omega L = 1/\pi \int V_D \sin \omega t \, d(\omega t)$$
The d.c. voltage across the whole diode is
$$V_0 = V_{D0} + I_0 R_0$$
Figure 9 shows the appropriate values of frequency and load resistance for the most power output. It is found that $\omega L/(R_L - R_0) = 2.9$ independently of k. If C_0 varies during the cycle then this value will be different.

The efficiency depends strongly on k. At very high applied voltages it reaches a limiting value
$$\eta = \frac{0.172(1 - k)}{1 - 0.828(1 - k)} \, 100\% \qquad \ldots\ldots(10)$$
The efficiency is thus increased by having a low value of k or a high peak/valley ratio in the material.

The need for a high impedance at the harmonics can also be shown mathematically. There can be no power lost in the domain's capacitance so that $\sum P_n = 0$ where P_n is the power flow into the capacitance at frequency $n\omega$. P_n is found by Fourier analysis of the domain's voltage and the current I_c through the domain's capacitance. The domain starts and finishes with zero charge stored in it so that
$$I_{C0} = (\omega/2\pi) \int I_c \, dt = (\omega/2\pi) \left| q \right|_0^{\omega t = T_0} = 0$$

Thus paradoxically there is zero d.c. power fed into the domain's capacitance, although there is power fed at d.c. into the circuit as a whole. P_0 is zero and the domain's capacitance is the only source of power at ω. It follows that for substantial power out only at ω (P_1 negative), P_n must be positive or zero ($n > 1$). Hence there must be a strong harmonic content to the domain's voltage. Thus a high impedance is required across the domain's equivalent circuit in order to prevent this necessary harmonic voltage from generating power in the external circuit. The harmonic power flow into the capacitance must be recognized as a mathematical way of considering the problem rather than as a reality. The d.c. power is effectively turned into r.f. power by the periodical insertion and removal of the domain's capacitance. Thus the device is a self-pumped parametric oscillator.

If C_D is constant, Fig. 9 shows that the frequency decreases with increasing applied voltage. This is caused physically by the d.c. and r.f. impedance being related to the impedance $1/\omega C_D$. As the d.c. impedance increases with increasing applied voltage the relation can only be maintained by ω decreasing. In general C_D is not fixed. With the dipole domain,[22] the frequency decreases at a slower rate. However, the domain might start as an accumulation domain and then turn into a dipole domain. This could radically alter the performance of this mode.

High-order Multipliers without Idlers

By

S. V. JUDD, B.Sc.†

Summary: This paper describes a theoretical and experimental investigation carried out in order to provide design information about high-order multipliers without idlers which utilize the approximate step-function voltage-capacitance relationship of step-recovery diodes.

1. Introduction

Until the advent of the Read diode and the Gunn oscillator, the generation of power at microwave frequencies by solid-state devices has been realized by transistor oscillators operating in a fundamental or harmonic mode, or varactor diode multipliers driven at v.h.f. by solid-state generators. This latter technique is still the only method of generating appreciable amounts of c.w. power at frequencies up to X-band.

In a varactor multiplier use is made of the non-linear voltage-capacitance relationship of a semiconductor junction diode in order to achieve efficient frequency multiplication. The diodes are driven into forward conduction in order to increase the power handling capability, and the efficiency of multiplication is maintained provided that the input frequency is high enough to take advantage of the charge storage phenomenon associated with forward-biased junctions.

Until recently the design of solid-state sources has been restricted by the necessity of using chains of low-order stages in order to achieve the overall multiplication ratio. This is because the exponent in the voltage-capacitance law is 0·33 for a graded junction and 0·5 for an abrupt junction and the use of these types of diode in high-order idlerless stages results in poor efficiency. Diodes in which resistivity profiles have been optimized to realize high-order idlerless multiplier stages are now available. The voltage-capacitance relationship approaches a step function, and the minority carrier lifetime under forward bias has been increased.

Single, high-order multiplier stages are preferable to the conventional chains of low-order stages because they are less complicated from a circuit point of view and are usually smaller and weigh less. They are not capable of producing such high output power as a multiplier chain because all the loss occurs in a single diode. Within the limitations imposed by the diode parameters however, the high-order idlerless multiplier is an attractive proposition.

† The Plessey Company Ltd., Electronics Group, Roke Manor, Romsey, Hampshire.

Theoretical and experimental results of numerous investigations into the design of harmonic multipliers have appeared in the literature.[1-7] These analyses have considered a wide range of capacitance variation as a function of applied voltage, or in some cases[4,5] have described a switching mode of operation which is completely different from the conventional varactor mode.

In all the varactor mode analyses except those of the abrupt junction and stepwise doubler[3] the solutions have to be evaluated numerically. The methods of Burckhardt[6] and Scanlan and Laybourn[7] are generally applicable. Numerical results for an octupler were obtained by Burckhardt for the case in which γ, the exponent in the capacitance voltage relationship, is zero. This case corresponds to a piecewise linear relationship for which analytic solutions are possible for the calculation of efficiency and optimum impedence levels.[8]

It has been found experimentally (Fig. 1) that the variation of capacitance as a function of applied voltage for a step-recovery diode may be approximated by:

$$C_{(v)} = \frac{C_{(0)}}{\left(1 - \dfrac{v}{0 \cdot 6}\right)^{1/9}}$$

where $C_{(v)}$ is the dynamic capacitance, v is the applied voltage, and $C_{(0)}$ is the capacitance when $v = 0$.

The assumption of a piecewise linear approximation (Fig. 2) for this type of diode leads to a simple analysis of idlerless multipliers. Grayzel[10] has analysed the frequency doubler circuit for varying values of the exponent in the capacitance-voltage relationship of a hypothetical diode and has computed the efficiency and output power as a function of γ. His results show that there is little change in the performance of a doubler for γ less than 0·1, which should mean that the piecewise linear approximation can be justified at least for a doubler. An analytical solution has not been found possible for the power handling capability of the idlerless multiplier for orders greater than two. However, a simple computer program has been written which provides results for typical component values.
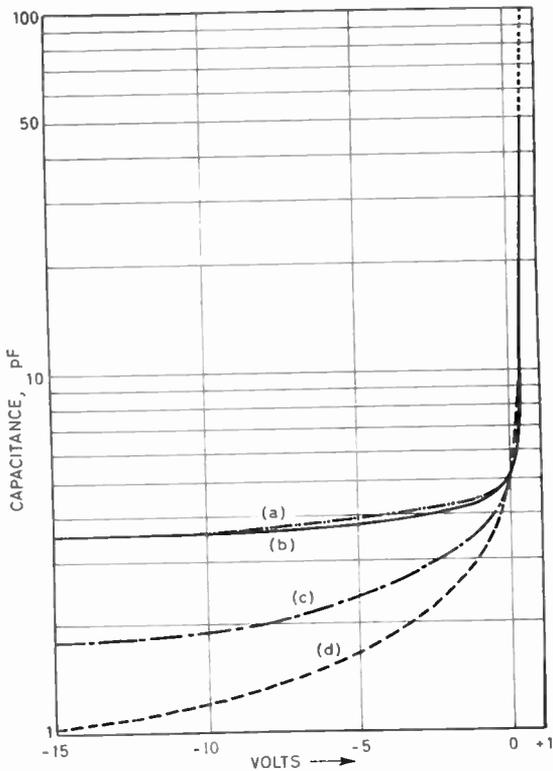
Fig. 1. Variation of capacitance with applied voltage for a
step-recovery diode.

(a) $C_{(v)} = C_{(0)}/(1 - v/0 \cdot 6)^{1/9}$

(b) measured characteristic (Hewlett Packard 0241).

(c) graded junction.

(d) abrupt junction.

Fig. 2. Idealized capacitance vs. voltage.
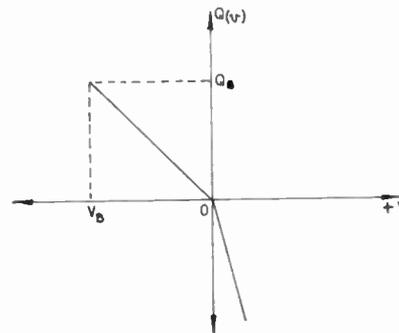
Fig. 3. Idealized charge vs. voltage.

Rafuse and Steinbrecher[8] have published the results of an analysis based upon an identical straight line approximation for the capacitance-voltage relationship in which all their equations are in closed form. The results of the analysis to be described here are in agreement with those of Rafuse and Steinbrecher where analytic solutions exist for both cases.

2. Principle of Operation

The voltage-capacitance relationship of a charge storage varactor or step-recovery diode is shown in Fig. 1 together with those of the graded and abrupt junction diodes. While low-order frequency multiplication with good efficiency may be realized with both graded and abrupt junction diodes operating in the reverse biased region, because of the approximate step-function characteristic of the step-recovery diode it may be seen that forward drive is necessary before satisfactory operation occurs.

The investigation may be separated into two sections. In Section 3, the theoretical aspects of circuit

design are considered and design parameters are established in terms of the diode characteristics, impedance matching and optimum drive level. In Section 4, utilizing the results obtained from Section 3, the realization of a practical example is described.

3. Theoretical Considerations

In order to simplify the analysis it will be assumed that the minority carrier lifetime is infinite and that the diode has an ideal step-function capacitance-voltage relationship in which C_{max} is very large so that $S_{min} = 1/C_{max} \to 0$ (Fig. 2). The charge-voltage characteristic is shown in Fig. 3.

The diode is used in a circuit of the type shown in Fig. 4. The filters are assumed to have zero loss at the frequencies indicated, and infinite loss elsewhere. In the case of idlerless multipliers only currents i_0 and i_n are allowed to flow. The maximum value of charge is q_B and we assume the ratio of fundamental to harmonic charge to be K. For maximum charge variation the alternating charge components must swing about zero.

Let the charge waveform be given by:

$$q = q_0 \sin(\omega_0 t + \theta_0) + q_n \sin(\omega_n t + \theta_n) \quad \ldots \ldots (1)$$

where $\omega_n = n\omega_0$. Then the current is given by:

$$i = \frac{dq}{dt} = q_0\omega_0\cos(\omega_0 t + \theta_0) + \omega_n q_n\cos(\omega_n t + \theta_n) \quad \ldots\ldots(2)$$

Putting eqn. (2) to zero gives the time at which the charge is a maximum, i.e.

$$q_0\omega_0\cos(\omega_0 t + \theta_0) = -q_n\omega_n\cos(\omega_n t + \theta_n)\ldots\ldots(3)$$

Substituting the value of $K = q_0/q_n$ in eqn. (3) gives:

$$\frac{K\omega_0}{\omega_n}\cos(\omega_0 t + \theta_0) = -\cos(\omega_n t + \theta_n) \quad \ldots\ldots(4)$$

Equation (4) is an nth order polynomial and in general will have to be solved numerically. If the solution $t = t'$ gives the time at which the total charge is a maximum, then:

$$q_B = q_0\left\{\sin(\omega_0 t' + \theta_0) + \frac{1}{K}\sin(\omega_n t' + \theta_n)\right\} \quad \ldots\ldots(5)$$

Assuming that $q = 0$ at $t = 0$ and that n is even, the elastance waveform will be a square waveform function of time given by:

$$S(t) = \frac{S_{max}}{2}\left[1 + \sum_{k=1}^{\infty}\frac{4}{\pi}\left\{\frac{1-(-1)^k}{2k}\right\}\sin k\omega_0 t\right] \quad \ldots\ldots(6)$$

and the diode voltage is given by[9]:

Fig. 4. Schematic diagram of multiplier.

The power input P_r to the diode due to current flowing at an angular frequency ω_r is given by:

$$P_r = \frac{\omega_0}{2\pi}\int_0^{2\pi/\omega_0} v(t)\, i_{\omega_r}(t)\, dt$$

where ω_0 is the fundamental angular frequency.

$$v(t) = \int S(t).i(t)\,dt \quad \ldots\ldots(7)$$

Substituting the values of $S(t)$ and $i(t)$ from eqn. (6) and eqn. (2) respectively in eqn. (7) gives

$$v(t) = \int \frac{S_{max}}{2}\left[1 + \sum_{k=1}^{\infty}\frac{4}{\pi}\left\{\frac{1-(-1)^k}{2k}\right\}\sin k\omega_0 t\right]\left[\sum_{r=0,n} q_r\omega_r\cos(\omega_r t + \theta_r)\right]dt \quad \ldots\ldots(8)$$

Therefore

$$v(t) = \int \frac{S_{max}}{2}\left[\sum_{r=0,n} q_r\omega_r\cos(\omega_r t + \theta_r)\right]dt + \int\left[\frac{S_{max}}{2}\sum_{k=1}^{\infty}\sum_{r=0,n}\frac{4}{\pi}\left(\frac{1-(-1)^k}{2k}\right)\sin k\omega_0 t\left[q_r\omega_r\cos(\omega_r t + \theta_r)\right]\right]dt$$

$$v(t) = \left\{\frac{S_{max}}{2}\sum_{r=0,n} q_r\sin(\omega_r t + \theta_r)\right\} - \frac{S_{max}}{2}\left[\sum_{k=1}^{\infty}\sum_{r=0,n}\left\{\frac{1-(-1)^k}{2}\right\}\frac{2}{\pi k}q_r\omega_r\left\{\frac{\cos[(k\omega_0+\omega_r)t+\theta_r]}{(k\omega_0+\omega_r)}\right.\right.$$

$$\left.\left. + \frac{\cos[(k\omega_0-\omega_r)t-\theta_r]}{(k\omega_0-\omega_r)}\right\}\right] + V_c \qquad k\omega_0 \neq \omega_r \quad \ldots(9)$$

where V_c is the d.c. voltage level obtained by inserting boundary conditions in eqn. (9).

The fundamental component of voltage is given by:

$$v_{\omega_0}(t) = \frac{S_{max}}{2}q_0\sin(\omega_0 t + \theta_0) - \frac{S_{max}}{\pi\omega_0}q_n\omega_n\left\{\frac{\cos(\omega_0 t - \theta_n)}{(n+1)} - \frac{\cos(\omega_0 t + \theta_n)}{(n-1)}\right\} \quad \ldots\ldots(10)$$

and the nth harmonic is given by:

$$v_{\omega_n}(t) = \frac{S_{max}}{2}q_n\sin(\omega_n t + \theta_n) - \frac{S_{max}}{n}q_0\omega_0\left\{\frac{\cos(\omega_n t + \theta_0)}{(n-1)} + \frac{\cos(\omega_0 t - \theta_n)}{(n+1)}\right\} \quad \ldots\ldots(11)$$

Hence the input power is given by:

$$P(\omega_0) = \frac{\omega_0}{2\pi} \int\limits_{0}^{2\pi/\omega_0} \left[\frac{S_{max}}{2} q_0 \sin(\omega_0 t + \theta_0) - \frac{S_{max}}{\pi\omega_0} q_n \omega_n \left\{ \frac{\cos(\omega_0 t - \theta_n)}{(n+1)} - \frac{\cos(\omega_0 t + \theta_n)}{(n-1)} \right\} \right] \left[q_0 \omega_0 \cos(\omega_0 t + \theta_0) \right] dt$$

Therefore

$$P(\omega_0) = \frac{\omega_n S_{max}}{2\pi} q_n q_0 \left\{ \frac{\cos(\theta_0 - \theta_n)}{(n-1)} - \frac{\cos(\theta_0 + \theta_n)}{(n+1)} \right\}$$

If $\theta_0 = 0$ the maximum value of $P(\omega_0)$ occurs when $\theta_n = (2s+1)\pi$, where s is an integer, i.e.

$$P(\omega_0)_{max} = \frac{\omega_n S_{max}}{\pi(n^2 - 1)} q_0 q_n \qquad \ldots\ldots(12)$$

In the same way the power output at angular frequency ω_n may be calculated:

$$P(\omega_n) = -\frac{\omega_0 S_{max}}{2\pi} q_n q_0 \left\{ \frac{\cos(\theta_0 + \theta_n)}{(n+1)} + \frac{\cos(\theta_0 - \theta_n)}{(n-1)} \right\}$$

Substituting $\theta_0 = 0$ and $\theta_n = (2s+1)\pi$ gives:

$$P(\omega_n)_{max} = -\frac{n\omega_0 S_{max}}{\pi(n^2-1)} q_0 q_n \qquad \ldots\ldots(13)$$

The diode acts as a source of power at the nth harmonic of the input frequency. These results are true only for a loss-free diode and in order to take into account practical devices we assume the diode is represented by a constant loss resistance r_s in series

Fig. 5. Theoretical model of diode.

with the variable junction capacitance (Fig. 5). This representation is only approximate, but measurements have shown it to be reasonably accurate. The equivalent circuit of a practical multiplier is shown in Fig. 6.

The efficiency of a multiplier having a conversion resistance r_c and a load resistance r_L is given by (see Appendix):

$$\eta = \frac{r_c}{r_c + r_s} \cdot \frac{r_L}{r_L + r_s} \qquad \ldots\ldots(14)$$

The power delivered to the load (assumed to be tuned to ω_n) is:

$$P_2 = \frac{n\omega_0 q_0 q_n}{\pi C_{min}(n^2 - 1)} = i_L^2 \frac{(r_L + r_s)}{2}$$

But $i_L^2 = (\omega_n q_n)^2$ hence,

$$\frac{n\omega_0 q_n q_0}{\pi C_{min}(n^2 - 1)} = \frac{\omega_n^2 q_n^2}{2} (r_L + r_s)$$

or

$$q_n = \frac{2q_0}{\omega_n \pi C_{min}(n^2 - 1)(r_L + r_s)} \qquad \ldots\ldots(15)$$

Also

$$r_c = \frac{2P_2}{i_1^2} = \frac{2P_2}{(\omega_0 q_0)^2} = \frac{4}{(r_L + r_s)\{\pi C_{min}\omega_0(n^2 - 1)\}^2} \qquad \ldots\ldots(16)$$

The cut-off frequency of the diode is defined as:

$$f_c = \frac{1}{2\pi C_{min} r_s}$$

Letting

$$m = \frac{\pi\omega_n(n^2 - 1)}{2\omega_c n}$$

and substituting in eqn. (16) gives:

$$r_c = \frac{r_s^2}{m^2(r_L + r_s)} \qquad \ldots\ldots(17)$$

Now from eqn. (14)

$$\eta = \frac{r_c r_L}{(r_L + r_s)(r_c + r_s)}$$
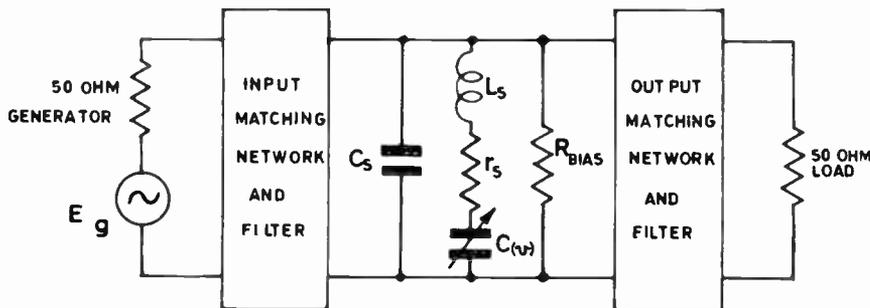
Fig. 6. Schematic diagram of a practical multiplier.

and substituting the value of r_c from eqn. (17) gives:

$$\eta = \frac{1}{(r_L+r_s)\left\{\dfrac{1}{r_L} + \dfrac{m^2}{r_L r_s}(r_L+r_s)\right\}} \quad \ldots\ldots(18)$$

The maximum value of η occurs when the denominator of eqn. (18) is a minimum, i.e.

$$\frac{\delta}{\delta r_L}\left[(r_L+r_s)\left\{\frac{1}{r_L} + \frac{m^2}{r_L r_s}(r_L+r_s)\right\}\right] = 0.$$

This occurs when:

$$r_L = \frac{r_s}{m}\sqrt{1+m^2} \quad \ldots\ldots(19)$$

Substituting this value in eqn. (18) gives the maximum efficiency as:

$$\eta_{max} = \frac{1}{\{m+\sqrt{1+m^2}\}^2} \quad \ldots\ldots(20)$$

From eqn. (15) we have:

$$K = \frac{q_0}{q_n} = n\{m+\sqrt{1+m^2}\} \quad \ldots\ldots(21)$$

under conditions of maximum efficiency. In order to calculate the power handling capability of the diode we have to solve the transcendental eqn. (4) with the phase angles $\theta_0 = 0$ and $\theta_n = (2s-1)\pi$ inserted, i.e.

$$K\frac{\omega_0}{\omega_n}\cos\omega_0 t = \cos\omega_n t \quad \ldots\ldots(22)$$

Since $\omega_n = n\omega_0$ and from eqn. (21)

$$K = n(m+\sqrt{1+m^2})$$

we can substitute in eqn. (22) to get:

$$a\cos\omega_0 t = \cos n\omega_0 t \quad \text{where} \quad a = m+\sqrt{1+m^2}$$

The solution to this equation has to be evaluated numerically. However, we can assume the value $t = t'$ in order to calculate q_0 and q_n and substitute actual values later.

$$q_0 = \frac{anq_B}{an\sin\omega_0 t' - \sin n\omega_0 t'} \quad \ldots\ldots(23)$$

and

$$q_n = \frac{q_B}{an\sin\omega_0 t' - \sin n\omega_0 t'} \quad \ldots\ldots(24)$$

Let $an\sin\omega_0 t' - \sin n\omega_0 t' = \alpha$, then:

$$q_0 = \frac{anq_B}{\alpha}; \qquad q_n = \frac{q_B}{\alpha}$$

From eqn. (12)

$$P(\omega_0) = \frac{n\omega_0 q_n q_0}{\pi C_{min}(n^2-1)}$$

and substituting the values of $q_0 q_n$ gives

$$P(\omega_0) = \frac{an^2\omega_0 q_B^2}{\alpha^2 \pi C_{min}(n^2-1)} = \frac{an^2\omega_0 V_B^2 C_{min}}{\pi\alpha^2(n^2-1)} \quad \ldots\ldots(25)$$

assuming that the input circuit is tuned to ω_0. Now

$$P(\omega_0) = \frac{1}{2}i_0^2(r_{in}-r_s) = \frac{1}{2}(r_{in}-r_s)q_0^2\omega_0^2$$

and substituting values for $P(\omega_0)$ and q_0 gives:

$$\frac{1}{2}(r_{in}-r_s)\frac{\omega_0^2 a^2 n^2 q_B^2}{\alpha^2} = \frac{an^2\omega_0 q_B^2}{\alpha^2\pi C_{min}(n^2-1)}$$

Therefore

$$r_{in} = r_s + r_s\cdot\frac{1}{am} = r_s\left(1 + \frac{1}{am}\right) \quad \ldots\ldots(26)$$

Although we have assumed that the minority carrier lifetime is infinite this is not true in practice. However, provided that the reciprocal of the input frequency is much less than the minority carrier lifetime τ the approximation may be justified.

Because of the finite minority carrier lifetime, self bias may be used.[8] From eqn. (1) with the condition $\theta_0 = 0$ and $\theta_n = (2s+1)\pi$ it may be seen that the charge waveform is symmetrical. The direct current flowing in a bias resistance R_b shunted across the diode is given by:

$$I_{d.c.} = \frac{\bar{q}^+}{\tau}$$

where $\bar{q}^+$ is the mean value of the positive half cycles of the charge waveform. The average voltage across the diode is given by:

$$V_c = S_{max}\bar{q}^+$$

Since $V_c = I_{d.c.}R_b$ we have:

$$R_b = S_{max}\tau \quad \ldots\ldots(27)$$

Apart from the solution of eqn. (22) we are now in a position to specify a procedure for the design of a multiplier. A computer program has been written to solve eqn. (22) numerically for a range of typical values of K and n.

From eqn. (25),

$$P(\omega_0) = \frac{an^2\omega_0 V_B^2 C_{min}}{\pi\alpha^2(n^2-1)}$$

From the value of

$$P_{in} = \frac{1}{2}i_0^2 r_{in} \quad \text{and} \quad P(\omega_0)$$

the input power for maximum efficiency may be calculated:

$$P_{in} = P(\omega_0) + \frac{i_0^2 r_s}{2} \quad \ldots\ldots(28)$$

Since

$$P_{out} = \eta P_{in}$$

then

$$P_{out} = \eta\left\{P(\omega_0) + \frac{i_0^2 r_s}{2}\right\}$$

Substitution of the values of i_0 from eqns. (2) and (23) and $P(\omega_0)$ from eqn. (25) gives:

$$P_{out} = \frac{V_B^2}{2\alpha^2 r_s}\left(\frac{\omega_n}{\omega_c}\right)^2\left[\frac{1+\eta}{1-\eta}\right] \quad \ldots\ldots(29)$$

Letting

$$\frac{V_B^2}{r_s} = P_{\text{norm}}$$

reduces eqn. (29) to:

$$\frac{P_{\text{out}}}{P_{\text{norm}}} = \frac{1}{2\alpha^2}\left(\frac{\omega_n}{\omega_c}\right)^2\left[\frac{1+\eta}{1-\eta}\right] \qquad \ldots\ldots(30)$$

A graph of eqn. (30) has been drawn to show the variation of efficiency and normalized output power as a function of (ω_n/ω_c) (Fig. 7) for several values of

Fig. 7. Variation of efficiency and normalized output power as a function of f_{out}/f_c.

the multiplication ratio. It is evident that the power handling capability of a diode falls rapidly with increase in the multiplication ratio for the optimum efficiency condition.

The results obtained for the doubler case are in agreement with those of Grayzel[10] in which γ tends to zero.

For high-order idlerless-multipliers it is essential to operate the circuits where the efficiency and power output approach the highest values.

4. Design Example

In order to verify the theoretical results a times-eight multiplier having an input frequency of 17 MHz was designed and constructed. The diode used was a Hewlett Packard type 0181. The following parameters were measured:

$$V_B = 80 \text{ V}$$
$$C_{\text{min}} = 5 \text{ pF}$$
$$r_s = 6 \text{ }\Omega$$
$$\tau = 120 \text{ ns}$$

The output frequency is 136 MHz, hence

$$\frac{\omega_n}{\omega_c} = 2\pi 136 \times 30 \times 10^{-6} = 0\cdot026$$

From

$$m = \frac{\pi}{2}\left(\frac{n^2-1}{n}\right)\cdot\left(\frac{\omega_n}{\omega_c}\right)$$

the numerical value of m is $0\cdot316$. The efficiency is given by:

$$\eta = \frac{1}{\{m+\sqrt{1+m^2}\}^2} = 53\%$$

$$r_{\text{in}} = r_s\left\{1 + \frac{1}{am}\right\} = 19\cdot5 \text{ }\Omega$$

$$r_L = \frac{r_s}{m}\sqrt{1+m^2} = 20 \text{ }\Omega$$

From eqn. (6) the average elastance is $\frac{1}{2}S_{\text{max}}$. The mean capacitance is therefore $2C_{\text{min}}$. These impedances are readily matched to 50 Ω source and load using standard techniques. The input and output circuits are shown in Figs. 8 and 9. A four-pole no-zero filter having a 10 MHz bandwidth was designed for the output in order to prevent unwanted current flowing in the output circuit. The input matching network introduced 2·5 dB of insertion loss, the output circuit 0·5 dB and the filter 0·6 dB. With 100 mW of input power the output at 136 MHz was 22 mW. The input power to the diode was 56 mW and the output from the diode 28 mW, an efficiency of 50% which is in good agreement with the calculated value of 53%. The optimum bias resistor was 30 kΩ compared to a calculated value of 24 kΩ.

From Fig. 7 the ratio $P_{\text{out}}/P_{\text{norm}}$ is found to be 8×10^{-6}, which gives the maximum value of input power as 17 mW. The experimental value of 56 mW is greater than the calculated value by a factor of 3. By increasing the series resistance of the diode it was found experimentally that the diode could be driven harder, although the power output remained constant, as would be expected from Fig. 7. No reason for the discrepancy between the experimental and theoretical values could be found. A circuit diagram of the complete multiplier is shown in Fig. 10.
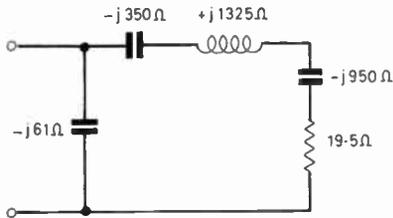
Fig. 8. 17 MHz input circuit.
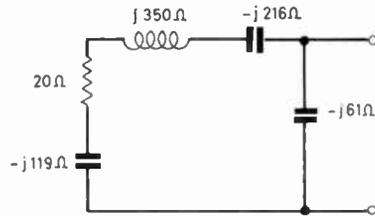
Fig. 9. 136 MHz output circuit.
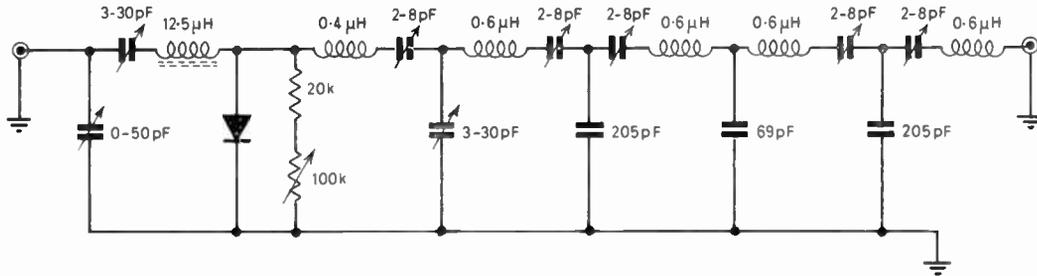
Fig. 10. Circuit diagram of the complete multiplier.

5. Conclusions

A simple analysis for the design of high-order idlerless multipliers using step recovery diodes has been described. The experimental results are in agreement with the theory, except for the power handling capability which in the particular design example is too high by a factor of 3.

The theory does not predict the instability or hysteresis encountered when the experimental circuit was being adjusted. However, it was found that once the adjustments had been completed the multiplier operated in a stable and reliable manner.

6. Acknowledgments

7. References

1. P. Penfield and R. P. Rafuse, 'Varactor Applications' (M.I.T. Press, Cambridge, Mass., U.S.A., 1962).

2. Cheng Lee Chao, 'Frequency doubler using varactors exhibiting "punch through" capacitance voltage behaviour', *Trans. Inst. Elect. Electronics Engrs on Microwave Theory and Techniques*, MTT-14, p. 43, January 1966. (Letters.)

3. R. P. Rafuse, 'Recent developments in parametric multipliers', Proceedings of National Electronics Conf., October 1963.

4. R. Hall, S. Hamilton and S. Krakauer, 'Impulse-shunt mode harmonic generation', Digest of Technical Papers, Solid-state Circuits Conference, 1966.

5. R. Thompson, 'Step-recovery-diode frequency multiplier', *Electronics Letters*, 2, pp. 117–89, March 1966.

6. C. B. Burckhardt, 'Analysis of varactor frequency multipliers for arbitrary capacitance variation and drive level', *Bell System Tech. J.*, 44, pp. 675–92, April 1965.

7. J. O. Scanlan and P. J. R. Laybourn, 'Large signal analysis of varactor harmonic generators without idlers', *Proc. Instn Elect. Engrs*, 112, pp. 1515–22, August 1965.

8. R. P. Rafuse and D. H. Steinbrecher, 'Harmonic multiplication with "punch through" varactors', Digest of Technical Papers, Solid-state Circuits Conference, 1966.

9. F. J. Hyde, 'Analysis of parametric amplifiers incorporating varactor diodes', *Proc. I.E.E.*, 110, pp. 1313–18, August 1963.

10. A. I. Grayzel, 'Greater doubler efficiency using varactor diodes with small values of gamma', *Proc. I.E.E.E.*, 53, p. 505, May 1965. (Letters.)

8. Appendix

The conversion resistance represents the ability of the diode to convert power at frequency f to nf. The equivalent circuits for the input and output frequencies are shown in Fig. 11 for tuned operation.

$$P_m = \frac{v_1^2}{8r_g} = \frac{v_1^2}{8(r_s + r_c)} \qquad \ldots\ldots(31)$$

where P_m is the available power from the generator and the circuit is matched.
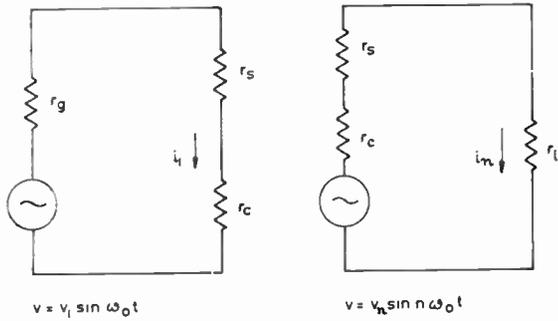
Fig. 11. Equivalent circuits of the diode at input and output frequencies for tuned operation.

From the definition of conversion resistance we have:

$$P_{nf} = \frac{i_1^2 r_c}{2} \qquad \dots\dots(32)$$

P_{nf} is the power at frequency nf generated by the diode. Also

$$i_1 = \frac{v_1}{2(r_s + r_c)} \qquad \dots\dots(33)$$

Under conditions of maximum power transfer, combining eqns. (31), (32) and (33) gives

$$P_{nf} = P_m \cdot \frac{r_c}{r_s + r_c} \qquad \dots\dots(34)$$

For the output circuit the power in the load is given by

$$P_L = \frac{i_n^2 r_L}{2} \qquad \dots\dots(35)$$

and

$$i_n = \frac{v_n}{2(r_L + r_s)} \qquad \dots\dots(36)$$

under conditions of maximum power transfer.

Also, for matched operation,

$$P_{nf} = \frac{v_n^2}{8(r_s + r_L)} \qquad \dots\dots(37$$

Combining eqns. (34), (35), (36) and (37) gives:

$$\eta = \frac{P_L}{P_m} = \frac{r_c}{r_c + r_s} \cdot \frac{r_L}{r_L + r_s} \qquad \dots\dots(38)$$

The Practical Design of Interdigital and Comb-line Filters

By

B. F. NICHOLSON,
A.Inst.P., C.Eng., A.M.I.E.R.E.†

Summary: The design procedures for interdigital and comb-line filters are summarized and the theory of these filters extended to include various aspects of design and performance which are not covered in the literature. The basic design equations are modified to give unified sets of expressions relating the resonator capacitances directly to the filter specification, and the theory of these filters is extended to include group delay, high-power breakdown effects and precise centre-frequency synthesis. In addition, the factors contributing to in-band dissipation loss are considered, and a method of predicting this loss in terms of the unloaded-Q of a slab-line resonator is given. Finally, a method of suppressing unwanted harmonic responses is outlined.

List of Symbols

A_a	attenuation in dB at frequency f_a
A_m	attenuation ripple in dB
b	ground-plane spacing
b'	'equivalent' coaxial outer diameter
c	velocity of electro-magnetic waves in air
C_f	fringing capacitance at open-circuited end of resonator
$\dfrac{C_k}{\varepsilon}$	normalized self-capacitance per unit length of the kth resonator
$\dfrac{C_{k,k+1}}{\varepsilon}$	normalized coupling capacitance per unit length between kth and $(k+1)$th resonators
C_p	parallel-plate capacitance at open-circuited end of resonator
C_t	total end capacitance $(C_f + C_p)$
d	rod diameter
f_a	frequency variable
f_0	mid-band frequency
f_1	lower band-edge frequency
f_2	upper band-edge frequency
g_k	kth ladder coefficient
h	admittance scaling factor
K	power coupling of factor
l	resonator length
n	total number of resonators
P	average power dissipated
Q_L	overall loaded Q-factor
Q_{Lk}	loaded Q-factor of the kth resonator
Q_U	mean unloaded Q-factor
Q_{Uk}	unloaded Q-factor of the kth resonator

r	inner radius of coaxial line
r'	normalized radius of inner of coaxial line
R	outer radius of coaxial line
S	complex frequency variable
s	surface-to-surface spacing of adjacent rods
W	maximum stored energy
Y_a	input/output characteristic admittance
Z_a	input/output characteristic impedance
Z_0	characteristic impedance of resonator section
ε	dielectric constant
ε_r	relative dielectric constant of the medium of propagation
λ_0	mid-band wavelength
μ	permeability of dielectric
ρ	resistivity of metallic boundaries
σ	real part of complex frequency variable
ϕ	phase angle
τ	group delay
ω	angular frequency
ω_0	mid-band angular frequency
Ω	normalized radian frequency variable

† The Marconi Company Limited, Research Laboratories, Great Baddow, Essex.

1. Introduction

The well-known comb-line and interdigital structures can be used as band-pass filters with characteristics, suitable for a wide variety of applications at u.h.f. and low microwave frequencies. They consist of arrays of parallel-coupled slab-line (or strip-line) resonators between common ground-planes and by suitable choice of the cross-sectional dimensions these filters can be designed to have pass-bands from less than 1% to an octave or more. The number of separate resonators in the array will define the attenua-
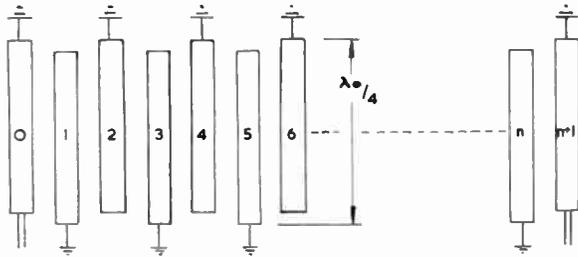
Fig. 1. Narrow-band interdigital structure.
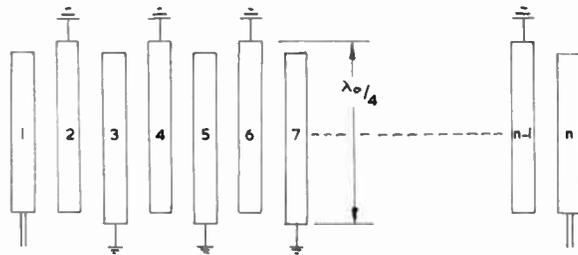
Fig. 2. Broad-band interdigital structure.
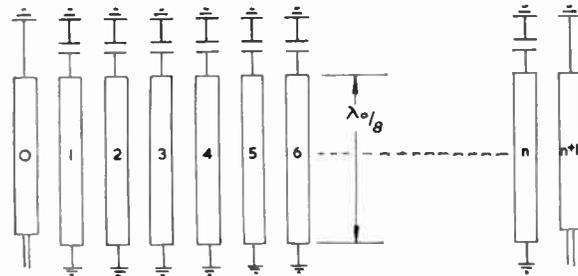
Fig. 3. Narrow-band comb-line structure.

tion rate in the stop-bands, for a given type of pass-band response. Figures 1–3 show diagrammatically the basic types of structure with each digit representing a slab-line resonator which is appropriately short-circuited, open-circuited or capacitively loaded at each end. Comb-line and interdigital filters find application in television equipment for Bands III, IV and V, radar, and communications systems in general, and since these are all applications for which stringent specifications are usually quoted it is imperative that the 'unknown quantities' in the practical realization of an idealized filter design can be pre-determined. Quantities such as in-band dissipation loss and group-delay variation can be very important and these, together with a number of other aspects of filter performance, are dealt with in the following sections.

2. Filter Design and Performance

2.1. Design Equations

The design procedures for interdigital and comb-line filters are based on two papers by Matthaei[1,2], and the subsequent realization as a physical structure is based on papers by Getsinger[3] and Cristal.[4] To design a particular filter it is necessary to refer to at least two of these papers, which can be rather tiresome and at first confusing, since different nomenclature is used in each. As an example the coupling capacitance between adjacent fingers is called ΔC by Getsinger, this becomes $C_{k,k+1}$ in Matthaei, and Cristal uses C_m for the same quantity. The actual design equations employ a number of intermediate quantities, so that the relationship between the filter parameters and the design information is not readily apparent, but by suitable rearrangement these equations can be simplified to give the coupling and self-capacitances directly as functions of the appropriate ladder coefficients and other specified data for a particular filter. Although the above points may be considered rather trivial, it is hoped that the presentation of the following design procedures will simplify the realization of these types of filter from any given specification.

As always, it is necessary to calculate the number of elements required for a given bandwidth, rejection loss, etc., and for a Chebyshev response this will be obtained from:

$$n = \frac{\cosh^{-1}\left(\frac{\text{antilog}_{10}\left(\frac{A_a}{10}\right)-1}{\text{antilog}_{10}\left(\frac{A_m}{10}\right)-1}\right)^{\frac{1}{2}}}{\cosh^{-1}\left|\frac{f_a-f_0}{f_2-f_1}\right|} \quad \ldots\ldots(1)$$

whereas for a Butterworth (maximally flat) response:

$$n = \frac{\log_{10}\left(\frac{\text{antilog}_{10}\left(\frac{A_a}{10}\right)-1}{\text{antilog}_{10}\left(\frac{A_m}{10}\right)-1}\right)}{2\log_{10} 2\left|\frac{f_a-f_0}{f_2-f_1}\right|} \quad \ldots\ldots(2)$$

n will usually be rounded up to the next highest integer but occasionally the next highest odd (or even) integer will be preferred for mechanical reasons. Equations (1) and (2) have been evaluated for several values of ripple attenuation (A_m) with rejection attenuation as a parameter and these are plotted in Fig. 4. Although Chebyshev and Butterworth filters are the types usually designed it should be stressed that comb-line and interdigital filters are not restricted to these two types of response. Any transfer function which can be realized from a simple low-pass proto-

(a) MAXIMALLY FLAT

(b) CHEBYSHEV
$A_m = 0.01$ dB
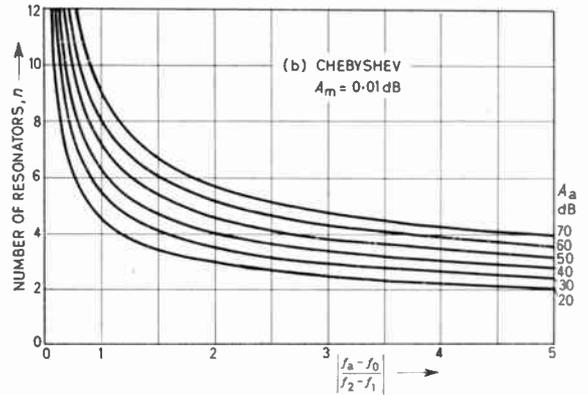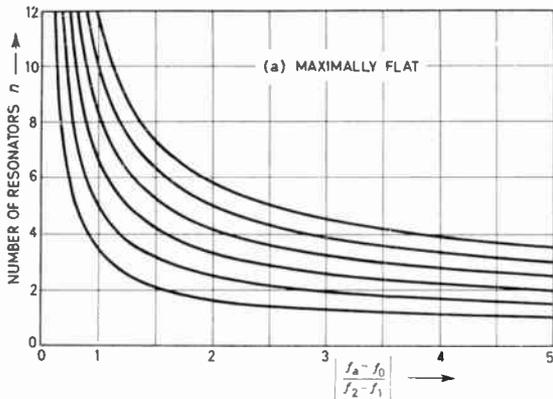
type ladder structure can be used as the basis of a comb-line or interdigital filter and several of these are shown diagrammatically in Fig. 5. These transfer characteristics can all be realized from a simple low-pass ladder network by suitable choice of the ladder coefficients. Having chosen the appropriate response function and the value of (n) to meet a particular filter specification, the prototype ladder coefficients (g_k) must be obtained. These are well tabulated[5,6] for Butterworth and Chebyshev responses and may be calculated for the other types shown in Fig. 5. By suitably rearranging and simplifying Matthaei's design equations the following sets of equations for the normalized capacitances are obtained.

Narrow-band interdigital filters:

$$\frac{C_0}{\varepsilon} = \frac{C_{n+1}}{\varepsilon} = \frac{120\pi}{\sqrt{\varepsilon_r}} Y_a \left[1 - \left(\frac{h}{g_1}\right)^{\frac{1}{2}} \right]$$

$$\frac{C_1}{\varepsilon} = \frac{C_n}{\varepsilon} = \frac{120\pi}{\sqrt{\varepsilon_r}} Y_A h \left[\frac{\tan\theta}{2} + \frac{1}{g_1} + \left(\frac{1}{g_1 g_2} + \frac{\tan^2\theta}{4}\right)^{\frac{1}{2}} - (g_1 g_2)^{-\frac{1}{2}} - (h g_1)^{-\frac{1}{2}} \right]$$

$$\frac{C_k}{\varepsilon} = \frac{120\pi}{\sqrt{\varepsilon_r}} Y_A h \left[\left(\frac{1}{g_{k-1} g_k} + \frac{\tan^2\theta}{4}\right)^{\frac{1}{2}} + \left(\frac{1}{g_k g_{k+1}} + \frac{\tan^2\theta}{4}\right)^{\frac{1}{2}} - (g_{k-1} g_k)^{-\frac{1}{2}} - (g_k g_{k+1})^{-\frac{1}{2}} \right]$$

$$\frac{C_{01}}{\varepsilon} = \frac{C_{n,n+1}}{\varepsilon} = \frac{120\pi}{\sqrt{\varepsilon_r}} Y_A \left(\frac{h}{g_1}\right)^{\frac{1}{2}}$$

$$\frac{C_{k,k+1}}{\varepsilon} = \frac{120\pi}{\sqrt{\varepsilon_r}} Y_A h (g_k g_{k+1})^{-\frac{1}{2}}$$

......(3)

where

$$\theta = \frac{\pi}{2}\left(1 - \frac{f_2 - f_1}{2f_0}\right)$$

(c) CHEBYSHEV
$A_m = 0.1$ dB

(d) CHEBYSHEV
$A_m = 0.25$ dB
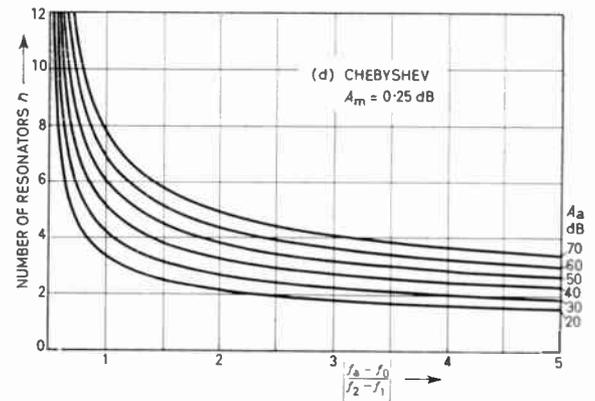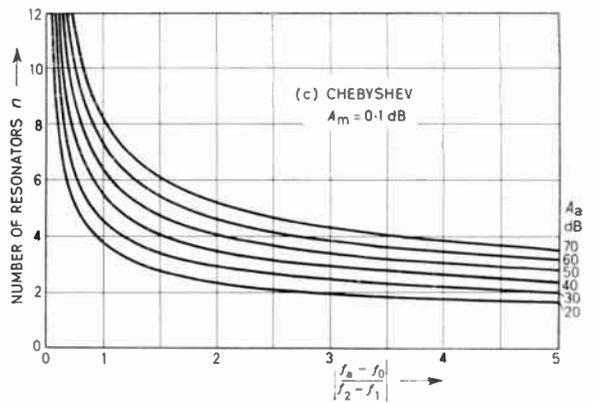
Fig. 4. Number of resonators, n, against frequency parameter for various degrees of attenuation.

and h is selected to satisfy the equation:

$$\frac{2C_{k-1,k}}{\varepsilon} + \frac{C_k}{\varepsilon} + \frac{2C_{k,k+1}}{\varepsilon} \simeq 5.4$$

for the middle of the n resonators in the filter. The value of h so chosen will produce a structure with maximum unloaded-Q for a given ground-plane spacing (refer to section on dissipation loss)
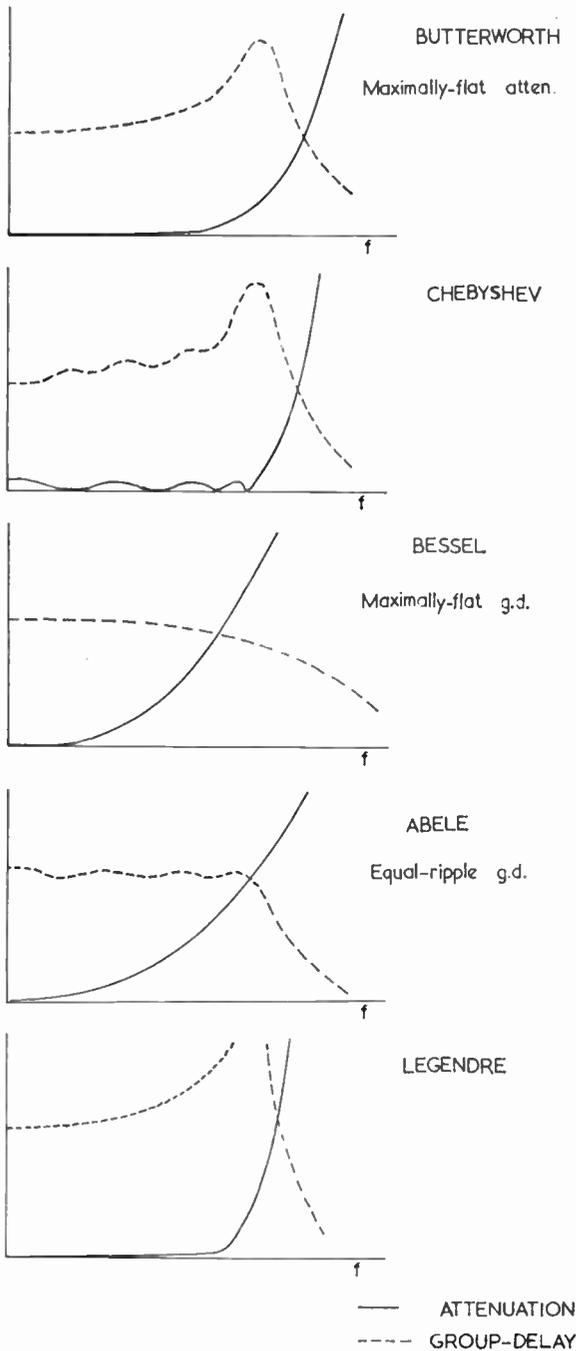
Fig. 5. Attenuation and group-delay responses.

BUTTERWORTH

Maximally-flat atten.

CHEBYSHEV

BESSEL

Maximally-flat g.d.

ABELE

Equal-ripple g.d.

LEGENDRE

—— ATTENUATION
---- GROUP-DELAY

Broad-band interdigital filters:

$$\frac{C_1}{\varepsilon} = \frac{C_n}{\varepsilon} = \frac{120\pi}{\sqrt{\varepsilon_r}} Y_A \left(\frac{1 + h^{\frac{1}{2}}}{g_1 \tan\theta} \right)$$

$$\frac{C_2}{\varepsilon} = \frac{C_{n-1}}{\varepsilon} = \frac{120\pi}{\sqrt{\varepsilon_r}} Y_A h g_2 \left[\frac{\tan\theta}{2} + \right.$$
$$\left. + \left(\frac{1}{g_2 g_3} + \frac{\tan^2\theta}{4} \right)^{\frac{1}{2}} - (g_2 g_3)^{-\frac{1}{2}} - \frac{h^{-\frac{1}{2}} - 1}{g_1 g_2 \tan\theta} \right]$$

$$\frac{C_k}{\varepsilon} = \frac{120\pi}{\sqrt{\varepsilon_r}} Y_A h g_2 \left[\left(\frac{1}{g_{k-1} g_k} + \frac{\tan^2\theta}{4} \right)^{\frac{1}{2}} + \right.$$
$$\left. + \left(\frac{1}{g_k g_{k+1}} + \frac{\tan^2\theta}{4} \right)^{\frac{1}{2}} - (g_{k-1} g_k)^{-\frac{1}{2}} - (g_k g_{k+1})^{-\frac{1}{2}} \right]$$

$$\frac{C_{12}}{\varepsilon} = \frac{C_{n-1,n}}{\varepsilon} = \frac{120\pi}{\sqrt{\varepsilon_r}} Y_A \frac{h^{\frac{1}{2}}}{g_1 \tan\theta}$$

$$\frac{C_{k,k+1}}{\varepsilon} = \frac{120\pi}{\sqrt{\varepsilon_r}} Y_A \frac{h g_2}{(g_k g_{k+1})^{\frac{1}{2}}} \qquad \ldots\ldots(4)$$

where θ and h are obtained as in the narrow-band case.

Comb-line filters with $\frac{\lambda}{8}$ resonators:

$$\frac{C_0}{\varepsilon} = \frac{C_{n+1}}{\varepsilon} = \frac{120\pi}{\sqrt{\varepsilon_r}} Y_A \left[1 - \left(\frac{0 \cdot 9(f_2 - f_1)}{g_1 f_0} \right)^{\frac{1}{2}} \right]$$

$$\frac{C_1}{\varepsilon} = \frac{C_n}{\varepsilon} = \frac{120\pi}{\sqrt{\varepsilon_r}} Y_A \left[\frac{0 \cdot 9(f_2 - f_1)}{g_1 f_0} - \frac{0 \cdot 9(f_2 - f_1)}{f_0 (g_1 g_2)^{\frac{1}{2}}} + \right.$$
$$\left. + 0 \cdot 7 - \left(\frac{0 \cdot 9(f_2 - f_1)}{g_1 f_0} \right)^{\frac{1}{2}} \right]$$

$$\frac{C_k}{\varepsilon} = \frac{120\pi}{\sqrt{\varepsilon_r}} Y_A \left(0 \cdot 7 - \frac{0 \cdot 9(f_2 - f_1)}{f_0 (g_{k-1} g_k)^{\frac{1}{2}}} - \frac{0 \cdot 9(f_2 - f_1)}{f_0 (g_k g_{k+1})^{\frac{1}{2}}} \right)$$

$$\frac{C_{01}}{\varepsilon} = \frac{C_{n,n+1}}{\varepsilon} = \frac{120\pi}{\sqrt{\varepsilon_r}} Y_A \left(\frac{0 \cdot 9(f_2 - f_1)}{g_1 f_0} \right)^{\frac{1}{2}}$$

$$\frac{C_{k,k+1}}{\varepsilon} = \frac{120\pi}{\sqrt{\varepsilon_r}} Y_A \frac{0 \cdot 9(f_2 - f_1)}{f_0 (g_k g_{k+1})^{\frac{1}{2}}} \qquad \ldots\ldots(5)$$

Evaluation of the above equations is straightforward but rather tedious, and computer programs have been written for each of the three types of filter. The programs can handle designs with up to 24

Fig. 6. Self and coupling capacitances.

(a)   (b)   (c)
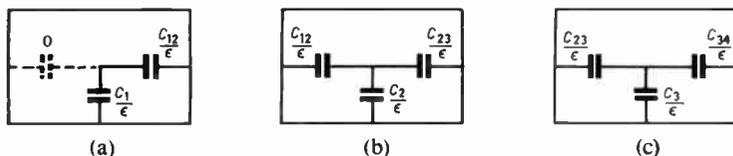
resonators, which should be more than adequate for all practical cases. The computing time is approximately $10 s + 5 s$ per resonator for the DEUCE computer (but these times will probably be reduced when the KDF9 computer is used). The programs require input data consisting of the filter specification and appropriate ladder coefficients, and produce output consisting of the input data plus values of C_k and $C_{k,k+1}$. For unusual passband ripples or very large numbers of resonators, for which the ladder coefficients may not be tabulated, an additional program is available to evaluate these. Having calculated a series of C_k's and $C_{k,k+1}$'s, the next step is to transform these into mechanical dimensions. This can be achieved by reference to the papers of Getsinger[3] or Cristal[4] for rods with rectangular or circular section respectively. The circular cross-section is attractive from the manufacturing point of view and a slightly

modified procedure based upon Cristal's work has been evolved. The proposed filter is sub-divided into separate resonators each with its associated capacitances, as shown in Fig. 6, and the graph of Fig. 7 is used to obtain s/b and d/b in the following manner. One of the sub-sections of Fig. 6 is chosen, say (b), then from the graph of Fig. 7, a value of d/b is selected such that the vertical co-ordinates of the intersections of this line with the lines having values corresponding to C_{23}/ε and C_{34}/ε, when added give the required value C_3/ε. In addition, the horizontal coordinates of the two intersections give normalized half-spacings to the two adjacent resonators. The remaining sub-sections are evaluated in a similar manner and the full normalized spacing for any pair of adjacent resonators is obtained by adding the two half-spacings for the appropriate gap. The above method is perhaps better explained by considering a particular example.
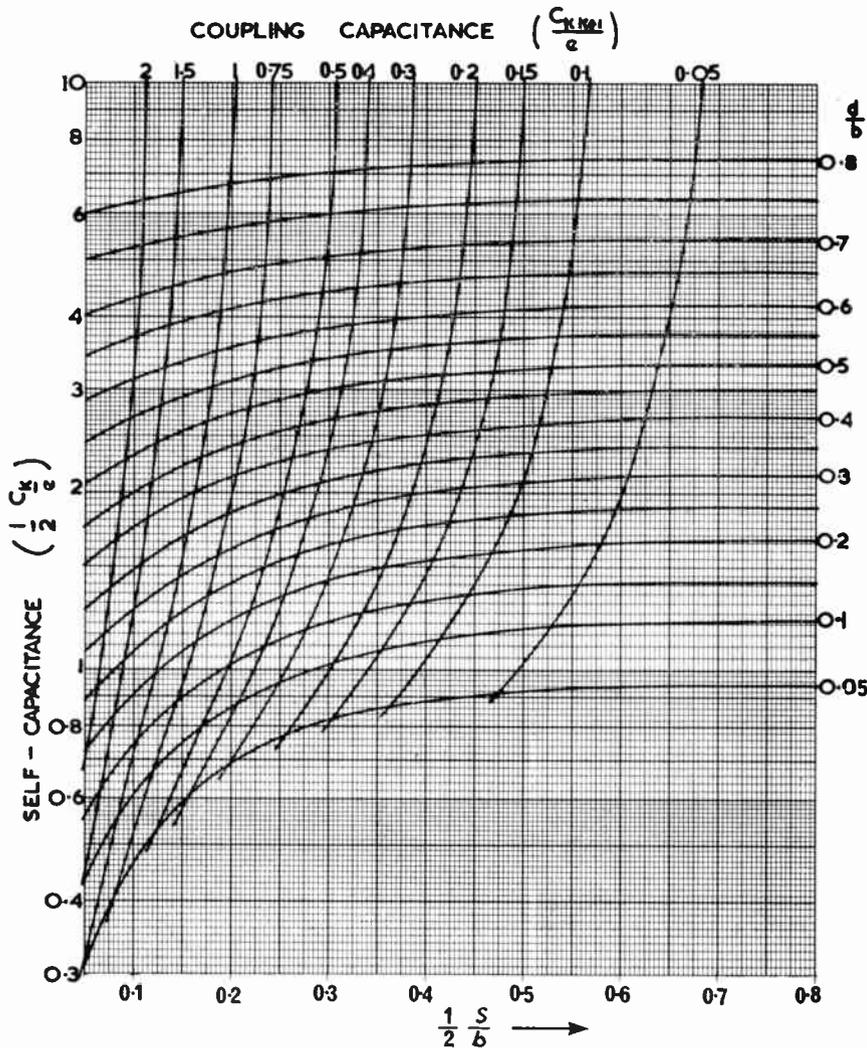
Fig. 7. Graph of rod diameters and spacings (after Cristal[4]).

In our example we will assume that the required capacitances are:

$$\frac{C_k}{\varepsilon} = 2 \cdot 87, \qquad \frac{C_{k-1,k}}{\varepsilon} = 0 \cdot 1, \qquad \frac{C_{k,k+1}}{\varepsilon} = 0 \cdot 5$$

By selecting on the graph the line $d/b = 0 \cdot 2$, the coordinates of the intersections of this line with the lines

$$\frac{C_{k-1,k}}{\varepsilon} = 0 \cdot 1 \quad \text{and} \quad \frac{C_{k,k+1}}{\varepsilon} = 0 \cdot 5$$

are found to be $0 \cdot 47$, $1 \cdot 60$ and $0 \cdot 22$, $1 \cdot 27$ respectively. Therefore by addition, $1 \cdot 60 + 1 \cdot 27 = 2 \cdot 87$ which is the required value for C_k/ε, and the normalized half-spacings to the adjacent resonators is $0 \cdot 47$ and $0 \cdot 22$. The evaluation of the other normalized dimensions is similarly straightforward with the exception of the terminating element. This element, shown in Fig. 6(a), has a neighbouring resonator on one side only and the coupling capacitance on the other side takes the value 0. In theory zero coupling capacitance will occur only when s/b is infinite, but in practice if we make $\frac{1}{2}s/b$ equal to $0 \cdot 8$ satisfactory results will be obtained since beyond this value the d/b curves are virtually horizontal. We take the right-hand vertical boundary on the graph of Fig. 7 as the zero coupling capacitance line and evaluate the diameter and spacing to the second resonator as above. All normalized cross-sectional dimensions can be evaluated in this way so that for a given ground-plane spacing the actual physical dimensions are easily obtained.

2.2. Resonant Length of Filter Elements

The resonators used in comb-line filters are $\lambda/8$ in length and are resonated by means of capacitive end-loading, whereas interdigital filters have resonators which are approximately $\lambda/4$ long, usually without end-loading. If the required design centre frequency is to be realized by a particular interdigital filter it is most important that the physical length of the resonators be accurately calculated since the facility of tuning will not generally be available. At present the accepted technique appears to be to separate the short-circuit planes by $\lambda/4$ at the design centre-frequency and to foreshorten the fingers by an amount based upon the designer's experience with some allowance for the geometric capacitance between the fingers and the end wall of the filter. Since fringing capacitance is frequently the dominant term in the expression for the total 'end-capacitance', this procedure is rather inexact and tends to shift the centre frequency of the filter by an amount which is often unacceptably large, particularly in the case of very narrow-band designs. As an example, the 10% bandwidth filter of Cristal,[4] when measured suffered a shift of $3 \cdot 8\%$ from the design centre frequency. It therefore seems apparent that the pass-bands of

filters with bandwidths less than 3% could easily be shifted beyond the desired band edges, and although the filter could be tuned by adjusting either the length of the finger or the capacitive gap at the open-circuited end, these methods are not always feasible since tuning mechanisms are undesirable when the filter is for a high-power application or when either dissipation loss or manufacturing costs must be minimized.

A rather more precise method of estimating the centre frequency can be derived by calculating the resonant frequency of each element, allowing for the capacitances associated with the open end of the finger, which are shown diagrammatically in Fig. 8.
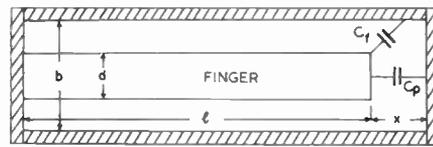
Fig. 8. Capacitances at open-circuit end of interdigital element.

Now the input impedance of a short-circuited length (l) of loss-less transmission line is given by:

$$Z_{in} = jZ_0 \tan\left(\frac{\omega l}{c}\right)$$

therefore a finger with lumped capacitance at the open-circuited end will be resonant when:

$$jZ_0 \tan\left(\frac{\omega_0 l}{c}\right) + \frac{1}{j\omega_0 C_t} = 0 \qquad \ldots\ldots(6)$$

where C_t is the total end-capacitance due to fringing and parallel plate capacitance effects. As far as is known, an exact expression for the fringing capacitance, C_f, in slab-line does not exist, but an approximation can be obtained from Whinnery et al.[7] and
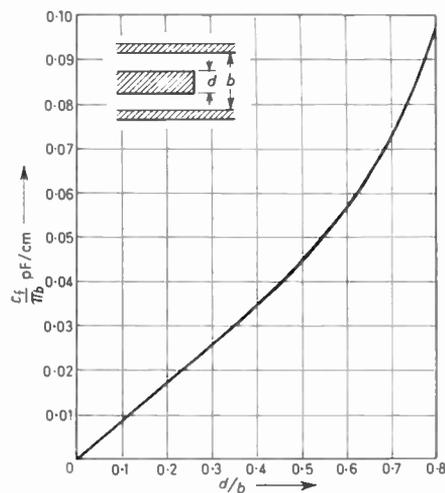
Fig. 9. Fringing capacitance.

this is shown graphically in Fig. 9. The dimension b' represents the 'equivalent' coaxial outer diameter, which may be obtained by the conformal transformation, $w = \tan z$. This mapping transforms a slab-line configuration with a nearly circular centre conductor in the z plane into a coaxial section in the w plane. The mean diameter of the inner conductor remains unchanged for typical d/b ratios, but the 'equivalent' outer diameter will be given by

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

The value of C_f may now be obtained using Fig. 9. C_p for air dielectric is given approximately by the expression

$$C_p = \frac{0.0885\pi d^2}{4x} \qquad \ldots\ldots(7)$$

where the units of capacitance will be in pF when the dimensions as shown in Figs. 8 and 9 are in cm. This method of arriving at C_t is rather inexact, but due to the proximity of the short-circuit plane, C_f will tend to be overestimated whereas C_p will always be underestimated. Thus to some extent C_t is self compensating and for dimensions commonly met in interdigital filter designs provides a reasonable estimation of the capacitive end-loading. The characteristic impedance (Z_0) in eqn. (6) will depend upon the value of the scaling factor used in the design equations, but this is usually chosen for optimum unloaded-Q which is assumed to occur when $Z_0 = 76$ ohms.

Assuming this value Z_0, if ω_0 and x can be specified, the filter may be designed by solving eqn. (6) for l. The above procedure was adopted to design a three-section, 0.1 dB ripple, Chebyshev filter with a bandwidth of 20 MHz centred upon 6.900 GHz. The ground-plane spacing was chosen to be 0.5 in which resulted in a finger diameter of 0.199 in and the spacing (x) was chosen to be 0.125 in. From Fig. 9 the fringing capacitance was found to be 0.132 pF and from eqn. (7) the parallel-plate capacitance was 0.055 pF, giving a total end-capacitance of 0.187 pF. By re-arranging eqn. (6):

$$l = \frac{c}{\omega_0} \arctan\left(\frac{1}{\omega_0 Z_0 C_t}\right) \qquad \ldots\ldots(8)$$

and hence the finger length could be calculated. The measured centre frequency of this filter was found to be 6.896 GHz, whereas scaling from the designs adopted by Matthaei[1] and Cristal[4] resulted in a filter having a centre frequency of 7.000 GHz. A number of existing filters in the frequency range 500 MHz to 7 GHz have been analysed by solving graphically eqn. (6), which is a transcendental equation in ω, for each mechanical structure. The centre frequencies calculated by this method would have re-

duced the discrepancies between design and measurement by at least 70% in all cases. As an example, the filter quoted by Cristal[4] had a design centre frequency of 1.500 GHz and by applying the above procedure a centre frequency of 1.543 GHz would be predicted, which is in close agreement with the measured centre frequency, 1.557 GHz.

It is interesting to note that to obtain a particular resonant frequency for a given end-plate separation, the length of each finger in any filter is a function of the diameter. Since the fingers in the centre of a filter are approximately equal in diameter they will all have the same resonant frequency. However, the fingers at the ends of the filter are usually of different diameters and ideally the finger lengths should be modified accordingly. This modification is not usually made and provides an alternative explanation of the slight adjustment in coupling (and hence Z_0) which always seems to be necessary between the end elements.

2.3. Group Delay

Any dispersive network will have a group- or time-delay which is dependent upon frequency and which may produce undesirable effects in a transmitted signal. Consider a band-pass filter which is used to transmit a signal composed of discrete pulses. If the Fourier components of the pulse suffer unequal delays, the phase relationship between the components will be disturbed and a distorted pulse will result. This effect will often be most marked with narrow-band filters or close to the band-edges of wider filters. It is therefore apparent that an accurate estimate of the group delay variation likely to be produced by a given filter is essential to the user of that filter. Group-delay is defined as

$$\tau = -\frac{d\phi}{d\omega} \qquad \ldots\ldots(9)$$

where ϕ is the phase angle in radians which in turn is given by[8]

$$\phi(j\omega) = -\arctan\frac{\omega-\omega_1}{-\sigma_1} - \arctan\frac{\omega-\omega_2}{-\sigma_2} \ldots -$$
$$-\arctan\frac{\omega-\omega_n}{-\sigma_n} \quad (10)$$

$$= -\sum_{k=1}^{n} \arctan\frac{\omega-\omega_k}{|\sigma_k|} \qquad \ldots\ldots(11)$$

where σ_k, ω_k are the co-ordinates of the pole locations in the complex frequency plane S. For a normalized low-pass filter with a cut-off frequency of 1 rad/s the group delay is given from eqns. (9) and (11) as:

$$\tau = \sum_{k=1}^{n} \frac{\sigma_k}{\sigma_k^2 + (\Omega-\omega_k)^2} \qquad \ldots\ldots(12)$$

For a Butterworth filter with n elements the poles of the response function all lie on a unit semi-circle and their co-ordinates are given by:

$$\left.\begin{aligned} |\sigma_k| &= \sin \frac{2k-1}{n} \frac{\pi}{2} \\ \omega_k &= \cos \frac{2k-1}{n} \frac{\pi}{2} \end{aligned}\right\} \quad \ldots\ldots(13(a))$$

For a Chebyshev filter the poles lie on an ellipse and are symmetrically placed with respect to the real and imaginary axes, their co-ordinates being given by:

$$\left.\begin{aligned} |\sigma_k| &= \sinh\left[\frac{1}{n}\operatorname{arc\,sinh}\frac{1}{\sqrt{\operatorname{antilog}_{10}\dfrac{A_m}{10}-1}}\right] \times \\ &\qquad\qquad\qquad\qquad \times \sin\frac{2k-1}{n}\frac{\pi}{2} \\ \omega_k &= \cosh\left[\frac{1}{n}\operatorname{arc\,sinh}\frac{1}{\sqrt{\operatorname{antilog}_{10}\dfrac{A_m}{10}-1}}\right] \times \\ &\qquad\qquad\qquad\qquad \times \cos\frac{2k-1}{n}\frac{\pi}{2} \end{aligned}\right\}$$
$$\ldots\ldots(13(b))$$

With interdigital and comb-line filters we are concerned only with band-pass devices, and the necessary denormalization for this case is:

$$\Omega = \frac{f_0}{f_2-f_1}\left|\frac{f}{f_0}-\frac{f_0}{f}\right| \qquad \ldots\ldots(14)$$

and the band-pass group delay is given by:

$$\tau = \frac{1+\left(\dfrac{f_0}{f}\right)^2}{2\pi(f_2-f_1)}\sum_{k=1}^{n}\frac{|\sigma_k|}{\sigma_k^2+(\Omega-\omega_k)^2} \quad \ldots\ldots(15)$$

where Ω may be obtained from eqn. (14) and σ_k and ω_k from eqns. (13(a)) or (13(b)), for Butterworth or Chebyshev responses respectively. The evaluation of eqn. (15) is straightforward but rather tedious, and a computer program has been written for the Chebyshev case. Typical results for a four-element filter with 0·5 dB attenuation ripple and bandwidth of 100 MHz at 0·9 GHz are plotted in Fig. 10. The asymmetry of the delay peaks near the band edges is due to the term $(f_0/f)^2$ in eqn. (15), but for narrow band designs the asymmetry is comparatively small and eqn. (15) can be simplified to

$$\tau = \frac{1}{\pi(f_2-f_1)}\sum_{k=1}^{n}\frac{|\sigma_k|}{\sigma_k^2+(\Omega-\omega_k)^2} \quad \ldots\ldots(16)$$

and the delay at band centre may also be approximately obtained from:
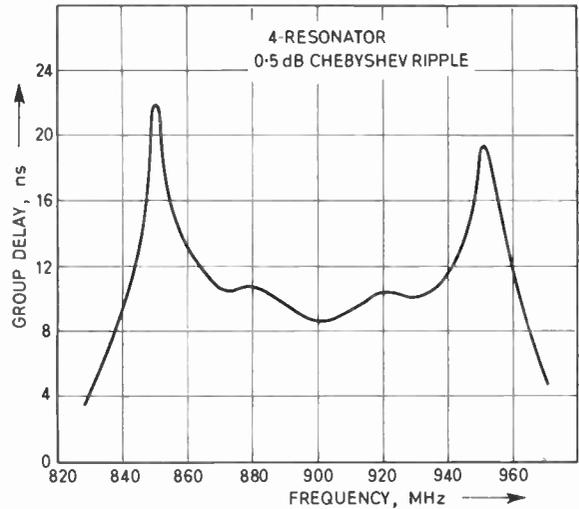
Fig. 10. Typical group-delay characteristic.

$$\tau_0 = \frac{1}{2\pi(f_2-f_1)}\sum_{k=1}^{n}g_k \qquad \ldots\ldots(17)$$

For very stringent group-delay requirements it will be necessary to employ one of the filter types shown in Fig. 5, which can give improved group delay performance, generally at the expense of out-of-band rejection.

2.4. Dissipation Loss[9]

The basic design equations for comb-line and interdigital filters make no allowance for the finite conductivity of the materials from which the filter is manufactured. The most obvious effect of this omission is an increase in the pass band attenuation of the filter, which can be quite considerable for narrow-band designs. A number of authors have covered the subject of dissipation loss in cascaded resonators fairly extensively[10-12] by calculating the dissipation loss in terms of the ratio of required filter Q, to the unloaded-Q available from the resonant elements, and this approach will be extended to cover the types of resonant structure encountered in the filters under consideration.

Consider the simple equivalent circuit of a resonant cavity shown in Fig. 11(a), in which R_s represents the resistive loss in the cavity and the two transformers provide coupling to the input and output transmission lines. These may each be replaced by an equivalent resistance R_0, as shown in Fig. 11(b), so that we are now able to define the loaded- and unloaded-Q-factors of the circuit:

$$Q_{Lk} = \frac{\omega_0 L}{2R_0+R_s} \qquad Q_{Uk} = \frac{\omega_0 L}{R_s}$$

At resonance the output voltage developed across R_0 will be given by:

$$V_{\text{out}} = V_{\text{in}} \frac{R_0}{2R_0 + R_s}$$

whereas the maximum available output voltage (when $R_s = 0$) would be $V_{\text{in}}/2$. Therefore the mid-band dissipation loss of a single resonant cavity with input and output coupling is given by

$$L_k(\text{dB}) = 20 \log_{10}\left(\frac{2R_0 + R_s}{2R_0}\right)$$

and replacing R_0 and R_s by the loaded- and un-loaded-Q's we obtain:

$$L_k(\text{dB}) = 20 \log_{10}\left(\frac{1}{1 - \dfrac{Q_{Lk}}{Q_{Uk}}}\right) \quad \ldots\ldots(18)$$

which for $Q_{Lk}/Q_{Uk} \ll 1$ reduces to:

$$L_k(\text{dB}) = 20 \times 0{\cdot}4343 \log_e\left(1 + \frac{Q_{Lk}}{Q_{Uk}}\right) \ldots\ldots(19)$$

and expanding

$$L_k(\text{dB}) = 20 \times 0{\cdot}4343 \left[\frac{Q_{Lk}}{Q_{Uk}} - \frac{1}{2}\left(\frac{Q_{Lk}}{Q_{Uk}}\right)^2 + \frac{1}{3}\left(\frac{Q_{Lk}}{Q_{Uk}}\right)^3 \cdots\right] \ldots\ldots(20)$$

Now the required loaded-Q of the kth resonator in a bandpass filter, in terms of its ladder coefficient, is:

$$Q_{Lk} = \frac{g_k}{2} Q_L$$

which when substituted in eqn. (20) gives

$$L_k(\text{dB}) = 4{\cdot}343 \left[g_k \frac{Q_L}{Q_{Uk}} - \frac{1}{4}\left(g_k \frac{Q_L}{Q_{Uk}}\right)^2 + \right.$$
$$\left. + \frac{1}{12}\left(g_k \frac{Q_L}{Q_{Uk}}\right)^3 \cdots\right] \ldots\ldots(21)$$

If we again apply the condition that $Q_L/Q_{Uk} \ll 1$, eqn. (21) simplifies to:

$$L_k(\text{dB}) = 4{\cdot}343 Q_L \frac{g_k}{Q_{Uk}} \quad \ldots\ldots(22)$$

Fig. 11. Resonator equivalent circuits.

Equation (22) is accurate within 5% for values of $Q_L/Q_{Uk} < 0{\cdot}1$. For a complete filter of n resonators the total mid-band dissipation loss will be given by:

$$L(\text{dB}) = 4{\cdot}343 Q_L \sum_{k=1}^{n} \frac{g_k}{Q_{Uk}} \quad \ldots\ldots(23)$$

In practice particularly for filters with a large number of resonators, the cross-sectional dimensions and hence the individual unloaded-Q-factors, will be similar, and a close approximation to the total loss will be obtained from:

$$L(\text{dB}) = 4{\cdot}343 \frac{Q_L}{Q_U} \sum_{k=1}^{n} g_k \quad \ldots\ldots(24)$$

The ladder coefficients have been summed for a number of commonly used Chebyshev ripples and also for a Butterworth response, with values of n up to 13 and are plotted in Fig. 12. (One interesting aspect of eqn. (24) is that the filter with minimum loss is not generally the one with fewest elements. If rejection at two specified frequencies is the prime requirement, then it is necessary to minimize the term $Q_L \sum_{k=1}^{n} g_k$, which occurs when $n = 6$ or 7.) By using eqn. (24) the problem of evaluating the dissipation in a particular filter has been reduced to that of calculating the unloaded-Q-factor of each resonator in the filter. The analysis given above is general for a cascade of n resonators which may take any suitable form. With comb-line and interdigital filters the resonators are one-eighth wavelength and quarter-wavelength fingers respectively between parallel ground-planes; the former being resonated by means of capacitive loading at the open-circuited ends of the fingers. As far as is known, exact expressions for the unloaded-Q of these slab-line resonators are not available, but a close approximation may be obtained by suitably modifying the expression for a coaxial resonator. The unloaded-Q-factor of a resonator of any arbitrary shape and size is defined as:

$$Q_U = 2\pi f_0 \frac{W}{P} \quad \ldots\ldots(25)$$

and for a quarter-wavelength coaxial cavity this becomes:

$$Q_U = \frac{240\pi^2 \log_e\left(\dfrac{R}{r}\right)}{\lambda_0 \sqrt{\pi f_0 \mu \rho}\left(\dfrac{1}{R} + \dfrac{1}{r}\right) + 8\sqrt{\pi f_0 \mu \rho} \log_e\left(\dfrac{R}{r}\right)} \ldots(26)$$

in which the two terms in the denominator represent the losses in (a) the fingers and ground-planes and (b) the end wall. Now the conformal mapping $w = \tan z$ (Fig. 13) transforms a slab-line cross-section in the z-plane into a coaxial cross-section in
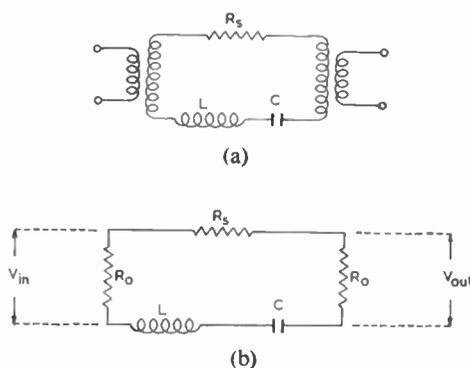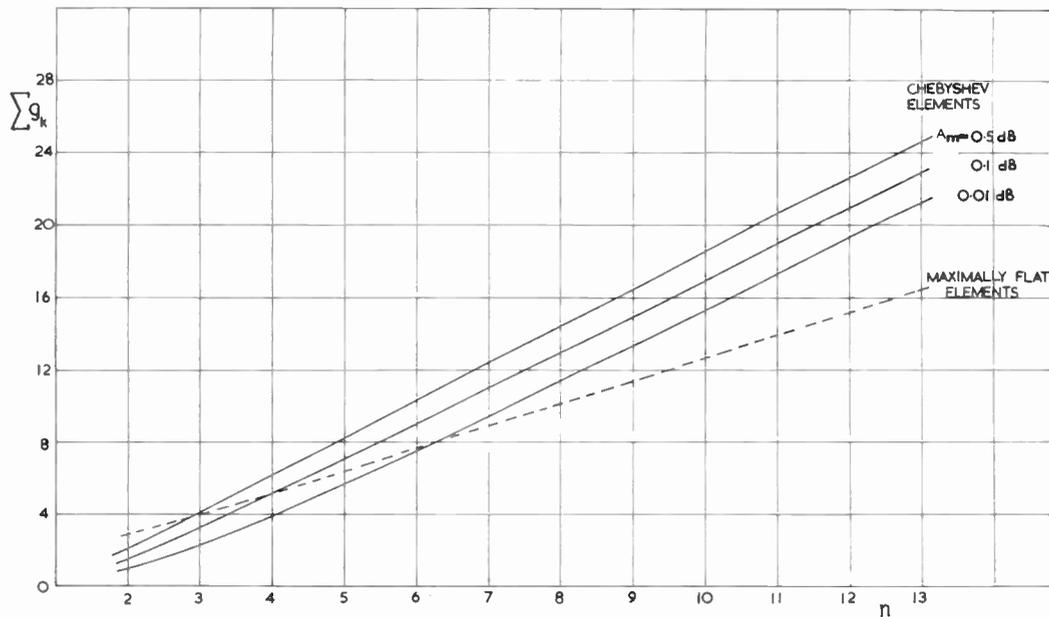
Fig. 12. Summation of ladder coefficients.

the w-plane. Although the exact transformation requires a slightly non-circular slab-line centre-conductor, for the structures under consideration this eccentricity will be very small and for all practical purposes may be ignored. If we apply this transformation to eqn. (26) we get for a slab-line resonator:

$$Q_U = \frac{240\pi^2 \log_e\left(\frac{1}{r'}\right)}{\frac{\pi\lambda_0}{2b}\sqrt{\pi f_0 \mu\rho}\left(1+\frac{1}{r'}\right) + 8\left(\frac{\lambda_0}{4l}\right)^2\sqrt{\pi f_0 \mu\rho}\log_e\left(\frac{1}{r'}\right)}$$
......(27)

The term $(\lambda_0/4l)^2$ has been introduced empirically into the second part of the denominator of eqn. (27) to satisfy the conditions for comb-line filters in which end-plate losses will have a greater effect. As mentioned above the cross-sectional dimensions of individual resonators in any filter are usually similar and it is sufficient to calculate the mean unloaded-Q, which will be obtained for $r' \simeq 0.29$. (The design equations in Section 1 are scaled by choosing h or Y_{Ak}/Y_A to give optimum Q_U and this occurs when $r' \simeq 0.29$.) Equation (27) has been evaluated for both the comb-line and interdigital structures with silver conductors and the results are plotted in Fig. 14 to show Q_U against frequency with b as a parameter. The results can be easily applied to conductors other than silver by dividing by the square-root of the ratio of the resistivities.

It must be emphasized that eqn. (27) gives the theoretical unloaded-Q of the slab-line resonator.

The practical value will always be less than that calculated above due to mechanical and chemical surface finish, lossy electrical contacts, porosity and impurities in plated metals, etc., and it is found that typically a multiplying factor of 0.8 is suitable for polished, high conductivity material at u.h.f. Another important factor is the value for the electrical resistivity (ρ) used in eqn. (27). Copper alloys and aluminium alloys have conductivities which can depend critically upon the quantities and nature of the constituent metals and it is important to select the correct value for ρ.
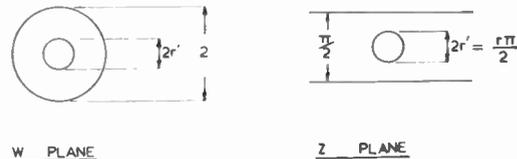
Fig. 13. Conformal mapping $w = \tan z$.

A three resonator comb-line filter was designed to have a 3.0 MHz pass-band centred on 600 MHz with a 0.1 dB Chebyshev ripple response, the ladder coefficients being $g_1 = g_3 = 1.0315$, $g_2 = 1.1474$. The filter had a ground-plane spacing of 1.0 in, 2.44 in fingers of 0.370 in diameter, and was manufactured from high-conductivity copper, (ρ) $\simeq 1.7 \times 10^{-6}$ ohm-cm and permeability (μ) $\simeq 1.257 \times 10^{-8}$ henry/cm. From eqn. (27) the theoretical unloaded-Q

was found to be 2550, giving an estimated practical unloaded-Q of 2040. By substituting this figure in eqn. (24) a value of $1 \cdot 36$ dB was arrived at for the mid-band dissipation loss. Figure 15 shows the pass-band loss predicted by the loss-less theory and by the dissipation analysis given above, together with the measured results. The close agreement which was obtained between the theory outlined above and the measured results is obviously dependent upon the empirical scaling factor from the theoretical Q_U to a practical Q_U, but it is felt that a figure of $0 \cdot 8$ is typical for filters of the type described above. In fact, values from $0 \cdot 7$ to $0 \cdot 8$ will produce excellent agreement for most filters in the u.h.f. and low microwave regions.

The derivation of Q_U above refers to resonators with fingers of circular cross-section, but a similar analysis could be used for fingers of rectangular section.

The theoretical loss which is plotted in Fig. 15 was obtained by merely adding the mid-band dissipation loss to the theoretical reflection loss predicted by a $0 \cdot 1$ dB Chebyshev response. In fact, the dissipation loss is itself a function of frequency and will increase towards the band edges. A measure of this increase can be obtained by introducing the normalized group-delay characteristic, when the dissipation loss as a function of frequency will be given by:

$$L(\omega) = L_0 \frac{\tau(\omega)}{\tau_0} \text{ (dB)} \qquad \ldots\ldots(28)$$

The peaks of the group-delay characteristic which occur near the band-edge are responsible for the 'rounding-off' of the edges which is often apparent in the responses of narrow-band filters.

2.5. Harmonic Suppression

As with any filter composed of transmission line elements, comb-line and interdigital filters are periodic

and will have harmonic responses at higher frequencies. These occur at 3 times and $4 \cdot 3$ times the fundamental response for interdigital and comb-line filters respectively and have approximately the same bandwidth. Although the attenuation of these harmonics by the filter is likely to be greater than the fundamental it is often necessary to obtain a greater level of suppression. One method of doing this is to cascade a comb-line with an interdigital structure so that the harmonic responses do not coincide, but a rather more attractive proposition would be to incorporate band-stop devices into the filter design.
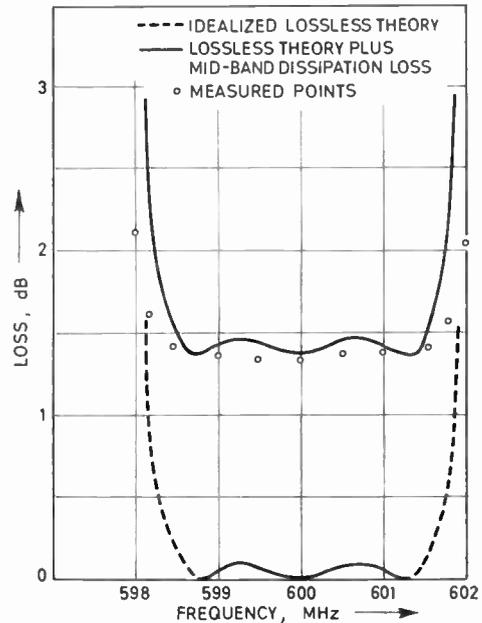
Fig. 15. Comb-line filter response.

Band-stop filters have been designed in slab-line,[13-16] the basic elements usually consisting of parallel or orthogonal coupled stub resonators, and the obvious way to incorporate these into a comb-line or interdigital filter would be to couple one or more resonators to each of the input/output coupling rods.

These shunt resonators would be somewhat less than a quarter-wavelength long at the harmonic frequency, open-circuited at one end and short circuited at the other, with a method of tuning. One possible arrangement is shown in Fig. 16 for a narrow-band interdigital filter in which two shunt resonators are coupled to the input and output lines. These would be tuned by means of capacitive screws to provide rejection of the unwanted harmonic, without significantly affecting the primary response.

Fig. 14. Unloaded-Q of slab-line resonator.
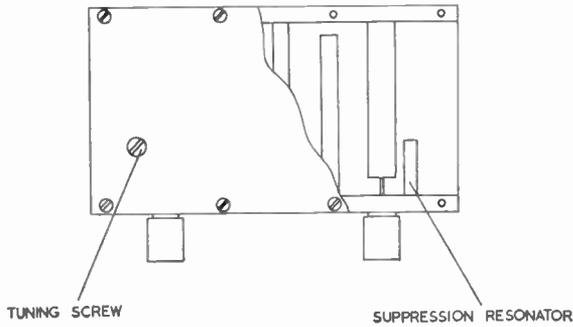
Fig. 16. Interdigital filter with band-stop elements.

2.6. *High-power Considerations*

One other important aspect of comb-line and inter-digital filter design is the maximum power which the devices can handle without overheating or voltage breakdown. Comb-line filters are obviously unsuitable for high peak-power applications since the small mechanical gaps associated with the capacitive end-loading will not withstand high voltages, but inter-digital filters can be used provided that the potential gradients in the structure do not exceed a certain critical value.

The idealized equivalent circuit of a lossless, single resonator with coupling to matched input and output transmission lines of equal impedance is shown in Fig. 17(a). By transforming the loading impedances and the voltage generator into the resonator as shown in Fig. 17(b) the loaded-Q of the circuit may be written, assuming that Z_0 is purely resistive, as:

$$Q_L \equiv \frac{2\pi f_0 L}{(K_{in} + K_{out})Z_a} \equiv \frac{1}{2\pi f_0 C(K_{in} + K_{out})Z_a} \quad \text{......(29)}$$

Now the maximum voltage due to a power input, P_0, which is developed across the inductance (or capacitance) at resonance, is given by:

$$V_{max} = \frac{(8 . P_0 . Z_a K_{in})^{\frac{1}{2}}}{(K_{in} + K_{out})Z_a} 2\pi f_0 L \quad \text{......(30)}$$

and by combining eqns. (29) and (30):

$$V_{max} = Q_L(8 . P_0 Z_a K_{in})^{\frac{1}{2}} \quad \text{......(31)}$$

The above results are for a single resonator, but we may apply them to one resonator of a chain if we assume that the remainder of the chain presents a match to the input and output of that resonator. This is obviously not strictly true, but for maximally-flat or Chebyshev response filters with low pass-band ripple should produce acceptable results. Now the relationship between the loaded-Q of the kth resonator and the overall loaded-Q of the filter has been given in Sect. 2.4 above so that we may write for the kth resonator:

$$V_{max} = \frac{g_k}{2} Q_L(8 . P_0 Z_a K_{in})^{\frac{1}{2}} \quad \text{......(32)}$$

where Q_L is now the overall loaded-Q of the entire filter. Cohn[17] has shown that the coupling coefficient between adjacent resonators is given by:

$$K_{k-1,k} = \frac{1}{Q_L(g_{k-1} g_k)^{\frac{1}{2}}}$$

and substituting this in eqn. (32) we obtain for the maximum voltage on the kth resonator:
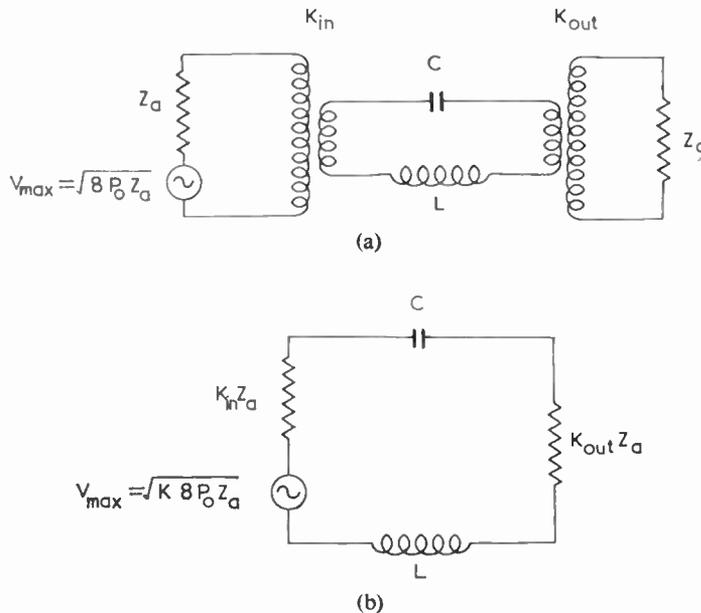
Fig. 17. Equivalent circuits of a lossless resonator.

$$V_{\max} = \left(4 \, _0 Z_a Q_L \sqrt{\frac{g_k}{g_{k-1}}} \right)^{\frac{1}{2}} \quad \ldots\ldots(33)$$

From eqn. (33), and an estimation of the maximum allowable potential gradient in the filter, the minimum mechanical spacings can be calculated. The limiting gradient for an air filled transmission line at microwave frequencies is usually taken to be 30 kV/cm under standard sea-level conditions of temperature, pressure and humidity, but this figure proves to be too optimistic for the interdigital structures under consideration. Breakdown will occur at the abrupt termination of the resonant finger and experimental evidence has shown a limiting gradient of 15–20 kV/cm to be more realistic for this geometry. Another parameter which requires some care in specifying is the power level (P_0) which the filter is required to handle, particularly with pulsed inputs. A 'rectangular' pulse is never exactly rectangular so that calculations of peak power based on the mean power level and the duty cycle should always be queried. In addition to general non-linearity, 'rectangular' pulses often have a superimposed overshoot spike at the leading edge and the power level to be used in eqn. (33) is the maximum power level reached at any time, with a suitable safety margin to allow for any random pulses of higher power.

The type of filter response shape which is most suitable for high-power applications is one for which the highest value of the ratio of any pair of adjacent ladder coefficients, (g_k/g_{k-1}) in eqn. (33) is a minimum. Obviously the 'equal element' filter is best in this respect, although certain equal ripple Chebyshev responses are only slightly inferior.

Since the voltage breakdown will occur between the open-circuited end of the finger and the ground-planes or end-plate it is essential to radius the ends of the fingers, and if capacitive tuning screws are used these should be of maximum possible diameter and also be radiused. One method of increasing the power handling capability of a particular filter is to fill the structure with pressurized freon. This gas has a much greater dielectric strength than air thereby being able to withstand great potential gradients, although the increase in manufacturing costs due to the complexity of the sealing arrangements is likely to be considerable.

The amount of overheating produced in any filter depends entirely upon the mean input power, the insertion loss and the rate at which the absorbed electrical energy can be dissipated as heat. The quantity which is of interest here is the mean as opposed to the peak-power, and the best way to minimize overheating is by reducing the in-band attenuation by the various methods described in Section 6. An additional refinement would be to incorporate support blocks manufactured from beryllium oxide into the finger structure. This material combines the thermal conductivity of aluminium with a dielectric constant of 6 and a loss tangent of 0·00045, so that excess heat could be removed from the fingers without appreciably affecting the electrical performance.

3. Conclusions

A complete design procedure for comb-line and interdigital filters of slab-line configuration has been presented. This includes a method for calculating the required resonator length for a given centre frequency which shows excellent agreement with measured results. A method for calculating the dissipation loss and group delay of a given structure is also presented which again shows good agreement with measured results. The effects of high power in a filter have also been covered and a method for the suppression of higher harmonic responses is suggested. It must be stressed that certain of the equations given above are only approximate and others will depend on the physical properties of the filter under consideration, e.g. the actual as opposed to the theoretical conductivity of the material in the case of dissipation loss. However, the author feels that this information should prove very useful to both the designer and user of these types of devices.

4. Acknowledgments

5. References

1. G. L. Matthaei, 'Interdigital band-pass filters', *Trans. Inst. Radio Engrs on Microwave Theory and Techniques*, MTT-10, p. 479, November 1962.

2. G. L. Matthaei, 'Comb-line band-pass filter of narrow or moderate bandwidth', *Microwave J.*, 6, p. 82, August 1963.

3. W. J. Getsinger, 'Coupled rectangular bars between parallel plates', *Trans. I.R.E.*, MTT-10, p. 65, January 1962.

4. E. G. Cristal, 'Coupled circular cylindrical rods between parallel ground-plates', *Trans. Inst. Elect. Electronics Engrs*, MTT-12, p. 428, July 1964.

5. 'The Microwave Engineers Handbook' (Horizon House, Dedham, Mass., U.S.A.).

6. G. L. Matthaei, *et al.* 'Microwave Filters, Impedance-Matching Networks, and Coupling Structures', (McGraw Hill, New York, 1964).

7. J. R. Whinnery, H. W. Jamieson, and T. E. Robbins, 'Coaxial-line discontinuities', *Proc. I.R.E.*, 32, p. 695, November 1944.

8. H. H. Schreiber, 'Phase and time delay of Butterworth and Tchebychev filters', *Microwaves*, 4, No. 3, p. 14, March 1965.

9. B. F. Nicholson, 'Dissipation loss in interdigital and comb-line filters', *Electronics Letters*, **2**, p. 89, March 1966.

10. J. J. Taub, 'Design of minimum loss band-pass filters', *Microwave J.*, 6, p. 67, November 1963.

11. L. Young, 'Some considerations in the design of narrow-band waveguide filters', *Trans. I.R.E.*, MTT-11, p. 522, November 1963.

12. S. B. Cohn, 'Dissipation loss in multiple coupled resonator filters', *Proc. I.R.E.*, **47**, p. 1342, August 1959.

13. B. M. Schiffman and G. L. Matthaei, 'Exact design of band-stop microwave filters', *Trans. I.E.E.E.*, **MTT-12**, p. 6, January 1964.

14. B. M. Schiffman, 'A harmonic rejection filter designed by an exact method', *Trans. I.E.E.E.*, MTT-12, p. 58, January 1964.

15. L. Young, G. L. Matthaei and E. M. T. Jones, 'Microwave band-stop filters with narrow stop-bands', *Trans. I.R.E.*, MTT-10, p. 416, November 1962.

16. E. G. Cristal, 'Addendum to "An exact method for synthesis of microwave band-stop filters"', *Trans. I.E.E.E.*, **MTT-12**, p. 369, May 1964. (Letters.)

17. S. B. Cohn, 'Direct-coupled-resonator filters', *Proc. I.R.E.*, **45**, p. 187, February 1957.

6. Appendix

Computer Program for the Design of Interdigital and Comb-line Filters

The design procedures for the three types of filter which are described in this paper are eminently suitable for use in conjunction with a digital computer since the equations for the calculation of the self and coupling capacitances of the resonators are mainly repetitive. This means that relatively simple programming loops can be used to remove much of the tedium from the design of these filters while producing results which are as accurate as the approximations in the design theory will allow. The programs which have been produced are written in ALGOL for the English Electric–Leo–Marconi KDF9 computer, and an outline of the integrated program for the three filter types which have been described is shown in the flow-chart of Fig. 18. It should be stressed that this is in no way an attempt at rigorous flow-charting but is merely included to give some idea of the structure of the program.

Input data which consists of the centre frequency pass-band, number of resonators, etc., together with an array of ladder coefficients is read into the computer, which calculates the self and coupling capacitance for each resonator up to the specified n. The mechanical dimensions of the filter must now be obtained and in order to do this the curves of Fig. 7 have been fitted with polynomial equations so that by using the appropriate capacitances and an iterative interpolation routine it is possible to converge on a normalized diameter and two half-spacings for each
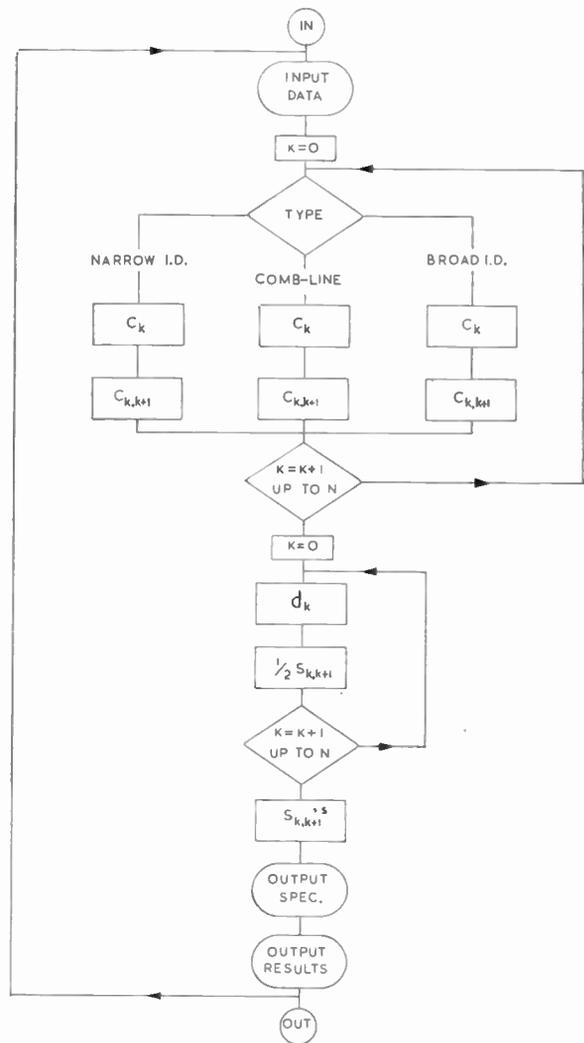
Fig. 18. Flow-chart for interdigital and comb-line filter program.

resonator in succession. These normalized dimensions can then be denormalized by the required ground-plane spacing, and the total spacing between adjacent resonators is obtained by adding the half-spacings of each. For the purpose of producing manufacturing drawings it is convenient to know the centre-to-centre spacings of the rods and this is obtained by adding the appropriate radii to each of the spacings which have been obtained above.

The output from the program consists of the input data (for reference) and the mechanical dimensions of the rods: diameters, spacings, etc., for each filter.

The Direction of Movement of F2 Region Perturbations

By

C. R. REDDI, M.Sc.(Tech.)†

AND

Professor B. R. RAO,

D.Sc., C.Eng., M.I.E.R.E.†

Summary: A new method is described for measuring the direction of movement of F2 region perturbations from the changes of phase path due to travelling ionospheric perturbations recorded at three closely-spaced aerials. The results of the analysis using this method for 151 perturbations are presented. It is found that the direction of movement of perturbations with quasi-periods of about 15 minutes and maximum height changes of about 2–3 km is within 230 to 250° to the east of north at a low-latitude station, Waltair. The direction of movement of these perturbations was found to be independent of the time of the day.

1. Introduction

Evidence for the existence of large scale irregularities in F2 region with horizontal dimensions of several hundred kilometres was obtained by Munro.[1,2,3] By comparing the virtual height records taken at different times and different places he evaluated their characteristics such as velocity of horizontal motion and direction, and lateral extent. Considerable evidence is now available to show that these large scale irregularities in ionization density are due to travelling wave-like perturbations, or the so called travelling ionospheric disturbances (t.i.d.s).[2,4]

Many experimental methods have since been used for the study of t.i.d.s, the most extensive being the spaced transmitter technique.[2,3,5] The signals from artificial earth satellites have been used by Titheridge[6] to obtain the lateral extent of the perturbations. The Doppler technique has been used by some investigators,[7] to find the north–south component of velocity.

In the present paper the effect of the t.i.d.s on the direction of arrival of the downcoming wave is investigated and a new method of measuring the direction of movement of the perturbation is presented.

2. Principle of the Method

Figure 1 shows the vertical and horizontal cross-sections of the surface of constant ionic density distorted in the three dimensions by the disturbance.‡ The horizontal section of an isotropic disturbance representing the contour of constant phase height, is shown in the figure as a circle. The movement is assumed to be entirely horizontal, along the direction indicated.

---

† Ionosphere Research Laboratories, Department of Physics, Andhra University, Waltair, India.

‡ The word 'disturbance' shall be used to represent the distorted iso-ionic surface.
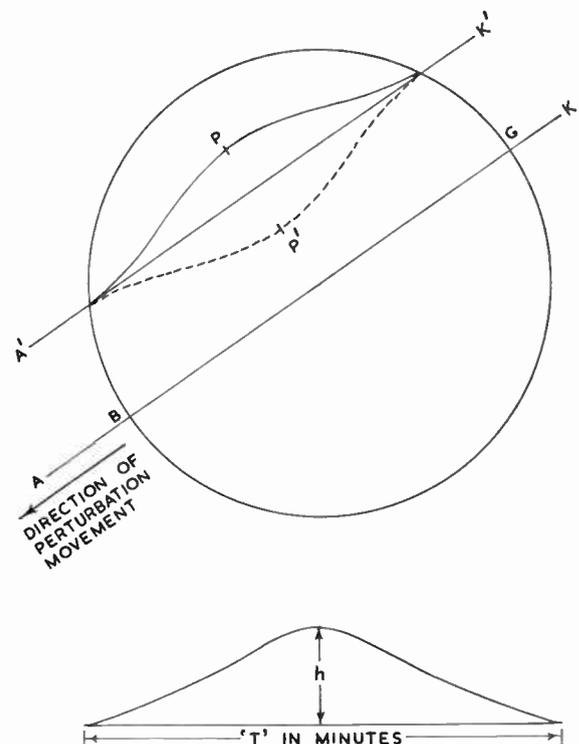
Fig. 1. Vertical and horizontal cross-sections of the surface of constant ionic density showing the direction of disturbance movement.

We shall consider pulses of radio waves radiated from an antenna at the ground, and received at the same point after reflection from the disturbance. The ray-theory shows that the reflected ray is always perpendicular to the reflecting surface so that the horizontal movement of the disturbance produces a gradual increase and decrease of phase height of reflection accompanied by changes in the direction of arrival.[8]

If the disturbance moves directly overhead along the direction indicated in Fig. 1, so that the receiving points on the ground lie along the line AK, the reflected ray is always confined to the vertical plane through AK. This means that the points of reflection on the disturbance, seen in the plan view, also lie along AK; the result is that the direction of arrival of the received signal undergoes gradual changes in the elevation angle of arrival only, the azimuth angle of arrival remaining constant. In this case the direction of movement of the disturbance is along AK.

If the movement of the disturbance is such that the receiving points on the ground lie along the line A′K′ the reflected rays would not be confined to the vertical plane passing through A′K′. The reflection points on the disturbance would lie along A′PK′. In this case the direction of arrival of the received signal undergoes gradual change both in elevation as well as azimuth angles. As the disturbance moves the phase height of reflection gradually increases and decreases, the change over taking place as the reflection point traverses the region around P. Exactly at the point P the rate of change of phase path is zero, since, the reflecting surface of the disturbance at P, although inclined to the vertical moves parallel to itself horizontally, it can be concluded that the direction of movement of the disturbance is normal to the azimuth angle of arrival at the time when there is a reversal in phase height changes observed on the ground.

The above discussion is valid even when the disturbance is convex, in which case the points of reflection on the surface will be along the dotted line A′P′K′.

3. Method of Analysis

Figure 2 shows the arrangement of the receiving antenna system. The eastern and northern antennas A and B are at a distance d from the central antenna at O. At any instant of time, PO is the direction of arrival of the received plane wave front, with an elevation angle of θ and an azimuth angle of ϕ. If ψ_0, ψ_A and ψ_B are the instantaneous phases of the signal received at O, A and B respectively, then it can be shown that

$$\psi_A - \psi_0 = -\frac{2\pi d}{\lambda} \cos\theta \cos\phi \qquad ......(1)$$

$$\psi_B - \psi_0 = -\frac{2\pi d}{\lambda} \cos\theta \sin\phi \qquad ......(2)$$

where λ is the wavelength of the operating frequency.
Hence

$$\tan\phi = \frac{\psi_B - \psi_0}{\psi_A - \psi_0} \qquad ......(3)$$

In the experimental records of travelling ionospheric disturbances, the slope of the phase path
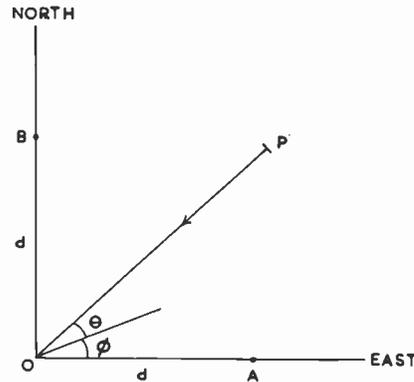
Fig. 2. Arrangement of receiving antenna system.

traces is found to be fairly constant, indicating a constant rate of phase path change for intervals of time of the order of a few tens of seconds excluding the intervals when there is a reversal in the phase path change. Hence it can be seen that the phase of the signal received at A and B would have the same value as that at O with time delays T_A and T_B depending on the values of θ and ϕ and the rate of phase path change. These time delays T_A and T_B are directly proportional to $(\psi_A - \psi_0)$ and $(\psi_B - \psi_0)$ respectively. Hence, from eqn. (3), the azimuth angle of arrival can be calculated as

$$\tan\phi = \frac{T_B}{T_A} \qquad ......(4)$$

Depending on the sign of T_A and T_B and the sense of the phase path change, whether increasing or decreasing, the exact azimuth angle of arrival can be calculated. T_A and T_B would reverse their sign as the sense of phase path change reverses sign. For all the disturbances presented in this paper the maximum height h (Fig. 1) is 3–5 km and the period T is about 15 min. Assuming a reasonable value of 8–10 km per min for the speed of the disturbance, the horizontal extent of the disturbance can be found to be about 100 km. As the disturbances are very shallow, no sharp changes in the direction of arrival are expected.

It is not possible to measure accurately the azimuth angle of arrival at the phase path reversal by any simple measuring procedure. The average of the azimuth angles on either sides of the reversal can be taken as a fairly accurate estimate of its value at the point of reversal, and using this value, the direction of movement of the perturbation can be evaluated.

From the above discussion, it can be seen that there is an ambiguity of 180°, in estimating the direction of travel of the disturbance. However, this ambiguity can be very easily resolved, from the sense of the

north-component of the velocity, which can be obtained from the sign of the time-delay between the reversals in the phase path change observed on the ordinary and extraordinary components. It is observed that the reversals in the phase height changes of the ordinary and extraordinary waves due to the travelling disturbances occur with a time-delay of about 1–2 min, the actual value being different from disturbance to disturbance. This must be due to the actual spatial separation of the reflection points of the ordinary and extraordinary rays. It is known from the magneto-ionic theory[9] that, due to the presence of the Earth's magnetic field, the magneto-ionic splitting of the radio wave occurs, and for a plane stratified ionosphere, the ordinary and extraordinary rays diverge from the vertical in opposite directions in the Earth's magnetic meridian plane. The ordinary ray deviates to the nearest pole, and the extraordinary ray deviates towards the Equator for propagations in both the hemispheres. Thus the reflection points for the ordinary and extraordinary rays are spaced both horizontally and vertically. Reasonable estimates[10] of the horizontal separation between the reflection points would be about 30 km, whereas the vertical true height separation between the two reflection points would be of the order of about 10–15 km. It is known that the phase of the travelling disturbance lags as the height decreases,[3] the tilt-angle of the front being in the range of 20–30° from the vertical. Assuming a reasonable value for the tilt-angle of the front of the travelling disturbance to be about 30° from the vertical, it can be seen that the horizontal separation of crests or troughs on a t.i.d.s for a vertical separation of 10 km is about 5 km which is much less than the horizontal separation of 30 km between the reflection points of O and X components and as such there is no possibility of the sense of the direction of movement being changed by the tilt of the front of the travelling disturbance. It can be therefore clearly seen, that

although the time-delay between the reversals in phase height changes at the two components is affected by the changes in the directions of arrival, and the tilt in the disturbance wave front, the sense of the time-delay can be taken as a reliable indication for the sense of the north–south component of the velocity of the disturbance. The reversal occurring earlier on the ordinary ray compared to the extraordinary ray indicates, that the north–south component of velocity is towards south.[10]

4. Experimental Technique

The essential principle of recording the phase path variations, used in the present investigation is similar to Findlay's original technique of comparing the phase of the reflected signal from the ionosphere against a local reference oscillator, phase coherent with the transmitted pulse. Changes in the phase of the reflected signal produce movement of the beat pattern, produced after phase comparison, detection, and further amplification along the length of the time-base displaying it.

The block diagram of the equipment is shown in Fig. 3. In contrast to Findlay's experiment,[11] the phase comparison in the present experiment is made at the intermediate frequency of 455 kHz to suit some other experiments planned with this equipment.

The three-way electronic switch, incorporating three flip-flop multivibrators connected to form a ring counter,[12] and three gated amplifiers, operated by the ring counter in order, enables recording the phase path changes simultaneously on three aerials. The three aerials used are horizontal dipoles, situated at the corners of a right angled triangle (see Fig. 2). The eastern and the northern aerials A and B are separated from the central aerial at O by 108 m. The dipoles have their lengths along the east–west direction, to facilitate the reception of the extraordinary compo-
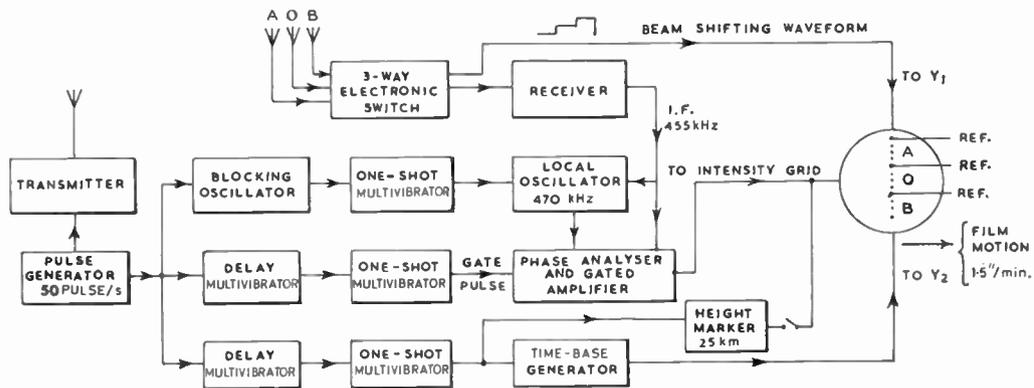
Fig. 3. Block diagram of the spaced aerial phase path recording equipment.

nent in sufficient strength, and are aligned to be perfectly parallel using distant aligning poles. An electronic gate synchronized with the transmitter, with a controlled delay isolates the required echoes. Also the beginning of the gate-pulse gives at the final output a very sharp spike, producing an intensity dot under intensity modulation display, fixed in position on the time-base, and is used as reference for the measurement of the instantaneous relative phase of the signal at the three aerials.

The height-marker circuit generates sharp spikes at 3 kHz. These spikes are used for the group height measurements. The height-marker circuit is made inoperative during the experiments presented in this paper, to avoid overlapping. A typical record is shown in Fig. 4, in which the three steady traces are the reference lines for the absolute phase measurements at the three aerials. The phase path variations across the width of the received pulse follow immediately below the reference trace of each antenna. Positive slope of the phase traces indicates increasing phase path, and negative slope indicates decreasing phase path. Both the ordinary and extraordinary components are always gated for recording to provide a check on the reliability of the measurement.

In the phase path records the position of the phase trace with respect to its reference line for any one of the three receiving antennae indicates the instantaneous phase of the received signal.

To measure T_A and T_B the time delays after which the phase of the signal received at A and B is the same as that at O the following procedure has been adopted. The record is magnified ten times and is projected on a graph sheet. On a separate small strip of paper with a vertical line, one of the phase traces along with the steady reference line of the central aerial, immediately before or after the reversal are traced carefully from the projected image. The phase trace and the reference line on the small paper are compared for exact coincidence with the corresponding traces of the

other two aerials, in the projected image on the graph sheet. T_A and T_B are measured as the shift in the position of the vertical line on the small paper against the graph sheet below. The diurnal changes in the phase height of the F2 layer reflected level would introduce some error in the measurement of the direction of movement of the perturbations by the above method as the time of reversal will not coincide with the time when the reflection point crosses P. The change of phase path due to the diurnal layer height changes added to the change of phase height due to the travelling disturbances makes the time of reversal shifted by an interval depending upon the two rates of change of phase paths. No reversal in the phase path would be recorded, if the rate of change of phase path due to layer movements is more than that due to the travelling disturbance. The reversal would be largely asymmetric if the rate of change of phase path due to the layer movement is considerable, when compared to the rate of change of phase path due to the moving disturbance. If the changes in phase path around reversal point are symmetrical, it can be said that the changes in the phase path recorded on the ground are mostly due to the t.i.d.

T_A and T_B are directly proportional to $\psi_A - \psi_0$ and $\psi_B - \psi_0$, and their magnitude and sign depend on the rate and sense of the phase path change. The measurement of the azimuth angle of arrival as the ratio T_B/T_A is valid even where vertical movement of the layer is superposed on a t.i.d. The important requisite of this method is that the rate of change of phase path should be constant (which is indicated by the constant slope of the phase trace).

5. Experimental Results

The spaced aerial phase path records of F2 region echoes on 5·6 MHz have been made from November 1965 through March 1966. The recording was mainly during daylight hours (0600–1800 hr) and each record is of 1–2 hr duration. The phase height of reflection
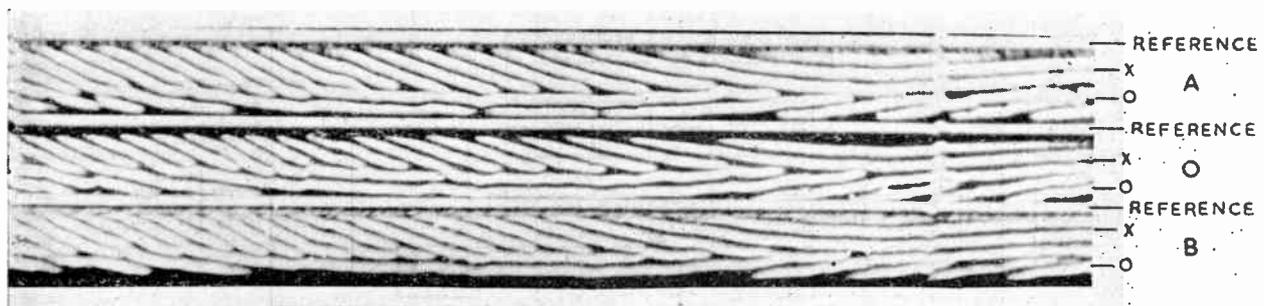
Fig. 4. Typical spaced aerial phase path record showing the occurrence of the reversal in the phase path changes on ordinary and extraordinary components with a time delay of about 40 seconds.
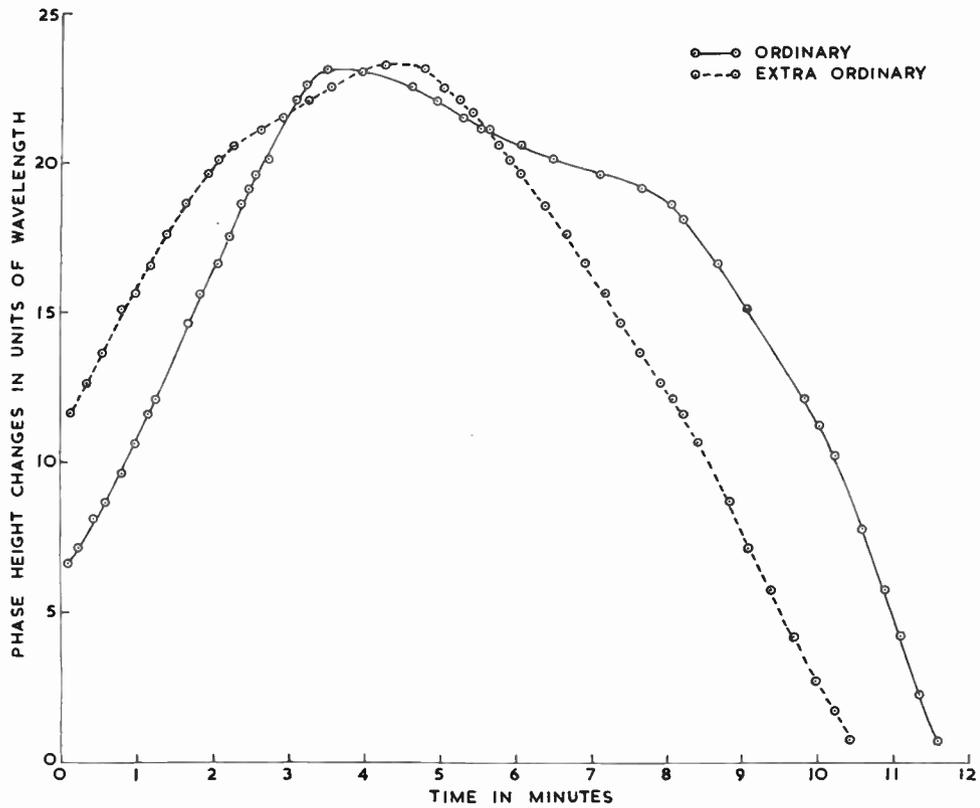
Fig. 5. Typical phase path variations of a disturbance on ordinary and extraordinary components. (Frequency = 5·6 MHz. The zero time phase paths are arbitrary.)
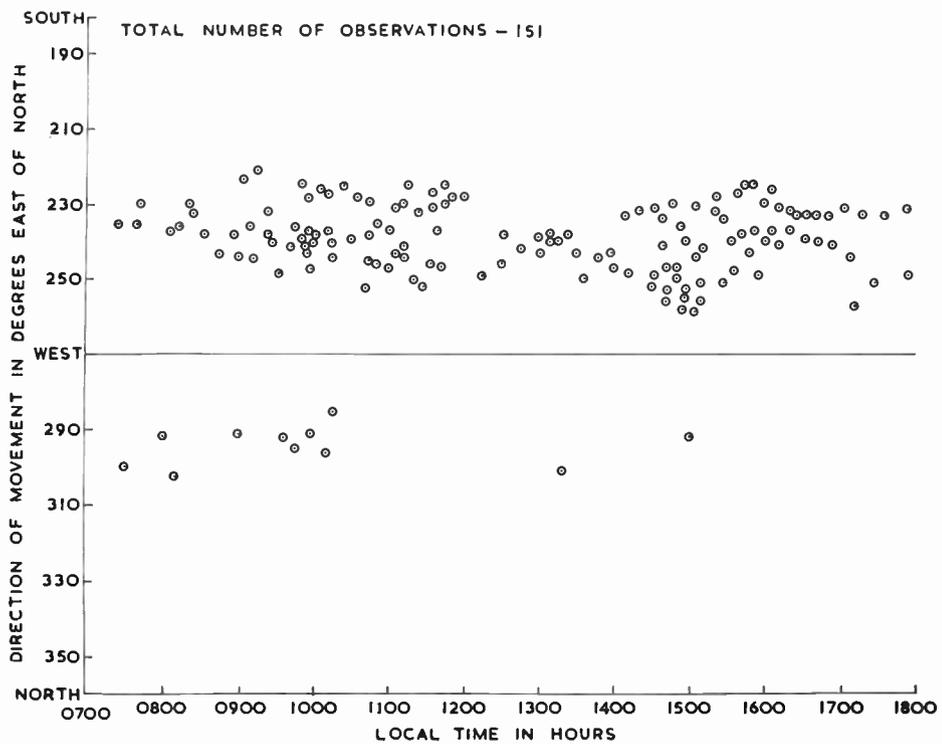
Fig. 6. The mass plot showing the direction of movement of the disturbances against local time.

of the fixed frequency pulsed signals from the F2 region was found to show quasi-periodic increases and decreases with amplitudes of about 2–3 km; and period about 15 min. Typical phase path variations due to disturbance recorded on the ordinary and the extraordinary components are shown in Fig. 5. These disturbances occur infrequently during night. It is proposed to measure the direction of movement of the

A plot of the direction of movement of the disturbances against the local time is shown in Fig. 6. Each observation is an average of the direction obtained from the ordinary and extraordinary components, which always agreed to within 5°.

In general it is observed that the frequency of occurrence of the disturbances around the noon is comparatively less. However, the density of points
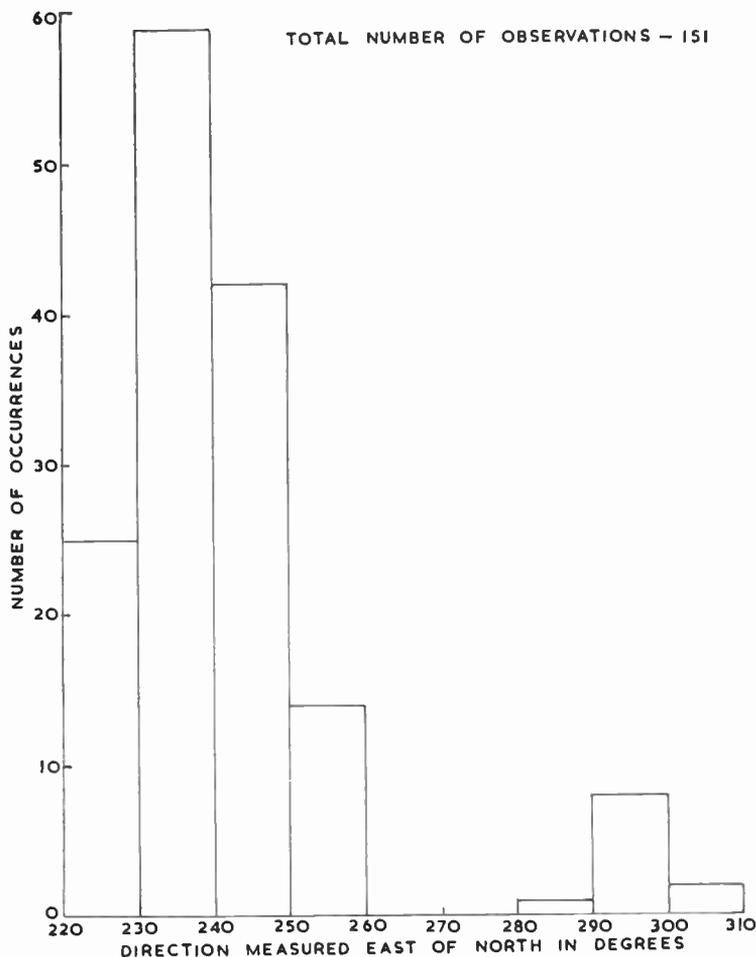
Fig. 7. Histogram showing the frequency of occurrence of the direction of movement of the travelling disturbances.

disturbances by the method described in the previous sections. Almost all the records during 0600–0700 hr and 1700–1800 hr were found to be not suitable for analysis, because the rapid phase height changes during these hours superposed on the phase path changes due to the disturbances did not give smooth, symmetrical reversals. Records effected by the presence of sporadic-E layer giving partial reflections were also rejected. A total of 100 hr of records were available giving 151 individual disturbances.

against the local time shown in Fig. 6 is by no means indicative of the relative frequency of occurrence.

Figure 7 is the histogram showing the frequency of occurrence of the direction of movement of the disturbances. From Figs. 6 and 7 it may be seen that the general direction of movement of the disturbances is within 230–250° east of north. It is observed that these disturbances do not, in general, travel strictly in the same direction even for periods of the order of one hour. In most of the records when three

or four disturbances were recorded during an hour the variation in the direction of travel from one disturbance to another was within 20°, whereas the direction calculated for the same disturbances from the ordinary and extraordinary components agrees within 5°.

As the extent of variation of the direction of movement within an hour is of the same order as that from hour-to-hour, during a day, it can be inferred that there is no significant diurnal variation in the direction of movement of t.i.d.s. However, previous investigators[2,3,5] at high latitudes have reported significant diurnal variation. The t.i.d.s have not been studied at any low latitude station, and as such the results of the present investigation cannot be compared with the available results at high latitudes.

In the above discussion, the effect of the disturbance at the reflection level has been considered. Fooks[13] has shown that the phase path variations observed on the ground correspond mostly to the variation at the reflection level.

The present method has the advantage, that disturbance with small vertical amplitudes can be detected. The technique used by Munro[2,3] would detect only large-scale disturbances producing the characteristic anomalies, or large changes in group height. It is worthwhile investigating the t.i.d.s using phase path recorders with a large separation of the order of 1 or 2 km.

6. Acknowledgment

7. References

1. G. H. Munro, 'Short period variations in the ionosphere', *Nature*, **163**, p. 812, 1949.

2. G. H. Munro, 'Travelling disturbances in the ionosphere', *Proc. Roy. Soc.*, **202**, p. 208, 7th July 1950.

3. G. H. Munro, 'Reflexions from irregularities in the ionosphere', *Proc. Roy. Soc.*, **A219**, p. 447, 7th October 1953.

4. C. O. Hines, 'Motions in the ionosphere', *Proc. Inst. Radio Engrs*, **47**, p. 176, February 1959.

5. R. E. Price, 'Travelling disturbances in the ionosphere', *Nature*, **172**, p. 115, 18th July 1953. (Letters.)

6. J. E. Titheridge, 'Large scale irregularities in the ionosphere', *J. Geophys. Res.*, **68**, p. 3399, 1st June 1963.

7. T. B. Jones and L. C. Wand, 'Doppler studies of complex reflections produced by travelling ionospheric disturbances', *J. Atmos. Terrestr. Phys*, **27**, p. 1111, June 1965.

8. E. N. Bramley, 'Direction finding studies of large scale ionospheric irregularities', *Proc. Roy. Soc.*, **A220**, p. 39, 1955.

9. K. G. Budden, 'Radiowaves in the Ionosphere', p. 246 (Cambridge University Press, London, 1961).

10. G. H. Munro and L. H. Heisler, 'Divergence of radiowaves in the ionospheres', *Aust. J. Phys.*, **9**, No. 3, p. 343, September 1956.

11. J. W. Findlay, 'The phase and group paths of radiowaves returned from region E of the ionosphere', *J. Atmos. Terrestr. Phys*, **1**, p. 353, No. 5–6, 1951.

12. N. A. Moerman, 'Four channel electronic switch', *Electronics*, **19**, p. 150, April 1946.

13. G. F. Fooks, 'Ionospheric irregularities and the phase path of radiowaves', *J. Atmos. Terrestr. Phys*, **24**, p. 937, November 1962.

The Equivalent Circuit of Two-port Antennae with an Application to Magic-T Systems

By

Professor J. VAN BLADEL, Ph.D.†

Summary: An equivalent circuit is presented for a two-port antenna immersed in the field of an external source. The circuit consists of admittances and current generators, and the value of the latter is particularly simple when the irradiating wave is a plane wave. The existence of a plane of symmetry also simplifies matters, particularly when the two output ports are connected to the series arms of a magic-T.

List of Symbols

$\bar{E}_i, \bar{H}_i$ incident fields

$\bar{E}_t, \bar{H}$ transverse fields

$\bar{F}$ radiation vector

I_g generator current

$\bar{J}$ volume current density

J_s surface current density

R_c characteristic resistance

$\bar{u}_a$ unit vector in direction 'a'

Y_a radiation admittance

Y_c characteristic admittance

Y_m mutual admittance

μ_0 permeability of free space
$$(4\pi \times 10^{-7} \text{ henry m}^{-1})$$

ω angular frequency

1. Introduction

Two typical magic-T antennae are shown in Figs. 1(a) and (b). In the first example, the magic-T is used to centre-feed a slotted antenna. In the second example, the magic-T is connected to a radiating system through a length of waveguide, possibly with a view toward avoiding mechanical obstacles. In both cases the antennae are symmetric, and 'sum' and 'difference' signals appear in the T_3 and T_4 arms respectively. These antennae are therefore suitable for monopulse applications. Both aerial systems are particular examples of a more general structure shown in Fig. 1(c), and where the 'two-port' antenna connected to terminal planes T_1 and T_2 contains dielectrics, absorbers, radiating apertures A_1 to A_4 and a reflector R. It is our purpose to derive formulae for signals picked up in T_3 and T_4 by the antenna when
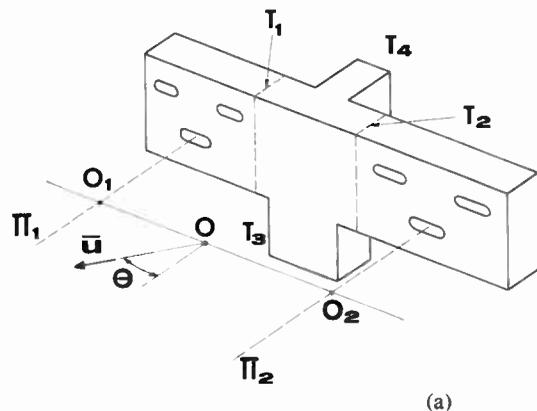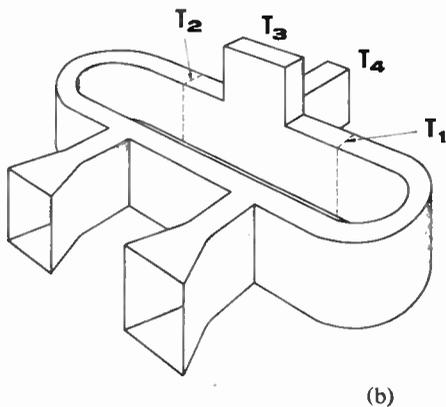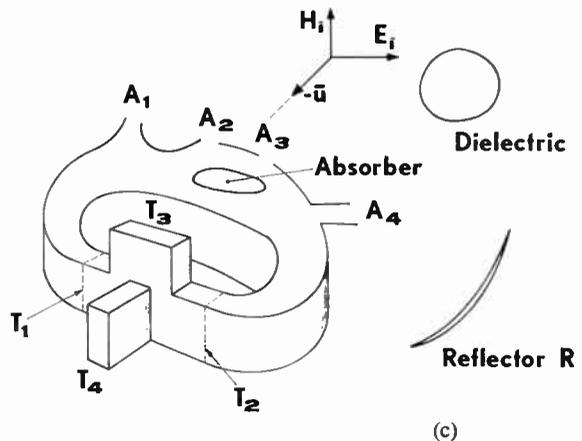
Fig. 1. Typical magic-T antennae.

(a)

(b)

(c)

† Laboratory for Electromagnetism and Acoustics, University of Ghent, Belgium.

the latter is excited by an incident plane wave $(\bar{E}_i, \bar{H}_i)$ propagating along the unit vector $-\bar{u}$. No anisotropic media are included, but the generalization to this case is quite straightforward.[1]

2. Equivalent Circuit of a One-port Antenna

We first summarize certain fundamental results valid for a one-port antenna of the type shown in Fig. 2. The antenna is connected to a receiver through a length of waveguide. We shall assume that the guide supports its lowest mode only,[2] and that S is sufficiently far from discontinuities to ensure negligible contributions from the damped modes. Under these conditions the following relationships hold for the transverse electric and magnetic fields:

$$\bar{E}_t = V\bar{\alpha}$$
$$\bar{H}_t \times \bar{u}_n = I\bar{\alpha} \qquad \ldots\ldots(1)$$

The eigenvector $\bar{\alpha}$, characteristic of the structure of the propagated mode, is real and normalized, by which is meant that

$$\iint\limits_S |\bar{\alpha}|^2 \, dS = 1$$

It is a function of the transverse co-ordinates in terminal plane S, and is seen to have the dimension m^{-1}. It then follows that the constants V and I have the dimensions of a voltage and a current. An application of Poynting's theorem shows that the average power flowing along $\bar{u}_n$ into the receiver is given by $\frac{1}{2}\,\mathrm{Re}\,(VI^*)$. The circuit aspect can be emphasized further by noticing that the field in the shielded receiver is generated by the tangential component of $\bar{E}$ on S. In particular, the value of $\bar{H} \times \bar{u}_n$ on the receiver side of S is given by a relationship of the form

$$I = Y_{rec}V \qquad \ldots\ldots(2)$$

where the admittance Y_{rec} is the ratio between $\bar{H} \times \bar{u}_n$ and $\bar{E}_t$ ($\bar{u}_n$ points inside the region). On the antenna side of S, the fields are due to the combined effects of the tangential component of $\bar{E}$ and of the volume currents $\bar{J}$, which represent the primary sources (a transmitting antenna for example). On S, in particular, the transverse magnetic field satisfies a relationship of the form

$$\bar{H}_t \times \bar{u}_n = (-Y_a V + I_g)\bar{\alpha} = I\bar{\alpha}$$

Putting $V = 0$ in this equation gives $\bar{H} \times \bar{u}_n = I_g\bar{\alpha}$. The physical interpretation of $I_g\bar{\alpha}$ is now obvious: it is the surface current density $\bar{J}_S$ which is induced on S by the source-currents $\bar{J}$ when S is short-circuited. I_g therefore represents a generator current. The admittance Y_a is the radiation admittance of the antenna, i.e. the ratio between $\bar{H}_t \times (-\bar{u}_n)$ and $\bar{E}_t$ as observed to the right of S in the absence of sources (i.e. for $\bar{J}$ equal to zero). We notice that the conductance g_a consists of two parts, one associated with
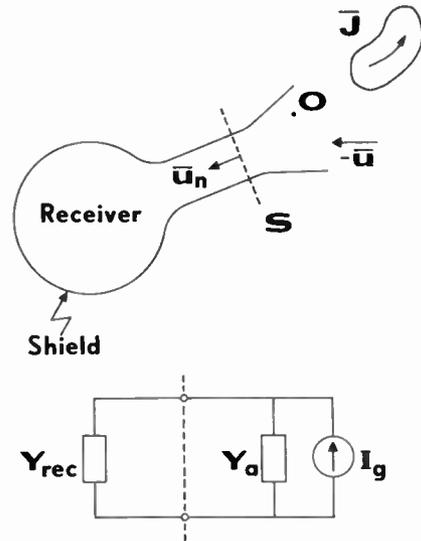
Fig. 2. A one-port antenna with equivalent circuit.

the power dissipated in the absorbers surrounding the antenna, and another associated with the radiated power. The negative sign in front of Y_a is justified by the fact that $-\bar{u}_n$ is the unit vector pointing inside the region. The circuit equation describing the behaviour of the antenna is obtained by equating the values of $\bar{H}_t \times \bar{u}_n$ on both sides of S. This yields

$$I_g - Y_a V = Y_{rec} V \qquad \ldots\ldots(3)$$

from which V can be found, I_g being given. In general, the determination of I_g is a major problem, but considerable simplification occurs when the antenna is located within the 'radiation region' of the sources. Under these conditions, the incident wave can be assimilated to a plane wave, and the response of the antenna is connected to its radiating properties by the reciprocity theorem. More precisely, assume that the tangential electric field on S is $V\bar{\alpha}$. The resulting radiation field in the direction $\bar{u}$ is of the form

$$\bar{E} = V \frac{e^{-jkR}}{R} \bar{F}(\bar{u})$$

with respect to an origin O. If a plane wave whose electric vector is $\bar{E}_i$ in O is incident on the antenna, it can be shown[1] that a short-circuit current

$$I_g = \frac{4\pi}{j\omega\mu_0} (\bar{F} \cdot \bar{E}_i) \qquad \ldots\ldots(4)$$

appears on S. It is to be noticed that I_g, Y_a and $\bar{F}$ are defined (and are to be measured) with respect to the structure *as given*, i.e. in the presence of the reflectors, dielectrics, absorbers and shielded receivers. We also remark that $\bar{E}_i$ is the undisturbed field, i.e. the field produced by the sources in the absence of the antenna.

3. Equivalent Circuit of a Two-port Antenna

The extension of these notions to a two-port antenna is straightforward (Fig. 3). In each waveguide arm we introduce quantities V, I and $\bar{\alpha}$. The following equations then hold on the antenna side of the cross-sections:

$$\bar{H} \times \bar{u}_{n1} = (I_{g1} - Y_{a1} V_1 - Y_m V_2)\bar{\alpha}_1 = I_1 \bar{\alpha}_1$$
$$\bar{H} \times \bar{u}_{n2} = (I_{g2} - Y_m V_1 - Y_{a2} V_2)\bar{\alpha}_2 = I_2 \bar{\alpha}_2 \quad \ldots(5)$$

We have assumed that the media are reciprocal, hence that a common value Y_m exists for Y_{12} and Y_{21}. There results the equivalent circuit shown in Fig. 3. The various terms have the following physical meaning:

(1) $I_{g1}\bar{\alpha}_1$ and $I_{g2}\bar{\alpha}_2$ are the current densities appearing on S_1 and S_2 when these surfaces are *simultaneously* short-circuited.

(2) Y_{a1} and Y_{a2} are the input admittances of respectively S_1 and S_2 when the other *terminal surface is short-circuited*. In Fig. 1(a), for example, Y_{a1} is the input admittance seen in T_1 when all slots are open, but T_2 is short-circuited.

(3) The short-circuit current produced on S_2 by a tangential electric field $V_1\bar{\alpha}_1$ on S is $-Y_m V_1 \bar{\alpha}_2$. In the case of Fig. 1(a), for example, the external coupling between the slots is known[3] to be negligible when the slot separation exceeds $\lambda/2$. For such case, the mutual admittance Y_m between T_1 and T_2 can be set equal to zero.

When the antenna system is immersed in a plane wave, the value of I_{g1} is

$$\frac{4\pi}{j\omega\mu_0}(\bar{F}_1 . \bar{E}_{i1})$$

where $V_1\bar{F}_1$ is the radiation vector produced by a tangential electric field $V_1\bar{\alpha}_1$ on S_1, S_2 being short-circuited. This formula is valid for a given phase centre O_1, and $\bar{E}_{i1}$ is the value of the incident electric field at that point. Similarly, the value of I_{g2} is

$$\frac{4\pi}{j\omega\mu_0}(\bar{F}_2 . \bar{E}_{i2})$$

where $\bar{F}_2$ and $\bar{E}_{i2}$ are defined with respect to a phase centre O_2. The choice of O_1 and O_2 is arbitrary, and is governed by the problem at hand. It is often convenient to locate O_1 and O_2 at the same point O.

It is to be noticed that the elements of the quadrupole are defined in the presence of a given 'receiver quadrupole'. In the case of Fig. 1(c), for example, the magic-T must be present and connected, for the currents induced in its external surface can have an influence on the radiation patterns and the input admittances of the antenna.
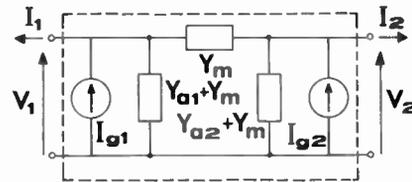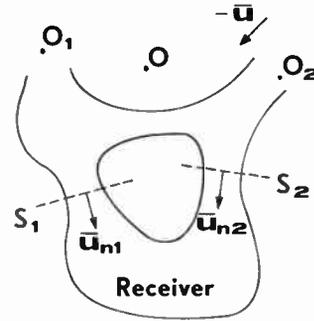
Fig. 3. A two-port antenna with equivalent circuit.

4. The Quadrupole Antenna Connected to a Magic-T[5]

Let us assume that the terminal surfaces S_1 and S_2 are connected to a magic-T, as shown in Fig 1(c). The terminal surfaces T_1 and T_2 are symmetrically located, and arms 3 and 4 are loaded with matched detectors. For such a case, no reflections are experienced in arms 1 and 2. With $\bar{u}_{n1}$ and $\bar{u}_{n2}$ pointing inside the T,

$$\bar{H}_1 \times \bar{u}_{n1} = Y_c \bar{E}_{t1}$$
$$\bar{H}_2 \times \bar{u}_{n2} = Y_c \bar{E}_{t2}$$

where $Y_c = 1/R_c$ is the characteristic admittance of the mode. The admittance matrix of the magic-T is seen to be

$$\begin{pmatrix} I_1 \\ I_2 \end{pmatrix} = \begin{pmatrix} Y_c & 0 \\ 0 & Y_c \end{pmatrix} . \begin{pmatrix} V_1 \\ V_2 \end{pmatrix} \quad \ldots\ldots(6)$$

In order to evaluate V_1 and V_2, we use the equivalent circuit of Fig. 3, where ports 1 and 2 are loaded with Y_c. The signals picked up in arms 3 and 4 to within a phase angle which depends on the choice of the terminal planes are:

$$V_3 = \frac{1}{\sqrt{2}}(V_1 + V_2)$$

$$V_4 = \frac{1}{\sqrt{2}}(V_1 - V_2)$$

The absorbed powers are

$$\tfrac{1}{4}Y_c|V_1 + V_2|^2 = \tfrac{1}{4}R_c|I_1 + I_2|^2$$

in arm 3, and

$$\tfrac{1}{4}Y_c|V_1 - V_2|^2 = \tfrac{1}{4}R_c|I_1 - I_2|^2$$

in arm 4. In terms of the admittances, and from eqn. (4), the signals take the form

$$V_3 = \frac{4\pi}{j\sqrt{2}\omega\mu_0[(Y_c+Y_{a1})(Y_c+Y_{a2})-Y_m^2]} \times$$
$$\times [(Y_c+Y_{a2}-Y_m)(\bar{F}_1.\bar{E}_{i1})+(Y_c+Y_{a1}-Y_m)(\bar{F}_2.\bar{E}_{i2})]$$
$$\qquad \ldots\ldots(7)$$

and

$$V_4 = \frac{4\pi}{j\sqrt{2}\omega\mu_0[(Y_c+Y_{a1})(Y_c+Y_{a2})-Y_m^2]} \times$$
$$\times [(Y_c+Y_{a2}+Y_m)(\bar{F}_1.\bar{E}_{i1})-(Y_c+Y_{a1}+Y_m)(\bar{F}_2.\bar{E}_{i2})]$$
$$\qquad \ldots\ldots(8)$$

For uncoupled, matched antennae

$$(Y_{a1} = Y_{a2} = Y_c, Y_m = 0)$$

the equations become

$$V_3 = \frac{\sqrt{2}\pi}{j\omega\mu_0 Y_c}[(\bar{F}_1.\bar{E}_{i1})+(\bar{F}_2.\bar{E}_{i2})] \quad \ldots\ldots(9)$$

$$V_4 = \frac{\sqrt{2}\pi}{j\omega\mu_0 Y_c}[(\bar{F}_1.\bar{E}_{i1})-(\bar{F}_2.\bar{E}_{i2})] \quad \ldots\ldots(10)$$

The outputs are seen to be proportional to the sum and difference of the outputs of the individual ports. These properties can be utilized to 'shape' the response of the magic-T antenna, for example to obtain zero in arm 4 for a given direction in space. Any deviation from this direction produces an 'error' signal, which can be used in a feedback loop to re-orient the antenna in the desired fashion. The two halves of the slot antenna shown in Fig. 1(a) can be used for that purpose. In general, the effects are particularly interesting when the antenna system is symmetric with respect to the plane of symmetry of the magic-T.

5. Symmetry Considerations

The value of a term such as $(\bar{F}_1.\bar{E}_{i1})$ is independent of the choice of the origin O. If we displace the origin from O_1 to O, we obtain new vectors

$$\bar{F}_1' = \bar{F}_1\, e^{-jk(\bar{u}.\overline{O_1O})}$$

and

$$\bar{E}_{i1}' = \bar{E}_{i1}\, e^{+jk(\bar{u}.\overline{O_1O})}$$

and the product $(\bar{F}_1.\bar{E}_{i1})$ is seen to be invariant. If O_1 and O_2 are chosen in O, eqns. (7) to (10) become simpler because $\bar{E}_{i1}' = \bar{E}_{i2}'$. Assuming, then, that $\bar{F}_1'$ and $\bar{F}_2'$ have been defined with respect to a common origin O, the radiation vector produced by tangential fields $V_1\bar{\alpha}_1$ in S_1 and $V_2\bar{\alpha}_2$ in S_2 is seen to be

$$\bar{F} = V_1\bar{F}_1' + V_2\bar{F}_2'$$

These considerations are particularly relevant when the antenna structure has a plane of symmetry π (Fig. 4). The eigenvectors $\bar{\alpha}_1$ and $\bar{\alpha}_2$ in the input guides are symmetric with respect to π, and $Y_{a1} = Y_{a2} = Y_a$. If we locate O in π, the radiation

vectors $\bar{F}_1'(\bar{u})$ and $\bar{F}_2'(\bar{u})$ relative to symmetric directions $\bar{u}$ and $\bar{u}'$ are themselves symmetric with respect to π. It is interesting to consider two modes of operation of the quadrupole antenna:

5.1. The Symmetric Mode

It corresponds to $V_1 = V_2 = 1$. The $\bar{E}$ and $\bar{H}$ fields are respectively symmetric and anti-symmetric with respect to π. The radiation pattern in a direction $\bar{u}$ is

$$\bar{F}_s = \bar{F}_1'(\bar{u}) + \bar{F}_2'(\bar{u}) = \bar{F}_1'(\bar{u}) + \bar{F}_1'(\bar{u}')$$

and is symmetric with respect to π. The input admittance is $Y_s = Y_a + Y_m$. Fields and admittances have the values which obtain when π is made into a magnetic wall ($\bar{H}$ perpendicular to π), and only one half of the antenna system is kept.

5.2. The Anti-symmetric Mode

Here $V_1 = 1$, $V_2 = -1$, $\bar{E}$ and $\bar{H}$ are respectively anti-symmetric and symmetric with respect to π, and the radiation pattern $\bar{F}_{as} = \bar{F}_1'(\bar{u}) - \bar{F}_1'(\bar{u}')$ is anti-symmetric with respect to π. The input admittance is $Y_{as} = Y_a - Y_m$. Fields and admittances are not disturbed if π is short-circuited. With these definitions and symbols, eqns. (7) and (8) become

$$V_3 = \frac{4\pi}{j\omega\mu_0\sqrt{2}(Y_s+Y_c)}(\bar{E}_i'.\bar{F}_s) \quad \ldots\ldots(11)$$

$$V_4 = \frac{4\pi}{j\omega\mu_0\sqrt{2}(Y_{as}+Y_c)}(\bar{E}_i'.\bar{F}_{as}) \quad \ldots\ldots(12)$$

The signal in branch 4 vanishes when the incident direction $-\bar{u}$ is in the π plane. A comparison with eqn. (3) shows that the antenna system has an admittance Y_s and a radiation vector $\frac{1}{\sqrt{2}}\bar{F}_s$, while the current generator, when seen from arm 3 is:

$$\left[\frac{4\pi}{j\omega\mu_0}\left(\bar{E}_i'.\frac{\bar{F}_s}{\sqrt{2}}\right)\right]$$

A similar remark can be made for arm 4 and the anti-symmetric quantities.[4]
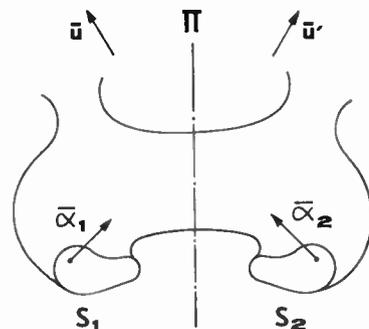
Fig. 4. Symmetrical two-port antenna.

Additional simplifications obtain when, as in Fig. 1(a), the two halves of the antenna system radiate symmetrically with respect to planes π_1 and π_2. This property is often encountered in interferometric systems. Under such circumstances, $\bar{F}_1(\bar{u}) = \bar{F}_1(\bar{u}')$ if the phase centre O_1 is located in π_1. It follows that

$$\bar{F}_s(\bar{u}) = \bar{F}_1(\bar{u})\, e^{-jk\bar{u}.\overline{O_1O}} + \bar{F}_1(\bar{u})\, e^{-jk\bar{u}.\overline{O_2O}}$$
$$= \bar{F}_1(\bar{u})\, 2\cos(k\sin\theta)$$

$$\bar{F}_{as}(\bar{u}) = \bar{F}_1(\bar{u})\, e^{jk\bar{u}.\overline{O_1O}} - \bar{F}_1(\bar{u})\, e^{-jk\bar{u}.\overline{O_2O}}$$
$$= \bar{F}_1(\bar{u})\, 2j\sin(k\sin\theta)$$

6. Conclusions

Antenna systems which are fed by two input ports are often encountered in practice, in particular when interferometric or mono-pulse applications are considered. The equivalent circuit shown in Fig. 3 is the answer to the study of such systems, and in particular to the logical determination of the parameters of the antenna by a minimum number of measurements. The method which has been presented here can easily be generalized to N-port antennas and anisotropic systems. In each case a precise definition of the meaning of parameters such as the source current, the mutual admittances, etc., is mandatory for their practical determination. One *must* short-circuit what should be short-circuited, and connect what should be connected. The text is emphatic in this respect.

7. References and Bibliography

1. J. Van Bladel, 'A generalized reciprocity theorem for radiating apertures', *Archiv der Elektrischen Übertragung*, **20**, pp. 447–450, August 1966.

2. For a generalization to multi-mode antennas, see J. Van Bladel, 'The matrix formulation of scattering problems', *Trans. Inst. Elect. Electronics Engrs on Microwave Theory and Techniques*, MTT–14, pp. 130–135, March 1966.

3. A. F. Kay and A. J. Simmons, 'Mutual coupling of shunt slots', *Trans. Inst. Radio Engrs on Antennas and Propagation*, AP–8, pp. 389–400, July 1960.

4. The present formulae generalize results obtained by C. B. Watts, 'Simultaneous radiation of odd and even patterns by a linear array', *Proc. I.R.E.*, **40**, pp. 1236–9, October 1952.

5. A description of a practical magic-T antenna has been published since the present paper was completed. See T. Takeshima, 'A slot array antenna for monopulse tracking radar', *Microwave J.*, **9**, pp. 63–5, December 1966.

STANDARD FREQUENCY TRANSMISSIONS

(Communication from the National Physical Laboratory)

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

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

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