FOUNDED 1925
INCORPORATED BY
ROYAL CHARTER 1961

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

THE RADIO AND ELECTRONIC ENGINEER

The Journal of the Institution of Electronic and Radio Engineers

VOLUME 36 No. 2 AUGUST 1968

Astronomy, Space and Radio

As a branch of science, radio astronomy is barely twenty years old: space research is even younger. Progress in electronic technology has enabled both to develop at an almost incredible rate and they are now major scientific disciplines involving expenditure of much money, industrial effort and intellectual activity.

In Great Britain the Science Research Council has set up the Astronomy, Space and Radio Board to co-ordinate the work in these related fields which is being carried out in the Universities and in Government research establishments, and a statement of the policy and programmes of the Board has recently been prepared to provide the S.R.C. with information on which to base its plans for 1969–74.† The greater part of the published document presents the scientific interests of the Board especially in their relation to the classical aim of the astronomer—whether he uses radio or optical waves—namely, the investigation of the whole of the Universe, and in particular its origins. This part can be recommended as a good and clear introduction to present cosmological theories.

The second, abridged part of the Report discusses satellite and rocket programmes as well as astronomical investigations into distant radio sources, cosmic rays, x-rays, and ultra-violet and infra-red radiation. Clearly, future levels of resources will limit the extent to which plans can be implemented and priority is proposed for British sounding rockets, for a successor to the *Ariel III* satellite, for large optical telescopes in England and Australia, and for the three-mile array of fully steerable dishes at Cambridge University's radio astronomy observatory. The Radio and Space Research Station is to continue exploitation of its steerable aerial at Chilbolton and to expand work in the millimetre wave-band—of considerable potential value for both space and terrestrial communications.

Although it is often believed that much of this research into the nature of space and the Universe is purely academic, the techniques and instrumentation are in the forefront of contemporary industrial interest. Outstanding examples can be found not only in space research, but also in radio and optical astronomy, and technological advances which have been made so far involve, for instance, low-noise parametric amplifiers, solid-state multipliers, techniques for steering paraboloids, and improved automatic control techniques as a result of attitude stabilization for the *Skylark* rocket.

Apart from straight technology, the Board believes that it can help to meet national needs by fostering a steady supply of postgraduate students trained in techniques which have a direct relevance to the requirements of modern industrial research and development. The radio astronomy schools have good records of providing government and industry with people trained in a variety of modern techniques. Space research is also very suited to the training of research students provided there is sufficient continuity and an adequate number of overlapping projects in a university department to allow Ph.D. work. These conditions are met in the large groups where young research workers are also brought into close contact with firms concerned with producing equipment to the exacting specifications for space flight. The work therefore produces individuals with the broad and flexible approach needed for employment in modern industry.

The 'fall-out', both in technology and in vocational training, from the work of the Astronomy, Space and Radio Board, is clearly a justification for continuation of these programmes.

F. W. S.

† 'The scientific interests of the Astronomy, Space and Radio Board and its present and planned programmes'. A report by the Board to the Science Research Council, July 1968.

INSTITUTION NOTICES

Standing Committee of Kindred Societies

Collaboration over a wide range of disciplines should result from the recent decision by four learned Societies to form a 'Standing Committee' to co-ordinate activities arising from their common interests. The participating societies are the Institution of Electronic and Radio Engineers, the Institution of Electrical Engineers, the Institute of Physics and the Physical Society, and the Institute of Mathematics and its Applications.

The Committee's first act has been to put on a regular footing the arrangements whereby members of each of the four societies may attend the meetings of the others. All technical meetings of each society (for which no fee is charged to members) will thus in future be freely open to all members of the other three societies.

At the same time the Committee has reached an agreement in connection with conferences. When one of the four societies proposes to hold a technical Conference which embraces the objects or normal activities of any of the other three societies, it will invite the relevant society or societies to be joint sponsors of the conference. The societies accepting such an invitation will join in organizing the programme, normally by appointing one or more members to the programme committee and their members will be admitted to the conference on the same terms as those of the originating society. (Conferences on technical or non-technical matters which are solely appropriate to members of one Society will not be covered by this arrangement.)

Co-operation of this kind has been taking place from time to time and with growing frequency in the past, but the new agreement will embrace all technical meetings and conferences arranged by these societies in future. The Committee will continue to consider ways in which the four societies may work together to provide improved services to their members and to their professions.

The Committee's Chairman is Mr. J. A. Ratcliffe, C.B., C.B.E., C.Eng., F.R.S., who is Immediate Past President of the Institution of Electrical Engineers, a Past President of the Physical Society and a Fellow of the I.E.R.E. Representation of the Institutions on the Standing Committee is by the President (or nominee) and the Executive Secretary of each. The I.E.R.E. is represented by the President, Major-General Sir Leonard Atkinson, K.B.E., and by the Director, Mr. Graham D. Clifford, C.M.G.

Conference on Lasers and Opto-electronics

The I.E.R.E. is organizing a three-day Conference on Lasers and Opto-electronics to be held at the University of Southampton on 25th to 27th March 1969. Further details of its scope will be given in the next issue of the *Journal*, but in the meanwhile offers of papers within the broad field of the Conference are invited by the Organizing Committee. The Organizing Committee is under the chairmanship of Professor W. A. Gambling, D.Sc., C.Eng., F.I.E.R.E., and includes representatives of the Institute of Physics and The Physical Society, the Institution of Electrical Engineers, the Institute of Electrical and Electronics Engineers and the University of Southampton.

Electronic Weighing Conference

Because the range of modern electronic weighing techniques is expanding rapidly, particularly load cell weighing, the I.E.R.E. has organized a two-day Conference to be held on Wednesday, 30th October–Thursday, 31st October 1968, at the Middlesex Hospital Medical School, Cleveland Street, London, W.1. Some sixteen papers on many aspects of this important subject will be presented and discussed.

Coverage will include the nature and performance of load cells, the achievement of Board of Trade weighing accuracies, weighing of vehicles in motion, check weighing of packages, belt weighing and crane weighing. There will be papers covering digital, solid-state, and telemetering techniques; also the problems and experiences in special environments such as the steel industry.

It is intended that the Conference will serve to disseminate present practice and provide a forum for discussion between designers and users. There is an increasing awareness that, due to modern digital techniques, weighing devices can tie in with automatic process control and computers.

Further information and registration forms can be obtained on application to the Conference Registrar, I.E.R.E., 8–9 Bedford Square, London, W.C.1. (Telephone: 01-580 8443, extension 3).

Northern Ireland Section

The Council of the Institution approved the establishment of a Northern Ireland Section, based initially on Belfast. A Committee of eight members under the Chairmanship of Mr. J. D. McEwan, B.Sc., C.Eng., M.I.E.R.E., has been formed, and the Honorary Secretary, from whom further information may be obtained, is Mr. J. T. Attridge, C.Eng., M.I.E.R.E., at 33 Kilmakee Park, Belfast BT5 7QY.

Quebec Section

At the Annual General Meeting of the Quebec Section, held in Montreal on 29th April, 1968, Mr. W. J. Willetts, C.Eng., M.I.E.R.E., was re-elected Chairman and Mr. R. C. Caston (Associate) was elected Honorary Secretary.

Offers of papers suitable for presentation at future Section meetings will be welcomed and should be sent to the Honorary Secretary, whose address is: Suite 600, 90 Beaubien West, Montreal, P.Q.

Correction

The following two corrections should be made in the 'Discussion on Gain and Stability of Transistor Amplifiers' published in the February 1968 issue of *The Radio and Electronic Engineer:*

Page 113, Col. 2, line 3, the expression for g_{max} should read:

$$g_{max} = \frac{1}{s} \cdot \left| \frac{p_{21}}{p_{12}} \right|$$

Page 114, Col. 1, the value of y_{21} *should read* 70 mmho (*not* 70 μmho).

(*Other announcements will be found on page 82.*

U.D.C. 621.376.23

Experimental Tests of the Theory of an Amplitude-limited Correlation Detector

By

M. J. YERBURY,
B.Sc.(Eng.), Ph.D. (*Graduate*)†

Summary: A sensitive correlation detector has been constructed in which the input signal and noise voltages are amplitude-limited. The correlator is described and experimental results are given in support of the theory. A method of analysing the data from the correlator to obtain the maximum output signal/noise ratio is discussed for the case where the signal and noise voltages depend on slowly varying functions of time. Particular reference is made to the analysis of data from observations with the stellar intensity interferometer for which the correlator was specifically designed. It is concluded that the dual-channel amplitude-limited correlator is stable and that it can be used to obtain accurate, meaningful and linear results despite the non-linear operations involved.

List of Symbols

ΔC_r	correlation expected in the rth 100-second cycle from the multiplier-correlator
$\Delta C_r'$	correlation measured in the rth 100-second cycle with the amplitude-limited correlator
i_i	total phototube anode current in channel i
i_i^*	phototube anode current due to starlight alone in channel i
$I_i\sqrt{\frac{2}{\pi}}$	low-level transfer function of the limiter in channel i
T	total integration time
$\frac{1}{\sqrt{\alpha}}$	distortion factor
$\psi_{12}(\tau)$	cross-correlation function of the undistorted outputs of channels 1 and 2
$\psi_{L_1,L_2}(\tau)$	cross-correlation function of the outputs of the limiters in channels 1 and 2
$\psi_{ii}(0)$	mean-square value of the background noise in the undistorted output of channel i
$\rho_{12}(\tau) = \frac{\psi_{12}(\tau)}{\sqrt{\mu_{11}\mu_{22}}}$	
σ_i	r.m.s. value of the limiter in channel i
$\sigma_T = [\sigma_T'^2 + n\sigma_d^2]^{\frac{1}{2}}$	effective noise output of the amplitude-limited correlator at time T
τ	time delay between the two input channels
$\mu_{ii} = \psi_{ii}(0) + \psi_{12}(0)$	
σ_d	drift uncertainty

† Formerly at the Cornell–Sydney University Astronomy Centre, School of Physics, University of Sydney; now at the Center for Radiophysics and Space Research, Cornell University, Ithaca, N.Y.

1. Introduction

In an earlier paper[1] the theory of a sensitive correlation detector in which the input signal and noise voltages were amplitude-limited was given. One advantage claimed for this technique was that instability of the correlator zero datum appearing in the form of a steady, unidirectional drift in the output voltage, should be considerably reduced. The work described was part of an investigation into the possibility of improving the existing correlation measuring equipment in the optical stellar interferometer[2] at Narrabri Observatory, New South Wales. Since this equipment is in almost constant operation, a separate correlator had to be constructed and this has been used to test the theory of dual-channel amplitude limiting when applied to a sensitive correlation detector. The final part of the investigation is presented in this paper.

2. The Stellar Interferometer

This unique instrument, a brief description of which follows, has been applied during the last few years to the measurement of the angular diameters of 15 stars.[3]

Two composite mirrors, roughly 6·5 metres in diameter and mounted on trucks separated by the interferometer baseline, are guided round a circular track 188 metres in diameter. Light from a star is collected on the cathode of a photomultiplier tube placed at the focus of each of the two reflectors and the radio frequency noise at the output of each tube is conveyed by coaxial cable, supported by a steel catenary, to a control building situated near the centre of the track. This building houses the multiplier-correlator which accepts the noise outputs of the two phototubes (in a bandwidth of 10–110 MHz) and measures the correlation between them.

The interferometer determines the correlation between the fluctuations in the *intensity* of the light originating at a star and falling on the photocathodes. Each photocathode surface may be considered to behave as a square-law detector[4] which produces beats between the Fourier components of the incident radiation field. Because of this square-law relationship, the intensity interferometer is relatively insensitive compared to the Michelson type of interferometer: the effective signal/noise ratio at the output of each phototube is typically less than −50 dB and at the output of the multiplying element of the correlator it is less than −100 dB. Unusually stringent limits are therefore set to the tolerable instability of the correlator.

A detailed description of the original equipment has been given elsewhere.[5] The new correlator differs from this in several respects, in particular it uses the technique of dual-channel amplitude limiting and, to achieve high reliability, has been constructed entirely from solid-state components.

3. Brief Description of the New Correlator

3.1. *The System*

A block diagram of the correlator is given in Fig. 1 where the two channels have been labelled A and B respectively.

The voltage in channel B is phase-inverted 5000 times per second by an electronic phase switch within the correlator so that when the noise from the phototubes is partially correlated, the multiplier produces a 5 kHz square-wave at its output in addition to the random noise. This 5 kHz signal is a measure of the correlation which in this way is separated from the d.c. terms in the multiplier and precludes the need for high-gain d.c. amplifiers with their attendant problems of instability.

A similar phase switch, in this case situated close to the phototube of the channel A reflector, inverts the phase of the noise once in 20 seconds. This renders any radio frequency coupling between the coaxial cables ineffectual; it ensures that external radio signals and noise intercepted by the cables do not appear as correlation, and eliminates the effect on the measured correlation of drift in the integrator output voltage.

Differences between the gains of the two sides of each phase switch result in amplitude modulation of the noise voltages at the switching frequencies. This modulation is detected by the non-linearities in the multiplier and is a major contributory factor to correlator instability.[1] It is possible, with these phase switches, to reduce the depth of modulation to about 1% but this figure depends on the width and shape of the spectral density of the corresponding input noise voltage which in turn, is a critical function of the phototube operating conditions. The amplitude-limiting technique was therefore chosen as perhaps the best means of reducing the effects of amplitude modulation on the correlator stability.

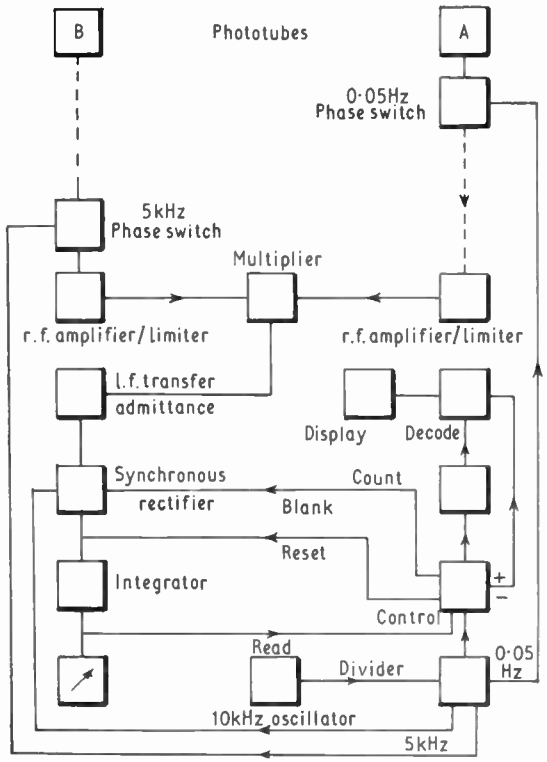


Fig. 1. Block diagram of the dual-channel amplitude-limited correlator.

Each channel contains a high-gain, wideband r.f. amplifier which precedes the first of two amplitude-limiting stages. The amplitude-limited voltages from the two channels are combined in a multiplier and its low-frequency output is converted into a current which flows through the 5 kHz synchronous rectifier and into the integrator. The signal input to the integrator is therefore a direct current which reverses its sign every 10 seconds; in addition a large noise current is present. At the end of this interval, the integrator is reset to zero by a standard input current of suitable sign. During the reset, a gate is opened admitting a train of pulses to one of two counters. These respectively accumulate positive and negative correlation as selected by combining the sign of the integrator output voltage and the state of the slower phase switch. Every 100 seconds the readings in the counters are transferred to the first two columns of a printer which also takes their difference and prints

this in a third column in the form of a running total. This represents the total integrated correlation since the beginning of the observation. In the remaining two columns, the results of integrating the phototube anode currents during the previous 100-second cycle are printed. These are required in the reduction of the data as illustrated in Section 5.

3.2. *The R.F. Amplifiers and Limiters*

The total attenuation of each phase switch and coaxial cable is about 8 dB at 20 MHz, increasing to 12 dB at 100 MHz. If it is assumed that the spectral density of the noise from each phototube has a rectangular shape then the frequency responses of the two input channels should also be rectangular for optimum system performance. Compensation for the cable losses is achieved by including an equalizing circuit within each r.f. amplifier. The resultant low-level frequency-response characteristics of the amplifiers, up to and including the first limiter stages, are shown in Fig. 2.

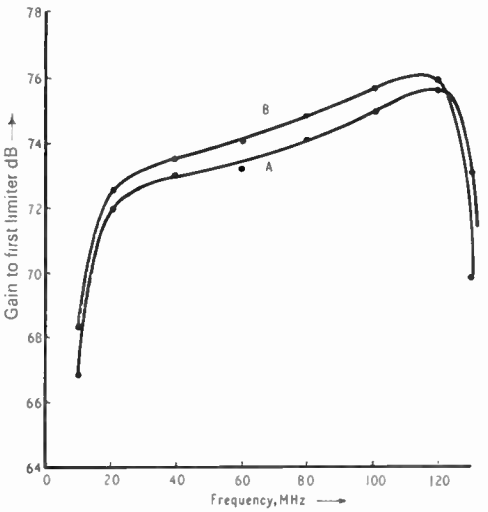
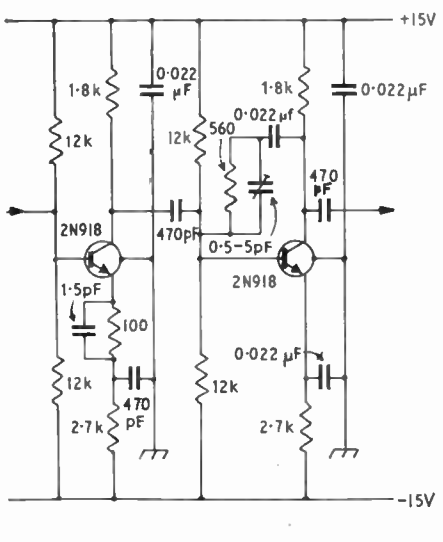


Fig. 2. Low-level frequency response characteristics of the r.f. amplifiers up to and including the first limiting stages.

The mismatch design procedure[6] was adopted for the r.f. amplifiers because of the advantages of low interaction between stages and the fact that two substantially identical amplifiers were required.

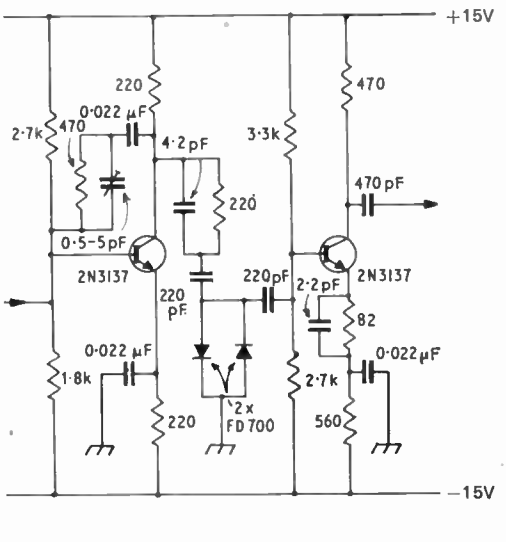
Each amplifier comprises six series-shunt feedback pairs in cascade. The fifth and sixth feedback pairs are used to drive the second and first limiters respectively. Typical circuit diagrams of a feedback pair and an amplitude-limiting stage are given in Figs. 3(a) and (b), the former having a gain of about 14 dB and a bandwidth of 260 MHz. The limiter diodes are matched pairs capable of switching times of about 750 ps. Their characteristics when connected as a limiter and driven from a 220 Ω resistive source, are shown in Fig. 4 for comparison with the theoretical model[7] used in the analysis.

From this figure the r.m.s. value of the limiter, defined as the input voltage at which the slope of the limiter characteristic has its maximum rate of change, was chosen to be 0·7 V.



(a)

(a) Typical feedback pair used in each r.f. amplifier.



(b)

(b) An amplitude-limiting stage.

Fig. 3.

3.3. *The Multiplier*

A description of the operating principle and construction of the multiplier has already been published.[8] A more advanced form has since been developed for use in the correlators at Narrabri Observatory. The response of this multiplier to r.f. inputs falls by 5% of its midband value at 470 MHz and the noise due to 'self detection' at each input can be reduced by careful adjustment, to less than 1% of the product term.

3.4. *The L.F. Transfer Admittance and 5 kHz Synchronous Rectifier*

The output of the multiplier is restricted to a bandwidth of 200 Hz–30 kHz which causes a loss of about 3% or 0·13 dB in the correlator output signal/noise ratio. However, this attenuates most of the undesirable higher harmonics of the 5 kHz switching wave for which the r.f. amplifiers have significant gain.

Synchronous rectification is achieved by using two insulated-gate field effect transistors in a shunt-series chopper configuration. These are connected between the input of the integrator and the output of the transfer admittance which can deliver up to 0·5 mA from a source impedance of 2 MΩ through a series coupling capacitor. The 5 kHz reference wave switches the field effect transistors alternately so that during one half-cycle current flows directly into the integrator and during the next, it is bypassed to earth. All excess charge stored in the series capacitor is eventually released to the integrator. Any relative delay between the signal into the synchronous rectifier and the reference wave causes a nearly proportionate loss of output signal/noise ratio. This is prevented by means of an adjustable delay circuit for the 5 kHz reference wave.

4. Preliminary Tests

Five tests were made with the correlator installed at the Observatory, four of which, described in this section, were direct experimental tests of the theory.

Initially, however, it was necessary to devise some means of introducing a known amount of correlated noise into each channel together with the uncorrelated noise background. The equivalent effect was produced with a cross-coupling network in which a coaxial T-junction was placed in each coaxial cable close to the respective phototube. The free ports of the junctions were connected to 470 Ω 'buffers' which were then joined by coaxial attenuators. Using this network it was possible to obtain a check on the performance of the equipment under typical operating conditions and this method was therefore adopted as a system calibration procedure.

Measurements were made of the power outputs of the phototubes as functions of their anode currents and voltages. The r.m.s. noise into a 75 Ω load was then determined and using the measured gains of the two channels, the degree of amplitude distortion was calculated.

It is useful to introduce here the term 'distortion factor' which will be defined as

$$\frac{\sigma}{\sigma_i} = \frac{1}{\sqrt{\alpha}}$$

$$= \frac{\text{r.m.s. noise at the limiter stage in the absence of limiting}}{\text{r.m.s. value of the limiter}} \qquad \text{.....(1)}$$

For the phototubes currently in use at the Observatory, the r.m.s. noise output is about 2 mV in a bandwidth of 70 MHz when the anode current is 100 μA and the anode voltage is 1·76 kV. The overall midband gain of each channel is approximately 64 dB so that under these conditions and

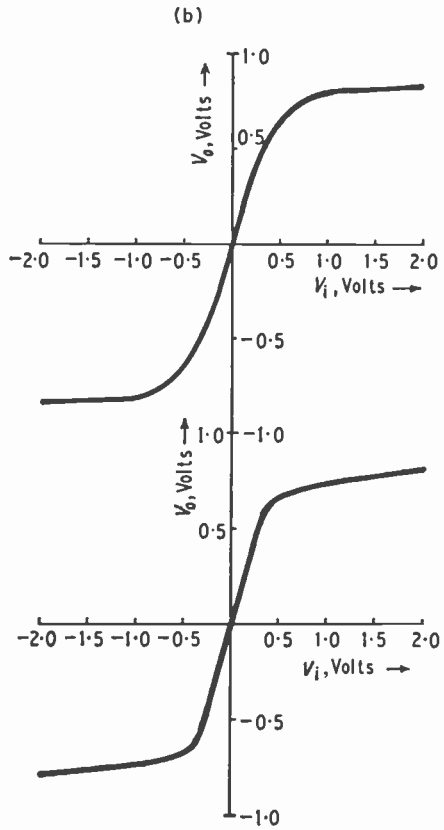


Fig. 4. The theoretical limiter characteristic (upper curve) given by

$$I = I_t \sigma_t \operatorname{erf} \frac{V_i}{\sigma_t \sqrt{2}} \qquad \text{(after Baum[7])}$$

The lower curve shows the resulting characteristic of the limiter diodes when driven from a 220 Ω source resistance.

Note added in proof: The values of V_i on the lower curve should be doubled, i.e. to read −4·0, −3·0, −2·0, −1·0, 1·0, 2·0, 3·0, 4·0.

assuming $\sigma_i = 0{\cdot}7$ V, the *nominal* distortion factor would be 4·5.

The exact values of the distortion factors are uncertain for several reasons. Firstly, the r.m.s. value of the limiters σ_i, is estimated by inspection of the limiter characteristic which does not conform exactly to the theoretical shape (Fig. 4). In addition, it is necessary to know what the r.m.s. noise level would be at the limiter stage in the absence of amplitude limiting and this requires knowledge of the shape and magnitude of both the spectral density of the phototube noise and the overall frequency response of the r.f. circuits up to the first limiter in each channel. This response is not easy to measure accurately because the amplifier noise alone is sufficient to partially-saturate the limiter and introduces an uncertainty into the low-level gain measurements.

4.1. *Signal Output from the Correlator*

In the earlier paper[1] it was shown that when the input signal/noise ratio is much less than unity the correlation function at the output of the dual channel amplitude-limited correlator is given approximately by

$$\psi_{L_1,L_2}(\tau) \simeq \psi_{12}(\tau) \cdot \frac{2}{\pi} I_1 I_2 \left[\frac{\alpha'\beta'}{(1+\alpha')(1+\beta')}\right]^{\frac{1}{2}} \quad \ldots(2)$$

where

$$\alpha' = \frac{\sigma_1^2}{\psi_{11}(0)+\psi_{12}(0)} \quad \text{and} \quad \beta' = \frac{\sigma_2^2}{\psi_{22}(0)+\psi_{12}(0)}$$

Equation (2) can be written as

$$\psi_{L_1,L_2}(\tau) = \frac{K'}{\sqrt{(1+\alpha')(1+\beta')}} \quad \ldots\ldots(3)$$

where

$$K' = \frac{2}{\pi}\sigma_1\sigma_2 I_1 I_2 \frac{\psi_{12}(\tau)}{\sqrt{\psi_{11}(0)\psi_{22}(0)}} \quad \ldots\ldots(4)$$

and is proportional to the effective input signal/noise ratio when it is small and when sources of noise other than the phototube shot noise are ignored.

The purpose of the test now described was to examine the validity of eqn. (3) for the special case of equal input levels, that is when $\alpha' \simeq \beta' = \alpha$, for which eqn. (3) becomes

$$\psi_{L_1,L_2}(\tau) = \frac{K'}{(1+\alpha)} \quad \ldots\ldots(5)$$

The photocathodes were illuminated by separate, adjustable sources of uncorrelated light so as to produce equal anode currents at a fixed anode voltage of 1·76 kV. For each setting of the light intensity and hence the noise level, the correlator output and the mean phototube anode currents were recorded. The cross-coupling network ensured that the effective input signal/noise ratio was maintained constant throughout the test.

Neither the phototubes nor the frequency responses of the two channels are identical but their differences are small enough in this case to permit the geometrical mean of the distortion factors to be used as the nominal value of $1/\sqrt{\alpha}$. This was calculated by using the relation $1/\sqrt{\alpha} = 0{\cdot}43\sqrt{i}$, where i is the geometrical mean of the phototube anode currents in microamperes and the factor 0·43 is derived from measurements of the phototube power outputs, the gains of the two channels and on the assumption that $\sigma_i = 0{\cdot}7$ V.

The theoretical curve of the normalized output signal as a function of $1/\sqrt{\alpha}$ has an asymptote at

$$\frac{\psi_{L_1,L_2}(\tau)}{K'} = 1$$

The corresponding asymptotic value of the experimental curve is not known and in view of the direct dependence of the distortion factors on a number of inexact quantities a least-squares fit was applied to the experimental points. This involved normalizing the results to an asymptotic value of 40·23 V in 100 seconds and increasing the nominal distortion factors by 20%. These experimental points are plotted in Fig. 5 for comparison with the theoretical curve

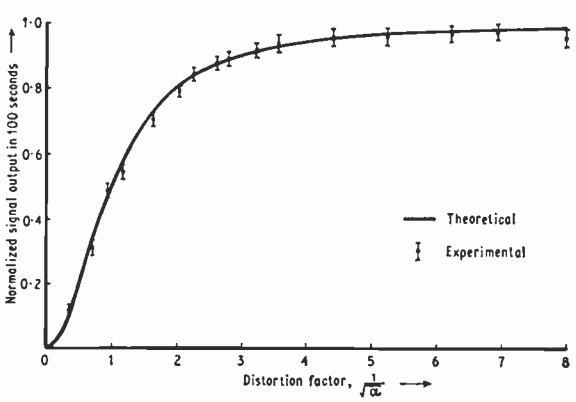


Fig. 5. The signal output of the dual-channel amplitude-limited correlator as a function of the equal r.m.s. noise input levels, when the input signal/noise ratio is constant.

where there is generally very good agreement. The departure from the curve of the point corresponding to the highest input level, is caused by overdriven stages in the r.f. amplifiers.

It can be shown that the presence of excess noise, due for example to the r.f. amplifiers, increases the effective r.m.s. value of the limiter in the denominator of eqn. (5). This is another way of expressing the fact that the gains of the r.f. amplifiers for very small

input levels are lower than predicted because the internally generated noise is sufficient to partially-saturate the limiters. It follows that the effect of this noise and of the loading of the inputs by the cross-coupling network have been allowed for in Fig. 5.

4.2. *Noise Output from the Correlator*

The theory shows that when the input levels are equal the noise output of the dual-channel amplitude-limited correlator can be written in the form

$$\sigma'_T = K'' \left[\sum_1^{\infty} \frac{1}{n^2} \left(\frac{1}{1+\alpha} \right)^{2n} \right]^{\frac{1}{2}} \qquad \ldots\ldots(6)$$

where

$$K'' = \frac{\sigma_1 \sigma_2 I_1 I_2}{\pi} \sqrt{\frac{2T}{B}} \qquad \ldots\ldots(7)$$

and B is the mutual bandwidth of the input noise voltages which are taken to have rectangular spectral densities.

In this test the correlator was run for a preselected time with fixed, equal input levels of uncorrelated noise. The r.m.s. noise per cycle of operation ($\sim$ 100 seconds) and the drift per cycle were then estimated from the numerical data and the experiment was repeated until a wide range of input levels had been covered. The normalized output noise σ'_T/K'' was then obtained as a function of $1/\sqrt{\alpha}$ and plotted for comparison with the theoretical curve.

The time required to obtain an estimate of the noise and the drift increases as the inverse square of the desired accuracy. The number of cycles associated with an uncertainty of $\pm p\%$ in an estimate of the r.m.s. noise for one cycle is

$$n = \frac{5000}{p^2} \qquad \ldots\ldots(8)$$

so that for an uncertainty of $\pm 1\%$ the correlator would have to be run for 143 hours. Apart from the obvious impracticability, such a run would be unlikely to yield the desired accuracy due to extraneous noise and the probable existence of diurnal, periodic drifts. Therefore runs lasting approximately 13 hours were used, except in the measurement of the noise due to the r.f. amplifiers which lasted for 3 hours. The uncertainties were then reduced to about $\pm 3{\cdot}5\%$, except in the latter case where the uncertainty was $\pm 7\%$.

In Fig. 6 the experimental points have been plotted as a function of $1/\sqrt{\alpha}$ for comparison with the theoretical curve which has an asymptote at $\sigma'_T/K'' = 1{\cdot}282$. Since the asymptotic value of the experimental set of results is not known, they were normalized to 1·18 V in 100 seconds. This figure was obtained by applying a least-squares fit using the vertical co-ordinates of all the points excluding the point due to the r.f. amplifier noise. This has been plotted against a distortion factor of zero in order to show the trend of the curve caused by the presence of excess noise not included in the calculation of the distortion factors. All the nominal distortion factors have been increased by 32% in Fig. 6 to account for the 20% increase derived from Fig. 5 and an extra 12% due to the fact that the cross-coupling network was removed for this test and hence did not load the input channels.

The agreement between the theoretical shape of the curve and the experimental points in Fig. 6 is quite good when it is considered that the true uncertainties in the latter are probably higher than those

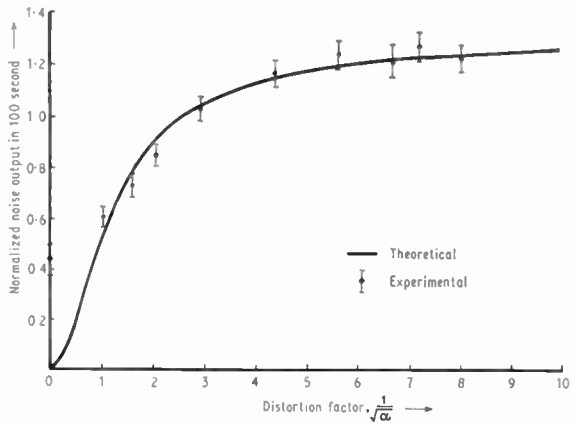


Fig. 6. The noise output of the dual-channel amplitude-limited correlator as a function of the equal r.m.s. noise input levels.

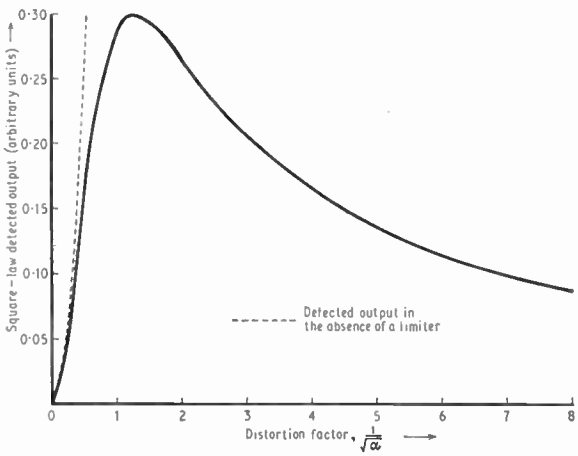


Fig. 7. The average output of a limiter after square-law detection and synchronous rectification. The input is Gaussian noise with a small depth of square-wave amplitude modulation.

shown due to extraneous noise and slight changes in the equipment performance during the period of the test.

Table 1 shows estimates of the drift in the correlator output voltage and their dependence on the input noise level. The uncertainties in the results are large because the time needed to obtain a given accuracy is more than twice that required for the same accuracy in a noise estimate.

Table 1

Drift at the output of the dual-channel amplitude-limited correlator as a function of the equal r.m.s. noise input levels.

Distortion factor	Drift in 100 s	Distortion factor	Drift in 100 s
R.f. amplifier noise	$-0{\cdot}07 \pm 0{\cdot}05$ V	4·40	$+0{\cdot}01 \pm 0{\cdot}07$ V
1·06	$-0{\cdot}02 \pm 0{\cdot}04$	5·67	$-0{\cdot}01 \pm 0{\cdot}07$
1·63	$-0{\cdot}11 \pm 0{\cdot}05$	6·72	$+0{\cdot}09 \pm 0{\cdot}08$
2·11	$-0{\cdot}11 \pm 0{\cdot}05$	7·21	$-0{\cdot}01 \pm 0{\cdot}07$
2·96	$-0{\cdot}17 \pm 0{\cdot}06$	8·05	$+0{\cdot}03 \pm 0{\cdot}07$

The results suggest that the relation between the drift and the input level is possibly similar to that between the square-law detected output of a limiter and the input level, the graph of which appears in Fig. 7. A drift characteristic proportional to the 2nd or even 3rd power of the square-law detected output might be expected, depending upon the exact mechanism of drift generation.

4.3. *Linearity of the Dual-channel Amplitude-limited Correlator*

It is shown in the Appendix that for a correlator where some form of amplitude limiting has been applied to both input voltages, the output signal is practically proportional to that from a multiplier-correlator if the input signal/noise ratio is much less than unity.

In order to test this point experimentally the cross-coupling network was used to inject a small amount of correlated noise into each channel. The degree of coupling was varied by inserting different coaxial attenuators into the network while the phototube anode currents and voltages were kept constant at 120 μA and 1·76 kV respectively. For each setting of the input signal/noise ratio, the correlator was run for a time sufficient to estimate the output signal with reasonable accuracy. The connection of several attenuators in series produced an extra relative delay between the correlated components which was partially compensated by the use of non-attenuating cable lengths. The length of the attenuating section was nevertheless measured and a small final correction applied where necessary by using a calibration/cross-coupling delay curve. In consequence, the uncertainties in the delay-corrected results were increased.

The effective input signal/noise ratio was taken to be the negative of half the total attenuation in the cross-coupling network (in dB) and the measured output signal/noise ratio was expressed as $10 \log_{10} (S/N)_{out}$ dB. The results are plotted in Fig. 8; the slope of the graph is unity if the correlator is linear. The figure shows that over a wide range of input signal/noise ratios and within the experimental errors, the dual-channel amplitude-limited correlator is linear.

4.4. *Comparison with the Multiplier-correlator*

Since the absolute scales of the experimental and theoretical results have not been compared, an accurate determination of the signal/noise ratio obtained under a given set of conditions was made with the amplitude-limited correlator and the multiplier-correlator. The latter instrument, situated at the Observatory and in its latest state of development,

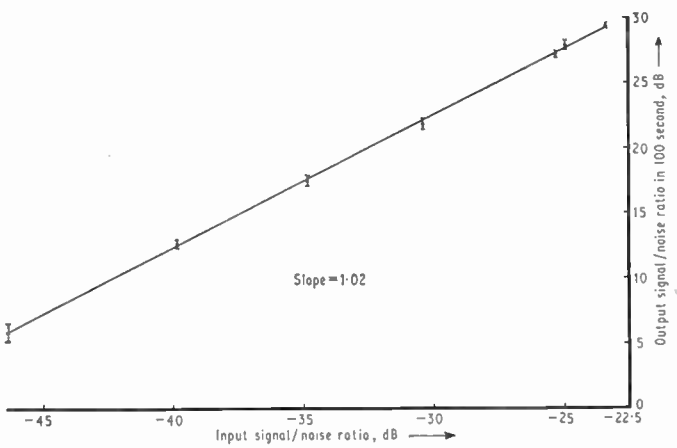


Fig. 8. The linear relationship between the input signal/noise ratio of the dual-channel amplitude-limited correlator and its output signal/noise ratio.

was taken to be close to the ideal multiplier-correlator in terms of the directly obtainable output signal/noise ratio, that is excluding such effects as long-term drift and unreliability.

The cross-coupling network was used to produce an effective input signal/noise ratio of about −40 dB with phototube anode currents and voltages of 120 μA and 1·76 kV respectively. Under these conditions, each correlator was run for 30 cycles prior to a noise-measurement run lasting 450 cycles and again for 30 cycles afterwards.

From these 60 cycles in which the signal/noise ratio was relatively high, an accurate estimate of the mean output signal was obtained.

The noise output of each correlator was estimated from the 450-cycle run. This was made with the cross-coupling network still loading the two input channels but with the coupling section broken and terminated to exclude interference. Since the phototube anode currents varied slightly during the signal and noise runs, the mean anode current product was determined in both cases and the results were normalized to allow for the small differences between the conditions of the tests on the two correlators. Referring to the multiplier-correlator as correlator-1, the output signal/noise ratio achieved in 100 seconds was

$$\left(\frac{S}{N}\right)_1 = 28{\cdot}2 \pm 1{\cdot}4 \qquad \text{.....(9)}$$

and for correlator-2 the output signal/noise ratio in 100 seconds was

$$\left(\frac{S}{N}\right)_2 = 20{\cdot}9 \pm 1{\cdot}0 \qquad \text{.....(10)}$$

The loss of output signal/noise ratio compared with the multiplier-correlator is given by

$$F = -10\log_{10}\frac{\left(\frac{S}{N}\right)_2}{\left(\frac{S}{N}\right)_1} \text{ dB} \qquad \text{.....(11)}$$

and was found to be

$$F = 1{\cdot}30 \pm 0{\cdot}29 \text{ dB} \qquad \text{.....(12)}$$

The expected value for F under the conditions of this test where the distortion factors were roughly 5·7, is 0·90 dB. There is therefore an excess loss of 0·40 ± 0·29 dB.

Two points are worth noting. Firstly, the true uncertainty in this figure is probably higher due to extraneous noise in both correlators. Secondly, part of the excess loss is explainable since, as pointed out in Section 3.4, there is a known loss of about 0·13 dB in the output signal/noise ratio due to restriction of the l.f. bandwidth in correlator-2. The corresponding figure for correlator-1 is probably less than 0·04 dB. Other possible causes of this extra loss are discussed in the Conclusions section.

5. Comparative Observations of a Star

To complete the series of tests on the dual-channel amplitude-limited correlator, some observations of a star were made between 10th and 19th October 1967.

The intention was to 'interleave' the observations by using each correlator on alternate nights; in this way it was hoped to be able to compare their performances under nearly identical conditions of atmospheric extinction and background light. However, due to poor weather it was not possible to keep to this plan and in fact conditions changed considerably over the observing period. As a result, the method of analysing the data was quite rigorously tested.

5.1. *The Method of Analysing the Data*

The correlation measured by the multiplier-correlator is proportional to the square of the intensity of the light from a star and is therefore proportional to $i_1^* i_2^*$. However, the phototube shot-noise is proportional to the square-root of the intensity of the total amount of light falling on the photocathode and hence the noise output from the multiplier-correlator is proportional to $\sqrt{(i_1 i_2)}$. The anode currents vary throughout the observations for a variety of reasons including tracking errors at the reflectors, cloud, atmospheric extinction, atmospheric scintillations and changing background light. To record these variations the currents are integrated during each 100-second cycle and the results are printed out at the end of the cycle. At suitable times during the night, the reflectors are pointed to an area of the sky close to the star and the integrated anode currents due to the background light are recorded. It is then possible to distinguish between i_i^* and i_i.

Finally, when the data from several nights of observations are to be combined, the dependence of the measured correlation on the anode currents is removed by normalization and a weighted mean correlation is determined by accounting for the uncertainties in the individual sets of results.

The signal output from the amplitude-limited correlator in one cycle of an n-cycle observation is, from eqn. (2)

$$\Delta C_r' = \Delta C_r \frac{kF_c}{\sqrt{i_1 i_2}} \qquad \text{.....(13)}$$

where k is a constant and F_c is the equipment calibration figure determined immediately before and after the observation by measuring the mean signal output of the correlator with the cross-coupling network.

It is required to estimate ΔC_r from the set of measurements $\Delta C_r' \pm \sigma_T'/\sqrt{n}$, where σ_T' is the observed noise per cycle. Firstly, a new set of readings

$$\Delta C_r \pm \frac{\sigma_r}{\sqrt{n}} = \frac{\Delta C_r' \sqrt{i_1 i_2}}{kF_c} \pm \frac{\sigma_T' \sqrt{i_1 i_2}}{kF_c \sqrt{n}} \quad \ldots\ldots(14)$$

are formed and are then normalized to their anode current dependence to give on the right-hand side of the above equation:

$$\Delta C_r' \frac{\sqrt{i_1 i_2}}{i_1^* i_2^*} \frac{1}{kF_c} \pm \frac{\sigma_T'}{\sqrt{n}} \frac{\sqrt{i_1 i_2}}{i_1^* i_2^*} \frac{1}{kF_c}$$

The separate readings may then be summed after first weighting each by the inverse square of its uncertainty. Hence, the normalized weighted mean correlation becomes

$$C_1 = \frac{\sum_{r=1}^{n} (\Delta C_r' - d) \frac{F_c}{\sigma_T^2} \frac{i_1^* i_2^*}{\sqrt{i_1 i_2}}}{\sum_{r=1}^{n} \frac{kF_c^2}{\sigma_T^2} \left[\frac{i_1^* i_2^*}{\sqrt{i_1 i_2}}\right]^2} \quad \ldots\ldots(15)$$

where d is the mean drift per cycle which, with its uncertainty σ_d, represents the correlator zero datum. To measure this drift, separate sources of uncorrelated light are used to produce roughly the same input levels as those from the star. The correlator is then run both before and after the observation and an estimate of the drift is made. It is implied in eqn. (15) that the drift is independent of anode current variations. This is not true in the case of the dual-channel amplitude-limited correlator but the change in d during an observation should be small and hence unimportant, if $d \ll \Delta C_r'$. However, the uncertainty σ_d in the zero datum is not usually negligible; for an n-cycle observation with 'dummy runs' lasting a total of N cycles the correlator output noise is, in effect, increased by the factor $\sqrt{[1+(n/N)^2]}$ due to this uncertainty.

The observed noise is σ_T' per cycle and is almost independent of anode current variations. Because of this, the uncertainty in C_1 may be written

$$\sigma = \left[\frac{\sigma_T'^2}{n} + \sigma_d^2\right]^{\frac{1}{2}} \left[\frac{\sqrt{i_1 i_2}}{i_1^* i_2^*}\right]_{av} \frac{1}{kF_c} \quad \ldots\ldots(16)$$

Putting

$$A = \frac{i_1^* i_2^*}{\sqrt{i_1 i_2}} \quad \ldots\ldots(17)$$

the signal/noise ratio obtained in one cycle can, from eqns. (15) and (16), be written as:

$$\left(\frac{S}{N}\right)_2 = \frac{\sum_{r=1}^{n} (\Delta C_r' - d).A}{[\sigma_T'^2 + n\sigma_d^2]^{\frac{1}{2}} \sum_{r=1}^{n} A^2} \left[\left(\frac{1}{A}\right)_{av}\right]^{-1} \ldots(18)$$

5.2. *Results*

Observations were made on two nights with the amplitude-limited correlator and on three nights with the multiplier-correlator. The total observing times were about 14½ hours and 13 hours respectively.

Since the ultimate object was the comparison of the signal/noise ratios that would be achieved by the two correlators under identical conditions, the individual noise estimates were all scaled to a standard mean phototube anode current product. On this occasion the drift at the output of each correlator was taken to be the mean value of all the estimates made during the observing period. In this way it was possible to ignore the noise contributions due to the drift uncertainties and hence obtain a true comparison of the two systems.

The individual sets of data each corresponding to a night of observation with one correlator were weighted and combined into final estimates of the normalized mean correlation and the scaled output noise. These results appear in Table 2 from which the loss of output signal/noise ratio, given by eqn. (11) is

$$F = 1{\cdot}72 \pm 0{\cdot}46 \quad \text{dB} \quad \ldots\ldots(19)$$

Table 2

Results per 100 seconds obtained from a series of stellar observations using the amplitude-limited correlator (2) and the multiplier-correlator (1)

Correlator	Drift	Mean normalized correlation	Normalized output noise	Output signal/noise ratio
1	−0·03 V	5·70 ± 0·27	3·00 ± 0·16	1·90 ± 0·13
2	0	4·29 ± 0·22	3·37 ± 0·17	1·28 ± 0·09

The ratio of the anode currents during the stellar observations was $i_1/i_2 \simeq 1{\cdot}5$, the highest mean value of i_2 during an observation being 95 μA. As a result, the values of F obtained in this Section and in Section 4.4 cannot be combined into a single figure although it is considered probable that the loss incurred by the dual-channel amplitude-limited correlator is roughly the same over a small range of anode currents for which $i_1 i_2$ is constant.

The uncertainties quoted in the last column of Table 2 are not derived directly from those in the third and fourth columns because the latter are partially interdependent.

6. Conclusions

It has been shown that a practicable method for the measurement of the correlation between small signals in noise is to distort severely the signal and

noise voltages in both channels of the correlator by limiting their amplitude fluctuations. From these experiments it appears that the theory adequately describes the behaviour of the dual-channel amplitude-limited correlator although the relative scales of the results have not been established with much certainty. This is mainly because the process of achieving accurate estimates is very time-consuming particularly when it is necessary to measure the difference between two nearly equal quantities.

The values of the drift in the correlator output voltage given in Table 1 as a function of the input levels are very uncertain. However, the results of some more accurate work suggest that the amplitude-limited correlator described here is practically drift-free provided the distortion factors exceed roughly 3.

The measured losses of the output signal/noise ratio are greater than might be expected from the theory and could be due to a loss of signal, an increased noise output or a combination of both effects. Signal information can be lost in several ways but mainly occurs as a result of imperfections in the equipment. For example, each of the following conditions will produce a loss of output signal/noise ratio of about 10% or 0·45 dB:

a relative delay of 5 μs between the signal at the 5 kHz synchronous rectifier and the reference wave;

a relative delay of 1 ns between the signals in the two r.f. channels;

a mark/space ratio of 0·82 instead of unity in the 5 kHz phase-switching wave.

A further mechanism by which signal may be lost in the amplitude-limited correlator concerns the fact that unidirectional pulses, whose amplitude considerably exceeds the r.m.s. noise level of the Gaussian noise, are known to exist at the outputs of the phototubes. The combined noise applied to each limiter is therefore asymmetrical and the Gaussian noise which contains the required information is not symmetrically limited, the effect being enhanced if the limiting diodes are not identical. This results in a loss of output signal (as a theoretical analysis using a model for a limiter with bias will show) and, it is inferred, a loss of output signal/noise ratio.

An additional effect of these pulses is to reintroduce the phase-switching wave at the output of the limiter. In the case of the 5 kHz square-wave there is sufficient gain in the rest of the system to result in a considerable offset current flowing into the integrator. This current is directly dependent on variations in the pulse content of the noise which, in turn, is a function of the phototube anode current, gain and temperature. These quantities may be expected to vary during an observation and hence indirectly produce low-frequency noise at the correlator output.

With the multiplier-correlator, extra noise is introduced if the spectral densities of the multiplier input voltages are not rectangular in shape and signal is lost if the two r.f. bands do not coincide. In the case of the dual-channel amplitude-limited correlator, it is difficult to say how the noise is affected by changes from the ideal rectangular shapes. However, the signal output, given by eqn. (3), is affected through α' and β' only and hence is practically independent of such variations provided α' and β' are much less than unity and the responses remain parallel. The latter restriction implies no relative phase shifts.

When the voltage consisting of the sum of a signal and noise is applied to a limiter, beats occur between the signal and noise spectra and hence more noise is introduced. The correlator output signal/noise ratio is therefore less than that of the corresponding linear multiplier-correlator. Also, as the degree of distortion is increased, the spectral density spreads so that signal and noise terms appear at higher frequencies. It is therefore necessary, in order to avoid further loss of output signal/noise ratio, to ensure that the bandwidths of the circuits between the limiters and the multiplier inputs are as wide as practicable.

The main object of this work has been to produce a correlator which may be applied to the measurement of the correlation between small signals in noise with a reasonable chance of obtaining accurate results. To achieve this object it was clearly desirable to avoid impracticable idealizations in the mathematical theory so that the behaviour of the correlator under varying input conditions can be best understood. With a variety of input levels this correlator has successfully detected very small correlated components in noise and, at the same time, has exhibited a highly stable zero datum. The experimental results described here confirm, with tolerable accuracy, some of the theoretical work and show that this unusual technique can be used to obtain meaningful results despite the non-linear operations involved.

7. Acknowledgments

The author wishes to thank Professor R. Hanbury Brown for originally suggesting this line of research and for his permission to use the facilities of the stellar interferometer. He would also like to thank Dr. L. R. Allen for his comments on the draft of the paper and Messrs. S. M. Owens and J. M. Rome for their assistance during the stellar observations.

This work was supported by grants from the Office of Scientific Research of the U.S. Air Force under contracts (AF-AFOSR–63/66), the Australian Research Grants Committee, the Research Grants Committee of the University of Sydney and the Science Foundation for Physics within the University of Sydney.

8. References

1. Yerbury, M. J., 'Amplitude limiting applied to a sensitive correlation detector', *The Radio and Electronic Engineer*, **34**, No. 1, pp. 5–14, July 1967.
2. Hanbury Brown, R., 'The stellar interferometer at Narrabri Observatory', *Sky and Telescope*, **28**, No. 2, pp. 64–69, August 1964.
3. Hanbury Brown, R., Davis, J., Allen, L. R., and Rome, J. M., 'The stellar interferometer at Narrabri Observatory. Part II: The angular diameters of 15 stars', *Mon. Not. R. Astr. Soc.*, **137**, No. 4, pp. 375–417, 1967.
4. Hanbury Brown, R., and Twiss, R. Q., 'Interferometry of the intensity fluctuations in light. 1. Basic theory: the correlation between photons in coherent beams of radiation', *Proc. Roy. Soc.*, **A242**, pp. 300–324, 1957.
5. Hanbury Brown, R., and Browne, A., 'The stellar interferometer at Narrabri, Australia', *Philips Tech. Rev.*, **27**, No. 6, pp. 141–159, 1966.
6. Cherry, E. M., and Hooper, D. E. ,'The design of wide-band transistor feedback amplifiers', *Proc. Instn Elect. Engrs*, **110**, No. 2, pp. 375–389, February 1963.
7. Baum, R. F., 'The correlation function of smoothly limited Gaussian noise', *Trans. Inst. Radio Engrs on Information Theory*, **IT-3**, No. 3, pp. 193–197, September 1957.
8. Frater, R. H., 'Accurate wideband multiplier-square-law detector', *Rev. Sci. Instrum.*, **35**, No. 7, pp. 810–813, July 1964.
9. Price, R., 'A useful theorem for non-linear devices having Gaussian inputs', *Trans. I.R.E.*, **IT-4**, No. 2, pp. 69–72, June 1958.

9. Appendix: Linearity of the Correlator Output Signal

The signal output of the dual-channel amplitude-limited correlator has been determined[1] using the smooth limiting function.[7] When the input signal/noise ratio is small it was shown that the input cross-correlation function is practically proportional to the output cross-correlation function. To show that this is true when any type of amplitude-limiting function is employed it is first necessary to define the term 'limiter'. A broad definition of a limiter is that it is a device for which the slope of the transfer function tends to zero at the limiting levels $\pm D_i$. However, the effect of the limiter on Gaussian noise would be negligible if D_i exceeded about three times the r.m.s. levels and the term 'limiter' in a practical sense implies that the noise amplitude is limited to less than about three times its r.m.s. value.

Consider an amplitude-limiting function $f_i(V_i)$ where V_i is the sum of the Gaussian signal and noise voltages in channel i of the correlator. Let the function be specified by

$$f_i'(V_i) = 0 \quad \text{when} \quad |V_i| \geq D_i \qquad \text{.....(20)}$$

where the prime denotes differentiation with respect to V_i. Price's theorem[9] is then used to give

$$\frac{\partial \psi_{L_1,L_2}(\tau)}{\partial \psi_{12}(\tau)} \simeq \int_{-D_1}^{D_1} dV_1 \int_{-D_2}^{D_2} dV_2 f_1'(V_1) f_2'(V_2) P(V_1, V_2, \tau) \qquad \text{.....(21)}$$

where

$$P(V_1, V_2, \tau) = \frac{[1-\rho_{12}^2(\tau)]}{2\pi\sqrt{\mu_{11}\mu_{22}}} \exp\left[\frac{-\frac{V_1^2}{\mu_{11}} - \frac{V_2^2}{\mu_{22}} + \frac{2\rho_{12}(\tau)V_1V_2}{\sqrt{\mu_{11}\mu_{22}}}}{2[1-\rho_{12}^2(\tau)]}\right] \qquad \text{.....(22)}$$

is the joint probability density of V_1 and V_2.

Since it is assumed that the input signal/noise ratio is very small, then $\rho_{12}^2(\tau) \ll 1$ and eqn. (21) becomes

$$\frac{\partial \psi_{L_1,L_2}(\tau)}{\partial \psi_{12}(\tau)} \simeq \frac{1}{2\pi\sqrt{\mu_{11}\mu_{22}}} \int_{-D_1}^{D_1} dV_1 \int_{-D_2}^{D_2} dV_2 \times \times f_1'(V_1) f_2'(V_2) \exp\left[-\frac{V_1^2}{2\mu_{11}} - \frac{V_2^2}{2\mu_{22}}\right] \times \times \left[1 + \rho_{12}(\tau)\frac{V_1V_2}{\sqrt{\mu_{11}\mu_{22}}} + \ldots\right] \qquad (23)$$

If the limiters are symmetrical about the line of zero input voltage, only the even powers of $(V_1 V_2)$ in the expansion on the right of eqn. (23) contribute to the result. Since there is negligible contribution to the result of the integrations for $V_1 \geq D_1$ and $V_2 \geq D_2$, all terms except the first in the summation may be ignored if

$$\rho_{12}^2(\tau) \cdot \frac{D_1 D_2}{2\mu_{11}\mu_{22}} \ll 1 \qquad \text{.....(24)}$$

When eqn. (24) and the requirement $\rho_{12}^2(\tau) \ll 1$ are satisfied, eqn. (23) is reduced to the form

$$\psi_{L_1,L_2}(\tau) \simeq \psi_{12}(\tau)\overline{f_1'(X)} \cdot \overline{f_2'(Y)} \qquad \text{.....(25)}$$

This result is interesting since it shows that no matter what form the amplitude-limiting functions take, provided they are symmetrical and the above elementary conditions are satisfied, the cross-correlation function of their outputs is proportional to the cross-correlation function of their inputs.

Manuscript first received by the Institution on 29th January 1968 and in final form on 4th June 1968.
(Paper No. 1201/IC2.)

Institution Premiums and Awards

The Council of the Institution announces that the following awards are to be made for outstanding papers published in *The Radio and Electronic Engineer* during 1967.

CLERK MAXWELL PREMIUM

'New Thin-film Resistive Memory' by J. G. Simmons and R. R. Verderber (August).

P. PERRING THOMS PREMIUM

'The Development of a Pay-Television System and Equipment' by G. L. Hamburger (April).

HEINRICH HERTZ PREMIUM

'Mechanisms in Gunn Effect Microwave Oscillators' by J. E. Carroll (July).

A. F. BULGIN PREMIUM

'Automatic Recognition of Low-quality Printed Characters using Analogue Techniques' by J. R. Parks (August).

ARTHUR GAY PREMIUM

'The Preparation and Application of Tantalum Thin Film Passive Components' by R. Naylor and R. Fairbank (April).

LESLIE McMICHAEL PREMIUM

'Design of Simple Frequency-division-multiplex Communication Systems without Band-pass Filters, with particular reference to the Use of Constant-resistance Modulators' by Professor D. G. Tucker (November).

LORD BRABAZON AWARD

'Radar Signal Processing for Resolution beyond the Rayleigh Limit' by A. A. Ksienski and R. B. McGhee (September).

CHARLES BABBAGE AWARD

'The Design of a Magnetic Thin-film Store for Commercial Production' by R. S. Webley and A. T. Gibson (March).

LORD RUTHERFORD AWARD

'Field-effect Studies on Indium Antimonide Films' by C. Juhasz and Professor J. C. Anderson (April).

MARCONI AWARD

'Design and Control of a High Precision Electron Beam Machine' by J. G. Vaucher, S. F. Gourley, C. J. Hardy and G. R. Hoffman (February).

The award of the Bose Premium and the Mountbatten Premium is under consideration by the Council of the Indian Division who will make recommendations to the Institution's Council later this year.

The following Premiums and Awards are withheld as papers of high enough standard have not been published within their respective terms of reference: Rediffusion Television Premium, J. Langham Thompson Premium, Vladimir K. Zworykin Premium, Dr. Norman Partridge Memorial Premium and Hugh Brennan Premium.

Joint Conferences

International Broadcasting Convention

The Postmaster-General, the Rt. Hon. Roy Mason, will open the International Broadcasting Convention, which is to be held at Grosvenor House, London, from 9th to 13th September 1968. Over 100 papers are to be presented, of which a quarter will come from countries outside Great Britain. Survey papers will be given on the following subjects:

Transmitters—Dr. R. Busi (Radiotelevisione Italiana).

Recording—J. A. Flaherty (CBS Television Network).

Studio-complex problems—Dr. H. Schonfelder (Fernseh GmbH).

Television Systems—Professor R. Theile (Institut für Rundfunktechnik GmbH., Munich).

Receivers—P. L. Mothersole (Mullard Ltd.).

Propagation—K. Bernath (Enterprise des Postes, Telephones et Telegraphes Suisse).

The Convention is sponsored by the Electronic Engineering Association, the Institution of Electrical Engineers, the I.E.R.E., the Institute of Electrical and Electronics Engineers, the Royal Television Society and the Society of Motion Picture and Television Engineers. Further information and registration forms may be obtained from the Secretariat, International Broadcasting Convention, Savoy Place, Victoria Embankment, London, W.C.2, or from the I.E.R.E., 8–9 Bedford Square, London, W.C.1.

Conferences on Microwaves

The Institution of Electrical Engineers, together with the I.E.R.E. and the I.E.E.E., is organizing a European Microwave Conference which will be held in London from 8th to 12th September 1969. The Organizing Committee would welcome particularly papers on the following subjects:

Advances in Microwave Circuits: microwave networks and integrated circuits; filters and directional couplers; ferrite devices; delay lines and microwave acoustics; microwave and optical waveguide and waveguide components (excluding specific application to long distance telecommunications).

Microwave Antennas: inertia-less scanning adaptive aerials; active aerials.

Solid State Microwave Devices: low noise receiving and amplifying devices; power sources; control and switching devices; bulk effect devices.

Synopses of papers (about 500 words) should be submitted by 5th January 1969 to the Joint Conference Secretariat, I.E.E., Savoy Place, London, W.C.2.

An associated conference on an important aspect of microwaves will be held in the following week, namely 15th to 17th September, on 'Trunk Telecommunications by Guided Waves'. Further information on this Conference may also be obtained from the I.E.E.

U.D.C.621.37/9.002

High-density Soldered Interconnections

By

A. G. COZENS,
C.Eng., M.I.Mech.E., M.I.Prod.E.†

AND

J. E. TOMLIN,
B.Sc., A.C.T.(Birm.), A.I.M.†

Reprinted from the Proceedings of the Joint I.E.R.E.–I.E.E.–I.S.H.M. Conference on 'Thick Film Technology' held at Imperial College, London, on 8th to 9th April 1968.

Summary: Various interconnection techniques for use in conjunction with thin film planar devices are described. The special character of the substrate imposed a number of limitations on processing and assembly, the most restrictive being that all interconnections should be made at the lowest possible temperature, at the same time being capable of repair. The method of mounting integrated circuit chips directly on the 2 μm thick copper circuitry uses soldering techniques for both chip mounting and the connections from chip to copper circuitry. The connection system used to make ground and signal connections between adjacent planes and the special handling and soldering equipment produced for this work, and the processes used to manufacture connection straps and overlays are described. The metallurgical aspects of soldered interconnections are discussed, particularly with regard to the effects of short range diffusion which greatly influences both the metallurgical structure and the resultant mechanical integrity of joints in the size range covered.

1. Introduction

With the present trend towards the use of glass, metal and ceramic substrates for electronic sub-assemblies, the need for suitable interconnection schemes of appropriate density and reliability has increased. The work to be described was directed towards investigating the possibility of mounting integrated circuit chips directly on substrates of suitable size, and interconnecting these substrate assemblies. One important objective was to find the highest pitch density that could be achieved using conventional techniques.

A number of problems associated with the substrate material and the small physical size of the parts were met and overcome. Although this work was carried out on evaporated films, it is considered to be equally relevant to thicker films.

2. Substrate Design

The substrate itself consists of a copper base 2 mm thick, with an insulation layer of silicon monoxide 12 μm thick followed by a copper film 2 μm thick. Chromium is used at interfaces to provide adequate adhesion. The copper film is plated with 200 nm (2000 Å) of gold to the circuit pattern, and the exposed copper etched away. Figure 1 shows the etched substrate and a complete cross-section.

The assembly of the substrate required the mounting and connection of unencapsulated silicon integrated circuit chips, and the connection of a ground plane and conductor lines to adjoining substrates. This is illustrated in Fig. 2. The chosen connection techniques had to fulfil certain fundamental requirements:

(a) Substrate temperature to be held below 200°C at all times.

(b) All connections to be repairable, and all chips replaceable.

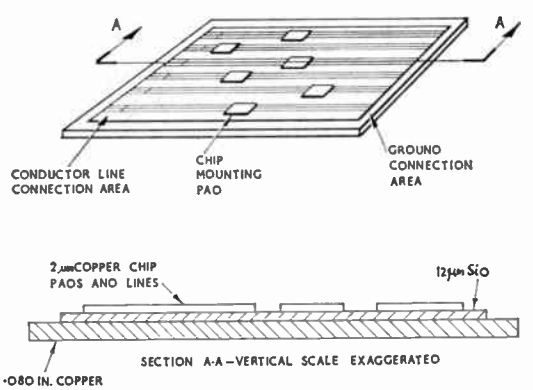


Fig. 1. Typical substrate used for interconnection experiments.

(c) The chips to be mounted so that all heat generated is rapidly dissipated to the substrate.

(d) Minimum impedance discontinuity between ground and conductors on substrate-to-substrate connections.

† IBM United Kingdom Laboratories Ltd., Hursley Park, Winchester.

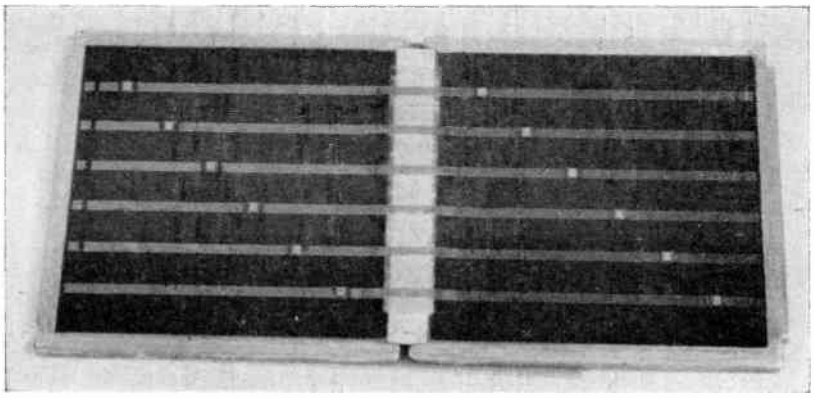


Fig. 2. Interconnected substrates—showing mounted chips, ground connection and conductor connections.

3. Mounting and Connection of Integrated Circuit Chips

The chips to be mounted were 1·5 mm (0·060 in) square with fifteen lead/tin solder pads 0·13 mm (0·005 in) diameter connected to the aluminium expanded contacts beneath through a protective film of 1 μm of glass (see Fig. 3). Thermal considerations previously mentioned dictated that the chips be mounted back down for best heat dissipation. Accordingly the back of the chip was metallized with successive layers of chromium, copper and gold so that a solderable surface was obtained. Thus the chip could be soldered back down giving good cooling and yet still be replaced should the need arise.

The method of soldering the chip to the substrate presented certain difficulties due to the heat-sink effect of the substrate, and additionally it was not possible to solder-plate either the substrate or the back of the chip. It was found that placing a fluxed solder preform between the chip and substrate and heating the chip by means of a 400°C nitrogen jet, resulted in a sound joint with a shear strength averaging 900 grammes. A substrate preheat of 150°C was necessary. A heated probe contacting the chip face was also found to yield satisfactory bonds, but the method was abandoned due to the risk of damage to the chip.

The connection between chip and substrate land

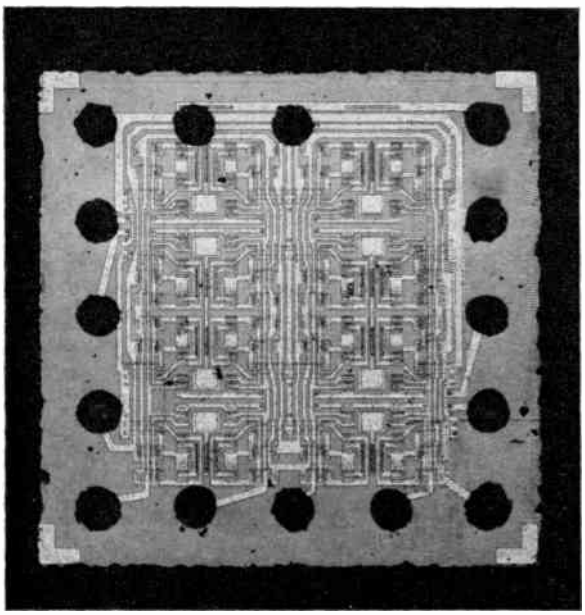


Fig. 3. Integrated circuit chip used for chip mounting experiments.

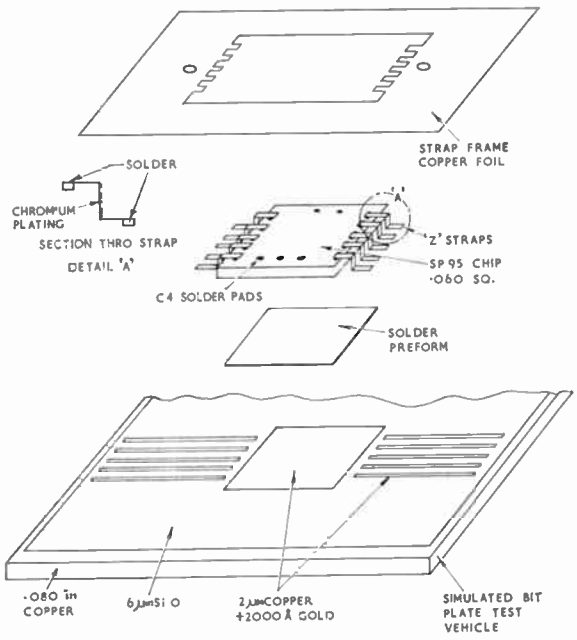


Fig. 4. Strap frame chip mounting concept.

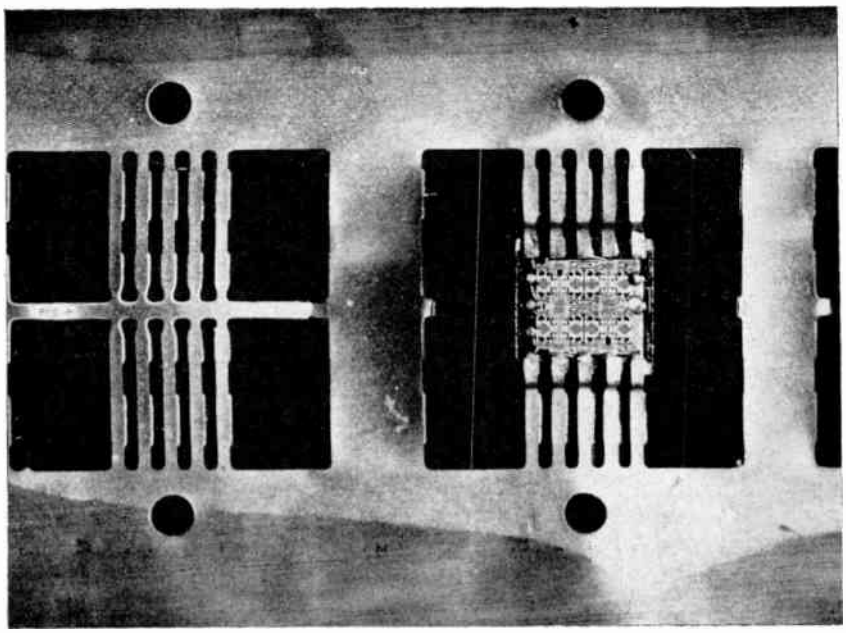


Fig. 5. Chip soldered into strap frame (× 12).

also had to be repairable, and therefore solder joints seemed to provide the most suitable system, again using substrate preheat of 150°C. The use of individual cranked solder plated wires was proposed, but although reasonably good solder joints were produced handling difficulties were too great. This resulted in the development of the lead frame or strap frame concept shown in Fig. 4. The lead frames are etched from 0·013 mm (0·0005 in) thick copper and plated with 60/40 tin lead solder and chromium (as a solder resist) in appropriate areas. These frames are punched and formed so that the necessary crank is produced in the straps to carry the connection from the pad level to the strap level. To assemble the chips on the strap frame, the chips are placed so that the solder pads on the chip contact the solder lands on the straps. The complete assembly is then passed through an oven, where the chip solder pads reflow on to the straps. During this operation the solder pads support and align the chip due to surface tension. The assembly now appears as shown in Fig. 5, and the chips complete with their straps must be removed from the frame prior to mounting. It will be seen that only 10 of the 15 pads on the chip receive straps. Due to the size and obvious fragility of the assembly two methods were closely studied—firstly a small and somewhat intricate press tool, and secondly electron-beam cutting on a Zeiss 3 kW electron-beam welder. In fact, both methods proved practical, but the electron-beam method shows better production potential (see Fig. 6).

When the chip/strap assembly is removed from the strap frame, it is picked up using a vacuum probe, lightly fluxed, and positioned over the appropriate lands on the substrate. The land which is to receive the back of the chip has the solder preform already placed upon it. The first operation is the soldering of the back of the chip to its land, and the assembly is heated by hot gas jet as described above, until the solder is seen to reflow. The solder joint between strap and chip also reflows during this operation, but

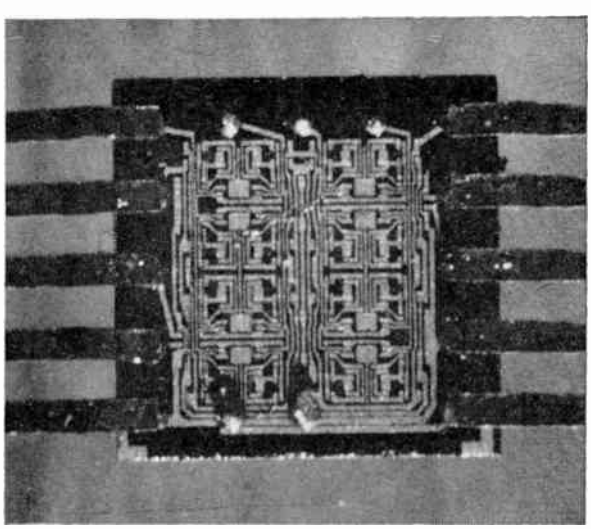


Fig. 6. Chip/strap assembly (× 35).

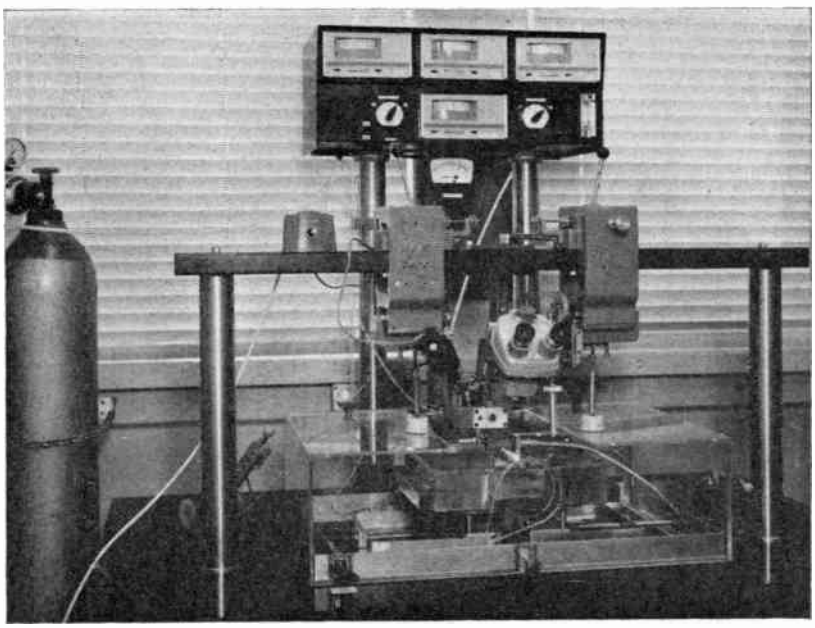


Fig. 7. Chip mounting machine.

again the effect of surface tension results in the solder holding its shape and position and yet at the same time allows strap movement to relieve stresses. The machine used for these mounting operations is illustrated in Fig. 7.

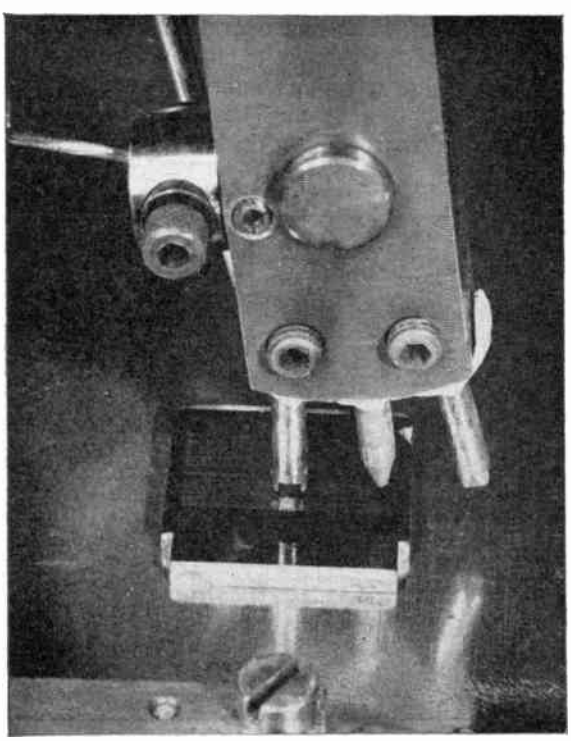


Fig. 8. Soldering chip/strap assembly to substrate.

Once the chip is soldered down, an arch-shaped probe is brought down to straddle the chip and hold the straps on two sides of the chip against the substrate lands (Fig. 8). The probe is heated during this operation to 350°C, but when the solder at the strap to substrate joints is seen to reflow, a jet of nitrogen cools the probe and the joints solidify, whereupon the probe is withdrawn.

4. Metallurgical Considerations—Chip Mounting

In joints of conventional size the quantity of intermetallic compounds produced during soldering has relatively little effect on joint integrity. However, high-density interconnection technology demands a much reduced joint size and as a consequence diffusion products begin to assume an importance approaching that given to them in silicon chip fabrication.

The 15 connector pads on the chip appear at the surface as a ball of solder 95/5 Pb/Sn. These are positioned on straps with solder pads of 60/40 Sn/Pb and the junctions are reflowed in a furnace designed to achieve a temperature of 340°C at the interface (Fig. 4). The 95/5 to 40/60 Pb/Sn solder joint is shown in Fig. 9. Attempts were made originally to use a simple wire or strap as a connection. Gold wires 0·025 mm (0·001 in) diameter could be continuously fed into and dissolved by molten 95/5 Pb/Sn solder, or, if held against the molten solder, the solder flowed along the gold surface. Gold-plated copper strap was also tried and successful joints were formed (Fig. 10), but these were eventually superseded by the solder pad strap, on which chromium was plated

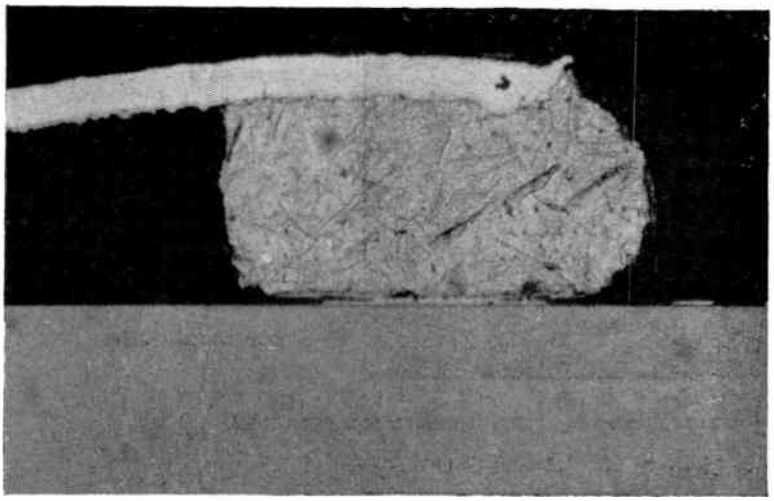


Fig. 9. Strap to chip pad solder joint (× 420).

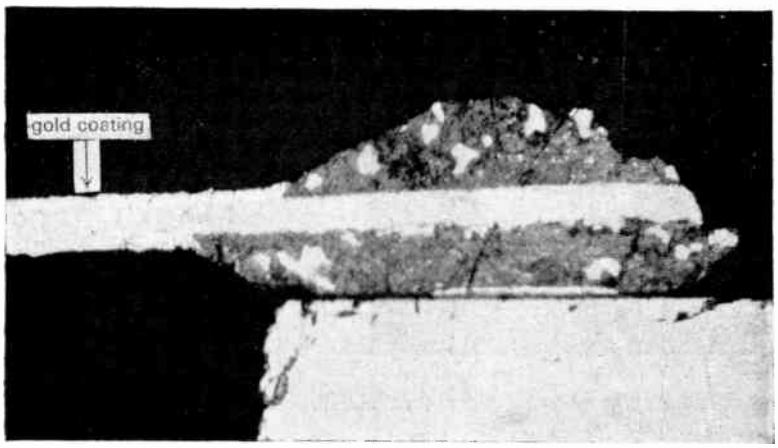


Fig. 10. Gold-plated strap to chip pad solder joint (× 420).

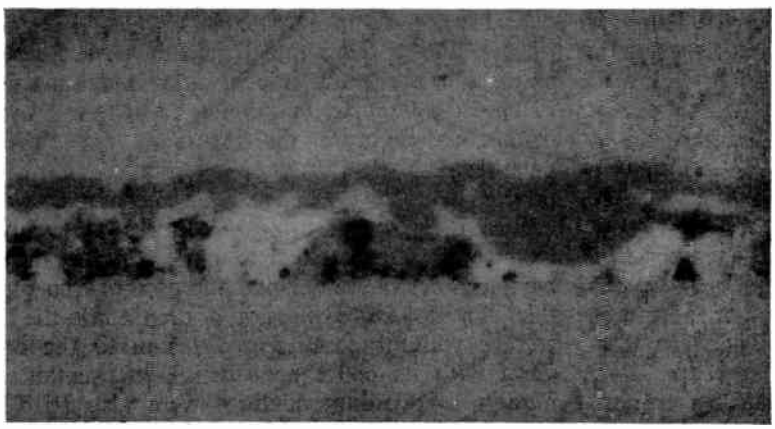


Fig. 11. Chip strap to substrate conductor joint (× 2400).

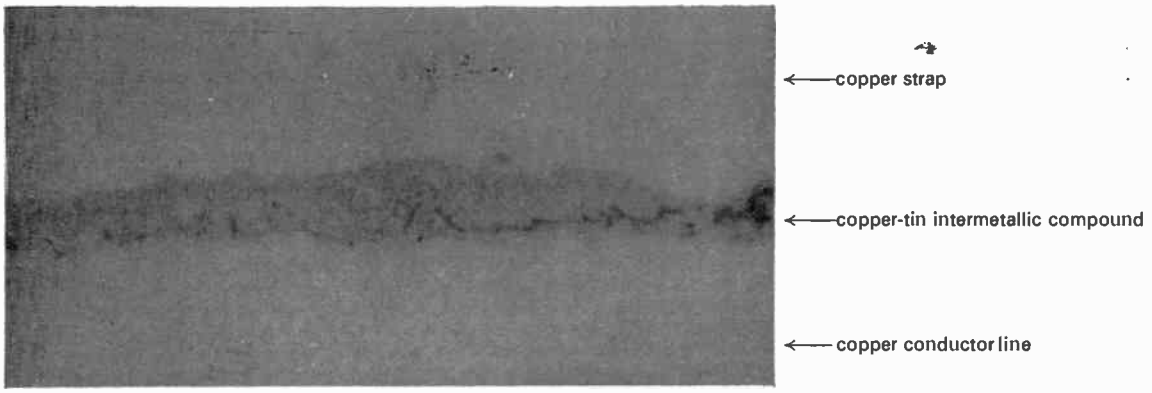


Fig. 12. Chip strap to substrate conductor joint after repair (× 2100).

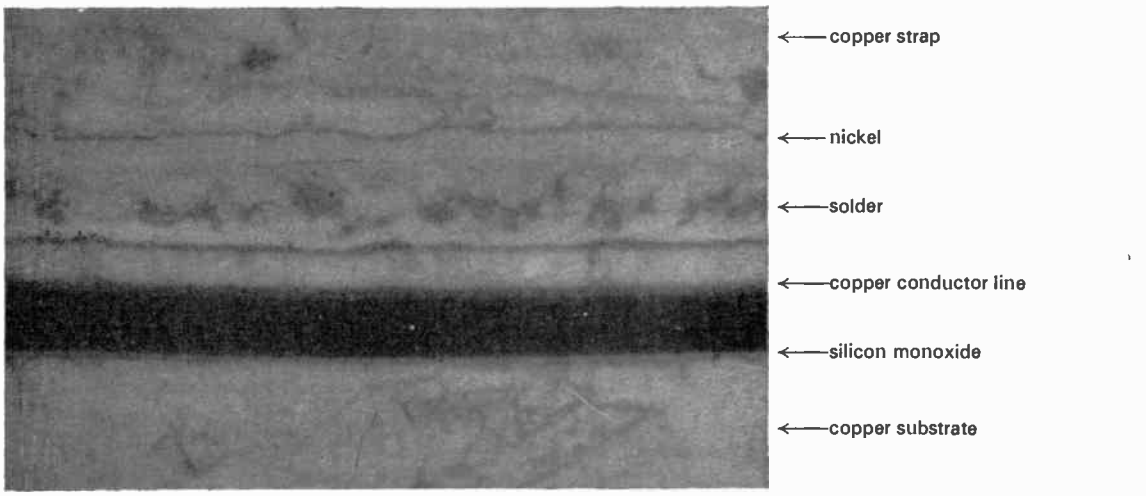


Fig. 13. Chip strap to substrate conductor joint showing effect of nickel diffusion barrier (× 2100).

in order to retain the solder on the areas chosen so that the risk of shorting the pad to the chip side or base was eliminated.

Due to temperature limitations imposed on the substrate, it was not feasible to solder the strap-to-substrate joint under the same controlled temperature and atmosphere conditions as used for the strap to chip joint. The heating system used relied upon a heated platen to raise the substrate temperature to 150°C and a hot probe on the top of the strap to raise the temperature of the junction to above the melting point of the solder at that point. In order to achieve this the probe was held at a temperature of approximately 350°C. It is, therefore, obvious that the joint has a temperature gradient across it, and the strap must be raised to a much higher temperature than the 183°C required for the nominally 60/40 Sn/Pb solder to become molten. The resultant joint is illustrated in Fig. 11. The amount of copper-tin compound constitutes a reliability hazard which is further exaggerated if replacement soldering is required. In fact, after a third replacement the joint consists almost entirely of intermetallic compound with very little solder remaining. (See Fig. 12.)

The higher diffusion rate at the strap side of the joint resulting from the higher temperature in the strap had to be counteracted by some means. As a balanced heat input could not be achieved because of the temperature limitation on the substrate, metallurgical solutions were investigated. Chromium proved to be too effective a barrier and de-wetting occurred, but nickel appeared to provide a suitable balance. It was decided to plate a nickel diffusion barrier under the solder-pad and as shown in Fig. 13, this permitted some diffusion of the tin into the nickel surface without permitting diffusion to the copper strap with its consequent brittle intermetallic formation. Palladium coatings were also tried and, although causing some loss of strength, appear to be promising for some situations.

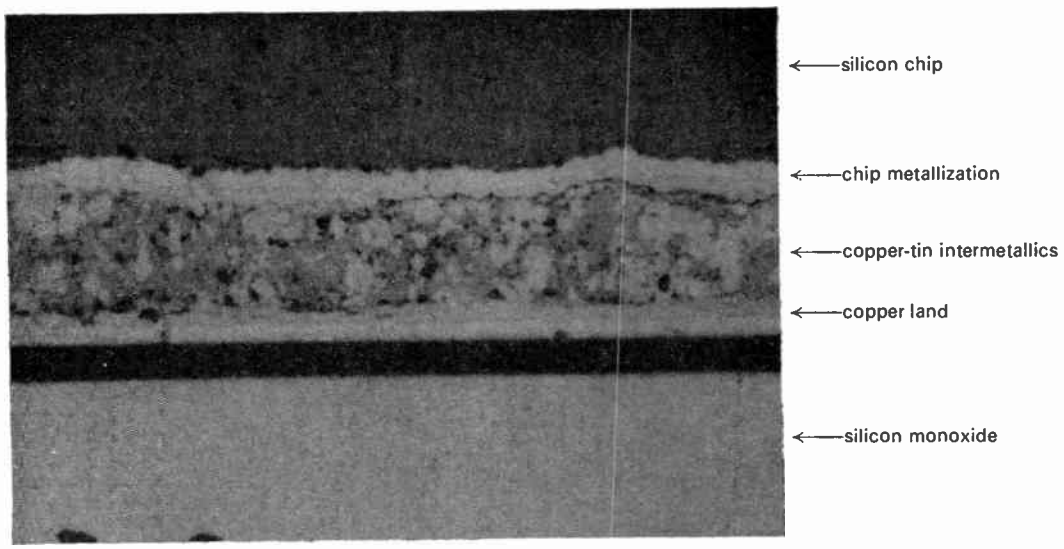


Fig. 14. Chip back to substrate pad joint (× 1200).

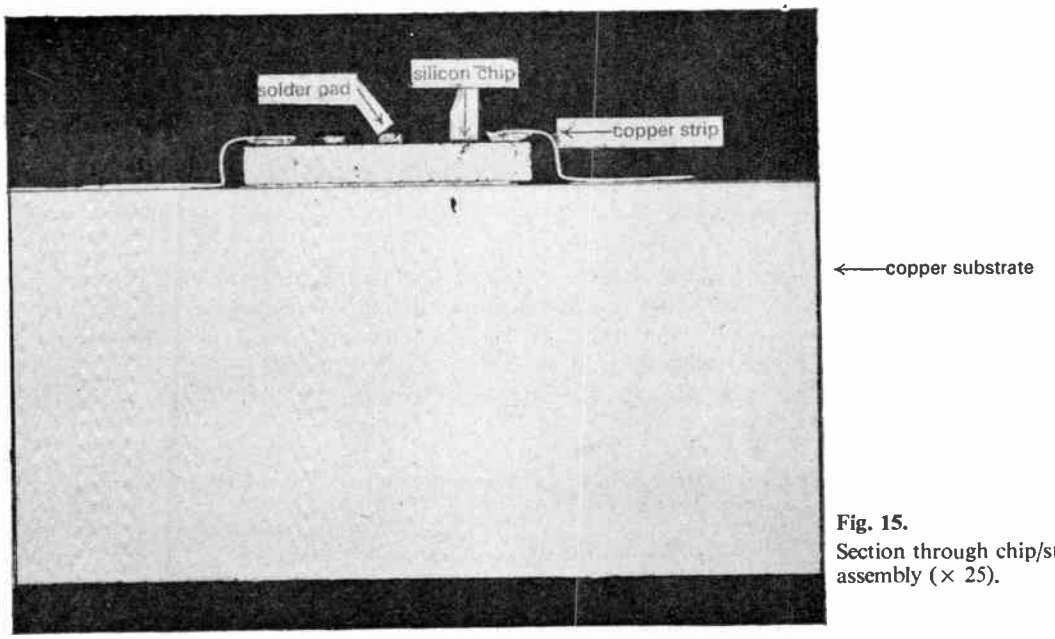


Fig. 15.
Section through chip/strap/substrate assembly (× 25).

The heat sources for the chip-substrate joint were originally identical to those used for the strap-substrate joint, but the resultant heat balance is somewhat different. The copper strap readily transfers heat from a heated probe to the joint but, as the thermal conductivity of the silicon chip is much lower, heat transfer to the solder is lower and slower. This results in the thermal gradient being much lower across the chip substrate joint and consequently the build-up of intermetallics is at a reduced level. (See Fig. 14.) Whilst this technique proved successful from a mechanical strength standpoint, it was found during electrical testing of the mounted chips that occasional mechanical damage to the active surface could occur. The heated probe was replaced by a hot gas jet which was of such a temperature and pressure that the chip was held in position whilst heat transfer took place. The heat input was slightly different to that resulting from the heated probe but the thermal gradient across the joint was similar and no great difference in diffusion rates was observed. A section through the assembled chip and connections is shown in Fig. 15.

To overcome the problems caused by surface oxidation of the copper conductors at the preheat temperature of 150°C, protective layers of gold or palladium are used. However, much concern has been expressed in the literature on the embrittling effect of gold on soldered joints,[1-4] and as the size of the joint decreases so the influence of any protective coating on the joint integrity increases. However, it is possible to use Foster's limiting criteria[4] as a guide in specifying the thickness of gold permissible as a protective coating. Palladium protective coatings do not give rise to embrittling effects in the same manner as gold, as the palladium does not dissolve in the solder at the same rate as gold.[3] Microprobe analysis shows the presence of an alloy layer of 70% Sn and 30% Pb remaining at the palladium-protected surface without any palladium being present in the solder in the joint. However, it must be noted that this is at the sacrifice of some strength.

Much difficulty was experienced in obtaining tin-lead solder of the correct composition, as this required great care in plating, particularly on small components. The usable life of parts after plating was found to be very short, seldom longer than 72 hours. A final coating of about 1 μm of electroless tin extends this period considerably.

4.1. *Joint Interaction*

Whilst development of techniques for each of the above joints could be carried out separately, eventually on assembling the whole package the effects of making one joint must affect and modify the other joints being made. The joint at the chip active face was between 95/5 Pb/Sn and 60/40 Sn/Pb, and the resultant alloy should melt at a higher temperature than the 60/40 Sn/Pb solder preform used under the chip or on the strap pad. However, when making the chip-substrate joint with the hot gas jet some melting of the chip strap joint occurred. Its action, fortunately, was beneficial in that it relieved stress in the pad to some extent and permitted the strap ends to settle on to the substrate prior to soldering. Similarly, when the strap ends were soldered to the circuit, there was sufficient heat-flow along the strap to cause softening at the chip joint and to cause stress-relief of the formed strap whilst under the strain of the soldering probe; the chromium plating prevented the pad-solder from spreading down the strap.

5. Substrate to Substrate Connections

The connections required from substrate to substrate consisted of a ground connection from copper base to copper base, and signal connection overlays 0·13 mm (0·005 in) wide on 0·25 mm (0·010 in) centres. These connections are illustrated in Fig. 16.

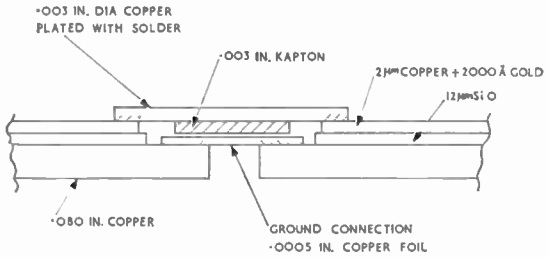


Fig. 16. Ground and conductor connections between substrates.

The ground connection between substrates is made using opposed carbon electrodes to reflow the solder plated on the copper foil. An alternating current of 200 A at 3 V provides enough energy to reflow a length of 12·7 mm × 3·2 mm (½ in × ⅛ in) wide. The

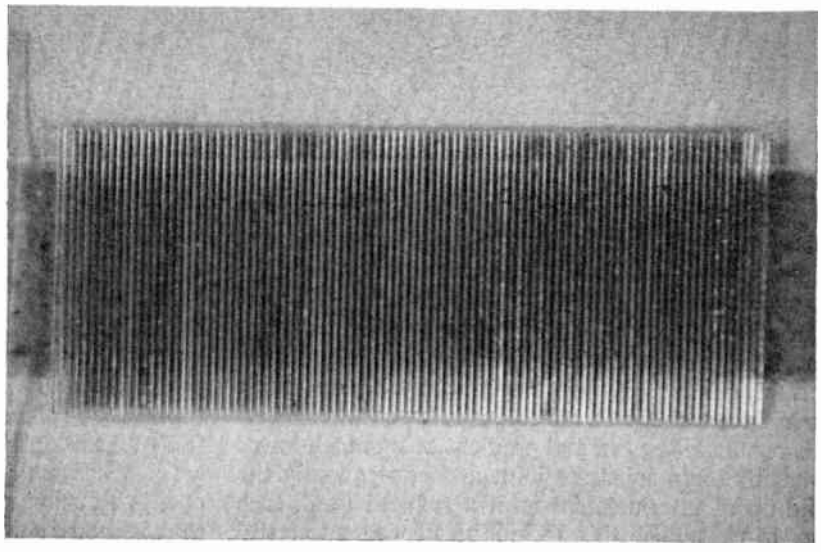


Fig. 17. Connection overlay.

electrodes are stepped along the substrate to provide a continuous seam.

The signal connection presented problems because of the high connection density and the heat sink effect of the substrate. Due to the temperature limitations on the substrate, and the necessity for repairability, a soldering technique was again chosen. Two overlay designs were investigated, one being an etched copper comb 0·013 mm (0·0005 in) thick plated with solder and having a bonded Kapton spacer to position and support the straps during soldering. The second design utilized round solder-plated copper wire wound on a former at the correct pitch and also bonded to a Kapton spacer. These overlays are illustrated in Fig. 17.

The method chosen to heat the solder and reflow the joint had to hold the strap or wire against the substrate, reflow the solder, and allow the solder to solidify without movement. A resistance wire formed to hold the strap or wire to the conductor and pulsed for 1·5 seconds at 80 A was found to reflow the solder adequately. During this process the resistance tip, which was formed from 0·13 mm (0·005 in) stainless steel, reached a temperature of approximately 500°C. This tip is illustrated in Fig. 18. No substrate heating was used during this operation.

The alignment of the conductors to the lines on the substrate was accomplished by hand with the aid of a microscope. The subsequent alignment of the resistance tip over each conductor was not easily and consistently achieved until the machine illustrated in Fig. 19 had been completed. This machine is arranged so that the resistance tip and its operating mechanism

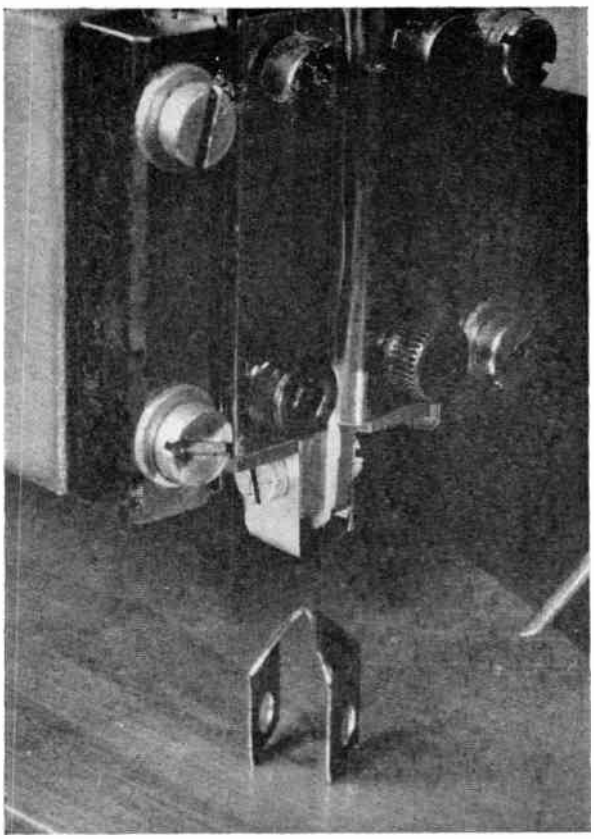
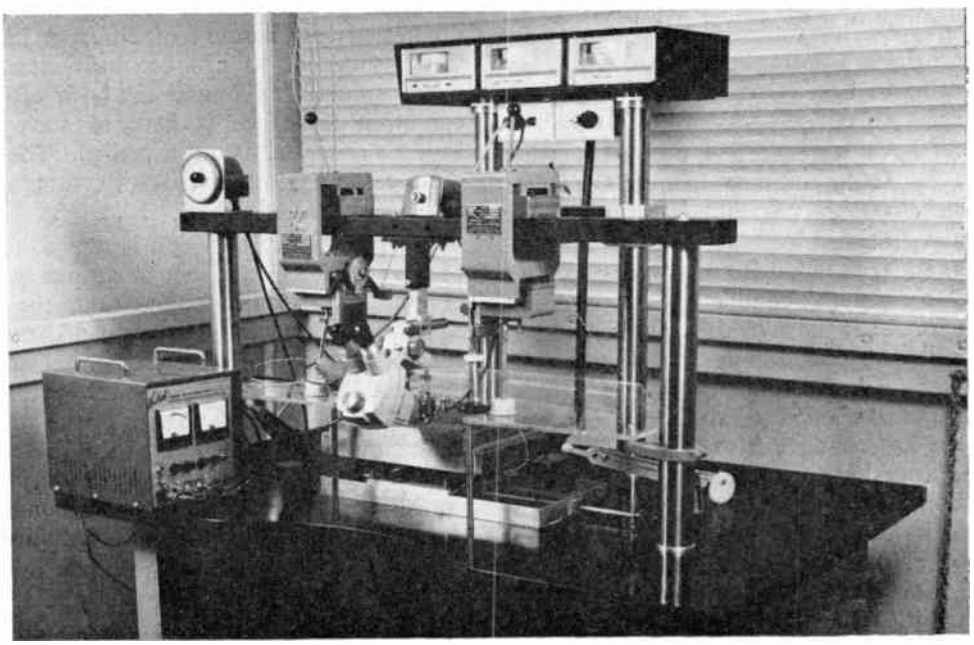


Fig. 18. Resistance tip used for soldering.

Fig. 19 (below). Overlay connection machine.



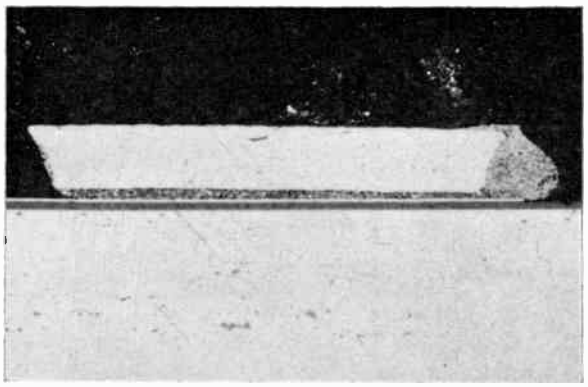


Fig. 20. Overlay joint using flat connection strap (× 300).

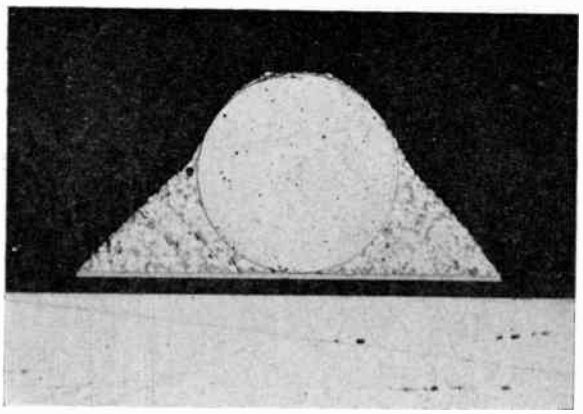


Fig. 21. Overlay joint using round wire (× 300).

is positioned by the left-hand micromanipulator, and the connection overlay, held on a vacuum pencil, is positioned using the right-hand micromanipulator. Resin-alcohol flux is painted on the substrate. When the overlay has been positioned by completing a few joints, the tip can be rapidly moved the $\frac{1}{4}$ mm from one joint to the next. One joint every 5 seconds can be achieved by a skilled operator.

Early experiments were carried out with etched straps which by nature of manufacture were rectangular in cross-section. With this method some difficulty was experienced in solder exuding from the joint (Fig. 20) under the load from the resistance tip, and shorting to adjoining conductors. In addition, the difficulty in obtaining a controlled thickness of solder resulted in undesirable intermetallics as mentioned later. The advantages of good planar rigidity for alignment afforded by the etched comb were felt to be important, but a superior joint was obtained using round copper wire (Fig. 21) and therefore this method was adopted. In this system the wire is plated with tin-lead throughout its length, and the heat and load of the resistance tip causes the wire itself to push down through the solder and bottom on the conductor. The solder continues to flow as more heat reaches the joint area and the effect of surface tension produces a neat, controlled fillet. Undesirable intermetallic compounds are minimized, and reproduceability is of a high order.

The design and manufacture of the overlays is worthy of mention as their quality and accuracy is extremely important. The wire used is plated with 60/40 tin lead solder 0·01 mm (0·0004 in) thick, and each batch is checked for plating composition and thickness. The carrier is a strip of Kapton 0·076 mm (0·003 in) thick coated with a thermosetting adhesive. Kapton is a polyimide with high resistance to shrinkage and degradation at elevated temperatures.

The wire is wound into accurately pitched notches on a jig holding the adhesive Kapton laminate, and the assembly is heated to 150°C, when the adhesive flows and fillets round the wire before curing. Pitch accuracy is maintained to within 0·05 mm (0·002 in) on an overlay 25·4 mm (1 in) long. Unfortunately, at the temperatures reached during the process rapid oxidation of the tin-lead occurs and the operation must be carried out in a vacuum oven or plated after curing.

6. Metallurgical Considerations—Jumper Connections

As with the joints on the chip, the effects of diffusion become important due to the size of the joint. The joints at the ends of the circuitry are made without preheating the substrate. The result of this is that the thermal gradient across the joint is even greater than that across the strap substrate joint from the chip. However, the use of a nickel barrier under the overlay solder pad was sufficient to restrict the tin diffusion to the nickel leaving the copper strap unaffected; the amount of diffusion into the nickel was insufficient to produce brittleness. The joint remains ductile as long as the layer of gold used for oxidation resistance is kept to such a level that Foster's criterion[4] is not exceeded. The use of palladium as an oxidation resistant coating also does not result in a brittle joint.

The problem of diffusion is diminished if a round wire is used in place of the flat strap as the solder mass becomes greater in volume than the intermetallic compounds formed. The ratio of brittle to ductile constituents is therefore decreased and thus the intermetallic compounds are more readily accommodated. In addition, the shape of the wire appears to permit more rapid melting of the solder in the joint area, thereby reducing the time during which the joint is at high temperature with a consequent reduction in intermetallic compound formation.

The shear strength of wire joints produced varies with the pad areas, but the pad size selected results in joint strengths of the same order as the wire strength.

7. Conclusions

It would appear that reliable soldered connections can be made at a density of ¼ mm pitch (100 connections per inch) and that the limitations on increasing this density are due to the ability to produce a system of making and placing the parts to be connected accurately enough rather than the soldering process itself. However, the importance of being able to plate solder of accurate composition and thickness, maintain an oxide-free surface prior to soldering and providing suitable solderable conductor surfaces, cannot be too highly stressed.

Adequate consideration must be given to the metallurgical aspects when attempting work on this scale. On normal size soldered connections there is sufficient volume of solder available to absorb undesirable intermetallics. However, when solder films have to be restricted to below about 0·025 mm to avoid shorts to adjacent conductors, then appropriate diffusion barriers and oxidation resistant surfaces must be provided. The use of solder at all may be questioned, but no other system has the advantages of low process temperature and good repairability, with minimal risk of damage to expensive substrates. The heat-sink effect of any substrate material should be considered in any joint design, and preheating of substrates avoided if the surface finish will not resist oxidation at temperatures above ambient.

In general, it appears that many electronic packages and devices are now so compact that their further development could be restricted by lack of suitable connection systems. Nowhere is this more apparent than in the semiconductor field, and any new advances may well originate in that industry.

8. Acknowledgment

The authors wish to express their appreciation for the contributions made by D. E. Hobbs, D. F. Rogers, T. W. S. Stenning and C. R. Tickner.

9. References

1. Powell, R., 'What's new in soldering?', *Electronic Products*, **8**, No. 7, p. 28, pp. 30–45, December 1965.
2. Slemmons, J. W. and Howell, J. R., 'Better bonding methods improve hybrid circuits', *Electronics*, **38**, No. 11, 24th May 1965.
3. Whitfield, J. and Cubbin, A. J., 'Experimental observations on the effect of gold and palladium on soldered joints', *A.T.E. J.*, **21**, No. 1, pp. 2–10, January 1965.
4. Foster, F. G., 'Gold-plated soldered joints', *Product Engng*, pp. 59–61, 19th August 1963.

Manuscript received by the Institution on 1st February 1968.
(Paper No. 1202/CC16.)

STANDARD FREQUENCY TRANSMISSIONS

(Communication from the National Physical Laboratory)

Deviations, in parts in 10^{10}, from nominal frequency for **July 1968**

July 1968	24-hour mean centred on 0300 U.T.			July 1968	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	− 300·0	+ 0·1	+ 0·1	17	− 299·9	+ 0·1	+ 0·1
2	− 299·7	+ 0·2	0	18	− 299·9	+ 0·1	+ 0·1
3	− 299·9	+ 0·2	0	19	− 299·9	+ 0·2	+ 0·1
4	− 299·9	+ 0·2	0	20	− 299·8	+ 0·1	+ 0·1
5	− 299·8	+ 0·2	+ 0·1	21	− 299·8	+ 0·2	+ 0·1
6	− 300·0	+ 0·1	0	22	− 299·8	+ 0·2	+ 0·2
7	− 299·9	+ 0·1	0	23	− 299·7	+ 0·3	+ 0·2
8	− 300·0	0	0	24	− 299·9	+ 0·2	+ 0·2
9	− 299·9	+ 0·1	0	25	− 299·8	+ 0·1	+ 0·1
10	− 299·9	+ 0·2	0	26	− 299·8	+ 0·2	+ 0·1
11	− 299·8	0	0	27	− 300·1	0	+ 0·2
12	− 300·0	+ 0·1	0	28	− 300·1	− 0·1	+ 0·1
13	− 299·9	0	+ 0·1	29	− 300·1	− 0·1	+ 0·2
14	− 299·9	+ 0·1	+ 0·1	30	− 300·1	− 0·1	+ 0·2
15	− 299·9	+ 0·1		31	− 300·1	− 0·1	+ 0·2
16	− 299·9	+ 0·1	+ 0·1				

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.

New Light—Visible and Invisible

The clamp of military security has operated perhaps more tightly on the publication of information on the techniques and uses of infra-red radiation than it has on many applications of electronics. It is therefore encouraging to record that some of the work in this field which has been carried out at the Signals Research and Development Establishment of the Ministry of Technology, at Christchurch, Hampshire, has now been 'de-classified', to use the military jargon. This Establishment which is mainly concerned with communications equipment for the British Army is now also concerning itself with the potentialities of fibre-optics communication, an application of opto-electronics which uses lasers—hence the title to this report. The Ministry is hoping that industrial applications will emerge from the military work.

Infra-red Applications

Two types of active infra-red devices were demonstrated recently at Christchurch to a representative of the *Journal*. The requirements of an active infra-red system are an infra-red source, which is a normal lamp with a filter which filters out the visible part of the spectrum, and a viewer, which consists of an optical system incorporating an image converter tube. Because the system relies on reflected radiation the relationship between the viewing distance and the source power is inverse fourth-power law. The filter used can be either a dyed plastic on glass or a coloured glass. The optical components of the viewer differ from conventional visible-spectrum optics in choice of glass, and in the blooming, which must cope with longer wavelengths than the visible, and is therefore thicker. The image converter tube consists of a photocathode, a conical anode and a phosphor screen, all enclosed in a glass envelope. The photocathode is made from caesium silver oxide, evaporated on to the front surface of the envelope. Radiation from the scene is focused on to the photocathode. The electrons emitted, as a result, are accelerated across the tube by a 12 kV potential and are focused by the conical anode to create a spatial replica of the photocathode scene pattern on the phosphor. The phosphor is looked at through an eye-piece, and the devices can be used indoors or out of doors for search or surveillance.

Far-infra-red devices, operating as heat-seeking heads in missiles, have been in use for some years. In the civil field, equipment for finding hot spots in cables has been made and it is now possible to make devices which can discriminate a 1 deg C difference between an object and its background. As a result it is possible to produce a device which will detect the presence of an animal, a vehicle or a human being. A remote alarm or counting device can be attached and thus the whole could be used as an intruder alarm, vehicle counter or to give evidence of the presence of animals in an area.

A more advanced type of far-infra-red device is a thermal imager. This incorporates one or more detectors and has a scanning system. The signal is applied to a cathode-ray tube and a television type of display is produced, forming pictures from heat contrasts rather than from colour and brightness contrasts. This type of system has also been employed in medical applications.†

Passing now to applications of infra-red to communications, standard pairs of binoculars have been modified to provide two-way voice communication up to ranges of 800 metres whilst retaining their normal function. A pulse modulated infra-red beam from a gallium arsenide lamp is used for transmitting from one half of the binoculars, and a silicon photodiode in the other half is used for receiving. Pulse frequency modulation at 20 kHz is employed, thus overcoming difficulties due to signal path variations (e.g. heat haze or hand shake).

Fibre-optics Communications

Until recently practical optical communication systems depended on propagation either through the atmosphere or through lens relay systems in large diameter pipes. However, the use of dielectric waveguides, in the form of glass fibres only 0·01 cm in diameter, is a most attractive alternative with many superior qualities.

A simple system using a short length of glass fibre and a gallium arsenide laser in a pulse-coded speech link operating in the near infra-red region was demonstrated at S.R.D.E. The glass fibre is inherently flexible and has high tensile strength; thus, if suitably protected from crushing, it can be used in much the same way as electrical wires are used at present, i.e. drawn along cable ducts, made up into multi-core cables, etc.

Clearly the fibre offers a considerably increased bandwidth over that of ordinary wire. In addition the small size allows, for example, 200 separate fibres to be packed into a diameter of only 2·5 mm. The glass is electrically insulating which suggests its usefulness where mutual interference with its surroundings may occur or where electrical isolation is required.

At this stage the successful implementation of optical fibre communication depends principally on the manufacture of fibre from very-low-loss glass. Optical glass at the moment has attenuation exceeding 200 dB/km whereas this application requires a minimum of 70 dB/km. Theory and experiment indicates that this is perfectly feasible and attempts are being made elsewhere to make suitable glass and fibre. Full exploitation of the system also requires increased efficiency and mean power of the source and improved detector performance.

According to their particular construction, glass fibres can support propagation of either single or multi-mode optical waves. Where very large bandwidth is not required, the multi-mode fibre is preferable because it allows simplification of the complete system. Some features of propagation in such fibres were also demonstrated, and micro-photographs illustrated imperfections in present samples. Typical performance figures for single-mode operation on a 1 μm diameter fibre clad by glass of 10% lower refractive index to 100 μm diameter was 10 GHz over 1 km; multi-mode fibre using a 80 μm diameter core with a cladding thickness of 5 μm gave 5 MHz bandwidth over 1 km.

† Lloyd Williams, K., Cade, C. M. and Goodwin, D. W., 'The electronic heat-camera in medical research,' *The Radio and Electronic Engineer*, **25**, No. 3, pp. 241–50. March 1963.

U.D.C. 534.88: 621.396.677

Correlation due to Inter-element Coupling in Line Arrays

By

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

Summary: In spite of the fact that inter-element coupling in a receiving array markedly affects the amplitude of the output signals when the beam is deflected from its normal axis, and also introduces correlation into otherwise-uncorrelated noise outputs from the various elements, yet its effect on signal/noise performance and on the beam-steering itself is small and usually negligible. The correlation introduced by inter-element coupling and the way it adds to correlation already existing in the acoustic field is examined and discussed in more detail. It is shown that it has important implications for a class of sonar systems in which signal and noise are distinguished by their inter-element correlation.

List of Symbols

N_T — number of elements in the array ($= 2N+1$)

v_n — output voltage of element n when the wave directly excites only the element n, the others being excited only through the mutual coupling between them and element n

V_n — r.m.s. value of v_n

v'_n — output voltage of element n when all elements are directly excited by wave

V'_n — r.m.s. value of v'_n

v_S, v_N — r.m.s. signal and noise voltages from one element when the wave directly excites only that one

v_{SO}, v_{NO} — ditto when there is no mutual coupling between elements

V_S, V_N — total r.m.s. signal and noise voltages from array

m_1 — mutual coupling coefficient representing the complex voltage transfer between adjacent elements of the array

m_2, m_3, etc. — ditto for elements separated by twice, three times, etc., the inter-element spacing

ϕ — phase shift increment per element to deflect beam to match angle of arrival of wave

ψ_q — correlation in applied field between elements spaced q inter-element spacings apart

ψ'_s — correlations between outputs of elements spaced s inter-element spacings apart

$\psi'_{s(n)}$ — ditto when the elements are numbered n and $n+s$ and their position is significant (e.g. in a short array)

ψ_m — correlation due to inter-element coupling when field is uncorrelated

λ — wavelength

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

1. Introduction

In a previous paper,[1] the effect of inter-element (or mutual) coupling on the signal/noise ratio in a uniform array was considered; it was assumed that a coherent unidirectional signal field is applied in the presence of a narrow-band noise field which is uncorrelated between one element of the array and another. The array is phased to suit the direction of the coherent signal wave. The conclusions drawn were that for an infinite array (whether planar or linear, although the analysis was that for a linear array), with equi-spaced non-directional identical elements, the signal/noise ratio is not altered at all by steering the beam or by the inter-element couplings; and that for a finite array the signal/noise ratio is altered, but only very little. In the infinite array the phasing required to steer the array to a particular signal direction is not affected by the coupling, although for the finite array it is affected. The inter-element coupling has a very marked effect on the amplitude of the signal when the beam of the array is steered off its normal axis.

The analysis given in the paper referred to was based on certain definitions which were stated to 'have a certain vagueness' so that the method was not rigorous. The working was, moreover, hard to relate to physical conceptions. The author now thinks that a much better and more rigorous analysis can be given which is at the same time much simpler to understand. It is outlined in Part 1 below. This method lends itself well to a calculation of correlations between the resultant outputs of the various elements of the array, and this is dealt with in Part 2. This calculation has a special importance in relation to sonar systems in which signal and background are

distinguished by the presence or absence of correlation between the outputs from adjacent elements.

A line array is assumed, of identical small elements uniformly spaced, with mutual coupling coefficients m_1, m_2, m_3, etc., where m_1 is the complex voltage transfer ratio between adjacent elements, m_2 is that between elements spaced by twice the inter-element spacing, etc. For ideal, loss-free elements (which are not mutually coupled electrically) these coefficients can be expressed in terms of the more familiar self-impedance Z_{11}, mutual impedances Z_{12}, Z_{13}, etc., and load impedance Z_L thus:

$$m_1 = Z_{12}/(Z_{11}+Z_L); \quad m_2 = Z_{13}/(Z_{11}+Z_L); \quad \text{etc.} \qquad \ldots\ldots(1)$$

Part 1. SIGNAL/NOISE RATIOS AND BEAM STEERING

2. The Infinite Array

Consider an infinite line array in which the centre element is numbered 0, elements to the right are numbered $+1$, $+2$, $+3$, etc., and elements to the left -1, -2, -3, etc., as shown in Fig. 1. The elements are connected to the output terminals via phase shifters with phase shifts given by $n\phi$ on element number n. Consider an incident wave to impinge only on element $+1$ and thus to excite directly only this element. (The fact that this would be difficult or impossible to achieve in practice is of no relevance here. It can in any case be easily simulated by electrical injection.) The other elements are excited, however, through the mutual couplings with element $+1$. This definition is different from that used in the paper referred to, where it is assumed that only the one element is excited at all. It overcomes the difficulties associated with that assumption.

Then we take the output voltage (before the phase shifter) of element $+1$ to be v_{+1}. The contribution to the output from element 0 is $v_{+1}m_1$, that from element -1 is $v_{+1}m_2 \exp(-j\phi)$, that from element $+2$ is $v_{+1}m_1 \exp(j2\phi)$, and so on; while that from element $+1$ itself is $v_{+1}\exp(j\phi)$. Clearly all the terms except the last-mentioned can be arranged in pairs, and since

$$\exp(jn\phi) + \exp(-jn\phi) = 2\cos n\phi,$$

we get a total output from the wave directly exciting only element $+1$ of

$$v_{+1}\exp(j\phi).(1+2m_1\cos\phi + \\ +2m_2\cos 2\phi + 2m_3\cos 3\phi + \ldots) \quad \ldots(2)$$

This makes it clear that the output voltage when only element $+1$ is directly excited varies according to the phasing provided for beam steering. Evidently if we consider any other element, say that numbered n, to be the only one directly excited, we obtain a similar result for the output, namely

$$v_n\exp(jn\phi).(1+2m_1\cos\phi + \\ +2m_2\cos 2\phi + 2m_3\cos 3\phi + \ldots) \quad \ldots(3)$$

If the signal is coherent, and comes from a direction for which the phase shifts are correct, then as we consider each element to be, in turn, the only one directly excited, all the contributions to the output are co-phasal. Thus, since the superposition theorem holds, the total output is the sum of these co-phasal outputs, i.e.

$$V_S = |1+2m_1\cos\phi+2m_2\cos 2\phi+\ldots|\sum v_S \qquad (4)$$

where

$$\sum v_S = \ldots + v_{-2} + v_{-1} + v_0 + v_{+1} + \ldots$$

when all the v are r.m.s. values and equal to v_S, and are co-phasal.

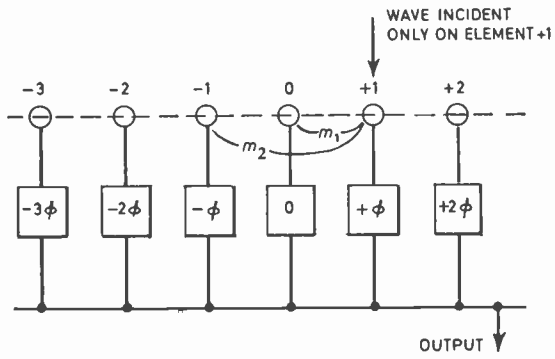


Fig. 1. The infinite array.

Clearly, the value of phase-shift required to steer the beam to a given direction is unaffected by the inter-element coupling, but the amplitude of the resultant signal is a function of coupling as well as of direction.

If the incident wave field is just noise, uncorrelated between the positions of one element and another, and with its bandwidth narrow enough for phase-shifts to be assumed constant over the band, we see that the successive contributions to the output as each element in turn is the only one to be directly excited by the wave are all uncorrelated, since the voltages v_0, v_{+1}, v_{+2}, etc., are all uncorrelated. The total output of noise is therefore given by the summation

of powers, so that the r.m.s. noise voltage is

$$V_N = |1 + 2m_1 \cos\phi + 2m_2 \cos 2\phi + \ldots| \cdot \sqrt{\sum v_N^2} \quad (5)$$

where

$$\sum v_N^2 = \ldots + v_{-2}^2 + v_{-1}^2 + v_0^2 + v_{+1}^2 + \ldots$$

when the v are all r.m.s. noise voltages v_N.

Clearly, if the total number of elements is N_T ($\rightarrow\infty$), the overall signal/noise power ratio is

$$\left(\frac{V_S}{V_N}\right)^2 = N_T\left(\frac{v_S}{v_N}\right)^2 = N_T\left(\frac{v_{SO}}{v_{NO}}\right)^2 \quad \ldots\ldots(6)$$

irrespective of the inter-element coupling and of the direction of beam-steering. Here v_{SO} and v_{NO} are the outputs which would be obtained from each element if mutual coupling were absent. This is the result obtained in the paper[1] and also by Mangulis.[2] It is clear that although this result will not hold if the noise is partially correlated between elements, as it usually will be in practice, yet the conclusion that the inter-element coupling has no effect on the signal/noise ratio will still apply.

It is worth noting that when all elements receive the incident wave, the output from each contains a contribution via inter-element coupling from all the others. The resultant output from element n is then v_n' which is different from v_n. When the incident wave is uncorrelated noise, the v_n are uncorrelated as discussed above; but the v_n are partially correlated because of the common components of their waveforms. This is examined in more detail for a two-element array in Section 5 of the original paper and in reference 3. It is calculated in more general terms in Part 2 of the present paper. It is therefore not so easy to calculate the signal/noise performance in terms of the resultant element voltages v_n; and this explains why we introduce the artifice of considering only one element to be directly excited at a time.

3. Finite Arrays

Obviously, if the array is not infinite, we no longer have the complete pairing of terms in eqns. (2) and (3) which led to the co-phasality of the various contributions, and the conclusions reached above are not properly valid. The finite array has a different and limited number of terms or phasors on each side of the directly-excited element, so that in addition to the symmetrical terms giving

$$\exp(+jn\phi) + \exp(-jn\phi) = 2\cos n\phi,$$

there will be remaining terms $\exp(+jr\phi)$ *or* $\exp(-jr\phi)$ on one side only which will cause the different resultants (as different elements are directly excited in turn) to be of different phase. This argument, of course, assumes that the value of ϕ is the same for a given beam deflection as in the infinite array, or in the finite array with no inter-element coupling.

Thus when a coherent signal wave is applied, the various contributions from the different elements do not add co-phasally and the total signal output voltage is therefore less than it would be if the addition were co-phasal. On the other hand, the noise contributions are uncorrelated as before, and the total noise r.m.s. voltage is not affected by the phase. It is clear therefore that the overall signal/noise ratio is less than it would be in the absence of inter-element coupling. Moreover, the signal is *not* maximized for a given direction by using the values of ϕ which are calculated on the basis of no coupling.

Even so, however, the discrepancies in signal/noise ratio are usually very small, and accounted for entirely by terms in m_1^2, m_2^2, $m_1 m_2$, etc., i.e. by second-order terms which will be negligible if the coupling is not large. This can be seen clearly by considering the case of a three-element array, in which only elements numbered -1, 0 and $+1$ are retained. The phasor diagrams for this case are shown in Fig. 2 with the simplifying (but unlikely) assumption that the arguments of m_1 and m_2 are zero. The lack of co-phasality as each element is in turn made the only one to be excited directly by the incident wave is clear. The resultant output from a coherent signal wave is

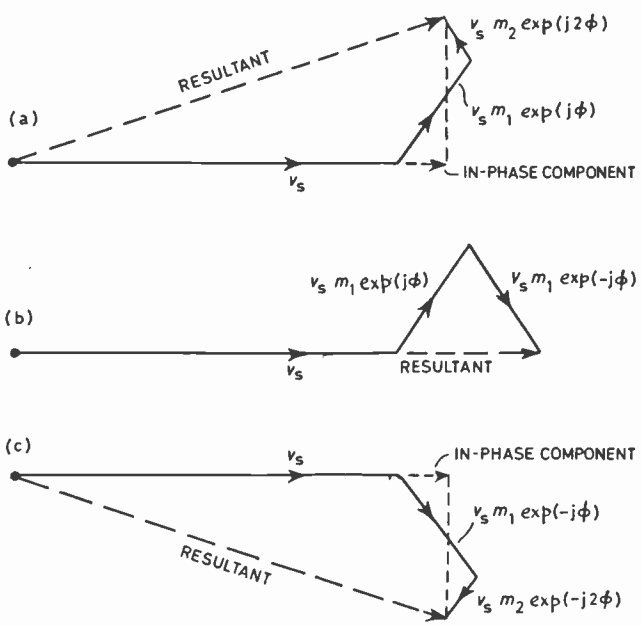


Fig. 2. Phasor diagram for three-element array.

the sum of the resultant of the centre diagram, and of the in-phase components of the other two. Now the relation of the resultant in phasor diagrams (a) or (c) to the in-phase component is that of a right-angled triangle with sides 1, x, and $\sqrt{(1+x^2)}$; i.e.

the hypotenuse is different from the longer of the other sides only by second-order quantities.

The signal output for a given direction can be maximized by an alteration in the value of the phase shifts such that the resultant phasors are brought into co-phasality. But even then it can be seen that the signal/noise ratio is reduced slightly by the mutual coupling, since the various resultants are not of equal amplitude, so that

$$[\sum v_S]^2/\sum v_N^2 < N_T.(v_{SO}/v_{NO})^2 \qquad \text{.....(7)}$$

4. Conclusions to Part 1

Although the results obtained in the previous paper[1] are still thought to be correct, a clearer yet more rigorous approach has been outlined here. In spite of the fact that inter-element coupling markedly affects the amplitude of the output signals when the beam is deflected from its normal axis, and also introduces correlation into otherwise-uncorrelated noise outputs from the various elements, yet its effect on signal/noise performance and on the beam-steering itself is small and usually negligible.

Part 2. CORRELATION DUE TO THE INTER-ELEMENT COUPLING

5. Calculation of Correlation

It has already been noted that when all elements of the array are excited by the incident wave there is a partial correlation between the outputs from the various elements due to the inter-element coupling. From the point of view of the factors considered in Part 1, namely signal/noise ratio and beam-steering, this is of little importance. There are, however, applications where this correlation is of fundamental importance. An example is a sonar system in which the degree of correlation between the outputs from each pair of adjacent elements is used as a criterion of whether it is signal or noise that is being received; such a criterion applied particularly to phase differences rather than to the amplitudes lends itself particularly well to a cheap system of electronic equipment using digital methods and microelectronic construction.[4,5] The significance of the effect of inter-element coupling in this application will be considered after the method of calculation has been set out. For this calculation, the symbols are used exactly as in Part 1, but the array is not considered to be infinite.

If the field is applied only to element r, then the output of element n is

$$m_{|n-r|}\,v_r \qquad \text{.....(8)}$$

Thus the total output from element n when the field is applied to all elements is

$$v'_n = \sum_{r=-N}^{r=+N} m_{|n-r|}\,v_r \qquad \text{.....(9)}$$

where

$$2N+1 = N_T$$

and

$$m_0 = 1$$

The total output from element $n+s$ when the field is applied to all elements is

$$v'_{n+s} = \sum_{r=-N}^{+N} m_{|n+s-r|}\,v_r \qquad \text{.....(10)}$$

The correlation between the outputs of elements n and $n+s$ is

$$\psi'_{s(n)} = \frac{\overline{v'_n.v'_{n+s}}}{V'_n.V'_{n+s}} \text{ where } V'_n \text{ is the r.m.s. value of } v_n \qquad \text{.....(11)}$$

$$= \frac{\left[\sum_r m_{|n-r|}.m_{|n+s-r|}\overline{v_r^2} + \sum_r \sum_{q=1}^{N} m_{|n-r|}.m_{|n+s-r\pm q|}\overline{v_r v_{r\pm q}}\right]}{\sqrt{\left\{\left(\sum_r m_{|n-r|}v_r\right)^2 \overline{\left(\sum_r m_{|n+s-r|}v_r\right)^2}\right\}}} \qquad \text{.....(12)}$$

Here it must be remembered that the m are all phasors and in the above process the product $m_x.m_y$ is the 'dot product' or scalar product, $|m_x||m_y|\cos\theta$, where θ is the difference between the arguments of m_x and m_y.

Now if the correlation in the applied field is ψ_q between elements spaced by q inter-element spacings, then

$$\psi_q = \frac{\overline{v_r v_{r\pm q}}}{V_r V_{r\pm q}} \qquad \text{.....(13)}$$

where V_r is the r.m.s. value of v_r and to a quite close approximation all the V are identical. Therefore let us normalize the system by putting $V_r = V_{r\pm q} = 1$. Then

$$\overline{v_r v_{r\pm q}} = \psi_q \quad \text{and} \quad \overline{v_r^2} = 1 \qquad \text{.....(14)}$$

The denominator of eqn. (12) is

$$\sqrt{\left\{\sum_r |m_{|n-r|}|^2\overline{v_r^2} + \sum_r \sum_q m_{|n-r|}.m_{|n-r\pm q|}\overline{v_r v_{r\pm q}}\right\}} \times \sqrt{\left\{\sum_r |m_{|n+s-r|}|^2\overline{v_r^2} + \sum_r \sum_q m_{|n+s-r|}.m_{|n+s-r\pm q|}\overline{v_r v_{r\pm q}}\right\}} \qquad \text{.....(15)}$$

$$= \sqrt{\left\{\sum_r |m_{|n-r|}|^2 + \sum_r \sum_q m_{|n-r|}.m_{|n-r\pm q|}\psi_q\right\}} \times \sqrt{\left\{\sum_r |m_{|n+s-r|}|^2 + \sum_r \sum_q m_{|n+s-r|}.m_{|n+s-r\pm q|}\psi_q\right\}} \qquad \text{.....(16)}$$

while the numerator of eqn. (12) becomes

$$\sum_r m_{|n-r|} \cdot m_{|n+s-r|} + \sum_r \sum_q m_{|n-r|} \cdot m_{|n+s-r\pm q|} \psi_q \quad (17)$$

Thus in expressing $\psi'_{s(n)}$ as the quotient of expressions (17) and (16) we have expressed the correlation between the outputs of two particular elements as a function only of the known coupling coefficients and the given correlation function of the applied field. Apart from the assumptions that all the V are identical and that m depends only on the spacing and not on the position in the array, this result seems rigorous.

In a long array, it may well be argued that the end effects may be neglected and that over most of the array the correlation $\psi'_{s(n)}$ is practically independent of n; and moreover that all the V' in eqn. (11) may be taken as equal. (Note that this last assumption is obviously a much coarser approximation than that of taking all the V to be equal.) In this case, n becomes an arbitrary choice and we may take it to be zero. We then have

$$\psi'_s = \frac{\sum_r m_{|r|} \cdot m_{|s-r|} + \sum_r \sum_q m_{|r|} \cdot m_{|s-r\pm q|} \psi_q}{\sum_r |m_{|r|}|^2 + \sum_r \sum_q m_{|r|} \cdot m_{|r\pm q|} \psi_q} \quad (18)$$

Although this formula has been derived only for the central part of a long array, yet it can be seen to apply also to a two-element array and to other special cases.

6. Two-element Array

For a two-element array, we put $r = 0$ or 1, $q = 1$ and $s = 1$ so that eqn. (18) becomes (if β is the argument of m_1):

$$\psi'_1 = \frac{2|m_1| \cos\beta + (1+|m_1|^2)\psi_1}{1+|m_1|^2 + 2|m_1| \cos\beta . \psi_1} \quad \ldots\ldots(19)$$

Defining ψ_m as the correlation introduced by mutual coupling when the field is uncorrelated, we have

$$\psi_m = \frac{2|m_1| \cos\beta}{1+|m_1|^2} \quad \ldots\ldots(20)$$

If $|m_1| < 1$ by a sufficient margin for $|m_1|^2 \ll 1$ to apply, then

$$\psi'_1 \simeq \frac{\psi_m + \psi_1}{1+\psi_m \psi_1} \quad \ldots\ldots(21)$$

and since $\psi_m \psi_1 \ll 1$ in most likely applications, we have

$$\psi'_1 \simeq \psi_m + \psi_1 \quad \ldots\ldots(22)$$

This will apply as a rough approximation even in long arrays.

7. Discussion of the Results in Relation to a Sonar System

As stated earlier, this work is very relevant to a particular type of sonar system in which correlation between elements is used as a criterion of signal versus noise. True isotropic noise could (in the absence of inter-element coupling) give $\psi_1 = 0$. If the received wave is entirely signal, then $\psi_1 = 1$ and from eqn. (21) we then have that $\psi'_1 = 1$ irrespective of ψ_m. The system then operates by determining whether the correlation is above or below a certain threshold.

In such a system, the 'noise' is very often reverberation, i.e. the back-scattering of one's own transmission from inhomogeneities in the body or boundaries of the sea. Such reverberation is, of course, not isotropic. In a typical system, where a wide sector is insonified and then searched by the high-resolution receiving system, the value of ψ_1 may well be around 0·2.

The significance of ψ_m is that it modifies the actual correlation between elements and so disturbs the criteria discussed above. However, it is not easy to put in numerical values, as the values of m are not usually known. Moreover, the elements are usually large in the sense that they occupy most (or all) of the space between their centres, and being also of finite width in the perpendicular direction, necessarily disturb the field to some extent. It is no easier to measure self and mutual impedances, although these may have more basic value as they are not dependent on the circuits connected to the transducers. Very few data have been published.

From calculations of mutual impedance made by Sherman[6] for a transducer array which was not too different from the linear kind which would be used in the system referred to, we can estimate that for an inter-element spacing of $0{\cdot}75\lambda$ (which is the spacing for which we quoted $\psi_1 \simeq 0{\cdot}2$ above), a typical transducer used around 50–100 kHz might have $m_1 \simeq 0{\cdot}07$ with $\beta = \pi$ radians. The correlation introduced by this mutual coupling is seen from eqn. (20) to be $\psi_m \simeq -0{\cdot}15$. The sign is significant here, so that the total correlation ψ_1 is reduced from 0·2 to only 0·05. At first sight it might seem that this must improve the performance of the sonar system, since the reverberation and noise background now has negligible correlation, so increasing the gap between signal and noise. Indeed, this conclusion might actually be true for a particular practical system; but it cannot be true of an ideal system, since the addition of more distortion (in this case the coupling between elements) cannot increase the information, and must usually decrease it.

The value of ψ_m is certain to be negative over some part of the range of element spacings between say, 0·3 and $1{\cdot}0\ \lambda$, and this is where the magnitude of coupling is relatively large. At larger spacings the sign of ψ_m becomes positive, and then alternates as the magnitude diminishes.

It is clear that the effect of inter-element coupling may often not be very large, but it must be taken into account if optimum performance is to be obtained.

8. References

1. Tucker, D. G., 'The effects of inter-element coupling in line arrays on signal-to-noise ratio and beam steering', *J. Sound Vib.*, **6**, pp. 143–151, 1967.
2. Mangulis, V., 'Signal-to-noise ratio for arrays in an uncorrelated noise field', *Trans. Inst. Elect. Electronics Engrs on Sonics and Ultrasonics*, **SU-12**, pp. 49–52, 1965.
3. Tucker, D. G., 'Spatial correlation of reverberation, correlation due to mutual coupling, and their implications for digital sonar', Proc. NATO Advanced Study Institute on Stochastic Processes in Underwater Sound Propagation, La Spezia, September 1967.
4. Nairn, D. 'Theoretical possibilities of a digital sonar system', I.E.R.E. Symposium on Signal Processing in Radar and Sonar Directional Systems, University of Birmingham, July 1964, Paper No. 20.
5. Griffiths, J. W. R., and Creasey, D. J., 'A digital sonar system', *J. Sci. Instrum.*, **43**, p. 534, 1966.
6. Sherman, C. H., 'Analysis of acoustic interactions in transducer arrays', *Trans. I.E.E.E.*, **SU-13**, No. 1, p. 9, March 1966.

Manuscript first received by the Institution on 19th March 1968 and in final form on 3rd June 1968.
(Paper No. 1203/AMMS 15.)

Letters to the Editor

The Commercial Engineer

SIR,

I was delighted with the content of your editorial of May 1968† and I would like to believe that the views expressed indicate a change of attitude on the part of our professional institutions towards sales engineers.

I have been employed in the electronics industry for more than twenty years and hold graduate qualifications for our two major institutions. My tactical error was, however, to join the sales force of a major radio communications company who, at that time, employed comparatively few chartered engineers. For all practical purposes the door to chartered status is closed to this type of applicant and after numerous enquiries I resigned myself to this fact.

I am now in a position which necessitates control of sales staff selling communications systems to the public utilities, major oil companies and a number of specialist users. The sales engineer requires consultancy skills to put at the disposal of the end user and expertise on the marketing research aspects of our profession. The need to improve the engineering status of our marketing is self-evident but there is reluctance on the part of the young engineer to jeopardize his professional future by accepting a position which badly needs his skills but is not recognized by the institutions as a path to Chartered status.

In my opinion the country's economic growth is adversely affected while this state of affairs exists.

T. W. RICKETT,
Graduate.

44 Pierce Lane,
Fulbourn, Cambridge.
28th June 1968.

SIR,

In undeveloped countries, such as Sabah, the Sales Engineer, although he is sometimes looked down upon by his more academic colleagues in research and development, has to play a vital role and play it in accordance with the highest ethical standards of our profession. Rarely does a client know what he wants except in general terms, and he must therefore place himself entirely in the hands of the sales engineer who in the end writes out the client's order for him.

This does not, as many engineers believe, call merely for super-salesmanship and a glib tongue. A good sales engineer has first to be a widely-versed technical expert, in every way *au fait* with the latest developments in an ever-expanding field of knowledge. All too often he is considered to be a mere jack of all trades or at the best a 'general practitioner'; although the good old G.P. in medicine and in engineering does most of the donkey work.

I think your editorial will greatly cheer and encourage all sales engineers, and as one of them I wish to express my sincere appreciation.

A. B. AVERY,
C.ENG., F.I.E.R.E.

c/o O'Connor's Ltd.,
P.O. Box 1197, Kepayan,
Kota Kinabalu, Sabah.
13th July 1968.

† 'The commercial engineer.'

U.D.C. 621.316.727

Pull-in Range of Phase-lock Circuits with Arbitrary Feedback Filter

By

R. LEONHARDT, Dipl., Ing.†

AND

H. H. FLEISCHMANN, Ph.D.‡

Summary: Existing methods for the theoretical determination of the capture range of selected automatic phase control loops, using a sin φ-phase comparator, are analysed. From the results of this analysis, an equivalent method to the one used by Goldstein in his treatment of loops with saw-tooth comparator is developed that can be applied to a wider variety of interesting loop filters. The method is based on the experimentally suggested hypothesis that all non-synchronized states of loops of most interest are asymptotically periodic. A function is derived which allows quick determination of the capture range of most stable a.p.c. loops containing one large integration time-constant in their filter branch. The method is directly applied to a few technically interesting cases, and very good agreement with experimental results is found.

1. Introduction

Automatic phase control (a.p.c.) loops in various modifications are used for many applications, such as oscillator synchronization, signal tracking, or simply noise suppression. Due to the basic characteristic of the phase comparator, their system equation generally is highly non-linear. For loops using a saw-tooth comparator, a closed-form theory of their non-linear characteristics, in particular of their acquisition behaviour, has been achieved by Goldstein[1] and was tested by Byrne.[2] However, no similar exact treatment has been reported so far on the more common loops using a sin φ-comparator. The behaviour of these loops has been extensively investigated for the synchronized state. In this case, closed expressions for various characteristics can be obtained by linearization and extensive theoretical data have been accumulated for a variety of conditions. In contrast, still comparatively few results have been published on the basically non-linear acquisition behaviour of these circuits. In this respect, the theoretical analysis is still limited to a few cases with rather special loop filters, the results being contained almost completely in References 3–8. The loop filters treated in these papers are probably those most frequently applied. However, in many applications these filters can be approximated only roughly, or the use of more complicated filters would at least seem advantageous for the achievement of wanted loop characteristics in the synchronized state. The pull-in behaviour of those more general loops has been investigated by Rey.[9] However, as shown in Section 3, his results are only applicable for a sufficiently small time-constant in the filter branch.

In this paper the pull-in behaviour of a.p.c. loops with one imperfect large time-constant integrator and additional networks in the filter branch of the loop is investigated, with particular emphasis on the pull-in range, assuming constant signal frequency. First, the approaches of the mentioned theoretical papers[3–9] are analysed. Following the conclusions from this analysis a semi-linear method is devised that allows additional networks to be added to a basic imperfect integrator in the filter loop. For this, only the frequency dependence of the phase shifts and the amplitude ratios of these additional filters need to be known. The mathematical foundation of the underlying assumptions is discussed. Using this method, an expression is derived that makes possible a direct determination of the pull-in range of these loops from the frequency characteristics of their filters. In the following section, this formula is then applied and the results illustrated for various loop filters, including the already known case of the imperfect integrator and the cases of an added RC-integrator and an added ideal delay line. These theoretical predictions are found to be in very good agreement with published experimental results of Gruen[4] and our own careful experiments. Finally the limitations for the applied method are assessed. It also appears possible to adapt this method to the determination of related characteristics of these loops.

2. Basic Assumptions and Equations

In the experimental literature, a variety of individual designs for a.p.c. loops is described. For theoretical analysis, however, most of them can be transformed to the block diagram shown in Fig. 1: a ' signal ' (with

† Rohde and Schwarz, Messgeraetebau, Munich, Germany.

‡ General Atomic Division of General Dynamics Corporation, John Jay Hopkins Laboratory for Pure and Applied Science, San Diego, California; now with Department of Applied Physics, Cornell University, Ithaca, N.Y. 14850, U.S.A.

the angular frequency† ω_1 and the phase φ_1) and the output of a voltage-controlled oscillator (v.c.o.) are fed into the phase discriminator. Its output voltage U_P, after passing through a linear filter with the characteristic $K_2 F(p)$, in turn controls the frequency of the mentioned oscillator.

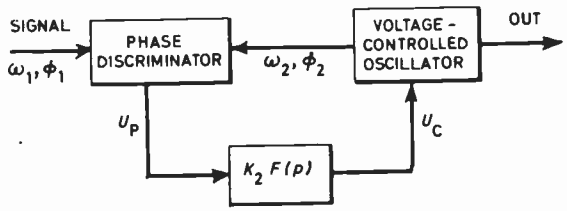


Fig. 1. Basic block diagram of automatic phase control loop.

For the following analysis it is assumed that:

(a) The frequency of the 'signal'

$$\omega_1 = \dot{\phi}_1 = \text{constant}$$

(b) The frequency ω_2 of the v.c.o. varies linearly with the controlling voltage around the centre frequency ω_0, which is obtained for $U_c = 0$,

$$\omega_2 = \omega_0 + \alpha U_c$$

(c) The output of the phase discriminator is proportional to the sine of the phase difference

$$\varphi = \varphi_1 - \varphi_2$$

between signal and oscillator:

$$U_p = K_1 \sin \varphi$$

and only the low beat frequency between signals and oscillator passes through the filter.

It has been shown[3-7] that under these conditions eqn. (1) can be derived for the phase difference φ between signal and oscillator:

$$\dot{\varphi} + KF(p) \sin \varphi = \delta\omega \qquad \text{.....(1)}$$

In this equation, $\delta\omega = \omega_1 - \omega_0$ denotes the difference between signal frequency and centre frequency of the v.c.o. $K = \alpha K_1 K_2$ is the frequency shift of the v.c.o. produced by the maximal output voltage of the phase discriminator, i.e. the maximal frequency shift of the v.c.o. in this circuit, and $F(p)$ is the filter transfer function in operational notation, with

$$F(0) = 1 \quad \text{and} \quad p \equiv \frac{\mathrm{d}}{\mathrm{d}t}$$

Equation (1) is the basic equation for the loop which is being investigated. For the synchronized state, i.e. $\dot{\varphi} = 0$, one obtains two equilibrium phase angles φ_1 and φ_2 with $\sin \varphi_{1,2} = \delta\omega/K$. Only one of these, namely, $\varphi_1 > \pi/2$, can be stable against small perturbations, whereas the second, with $\varphi_2 > \pi/2$, is principally unstable. By linearization of (1) around the 'stable' equilibrium point, φ_1, the small signal response function of the system in the synchronized state is obtained. For an investigation of the acquisition behaviour, however, the full eqn. (1) has to be used.

In these loops, the 'capture range' or 'pull-in range' is defined as maximum deviation of the signal frequency from the centre frequency of the undisturbed oscillator for which synchronization is achieved independently of the initial condition.

3. Analysis of Existing Approaches

A closed mathematical analysis of the system in the non-synchronized state can only be done for $F(p) \equiv 1$ (first-order loop). In this case synchronization is accomplished if and only if $\delta\omega \leqq K$. For all other filter functions $F(p)$, only approximate expressions can be given.

Preston and Tellier[3] investigated an a.p.c. loop with a simple RC-integration filter (Fig. 2(a)). In this case eqn. (1) becomes

$$\frac{\mathrm{d}^2\varphi}{\mathrm{d}t^2} + \frac{1}{\tau}\frac{\mathrm{d}\varphi}{\mathrm{d}t} + \frac{K}{\tau}\sin\varphi = \frac{\delta\omega}{\tau} \qquad \text{.....(2)}$$

By introduction of $y = \mathrm{d}\varphi/\mathrm{d}t$, they then convert this equation to

$$\tau y \frac{\mathrm{d}y}{\mathrm{d}\varphi} + y + K \sin\varphi = \delta\omega \qquad \text{.....(3)}$$

and integrate it graphically in the phase plane $y = y(\varphi)$. As a necessary and sufficient condition for synchronization to occur independently of the initial conditions of the system, they compute the trajectory starting with $y_0 \to 0$ from the unstable equilibrium phase φ_2 mentioned in Section 2. If this trajectory does not cross either of the phases $\varphi_2 \pm 2\pi$, then the system will always become synchronized.

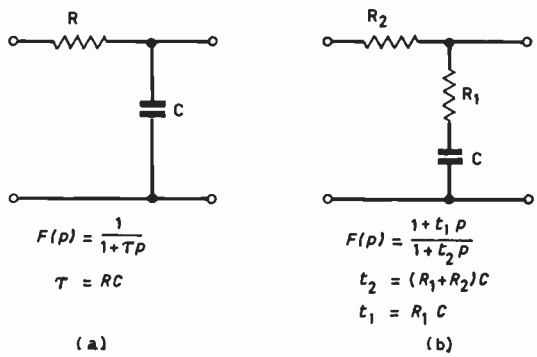


Fig. 2. Integrators and their definitions.
(a) Normal. (b) Imperfect.

† In this paper all frequencies are angular frequencies.

The validity of this criterion is evidenced by the equivalence of eqn. (2) to the equation of motion for a circular pendulum which is accelerated by a constant torque $\delta\omega/K$ and decelerated by a velocity-dependent 'friction' $(1/\tau)\mathrm{d}\varphi/\mathrm{d}t$. Synchronization, of course, is analogous to a stopping of the pendulum. For stopping to occur independently of the initial conditions, the energy of the pendulum lost by friction during one full 360° cycle has to be smaller than the energy added by the torque for all initial conditions. Due to the velocity dependence of the friction term in eqn. (2) the frictional losses are smallest if the pendulum just barely passes the unstable equilibrium point. Thus the criterion used is justified. Using this criterion and a perturbation approach for small $\delta\omega$ and $1/K\tau$ we computed the capture range Ω analytically. Employing a third-order approximation we obtain the relation

$$(K\tau)^{-\frac{1}{2}} = (\pi/4)(\Omega/K)/(1-0{\cdot}3(\Omega/K)^2) \quad \ldots\ldots(4)$$

which, for large $K\tau$, is very close to the relation

$$\frac{1}{K\tau} = 0{\cdot}64(\Omega/K)^2$$

derived graphically in Preston and Tellier's paper. But eqn. (4) should hold up to $K\tau \lesssim 1$.

This same criterion is applied by Gruen[4] for an a.p.c. loop with the imperfectly integrating filter shown in Fig. 2(b). His results, together with experimental points, are reproduced in Fig. 3, from his Fig. 7. They show good agreement for $Kt_2 < 100$, but fail for $Kt_2 > 100$. A case of failure of this criterion also is evident in the phase plot in Fig. 16 of Viterbi's paper,[7] where a stable non-synchronized state exists in spite of an apparent fulfilment of this criterion.

A good perceptual explanation for this failure can be obtained by means of Preston–Tellier's pendulum analogue. For this filter, eqn. (2) is replaced by

$$\frac{\mathrm{d}^2\varphi}{\mathrm{d}t^2} + \left[\left(\frac{t_1}{t_2}\right)K\cos\varphi + \frac{1}{t_2}\right]\frac{\mathrm{d}\varphi}{\mathrm{d}t} + \left(\frac{K}{t_2}\right)\sin\varphi = \delta\omega \quad \ldots\ldots(5)$$

in which an additional 'friction term'

$$(t_1/t_2)K\cos\varphi\,\mathrm{d}\varphi/\mathrm{d}t$$

appears. Computing again the 'energy' dissipated over one full cycle, one can see that this new term contributes only through a difference in the average of the 'velocity' $\mathrm{d}\varphi/\mathrm{d}t$ in the upper and lower halves of the cycle where the cosine has a different sign. Those differences are mainly due to the 'gravity' term $(K/t_2)\sin\varphi$ and become smaller with larger average velocity. Thus, as the later correct treatment brings out, the minimum of the 'frictional' energy dissipation no longer occurs at the trajectory which just barely passes over the unstable equilibrium phase angle, but at higher average velocities $\mathrm{d}\varphi/\mathrm{d}t$, i.e. higher average beat frequencies. Thus Gruen's

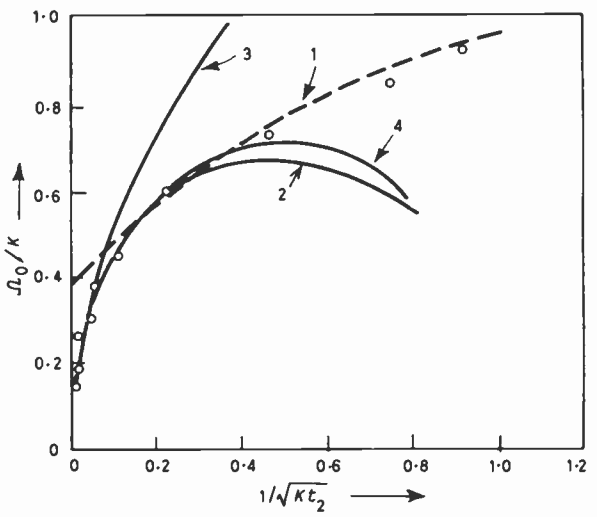


Fig. 3. Capture range with imperfect integrator for Gruen's parameter $\xi = (t_2/K)^2(1/t_2 + Kt_1/t_2) = 0{\cdot}5$

Experimental result of Gruen: ○—○. Theoretical predictions: Gruen, 1; Richman, 2; Viterbi, 3; Equation (10b), 4.

criterion can provide only an upper limit for the capture range, and it particularly breaks down for large values of Kt_2 where the described effects become prominent.

The criterion used by Rey,[9] in particular his condition (32) and the preceding, are similar to Gruen's argument, although somewhat more stringent. Correspondingly, his result (34) is a good approximation for sufficiently small values of Kt_2 or Kt_1, but it becomes invalid for large values of Kt_2 and Kt_1 where our treatment will give better results. This is evident in his Fig. 13 where the extrapolation for

$$\omega_n/K \equiv (Kt_2)^{\frac{1}{2}} \to 0$$

is incorrect. An analysis of his formula (34) shows that $\Omega/K \to 0{\cdot}27$ for $\omega_n/K \to 0$, but not $\Omega/K \to 0$.

Also Richman[6] and Viterbi[7] investigate the acquisition behaviour of these loops for fixed signal frequency. After transforming eqn. (5) into an equation similar to our eqn. (3), Viterbi first uses a phase plane analysis. From this he observes that for sufficiently large value of Kt_2 the relative fluctuations of the beat frequency between signal and oscillator within one beat period are small for many trajectories, including the 'stable asymptotic non-synchronized' trajectory. In the same cases, the relative change of the beat period from one beat period to the next is generally small. He takes these facts as basis for his analytic treatment. Richman, although somewhat different in his mathematical approach, has essentially the same starting point. The corresponding results of both papers are also included in Fig. 3. As is to be expected, they agree very well with the experimental results for large values of Kt_2. For

small values of Kt_2, however, the applied analytic approximations break down, and the theoretical results of Gruen agree better with the experiment. The larger deviations of Viterbi's results at this point are probably due to the assumption, in his eqn. (48), that $\omega(\varphi)$ is constant within a beat period.

As is already apparent from the pendulum analogue, the success of this approach for large Kt_2 is due to the fact that the decision over a final achievement of synchronization occurs at beat frequencies $\Delta\omega \ll 1/t_2$. Thus the filter strongly attenuates the a.c. output of the phase discriminator and the beat frequency fluctuates only little during one beat period. At the same time, however, due to the non-linearity of $\sin\varphi$, slow changes of the d.c. component of the oscillator control voltage can occur, i.e. the oscillator frequency drifts until it reaches either synchronization or a non-synchronized equilibrium state. This same basic idea will now be extended to a treatment of a.p.c. loops with more general filters which include one large integrating time-constant.

4. Derivation of General Formula

A direct extension of the mathematical techniques used by Richman and by Viterbi to a.p.c. loops with more complicated filter systems is hardly possible. In this case, the resulting differential equation for the phase angle φ is of higher than second-order, and the asymptotic behaviour of the solutions has to be investigated for all possible initial conditions. This makes, as Viterbi mentions, the graphical phase–plane approach very impractical. Also, the analytic techniques used by both authors are generally no longer applicable. In the following, therefore, a somewhat different viewpoint will be exploited. The capture range of an a.p.c. loop is generally defined by the condition that the loop, for all signal frequencies within this range, becomes stably synchronized finally with the signal, independently of the initial conditions of the system. For all other signal frequencies, at least for some initial conditions, the system will proceed into one of possibly more asymptotic non-synchronized states which will only very grossly reflect the initial conditions. Thus the defining condition for the capture range is equivalent to the non-existence of a stable non-synchronized state.

Viterbi's ' limit cycles ' correspond to such a state. In this case, due to the structure of (2), these states must be periodic in φ and also in time. However, the same is experimentally found true for a very wide variety of more general loop filters which include one large integrating time-constant. Based on this experimental evidence, in the following treatment, it will be assumed, for all investigated systems, that all non-synchronized states are simply periodic in time. Then, correspondingly, the capture range is determined as the signal frequency range for which simply time-periodic non-synchronized states do not exist.

Although the mathematical formalism is different, this criterion is fully equivalent to that applied by Goldstein[1] in his treatment of a.p.c. loops with saw-tooth comparator. His ' discontinuity point condition ' corresponds directly to our basic assumption. As in his case, no clean mathematical proof of our criterion can be given at this point. Thus, mathematically, this criterion only constitutes a necessary but not a sufficient condition for the capture range. In particular, disturbances might arise from multiple periodic solutions with non-rational frequency ratios, produced, for instance, by additional feedback oscillations. On the other hand, the existence of such solutions in systems with one single large time-constant does appear rather unlikely. For more than one large integration time-constant, the system is generally unstable anyhow. Thus we would expect our criterion to hold rather generally, in spite of its principal mathematical insufficiency.

Similar to Viterbi's results, also in many experiments with more complicated loop filters, only small variations of the beat frequency over a beat period are found, even for signal frequencies close to the capture range. This fact will also be used in the following derivation as a heuristic principle. In this case, however, a self-consistency check of the final results can be made.

To make maximum use of the already known behaviour of a loop with an imperfect integrating filter, the filter transfer function will be used in the operational form

$$F(p) = \frac{1+pt_1}{1+pt_2}\,T(p) \qquad \ldots\ldots(6)$$

In this equation, the splitting of $F(p)$ into the factor $(1+pt_1)/(1+pt_2)$ and the 'additional transfer function' $T(p)$ is somewhat arbitrary. The selection of both is not necessarily defined by the circuit. Both t_1 and t_2 may not be explicitly realized in the actual circuit, but may contain an aggregate of time-constants of the entire circuit. $T(p)$ is then only defined by

$$T(p) = F(p)(1+t_2\,p)/(1+pt_1)$$

However, it is felt that best results generally will be obtained if t_2, at least, very nearly agrees with the largest time-constant of the filter network. On the other hand, in the definition of the time-constant t_1, one is certainly considerably less bound. In this respect, one has probably only to make sure that the loop corresponding to the filter function

$$(1+pt_1)/(1+pt_2)$$

falls into the region of validity of the treatments by Richman or Viterbi.

Due to splitting (6), the basic differential equation of the system becomes

$$(1+pt_2)p\varphi+K(1+pt_1)T(p)\sin\varphi=\delta\omega \quad \text{.....(7)}$$

Here, in principle, $T(p)$ would have to be used in its operational notation. In this case, a new differential equation would be generated by each filter form and the corresponding difficulties already discussed would arise. Therefore, due to our intentions to investigate only periodic asymptotic solutions it seems more appropriate to use $T(p)$ in the form

$$T(p)\langle=\rangle a(\omega)\,e^{-j\varepsilon(\omega)} \quad \text{.....(8)}$$

where $a(\omega)$ designates the amplitude and $\varepsilon(\omega)$ the phase of the transfer function of the ' additional filter '. Using these definitions, we now investigate the asymptotic non-synchronized states of eqn. (7). The details of this treatment are given in the Appendix: due to our basic assumption, $d\varphi/dt$ and $\sin\varphi$ can be expanded into Fourier series with the periodicity of a basic beat frequency $\Delta\omega$. Separating eqn. (7) into the various harmonic components, a relation can be derived between $\Delta\omega$ and the difference $\delta\omega$ between the signal frequency from the free-running-oscillator centre frequency ω_0. As shown in the Appendix, one obtains an incomplete second approximation:

$$\delta\omega/K=R(\Delta\omega) \quad \text{.....(9)}$$

whereby

$$\begin{aligned}R(\Delta\omega)\equiv{}&\frac{\Delta\omega}{K}+\frac{1}{2}\cdot\frac{a_1K^2t_2^2}{Kt_2^3\Delta\omega^2+1}\times\\&\times\left\{t_1\Delta\omega\left(1+\frac{1}{t_1t_2\Delta\omega^2}\right)\cos\varepsilon_1-\right.\\&\left.-\left(1-\frac{t_1}{t_2}\right)\sin\varepsilon_1\right\}+\\&+(a_2-a_1)\frac{a_1^2t_1^2K^4}{16t_2^2\Delta\omega^4}\left[t_1\Delta\omega\cos\varepsilon_1-\sin\varepsilon_1\right]+\\&+\frac{a_1^3t_1^2K^4}{8t_2^2\Delta\omega^4}\left[t_1\Delta\omega\cos3\varepsilon_1-3\sin\varepsilon_1\right]-\\&-\frac{a_1^2a_2t_1^2K^4}{8t_2^2\Delta\omega^4}\left[t_1\Delta\omega\cos(2\varepsilon_1+\varepsilon_2)-\right.\\&\left.-(5/2)\sin(2\varepsilon_1+\varepsilon_2)\right]+0\left(\frac{1}{\Delta\omega^6}\right)\end{aligned}$$

.....(10a)

where a_1, ε_1 and a_2, ε_2 designate damping and phase shift of the ' additional filter ' at the frequencies $\Delta\omega$ and $2\Delta\omega$ respectively.

The accuracy of eqn. (10a) depends essentially on how fast the Fourier series converges and thus on the size of the frequency fluctuations during one beat period. As mentioned, experimentally these fluctuations are often small compared with $\Delta\omega$. A general evaluation cannot be given, but a good approximation of the neglected higher terms with $(1/\Delta\omega^6)$ is obtainable from the relative contributions of the terms with $(1/\Delta\omega^4)$. However, it should be noted that these last terms themselves are not fully complete in eqn. (10a), as mentioned in the Appendix, but they still give a useful approximation for most cases. The relative contributions of these terms can thus provide a self-consistency check.

In the region of the stable non-synchronized equilibria generally, the terms with $(1/\Delta\omega^4)$ are small compared with the lower terms. Thus for most cases, sufficient accuracy is obtained from the simplified function

$$R(\Delta\omega)=(\Delta\omega/K)+[t_1\Delta\omega\cos\varepsilon_1-\sin\varepsilon_1]\frac{a_1K}{2t_2\Delta\omega^2} \quad \text{.....(10b)}$$

which generally will be accurate within a few percent.

In the regions where the used approximations and basic assumptions are good, eqns. (9) and (10) make a connection between the frequency deviation $\delta\omega$ of the signal and the resulting asymptotic beat frequencies $\Delta\omega$ in a non-synchronized asymptotic state of the loop. In particular, from the discussion at the beginning of this section, it follows that the absolute minimum of the expressions (10a) or (10b) determines the minimal frequency deviation $\delta\omega$ of the signal for which a periodic non-synchronized state exists. Thus for the determination of the capture range of a given loop only the computation of the absolute minimum of (10a) or (10b) is required and the size of the correction terms with $(1/\Delta\omega^4)$ indicates the quality of the approximation.

The following section will present an illustration and a comparison of the theoretical predictions with experimental results for loops with a few selected ' additional filters '.

5. Application to Particular Networks

5.1. *Ideal Imperfectly Integrating Filter* [$T(p)\equiv1$]

This is the case extensively treated by Richman and Viterbi. Due to the close relationship of their principal ideas with our basic starting point, no essentially new results can be derived from eqns. (10) for this case. Thus in this section, only an illustration of our method and a comparison with these old results are given. The basic formula, eqn. (10b), simplifies to

$$R(\Delta\omega)=\Delta\omega/K+(t_1/2t_2)(K/\Delta\omega) \quad \text{.....(11)}$$

Here the second term on the right-hand side is the result of the ' velocity-dependent friction ' mentioned in Section 3. The general form of eqn. (11) is shown in Fig. 4. For each sufficiently large value of $\delta\omega/K$, two equilibrium states exist with the corresponding beat frequencies $\Delta\omega_1$ and $\Delta\omega_2$. By some arguments (also derivable from the pendulum analogue), it can be shown that only equilibrium states

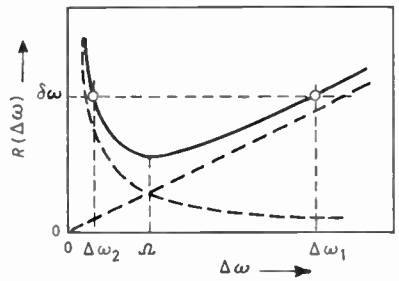


Fig. 4. Principal form of $R(\Delta\omega)$ for imperfect integrating filter.

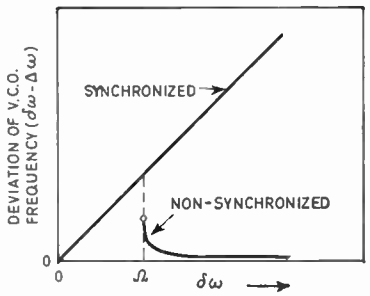


Fig. 5. Pulling of oscillator frequency by signal for the case of an imperfect integrator.

$\Delta\omega_1$ that correspond to a positive slope of the curve are stable, whereas the states $\Delta\omega_2$ with negative slope are always unstable. The states $\Delta\omega_1$ constitute the desired asymptotic states of the system for all initial conditions with $d\varphi/dt(t=0) \gtrsim \Delta\omega_1$, independently of the initial phase $\varphi(t=0)$.

As discussed in Section 4, the minimum of eqns. (10a) or (10b) defines the capture range Ω. Thus one obtains

$$\Omega = K(2t_1/t_2)^{\frac{1}{2}} \qquad \text{.....(12)}$$

The corresponding curve for Gruen's parameters is also shown in Fig. 3. As is to be expected, the results for large values of t_2 coincide with the results of Viterbi and Richman. For values of $\Omega \lesssim 0{\cdot}6K$, the results of eqn. (12) are in better agreement with the experimental results of Gruen than the older predictions and approach them within about 10%. This limit corresponds to the point where the absolute values of the higher terms of eqn. (10a) come into the order of magnitude of the lowest order terms.

From Fig. 4, it also can be seen that, for the stable non-synchronized equilibria, the beat frequencies $\Delta\omega_1$ are always smaller than the deviation $\delta\omega$ of the signal frequency. Thus even when no final synchronization takes place, the v.c.o. frequency is already pulled towards the signal frequency. This appearance becomes the stronger the closer the signal approaches the capture range. However, as indicated in Fig. 5, the asymptotic beat frequency goes steadily down only to $\Omega/2$ and then jumps to zero when the signal enters the capture range and does not approach zero steadily as Richman assumes in his Fig. 23(a). This is also very evident experimentally.

5.2. *Additional Independent RC-Integrator* $[T(p) = 1/(1+pt_3)]$

5.2.1. Theory

This case applies, for example, to the widespread custom of suppressing the r.f. output of the phase discriminator by means of a simple RC-integrator. Here the time-constant used is generally too small to affect the acquisition behaviour. But an additional integrator might also be used to improve the selectivity of the entire loop against signals just outside the normal frequency range. Also an additional term like this would occur for an i.f. tuned circuit when i.f. conversion is used in the loop. Generally, in all these cases, there is still some coupling to the imperfect integrator, but the idealization treated in this section is a good approximation.

Taking $T(p) = 1/(1+pt_3)$ we obtain

$$a(\omega) = (1+\omega^2 t_3^2)^{-\frac{1}{2}}$$
$$\cos\varepsilon(\omega) = (1+\omega^2 t_3^2)^{-\frac{1}{2}}$$
$$\sin\varepsilon(\omega) = \omega t_3/(1+\omega^2 t_3^2)^{\frac{1}{2}} \qquad \text{.....(13)}$$

Thus eqn. (10b) becomes

$$R(\Delta\omega) = \frac{\Delta\omega}{K} + \frac{t_1 - t_3}{2t_2} \times \frac{K}{\Delta\omega} \cdot \frac{1}{1+(\Delta\omega t_3)^2} \qquad \text{.....(14)}$$

From this, the general form of $R(\Delta\omega)$ shown in Fig. 6 follows. Here, Ω_0 designates the capture range obtained by the basic imperfectly integrating filter alone. The curve for $t_3 = 0$ applies to that same case. For $0 < t_3 \ll 1/\Omega_0$ the curves approach the asymptotic

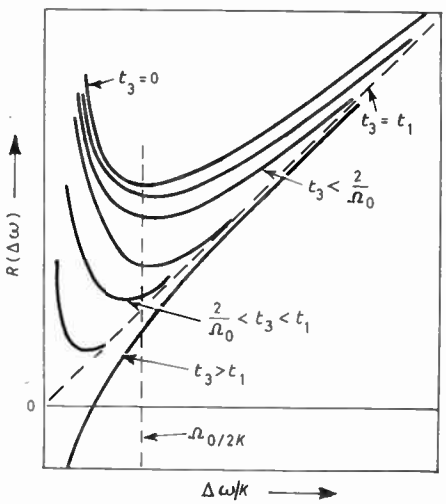


Fig. 6. Principal form of curves $R(\Delta\omega)$ for filter with imperfect integrator and additional integrator with time-constant t_3.

curve $R = \Delta\omega/K$ somewhat faster. In the region of interest around the minimum of the ($t_3 = 0$) curve, no essential changes occur. Thus only the pulling of the v.c.o. outside the capture range is decreased, but the capture range itself hardly changes. For t_3 of the order of $1/\Omega_0$, t_3 produces only some decrease in the factor multiplying $(K/\Delta\omega)$, but $1/[1+(\Delta\omega t_3)^2]$ may be treated as constant in the vicinity of the minimum. Thus one obtains, to a good approximation,

$$\frac{\Omega}{K} = \frac{\Omega_0}{K}\left(\frac{1-t_3/t_1}{1+(\Omega_0 t_3)^2}\right)^{\frac{1}{2}} \quad \text{for} \quad t_3 \lesssim \Omega_0 \quad \ldots\ldots(15)$$

In the indicated range, the accuracy of this formula should be better than a few percent, with the exact values of Ω being somewhat higher.

For $2/\Omega > t_3 > t_1$, the minimum of eqn. (14) must be computed. In Fig. 6, this case corresponds to the lower, but still concave, curves. Here the capture range and thus the frequencies important for the acquisition process, may become rather small. Therefore, the magnitude of the higher terms in eqn. (10a) may have to be investigated. For $t_3 = t_1$, the filter degenerates to a simple RC-integrator, in which case eqn. (14) fails to determine the capture range.

For $t_3 > t_1$, the synchronized state of the loop is unstable. The second term in eqn. (14) then becomes negative, and one obtains the convex curves of Fig. 6. Here in all cases $\delta\omega > \Delta\omega$, i.e., in the unsynchronized state, the v.c.o. is pushed away from the signal. Furthermore, eqn. (14) does not show a minimum. This means that no synchronization is predicted even for negative $\delta\omega/K$. If one moves the signal frequency from one side across the centre frequency ω_0 of the v.c.o., the actual v.c.o. frequency is pushed by the signal until the beat frequency becomes so small that the theory no longer applies.

5.2.2. Experimental test of the predictions

To test the predictions of eqn. (14), some careful measurements were carried out on a slightly modified Rohde and Schwarz Sideband Selector NZ1. The filter used is shown as an inset in Fig. 7. Mainly two parameter combinations were investigated:

(1) $K = 1{\cdot}5$ kHz $= 0{\cdot}95 \times 10^4$ rads^{-1}, $R_1 = 4$ kΩ, $R_2 = 220$ kΩ, $C_1 = 4$ μF;
i.e. $t_1 = 16$ ms, $t_2 = 0{\cdot}88$ s, and

(2) K, R_1, R_2 as above, $C_1 = 8$ μF;
i.e. $t_1 = 32$ ms, $t_2 = 1{\cdot}76$ s.

In both cases, C_2 was varied to change t_3. The d.c. attenuation by the divider 70 kΩ/10 kΩ is already included in the given value of K. In Fig. 7, the experimental results are confronted with the predictions from eqn. (14). Agreement is found generally within a few percent, which corresponds to the experimental accuracy.

The disappearance of the capture range for time-constants $t_3 \simeq t_1$ was well verified. As theoretically predicted, in this region very clearly a transition from pulling to pushing of the v.c.o. by the signal was evident. By careful operation it was possible, in this case, to increase the signal frequency from below the centre frequency ω_0 of the v.c.o. to above it without synchronization occurring. At the same time the v.c.o. was being pushed to an even higher frequency. Then after a further increase of the signal frequency the v.c.o. either became unstably synchronized or the v.c.o. frequency jumped into another non-synchronized state with a frequency below ω_0. In the last case the beat frequency was the same as if the signal had adopted the same frequency coming from the high-frequency side. The corresponding phenomena occurred when the signal frequency was lowered from above ω_0.

The explanation for these findings is evident from Fig. 8 which shows a complete plot of eqn. (14). The observed jump of the oscillator frequency corresponds to a jump between the unsynchronized states 1 and 2. Although the circumstances are somewhat different, this occurrence is similar to the ' swing over ' shown

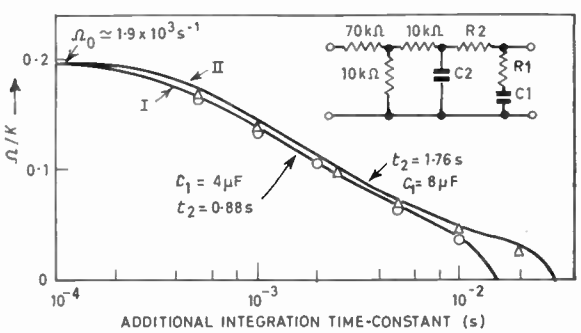


Fig. 7. Experimental and theoretical dependence of capture range on additional time-constant t_3.
Experimental points: case I, ○—○; case II, △—△.

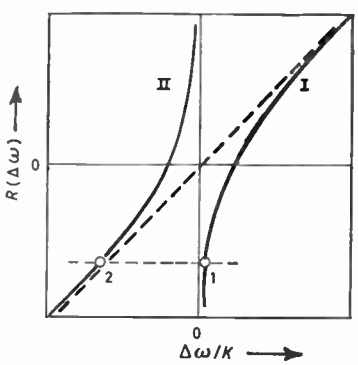


Fig. 8. Complete plot of $R(\Delta\omega)$ for $t_3 > t_1$.

by Viterbi's phase-plot (Fig. 11 of Ref. 7) where the oscillator frequency also runs across the centre frequency ω_0.

5.3. *Additional Delay Line* $T(p) = \exp(jp\tau)$†

For many low-pass filters, the transfer characteristics for frequencies smaller than the cut-off frequency are very similar to that of an ideal delay line. For our purpose, this applies mainly to the use of a series of decoupled RC-integrators or, correspondingly, a series of i.f. tuned circuits in the filter loop, when the total cut-off frequency is larger than the inverse of the resulting total low-frequency time delay. In view of these applications, a discussion of an a.p.c. loop with an ideal additional delay line in the feedback loop is given in the following.

In this case

$$a(\omega) = 1 \quad \text{and} \quad \varepsilon(\omega) = \tau\omega$$

where τ designates the corresponding delay time. Thus the function (10b) becomes

$$R(\Delta\omega) = \frac{\Delta\omega}{K} + (t_1 \Delta\omega \cos\tau\Delta\omega - \sin\tau\Delta\omega)\frac{K}{2\Delta\omega^2 t_2}$$

$$= \frac{\Delta\omega}{K} + (\cos\tau\Delta\omega) - \frac{\tau}{t_1}\cdot\frac{\sin\tau\Delta\omega}{\tau\Delta\omega}\cdot\frac{\Omega_0^2}{4K^2}\cdot\frac{K}{\Delta\omega}$$

where Ω_0 again means the capture range for $\tau = 0$. Also for this case no general analytic results can be derived for the capture range. Therefore our discussion will be limited to general features of the acquisition behaviour of these loops. For this purpose, it will be assumed for the moment that $\tau \ll t_1$, so that the second term in the bracket of eqn. (14) becomes negligible. Then eqn. (14) simplifies to

$$R(\Delta\omega) = \frac{\Delta\omega}{K} + \frac{\Omega_0^2}{4K^2}\cdot\frac{K}{\Delta\omega}\cos\tau\Delta\omega \quad \text{.....(16)}$$

The general form of (16) is shown in Fig. 9 for various delay times τ. In this figure, curve I corresponds to $\tau = 0$, i.e. to the curve of Fig. 4. As shown in curve II, for $\Omega_0\tau > 1$ the delay line introduces changes primarily in the asymptotic behaviour for $\Delta\omega > \Omega_0/2$. For increasing $\Delta\omega$, instead of a monotonic approach to the line $R \equiv \Delta\omega/K$, oscillations with decreasing amplitude are found around it. Thus when the signal frequency approaches the v.c.o. centre-frequency, one obtains a periodic change of attraction and repulsion of the oscillator frequency, instead of the monotonically increasing attraction discussed in Section 5.1. In the vicinity of $\Delta\omega/K = (\Omega_0/K)$, curve II is not very different from curve I, i.e. the capture range is only slightly decreased by the delay line.

For larger τ, the decrease of the capture range becomes more pronounced. Furthermore, the attraction of the v.c.o. frequency at the limit of the capture range decreases and changes to repulsion for $\tau\Omega_0 \geqq (2\pi)^{\frac{1}{2}}$ (curve III). At this limit, one finds $\Omega \simeq 0{\cdot}65\Omega_0$. As shown in curves IV and V, this repulsion can increase to the point where one has to move the signal frequency into the v.c.o. centre-frequency, or even across it, to obtain synchronization. This occurrence can be observed experimentally. In the sense of the word defined earlier in this case, the capture range is zero, since due to the negative branch of $R(\Delta\omega)$ (not shown in Fig. 9), a non-synchronized state exists for all signal frequencies. However, in contrast to the case applying to Fig. 8, synchronization always occurs when the signal frequency is slowly moved over the last minimum of Fig. 9. Therefore, this effect is sometimes called ' negative capture range '.

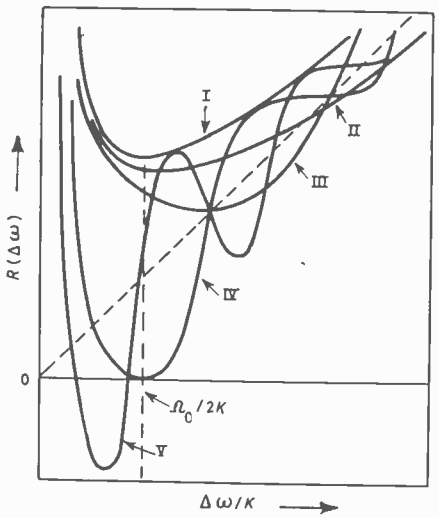


Fig. 9. Principal form of $R(\Delta\omega)$ for various additional delay times.

Furthermore, in this case, the oscillations of the function R are already large enough for secondary minima to occur. Experimentally, this results in a jumping of the oscillator and beat frequency when the signal frequency is continuously moved in the range of the oscillator centre-frequency, and may look like a sideband synchronization. This effect, too, is sometimes found experimentally.

In the above discussions the second term in the bracket of eqn. (15) has been neglected. For $\tau > t_1$, this term causes only an aggravation and slight shift of the effects described. For $\tau \gtrsim t_1$, similar to the effect described in Section 5.2, a change in sign of the bracket in eqn. (15) results for the limit $\Delta\omega \to 0$, i.e. the capture range vanishes. However, the approximations made in the derivation of (10a) are then no longer valid. On the other hand, the synchronized state is already unstable in this case, and thus the whole question is without interest.

† *Note added in proof:* Some of the results in Section 5.3 are similar to features described in the book 'Phase-lock Techniques', by F. M. Gardner (Wiley, New York, 1966).

For $\tau \lesssim 2/\Omega_0$ the capture range can be determined by a series expansion of the cosine and sine terms in (15). Computing the minimum in this approximation one obtains

$$\Omega = \Omega_0\left(1 - \frac{\Omega_0^2\tau^2}{8}\right)^{\frac{1}{2}}\left(1 - \frac{\tau}{t_1}\right)^{\frac{1}{2}} \quad \text{for} \quad \Omega_0\tau < 2 \qquad \text{.....(17)}$$

to an accuracy generally better than 5%. The decrease in capture range is about one-half the decrease in Section 5.2, for $\tau = t_3$. For $\tau\Omega_0 \lesssim 1$ the diminution of the capture range caused by the delay line is less than 10%.

6. Conclusions

In total, the method used, and in particular eqn. (10a) seem to provide a good way to determine the capture range of a wide variety of a.p.c. loops. The basic assumption of the existence only of singly periodic non-synchronized states seems at least to hold for all cases discussed in Section 5. However, we want to emphasize that this is also only additional experimental evidence. An investigation using the method of Benes[10] might show how far this assumption carries beyond that.

As indicated in Fig. 3, the approximations made in the derivation of eqn. (10a) will generally break down when the main integrating time-constant is too small or $\Omega_0 > \frac{1}{2}K$. A second limit for the validity of the approximations occurs when the capture range determined by (10a) comes close to the one defined by (4) when substituting t_2 for T, i.e. when the filter effectively becomes a pure RC-integrator. In Section 5 this limit was reached when $t_3 \simeq t_1$ and $\tau \simeq t_1$. Both these limits are built into the separation (6) of the filter function. Beyond these limits, phase-plane plots have to be used, and theoretical predictions become rather difficult. The same applies to a close investigation of the observed ' flip-over ' and of the ' negative capture range ' mentioned in Section 5.2 and 5.3, which both are, however, probably only of theoretical interest.

Regarding the ' pull-in time ' necessary for the transition to the synchronized state, no good estimates can be given at present. However, a derivation similar to the method used by Viterbi is being investigated and seems closely related to eqn. (10b).

7. Acknowledgments

We gratefully acknowledge very stimulating discussions with J. Hacks. We also want to thank S. Wagner for his helpful effort during the experiments and C. B. Wharton for his critical reading of the manuscript.

This work was supported in part by Rohde and Schwarz, Munich, Germany, and in small part by the joint General Atomic–Texas Atomic Energy Research Foundation programme on controlled thermonuclear reactions. Most of the work was carried out by the authors together when H. H. F. was affiliated to Rohde and Schwarz, Munich.

8. References

1. Goldstein, A. J., ' Analysis of the phase-controlled loop with a sawtooth comparator ', *Bell Syst. Tech. J.*, **41**, p. 603, 1962.

2. Byrne, C. J., ' Properties and design of the phase-controlled oscillator with a sawtooth comparator ', *Bell Syst. Tech. J.* **41**, p. 559, 1962.

3. Preston, G. W. and Tellier, J. C., ' The lock-in performance of an a.f.c. circuit ', *Proc. Inst. Radio Engrs*, **41**, p. 249, 1953.

4. Gruen, W., ' Theory of a.f.c. synchronization ', *Proc. I.R.E.*, **41**, p. 1043, 1953.

5. Jelonek, Z. J. and Cowan, C. I., ' Synchronized systems with time delay in the loop ', *Proc. Instn Elect. Engrs*, **104C**, p. 388, 1957. (I.E.E. Monograph No. 229R, March 1957.)

6. Richmann, D., ' Colour carrier reference phase synchronization accuracy in N.T.S.C. colour television ', *Proc. I.R.E.*, p. 106, **42**, 1954.

7. Viterbi, A. F., ' Acquisition and tracking behavior of phase-locked loops ', Symposium on Active Networks and Feedback Systems, 19th–21st April 1960.

8. Tausworthe, R. C., 'Theory and Practical Design of Phase-locked Receiver', JPL-Report 32–819, February 1960.

9. Rey, R. J., ' APC: theory and design ', *Proc. I.R.E.*, **48**, p. 1760, 1960.

10. Benes, V. E., ' Ultimately periodic solutions to a non-linear integro-differential equation ', *Bell Syst. Tech. J.*, **41**, p. 257, 1962.

9. Appendix

In the following, we want to show the derivations which led from (7) to (9), (10a) and (10b) in some more detail. For this purpose, we introduce the new terms and variable:

$$s = 1/(Kt_2)^{\frac{1}{2}}, \quad r = Kt_1/(Kt_2)^{\frac{1}{2}} \quad \text{and} \quad x = sKt$$

Thus eqn. (7) becomes

$$\frac{\mathrm{d}^2\varphi}{\mathrm{d}x^2} + s\frac{\mathrm{d}\varphi}{\mathrm{d}x} + rT'\frac{\mathrm{d}}{\mathrm{d}x}\sin\varphi + T'\sin\varphi = \frac{\delta\omega}{K} \qquad \text{.....(18)}$$

where, in operation form,

$$T'(\mathrm{d}/\mathrm{d}x) = T(sK \cdot \mathrm{d}/\mathrm{d}x)$$

Due to our hypothesis on the simple periodicity of the non-synchronized states, one can expand, after proper time normalization:

$$\varphi = px + \sum_1^\infty \alpha_m \sin mpx + \sum_1^\infty \beta_m \cos mpx \qquad \text{.....(19)}$$

Here, $p = \Delta\omega/sK$ is just the renormalized beat frequency. Abbreviating the two sums in (19) by S, one can then write

$$\sin\varphi = \sin px \cdot \cos S + \cos px \cdot \sin S$$

Then we assume $S \ll 1$ and neglect the higher than linear terms in the expansions of $\sin S$ and $\cos S$,

i.e. we put

$$\sin\varphi = \sin px + \cos px \cdot \sum_1^\infty \alpha_m \sin mpx + \cos px \cdot \sum_1^\infty \beta_m \cos mpx$$

This is multiplied by T' in the representation given in (8) and by

$$\mathrm{d}/\mathrm{d}x\, T' = T'\, \mathrm{d}/\mathrm{d}x$$

Abbreviating further

$$a_m = a(m\Delta\omega) \quad \text{and} \quad \varepsilon_m = \varepsilon(m\Delta\omega)$$

one obtains for the d.c. component of (18)

$$sp + a_0\beta_1/\beta_2 = \delta\omega/K \qquad \ldots\ldots(20)$$

Similarly the coefficients of $\sin px$ and $\cos px$ yield respectively

$$-p^2\alpha_1 - sp\beta_1 - rpa_1\left[\frac{\beta_2}{2}\cos\varepsilon_1 - \left(1+\frac{\alpha_2}{2}\right)\sin\varepsilon_1\right] + a_1\left[\left(1+\frac{\alpha_2}{2}\right)\cos\varepsilon_1 + \frac{\beta_2}{2}\sin\varepsilon_1\right] = 0 \qquad \ldots\ldots(21)$$

and

$$-p^2\beta_1 + sp\alpha_1 + rpa_1\left[\left(1+\frac{\alpha_2}{2}\right)\cos\varepsilon_1 + \frac{\beta_2}{2}\sin\varepsilon_1\right] + a_1\left[\frac{\beta_2}{2}\cos\varepsilon_1 - \left(1+\frac{\alpha_2}{2}\right)\sin\varepsilon_1\right] = 0 \qquad \ldots\ldots(22)$$

and the terms with $\sin mpx$ and $\cos mpx$ for $m \geqq 2$

$$0 = -m^2p^2\alpha_m - spm\beta_m - rmpa_m\left[\frac{\beta_{m-1}+\beta_{m+1}}{2}\cos\varepsilon_m - \frac{\alpha_{m-1}+\alpha_{m+1}}{2}\sin\varepsilon_m\right] + a_m\left[\frac{\alpha_{m-1}+\alpha_{m+1}}{2}\cos\varepsilon_m + \frac{\beta_{m-1}+\beta_{m+1}}{2}\sin\varepsilon_m\right] \qquad \ldots\ldots(23)$$

and

$$0 = -m^2p^2\alpha_m - spm\beta_m - rmpa_m\left[\frac{\beta_{m-1}+\beta_{m+1}}{2}\sin\varepsilon_m + \frac{\alpha_{m-1}+\alpha_{m+1}}{2}\cos\varepsilon_m\right] + a_m\left[\frac{\alpha_{m-1}+\alpha_{m+1}}{2}\sin\varepsilon_m - \frac{\beta_{m-1}+\beta_{m+1}}{2}\cos\varepsilon_m\right] \qquad \ldots\ldots(24)$$

In principle, from eqns. (21) and (22) the coefficients α_2 and β_2 can be computed for a given p as a function of α_1 and β_1 and of the various constants. Also, due to eqns. (23) and (24), the higher coefficients can be derived. Thus, neglecting eqn. (20), all coefficients α_m and β_m with $m \geqq 2$ are defined only as a function of α_1 and β_1 and of the systems constants s, r and $\delta\omega$. Therefore even after introducing eqn. (20), two variables, e.g. p and α_1, remain free. In principle we would have to check for which values of p and α_1 the resulting set of coefficients α_m and β_m would represent the coefficients of a converging Fourier series. However, due to the above assumption that $S \ll 1$, it follows from eqn. (18) that

$$|\alpha_{m+1}/\alpha_m| \simeq |\beta_{m+1}/\beta_m| \simeq (rp+1)/p^2$$

for large m. Thus a good approximation should generally be obtained by putting

$$\alpha_m = \beta_m = 0 \quad \text{for} \quad m \geqq N \qquad \ldots\ldots(25)$$

Also, from physical arguments good reasons for this truncation can be given.

Taking first $N = 2$ and eliminating α_1 and β_1 in eqns. (20–22), one obtains the wanted relation between $\delta\omega$ and p in the form

$$sp + \frac{a_0}{2}\cdot\frac{a_1}{p^2+s^2}\left[rp\left(1+\frac{s}{rp^2}\right)\cos\varepsilon_1 - (1-rs)\sin\varepsilon_1\right] \simeq sp + \frac{a_1 r}{2a_0 p}\cos\varepsilon_1 - \frac{a_1}{2a_0 p^2}\sin\varepsilon_1 \qquad \ldots\ldots(26)$$

Since $F(0) = 1$, we obtain $a_0 = 1$, and (26) becomes equal to (10b).

To obtain an estimate on the validity limitations of (26) one has to put $N = 3$. The derivations then become rather lengthy and cannot be given here. They result in the higher terms of (10a). In this case, however, the higher terms of the expansions of $\sin S$ and $\cos S$ also become important, and these correction terms are basically incomplete. However, in most instances, they should still give a good approximation of the necessary corrections.

A further criterion on the quality of the used approximations can be obtained from the magnitude of the first expansion coefficients, which to first order are

$$\alpha_1 = \frac{a_1}{p^2}(rp\sin\varepsilon_1 + \cos\varepsilon_1)$$

$$\beta_1 = \frac{a_1}{p^2}(rp\cos\varepsilon_1 - \sin\varepsilon_1)$$

and

$$\alpha_2 \simeq -\frac{a_1 a_2}{p^2} r^2 \cos(\varepsilon_1+\varepsilon_2)$$

$$\beta_2 \simeq \frac{a_1 a_2}{p^2} r^2 \sin(\varepsilon_1+\varepsilon_2)$$

Therefore the predictions of (10a) or (10b) should be rather good in cases where $r/p \gg 1$ in the vicinity of the minimum of (10b).

Manuscript first received 24th November 1967 and in final form on 16th April 1968. (Paper No. 1204/CC17.)

U.D.C. 621.372.826

Study of the *Q*-factor of a Surface-wave Resonator

By

K. P. ZACHARIA, Ph.D.†

AND

Professor
S. K. CHATTERJEE, M.Sc.‡

Summary: By applying field theory and transmission line theory, the *Q*-factor of a surface-wave resonator is derived. A new technique for determining the *Q*-factor is developed. Attenuation constants of Sommerfeld and Goubau surface-wave lines of different diameters are determined from a large number of measurements of *Q*-factor as a function of length of the resonator. The effect of tilt and rotation of one of the end-plates on the *Q*-factor is also studied. Radiation loss is evaluated. The paper also contains power-flow calculations and experimental verifications of the radial field spread for Sommerfeld and Goubau lines. Physical realizability of the Sommerfeld wave is confirmed.

1. Introduction

The investigations on electromagnetic wave propagation initiated by Sommerfeld[1-3] and Zenneck[4] and followed by Bouwkamp, Barlow and many others[5-33] have led to the modern concept of surface-waves and evolution of different types of surface-wave structures which can be used as waveguides or antennas depending on the nature of surface reactance. Chandler,[34] Scheibe, King and others[35-37] have contributed significantly to the development of surface-wave resonator technique. The present investigations on surface-wave resonator have been motivated with the object of: (i) studying the variation of the *Q*-factor with change of length of the resonator and with different surface wave lines such as bare copper wire and enamelled copper wire of different diameter; (ii) utilizing the above data to determine the attenuation constant; (iii) studying the effect of tilt of one of the end-plates on the *Q*-factor; (iv) calculating the percentage of surface-wave power which flows outside a radius equal to that of the resonator, when bare copper wires and enamelled copper wires of different diameters are used as surface-wave lines; (v) evaluating the radiation loss which occurs due to the finite size of the end-plates; (vi) estimating the minimum radius that the end-plates should have so as to make the radiation effects negligibly small; and (vii) determining the nature of variation of surface-wave field in the radial direction for different gauges of bare and enamelled copper wires.

2. Field Components

The solution of Maxwell's equations in cylindrical co-ordinates is expressed in terms of Bessel functions. For the case of surface-waves on cylindrical conductors, the boundary conditions require the use of Bessel functions of the 1st kind for the field components inside the conductor and the use of the Bessel functions of the 3rd kind, i.e. Hankel functions for the field components outside the conductor.

The mode of primary interest in cylindrical surface-wave transmission is the E_0-mode since all the other modes have very high attenuation. For bare conductor surface-wave lines, i.e. Sommerfeld lines having radius s (see Fig. 1) and immersed in air, the field components of the E_0-mode in cylindrical co-ordinate system (ρ, ϕ, z) are given by Barlow and Brown[13] as:

$$E_{z1} = B' \exp(-\gamma z) J_0(ju_1 \rho)$$

$$E_{\rho 1} = B' \frac{\gamma}{ju_1} \exp(-\gamma z) J_1(ju_1 \rho)$$

$$H_{\phi 1} = B' \left(\frac{\sigma_1 + j\omega\varepsilon_1}{ju_1}\right) \exp(-\gamma z) J_1(ju_1 \rho)$$

.....(1) when $\rho < s$

$$E_{z2} = B \exp(-\gamma z) H_0^{(1)}(ju_2 \rho)$$

$$E_{\rho 2} = B \frac{\gamma}{ju_2} \exp(-\gamma z) H_1^{(1)}(ju_2 \rho)$$

$$H_{\phi 2} = B \frac{\omega\varepsilon_0}{u_2} \exp(-\gamma z) H_1^{(1)}(ju_2 \rho)$$

.....(2) when $\rho > s$

where the symbols have the following meaning:

B' and B excitation constants, V/m

γ axial propagation constant $= \alpha + j\beta$

α attenuation constant, nepers/metre (N/m)

β phase constant, rad/m

u_1 radial propagation constant for the region inside the conductor $= a_1 + jb_1$

u_2 radial propagation constant for the region outside the conductor $= a_2 - jb_2$

ε_1 dielectric constant of the conductor, F/m

ε_0 dielectric constant of the region outside the conductor, i.e. air, F/m

ω angular frequency of the wave, rad/s.

† Formerly at the Indian Institute of Science, Bangalore; now at the Indian Institute of Technology, Bombay.

‡ Indian Institute of Science, Bangalore.

The time variation of each of these field components is according to the factor $\exp(j\omega t)$ and is omitted in eqns. (1 and 2) for the sake of convenience.

The surface-wave resonator is formed by terminating the two ends of a short length L of surface-wave line by conducting plates (see Fig. 1). The end-plates are assumed to be of infinite size in the

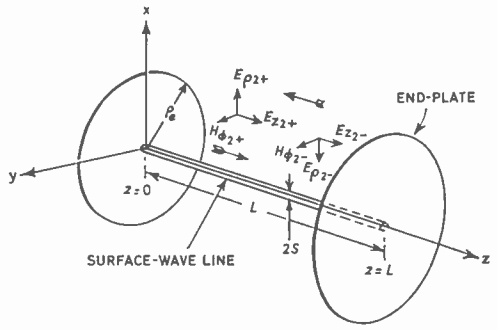


Fig. 1. Schematic diagram of the surface-wave resonator.

theory. Due to reflections taking place at the two end-plates, there will be waves travelling in the $+z$ and $-z$ directions inside the resonator. The standing waves that result due to the superposition of these waves are represented as follows:

$$\begin{aligned} E_{z2s} &= E_{z2+} + E_{z2-} \\ E_{\rho 2s} &= E_{\rho 2+} + E_{\rho 2-} \\ H_{\phi 2s} &= H_{\phi 2+} + H_{\phi 2-} \end{aligned} \qquad \ldots\ldots(3)$$

The subscript 's' represents the standing wave and the subscripts + and − indicate components of the wave travelling in the $+z$ and $-z$ directions respectively. Since the travelling waves are produced as a result of reflections at the end-plates, the following relations hold good.

$$\begin{aligned} E_{z2-} &= +E_{z2+} \\ E_{\rho 2-} &= -E_{\rho 2+} \\ H_{\phi 2-} &= +H_{\phi 2+} \end{aligned} \qquad \ldots\ldots(3a)$$

Using eqns. (2), (3) and (3a) and neglecting the attenuation term α in γ, since the length of surface wave line is small, the field components of the standing waves inside the resonator are

$$\begin{aligned} E_{z2s} &= 2BH_0^{(1)}(ju_2\rho)\cos\beta z \\ E_{\rho 2s} &= -j2B\frac{\beta}{u_2}H_1^{(1)}(ju_2\rho)\sin\beta z \\ H_{\phi 2s} &= 2B\frac{\omega\varepsilon_0}{u_2}H_1^{(1)}(ju_2\rho)\cos\beta z \end{aligned} \qquad \ldots\ldots(3b)$$

The field components of the standing wave given in eqn. (3b) satisfy the boundary conditions at the end-plates, i.e. $E_{\rho 2s} = 0$ at $z = 0$ and $z = L$ (see Fig. 1). For $E_{\rho 2s}$ to be zero at $z = 0$ and $z = L$, $\sin\beta L$ must be zero which requires that $\beta = n\pi/L$, where n is a positive integer indicating the number of half-cycle variations of the field components in the z direction. Substituting $\beta = n\pi/L$ in eqn. (3b) yields the following expressions for the field components of the resonant waves

$$\begin{aligned} E_{z2r} &= 2BH_0^{(1)}(ju_2\rho)\cos\left(\frac{n\pi z}{L}\right) \\ E_{\rho 2r} &= -j2B\frac{n\pi}{u_2 L}H_1^{(1)}(ju_2\rho)\sin\left(\frac{n\pi z}{L}\right) \\ H_{\phi 2r} &= 2B\frac{\omega\varepsilon_0}{u_2}H_1^{(1)}(ju_2\rho)\cos\left(\frac{n\pi z}{L}\right) \end{aligned} \qquad \ldots\ldots(4)$$

When the fields inside the surface-wave resonator are completely described by eqn. (4), the mode of oscillation is pure and is designated as the E_{0n} mode. If the frequency of excitation or the distance between the end-plates is so adjusted that eqns. (4) are satisfied, the cavity is said to be in resonance.

3. Q of the Resonator

3.1. *Field Theory Approach*

The quality factor, Q, of a resonator is defined as

$$Q = \omega\frac{W_E}{P} \quad \text{or} \quad \frac{W_M}{P}\omega \qquad \ldots\ldots(5)$$

where ω is the angular frequency at resonance, W_E and W_M represent the maximum energy stored in the electric and magnetic fields respectively inside the resonator, and P is the total power loss inside the resonator.

The maximum energy W stored in a magnetic field is given by the expression

$$W = \frac{\mu}{2}\oint_V |H_{max}|^2\,dV$$

where H_{max} is the maximum value of the magnetic field intensity, V is the volume of the resonator and μ is the permeability of the medium inside the volume considered. The maximum energy W_M stored inside the resonator is, therefore, given by

$$W_M = \frac{\mu_0}{2}\int_{\rho=s}^{\infty}\int_{\phi=0}^{2\pi}\int_{z=0}^{L}|H_{max}|^2\rho\,d\rho\,d\phi\,dz \qquad \ldots(5a)$$

where μ_0 is the permeability of air and $\rho\,d\rho\,d\phi\,dz$ represent an element of volume. Since the magnetic field possesses the azimuthal component only

$$|H_{max}|^2 = |H_{\phi 2r}|^2 \qquad \ldots\ldots(5b)$$

Using the relation

$$|H_{\phi 2r}|^2 = H_{\phi 2r}H_{\phi 2r}^* \qquad \ldots\ldots(5c)$$

where the asterisk indicates the complex conjugate, and evaluating the integral in eqn. (5a) as shown in Appendix 1, we get from eqns. (4), (5a), (5b) and (5c)

$$W_M = -\frac{2\pi\mu_0 B^2\omega^2\varepsilon_0^2 L}{u_2 u_2^*(u_2^{*2}-u_2^2)} \times \\ \times[-ju_2^* sH_1^{(1)}(ju_2 s)H_0^{(2)}(-ju_2^* s) - \\ -ju_2 sH_0^{(1)}(ju_2 s)H_1^{(2)}(-ju_2^* s)] \quad ...(6)$$

The total power loss P is the sum of the power loss P_e in the two end-plates and the power loss P_w in the surface-wave line. The power-flow P into a conductor is given by the relation.

$$p = \frac{1}{2\sqrt{2}}\sqrt{\frac{\omega\mu}{\sigma}}\int_S |H_{tan}|^2\,ds$$

where H_{tan} is the magnetic field intensity at the surface of the conductor, S represents an area normal to the direction of power-flow, ω is the angular frequency of the wave, and σ and μ are respectively the conductivity and permeability of the conducting medium. The power-flow into the two end-plates and the surface-wave line are, therefore, given respectively by the following expressions

$$P_e = 2\times\frac{1}{2\sqrt{2}}\sqrt{\frac{\omega\mu_e}{\sigma_e}}\int_{\rho=s}^{\infty}\int_{\phi=0}^{2\pi}|H_{tan}|_e^2\,\rho\,d\rho\,d\phi \\ P_w = \frac{1}{2\sqrt{2}}\sqrt{\frac{\omega\mu_1}{\sigma_1}}\int_{\phi=0}^{2\pi}\int_{z=0}^{L}|H_{tan}|_w^2\,s\,d\phi\,dz \quad ...(7)$$

where σ_1 = conductivity of the surface-wave line, $\Omega^{-1}\,m^{-1}$

σ_e = conductivity of the end-plates, $\Omega^{-1}\,m^{-1}$

μ_1 = permeability of the surface-wave line, H/m

and μ_e = permeability of the end-plates, H/m.

$\rho\,d\rho\,d\phi$ and $s\,d\phi\,dz$ represent elements of areas on the surface of the end-plates and the wire respectively. The tangential component of the magnetic field possesses the azimuthal component only and so

$$|H_{tan}|_e^2 = |H_{tan}|_w^2 = |H_{\phi 2r}|^2 \quad .....(7a)$$

Evaluating the integrals as shown in Appendix 1, we obtain from eqns. (4), (5c), (7) and (7a)

$$P_e = -\sqrt{\frac{\omega\mu_e}{2\sigma_e}}\frac{8\pi B^2\omega^2\varepsilon_0^2}{u_2 u_2^*(u_2^{*2}-u_2^2)} \times \\ \times[-ju_2^* sH_1^{(1)}(ju_2 s)H_0^{(2)}(-ju_2^* s) - \\ -ju_2 sH_0^{(1)}(ju_2 s)H_1^{(2)}(-ju_2^* s)]$$

$$P_w = \sqrt{\frac{\omega\mu_1}{2\sigma_1}}\frac{2\pi B^2\omega^2\varepsilon_0^2 Ls}{u_2 u_2^*}H_1^{(1)}(ju_2 s)H_1^{(2)}(-ju_2^* s) \quad ...(8)$$

Substituting eqns. (6), (8) in eqn. (5), an expression based on field theory for the Q of the resonator containing Sommerfeld line is obtained:

$$Q = \frac{\sqrt{\omega\mu_0}[A]}{\frac{4}{L\sqrt{2\sigma_e}}[A] - \frac{s(u_2^{*2}-u_2^2)}{\sqrt{2\sigma_1}}H_1^{(1)}(ju_2 s)H_1^{(2)}(-ju_2^* s)} \quad .....(9)$$

where

$$[A] = -ju_2^* sH_1^{(1)}(ju_2 s)H_0^{(2)}(-ju_2^* s) - \\ -ju_2 sH_0^{(1)}(ju_2 s)H_1^{(2)}(-ju_2^* s) \quad .....(9a)$$

In most of the practical cases, the magnitude of arguments of the Hankel functions in eqn. (9) is less than 0·05. Therefore, using the small argument approximations (see Appendix 2), the expression for Q (eqn. 9) reduces to

$$Q = \frac{\sqrt{\omega\mu_0\sigma_1\sigma_e}}{\frac{2\sqrt{2\sigma_1}}{L} - \frac{a_2 b_2\sqrt{2\sigma_e}}{s[n(a_2^2-b_2^2)+2a_2 b_2 m]}} \quad ...(10)$$

The values of a_2 and b_2 are obtained from the solution of the following characteristic equation[13]

$$\frac{u_2}{\omega\varepsilon_0}\frac{H_0^{(1)}(ju_2 s)}{H_1^{(1)}(ju_2 s)} = \frac{ju_1}{\sigma_1 + j\omega\varepsilon_1}\frac{J_0(ju_1 s)}{J_1(ju_1 s)} \quad .....(11)$$

This is solved by using small argument approximation $|ju_2 s| \ll 1$ for the Hankel functions and large argument approximation $|ju_1 s| \gg 1$ for the Bessel functions when the following values for a_2 and b_2 are obtained.

$$a_2 = \frac{1{\cdot}123|\xi|^{1/2}}{s}\cos\frac{\delta}{2} \quad .....(12)$$

$$b_2 = \frac{1{\cdot}123|\xi|^{1/2}}{s}\sin\frac{\delta}{2} \quad .....(13)$$

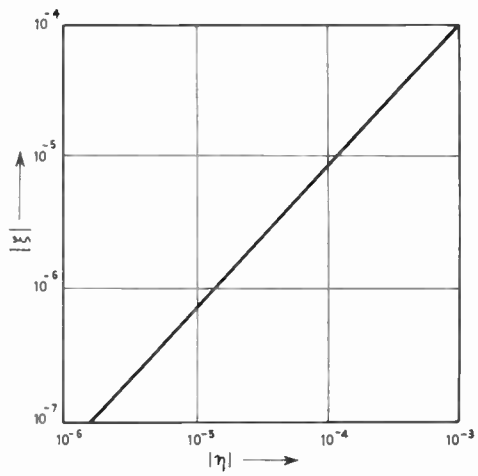


Fig. 2. Relation between $|\xi|$ and $|\eta|$ according to eqn. (15).

where

$$\delta = -\frac{\pi}{4}\left(1 - \frac{1}{\ln|\xi|+1}\right) \qquad \text{.....(14)}$$

$$-|\eta| = |\xi| \ln|\xi| \qquad \text{.....(15)}$$

and

$$|\eta| = \frac{24{\cdot}6 \times 10^{-14}}{\sqrt{\sigma_1}} f^{3/2} s \qquad \text{.....(16)}$$

The relation between $|\eta|$ and $|\xi|$ is plotted graphically in Fig. 2 for the convenience of computation.

3.2. *Transmission Line Theory Approach*

By adopting the same analytical approach as outlined by Barlow and Cullen,[38] the surface-wave resonator with its exciting and receiving loops (see Fig. 3) is reduced to an equivalent circuit (see Fig. 4) consisting of lumped constants Z_m, the impedance offered by each of the end-plates to the surface wave, a transmission line of length L and the coupling of the sending end and receiving end loops with the surface wave line represented respectively by M_s and M_r. Z_0 is the wave impedance of the surface-wave and $\gamma = \alpha + j\beta$ is the axial propagation constant.

Assuming M_s and M_r to be small

$$\begin{aligned} Z_s &= \frac{Z_m + Z_0 \tanh \gamma L}{1 + \frac{Z_m}{Z_0} \tanh \gamma L} \\ &\cong Z_m + Z_0 \tanh \gamma L \\ &= R_m + jR_m + Z_0 \tanh \gamma L \end{aligned} \qquad \text{.....(17)}$$

since

$$Z_m/Z_0 \ll 1$$

and

$$Z_m = R_m + jR_m = \sqrt{\frac{j\omega\mu_e}{\sigma_e}} \qquad \text{.....(18)}$$

where

$$R_m = \sqrt{\frac{\omega\mu_e}{2\sigma_e}} \qquad \text{.....(19)}$$

The voltage developed at the receiving loop is

$$e_r = iZ(\omega) \qquad \text{.....(20)}$$

where

$$Z(\omega) = \omega^2 M_s M_r / [2R_m + j2R_m + Z_0 \tanh(\alpha L + j\beta L)]$$

which is easily derived from the relations

$$e_s = -j\omega M_s i \qquad \text{.....(21)}$$

$$e_r = -j\omega M_r i_r \qquad \text{.....(22)}$$

$$i_r = e_s/[Z_m + Z_s] \qquad \text{.....(23)}$$

At resonance, $\omega = \omega_0$

$$2R_m + I_m[Z_0 \tanh(\alpha L + j\beta L)] = 0 \qquad \text{.....(24)}$$

which yields the condition for resonance as

$$2R_m + Z_0 \eta_0 = 0 \qquad \text{.....(25)}$$

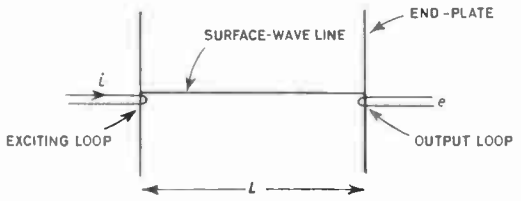


Fig. 3. Schematic representation of the surface-wave resonator.

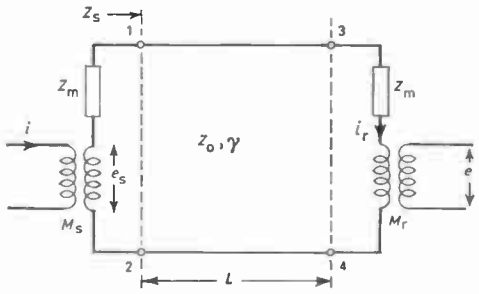


Fig. 4. Equivalent circuit representation of the resonator.

since at resonance

$$\beta L = n\pi + \eta_0 \qquad \text{.....(26)}$$

and αL, and βL being small

$$\tanh(\alpha L + j\beta L) \simeq \alpha L + j\eta_0 \qquad \text{.....(27)}$$

n is an integer and η_0 is small so that it satisfies eqn. (24). Hence, at resonance ($\omega = \omega_0$), $Z(\omega)$ reduces to

$$Z(\omega_0) = \frac{\omega_0^2 M_s M_r}{2R_m + Z_0 \alpha L} \qquad \text{.....(28)}$$

When ω changes from ω_0 by $\delta\omega$ the value of η will change by $\delta\eta$ from η_0 so that

$$Z(\omega_0 + \delta\omega) = \frac{\omega_0^2 M_s M_r}{2R_m + Z_0 \alpha L + jZ_0 \delta\eta} \qquad \text{.....(29)}$$

if

$$Z_0 \delta\eta = 2R_m + Z_0 \alpha L \qquad \text{.....(30)}$$

the response of the circuit will reduce to 0·707 of its maximum value. The half-power bandwidth is therefore

$$\Delta f_0 = 2\delta f = (v/\pi Z_0 L)(2R_m + Z_0 \alpha L) \qquad \text{.....(31)}$$

since

$$\delta\eta = \delta\beta L = \frac{2\pi L \delta f}{v} \qquad \text{.....(32)}$$

where v is the velocity of surface-wave propagation. Therefore,

$$\begin{aligned} Q &= \pi f_0 L Z_0 / [v(2R_m + Z_0 \alpha L)] \\ &= \frac{\pi Z_0}{\frac{2\lambda_g R_m}{L} + \lambda_g Z_0 \alpha} \end{aligned} \qquad \text{.....(33)}$$

where λ_g is the guide wavelength.

4. Attenuation Constants

4.1. *Sommerfeld Line*

The attenuation constant of Sommerfeld line is calculated from the relation[13]

$$\alpha = \frac{c}{\omega}(a_2 b_2) \qquad \text{.....(34)}$$

which is derived from

$$\gamma = \alpha + j\beta$$
$$u_2 = a_2 - jb_2$$
$$k_2^2 = \omega^2 \mu_0 \varepsilon_0 = -(\gamma^2 + u_2^2)$$

by assuming that in the microwave range,

$$k_2^2 \gg (a_2^2 - b_2^2)$$

and

$$a_2 \gg b_2$$

4.2. *Goubau Line*

The loss in a Sommerfeld line is the ohmic loss the wave suffers due to the finite conductivity of the wire, whereas, in a Goubau line, the loss is not only due to the finite conductivity of the line but also due to the dielectric loss of the coating. So the attenuation constant of Goubau line is $\alpha = \alpha_c + \alpha_i$, where α_c and α_i are the attenuation constants due to the above two causes respectively.[18]

$$\alpha_c \simeq \frac{1}{2a}\sqrt{\frac{\omega\varepsilon\mu_c}{2\mu\sigma_c}}\,\frac{1}{\ln \gamma' a' + 0{\cdot}38} \text{ N/m} \qquad \text{.....(35)}$$

$$\alpha_i \simeq \frac{\gamma'^2}{2K}\,\frac{\varepsilon}{\varepsilon_i - \varepsilon}\left(1 - \frac{0{\cdot}5}{\ln \gamma' a' + 0{\cdot}38}\right)\tan\delta \text{ N/m} \qquad \text{.....(36)}$$

The radial propagation factor is obtained from

$$G(\gamma' a') = -\left(\frac{\gamma' a'}{2\pi}\right)^2 \ln(0{\cdot}89\gamma' a') \qquad \text{.....(37)}$$

where

$$G(\gamma' a') = \frac{\ln(a'/a)}{\left(\frac{\varepsilon_i}{\varepsilon_i - \varepsilon}\right)\left(\frac{\lambda}{a'}\right)^2} \qquad \text{.....(38)}$$

in which $a' - a$ = thickness of the dielectric coating and ε_i is the permittivity of the dielectric coating. The function $G(\gamma' a')$ is represented graphically in Fig. 5.

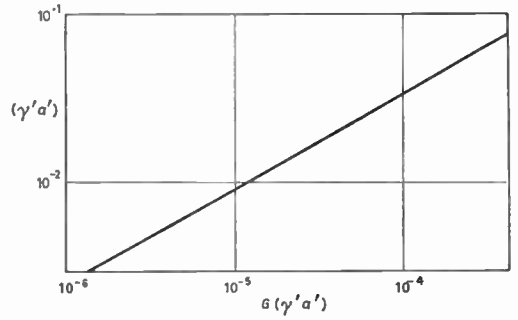


Fig. 5. Relation between $(\gamma' a')$ and $G(\gamma' a')$ according to eqn. (37).

5. Power-flow

In calculating the value of Q, the power lost has been assumed to be only due to the loss at the surface-wave line and loss at the end-plates. The radiation loss which may arise due to the finite size of the end-plates has not been taken into consideration. An estimate of the effectiveness of the end-plates as terminations for the surface-wave line is obtained from the power flowing outside a radius ρ_e corresponding to that of the end-plates.

5.1. *Sommerfeld Line*

The total power-flow N outside the surface-wave line is

$$N = \tfrac{1}{2}\mathrm{Re}\int_{\phi=0}^{2\pi}\int_{\rho=s}^{\infty} E_{\rho 2} H_{\phi 2}^{*} \rho \, d\rho \, d\phi \qquad \text{.....(39)}$$

which reduces to (see Appendix 3)

$$N = \mathrm{Re}\left\{\frac{\pi\omega\varepsilon_0 \gamma B^2}{ju_2 u_2^*(u_2^{*2} - u_2^2)}\left[ju_2^* s H_1^{(1)}(ju_2 s) H_0^{(2)}(-ju_2^* s) + ju_2 s H_0^{(1)}(ju_2 s) H_1^{(2)}(-ju_2^* s)\right]\right\}$$

$$\simeq \frac{2\omega\varepsilon_0 B^2 \beta}{\pi a_2 b_2 (a_2^2 + b_2^2)^2}\left[(b_2^2 - a_2^2)n - 2a_2 b_2 m\right] \qquad \text{.....(40)}$$

using small argument approximation (see Appendix 1) for the Hankel functions.

The energy-flow outside a radius ρ_e is

$$N_{\rho_e} = \tfrac{1}{2}\mathrm{Re}\int_{\phi=0}^{2\pi}\int_{\rho=\rho_e}^{\infty} E_{\rho 2} H_{\phi 2}^{*} \rho \, d\rho \, d\phi$$

$$= \mathrm{Re}\left\{\frac{\pi\omega\varepsilon_0 \gamma B^2}{ju_2 u_2^*(u_2^{*2} - u_2^2)}\left[ju_2^* \rho_e H_1^{(1)}(ju_2 \rho_e) H_0^{(2)}(-ju_2^* \rho_e) + ju_2 \rho_e H_0^{(1)}(ju_2 \rho_e) H_1^{(2)}(-ju_2^* \rho_e)\right]\right\} \qquad \text{...(41)}$$

The percentage of power-flow outside a radius ρ_e is

$$(N_{\rho_e}/N)\times 100 = 100\,\frac{\pi^2(a_2^2+b_2^2)\rho_e}{8\beta}\times\frac{\mathrm{Re}\,\{\gamma[ju_2^*H_1^{(1)}(ju_2\rho_e)H_0^{(2)}(-ju_2^*\rho_e)+ju_2H_0^{(1)}(ju_2\rho_e)H_1^{(2)}(-ju_2^*\rho_e)]\}}{[n(a_2^2-b_2^2)+2a_2b_2m]} \quad \ldots\ldots(42)$$

5.2. *Goubau Line*

Goubau[18] calculated the percentage power-flow outside the radius ρ_e ($a' < \rho_e < \infty$) by first calculating

$$N_{\rho_e} = \tfrac{1}{2}\mathrm{Re}\int_{\phi=0}^{2\pi}\int_{\rho_e}^{\infty}\rho E_{\rho 2}H_{\phi 2}^*\,d\rho\,d\phi$$

by using appropriate field components. The total energy-flow N outside the wire is obtained by extending the lower limit of ρ in the integral to the wire surface. The expression for the percentage power-flow $N_{\rho_e}/N\times 100$ is given by

$$\frac{N_{\rho_e}}{\rho}\times 100 = -\frac{2\pi^2\gamma'^2F(\gamma'\rho_e)}{\ln\gamma'a'+0{\cdot}38}\times 100 \quad \ldots\ldots(43)$$

where

$$F(\gamma'\rho_e) = (\gamma'\rho_e)^2\left\{-\frac{2}{\gamma'\rho_e}jH_0^{(1)}(j\gamma'\rho_e)H_1^{(1)}(j\gamma'\rho_e) - [H_0^{(1)}(j\gamma'\rho_e)]^2-[H_1^{(1)}(j\gamma'\rho_e)]^2\right\} \quad \ldots(44)$$

in which a' and γ' are the radius of the outer surface of the dielectric coating, and radial propagation constant respectively.

6. Numerical Calculations

6.1. *Q-factor of the Resonator with Sommerfeld Line*

The value for Q of the resonator containing bare copper Sommerfeld line has been calculated by two methods, namely, field theory and transmission line theory.

6.1.1. Field theory

The Sommerfeld line under consideration being bare copper wire and the end-plates being aluminium, the following values are used for the various parameters in eqn. (10):

$\omega = 2\pi\times 9{\cdot}5\times 10^9$ rad/s

$\mu_0 = 1{\cdot}2\,6\times 10^{-6}$ H/m

$\sigma_1 = 5{\cdot}8\times 10^7\ \Omega^{-1}\,\mathrm{m}^{-1}$

$\sigma_e = \sigma$ (aluminium) $= 3{\cdot}54\times 10^7\ \Omega^{-1}\,\mathrm{m}^{-1}$

Using these values, the expression for Q (eqn. 10) obtained by field theory approach reduces to

$$Q = \frac{10^6}{\dfrac{1{\cdot}74}{L}+K} \quad \ldots\ldots(45)$$

where

$$K = -\frac{0{\cdot}678a_2b_2}{s[n(a_2^2-b_2^2)+2a_2b_2m]} \quad \ldots\ldots(46)$$

The values of m, n, a_2 and b_2 are given in Appendix 2 and eqns. (12) and (13). The values of a_2 and b_2 have been calculated for different values of the radius s of the surface-wave line from $0{\cdot}5\times 10^{-4}$ metres to 15×10^{-4} metres and conductivity ranging from $0{\cdot}1\times 10^7$ to $7\times 10^7\ \Omega^{-1}\,\mathrm{m}^{-1}$. By substituting the values of a_2 and b_2 for the case of copper ($\sigma_1 = 5{\cdot}8\times 10^7\ \Omega^{-1}\,\mathrm{m}^{-1}$) the Q of the surface-wave resonator containing different lengths of bare copper Sommerfeld line has been calculated. Some of the values are shown in Table 1.

Table 1

Values of Q of the surface-wave resonator with bare copper wire

$f = 9500$ MHz; $\sigma_1 = 5{\cdot}8\times 10^7\ \Omega^{-1}\,\mathrm{m}^{-1}$; L is the length of the resonator in metres

$s\times 10^4$ metres	K	10150α	Q					
			$L = 0{\cdot}1$		$L = 0{\cdot}25$		$L = 0{\cdot}75$	
			Field theory	Transmission line theory	Field theory	Transmission line theory	Field theory	Transmission line theory
0·5	782	862	1250	1140	1270	1150	1280	1160
3	167	168	5440	5380	5750	5710	5920	5880
6	88·3	88·3	9430	9430	10460	10490	11040	11040
9	61·0	60·7	12760	12770	14710	14770	15800	15870
11	50·8	50·7	14660	14640	17300	17510	18830	18870

6.1.2. Transmission line theory

The Q of the resonator has also been calculated according to eqn. (33) derived on the basis of transmission line theory. For the case of bare copper surface-wave line and aluminium end-plates, the following numerical values hold good for various parameters in eqn. (33).

$$\omega = 2\pi \times 9{\cdot}5 \times 10^9 \text{ rad/s}$$

$$R_m = \sqrt{\frac{\omega\mu_0}{2\sigma_e}} = 3{\cdot}26 \times 10^{-2} \text{ ohms}$$

$$\varepsilon_0 = 8{\cdot}85 \times 10^{-12} \text{ F/m}$$

$$\mu_0 = 1{\cdot}26 \times 10^{-6} \text{ H/m}$$

$$\sigma_e = 3{\cdot}54 \times 10^7 \ \Omega^{-1}\,\text{m}^{-1}$$

$$\lambda_g \simeq 3{\cdot}14 \times 10^{-2} \text{ m}$$

$$Z_0 = E_{\rho 2}/H_{\phi 2} = \frac{2\pi}{\lambda_g \omega \varepsilon_0} \simeq 374 \text{ ohms}$$

Using these values eqn. (33) reduces to

$$Q = \frac{10^6}{\frac{1{\cdot}76}{L} + 10150\alpha} \qquad \text{.....(47)}$$

The values of α (Section 6.3) are obtained from eqn. (48). The values of Q calculated for some values of L and s are given in Table 1 for the sake of comparison. It can be seen that there is a slight difference between the values of Q obtained by the two different approaches. This divergence is due to the difference in the numerical values of the terms containing K and 10150α occurring in these equations. The difference is only of the order of the error involved in calculation. Figure 6 shows the mean value of K and 10150α plotted as a function of the radius of the wire.

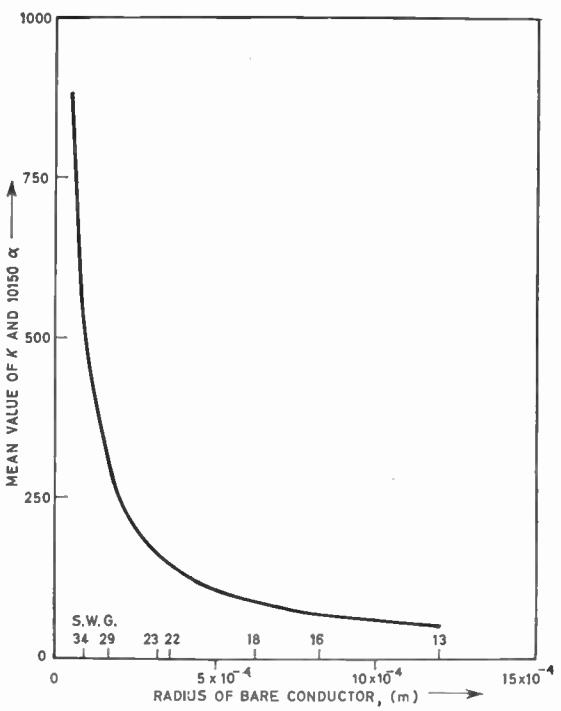


Fig. 6. Mean value of the terms K and 10150α as a function of radius of bare conductor.

6.2. *Q-factor of the Resonator with Goubau Line*

The value of Q for the surface-wave resonator with Goubau line has been calculated from eqn. (47), and using the value of $\alpha = \alpha_c + \alpha_i$ as given by the expressions (35 and 36) for different gauges of Paramex enamelled copper winding wires† Type-M (medium covering) and Type-X (extra thick covering). The values of α are calculated with the help of eqns. (4), (9), (50) (51) and Fig. 5. Some of the results are given in Table 2.

Table 2

Q values of the surface wave resonator with 'Paramex' enamelled copper wires

$f = 9500$ MHz; L is the length of the resonator in metres

Gauge of wire s.w.g.	10150α	Q		
		$L = 0{\cdot}1$	$L = 0{\cdot}25$	$L = 0{\cdot}75$
13 X	178	5100	5410	5560
14 X	289	3260	3380	3440
15 M	199	4610	4850	4980
18 X	397	2410	2480	2510
19 M	310	3050	3160	3210
24 M	445	2160	2210	2240

Theoretical values of Q obtained for Sommerfeld and Harms–Goubau lines are plotted as a function of length of the resonator in Figs. 23 and 24 respectively for the sake of comparison with the values of Q obtained experimentally.

6.3. *Attenuation Constant of Sommerfeld Line*

The attenuation constant of Sommerfeld line at a frequency of 9500 MHz is from eqn. (34):

$$\alpha = 0{\cdot}503(a_2 b_2) \qquad \text{.....(48)}$$

The values of a_2 and b_2 have been calculated for different values of the radius s varying from $0{\cdot}5 \times 10^{-4}$ metre to 15×10^{-4} metre and conductivity σ_1 ranging from $0{\cdot}1 \times 10^7$ to $7 \times 10^7\ \Omega^{-1}\,\text{m}^{-1}$.

† 'Paramex Winding Wires Catalogue', The India Cable Co. Ltd., Publication No. 571, pages 11, 18–21 (Bombay, 1957).

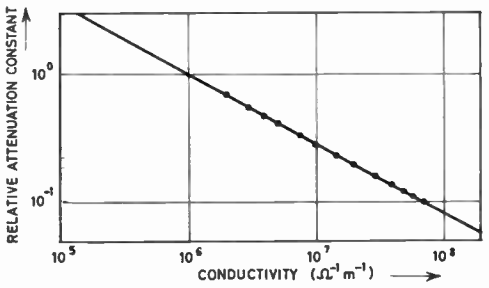


Fig. 7. Log–log plot of relative attenuation constant as a function of conductivity of the wire.

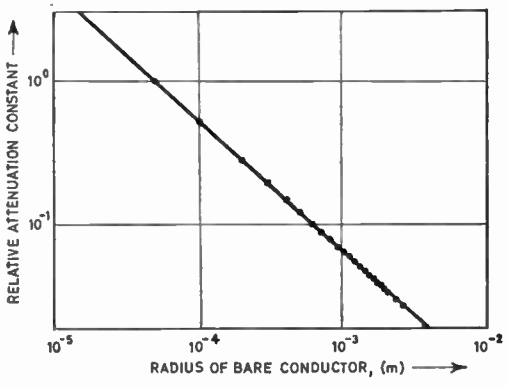


Fig. 8. Log–log plot of relative attenuation constant as a function of radius of the wire.

The variation of attenuation constants with conductivity and radius of the wire is represented graphically (see Fig. 7 and Fig. 8). In the first case (see Fig. 7) the attenuation constant is normalized with respect to the attenuation constant of wire having conductivity $0{\cdot}1 \times 10^7\ \Omega^{-1}\ m^{-1}$. In the second case (see Fig. 8) the attenuation constant is normalized with respect to the attenuation constant of wire having radius equal to $0{\cdot}5 \times 10^{-4}$ metre.

6.4. *Attenuation Constant of Goubau Line*

Using the following values for Paramex wire

$$\sigma_c = 5{\cdot}8 \times 10^7\ \Omega^{-1}\ m^{-1}$$
$$\mu_c = 1{\cdot}65 \times 10^{-6}\ H/m$$
$$\varepsilon = 8{\cdot}85 \times 10^{-12}\ F/m$$
$$\tan\delta = 0{\cdot}027$$
$$\lambda = 3{\cdot}2 \times 10^{-2}\ m$$

the following values for α_c, α_i and $G(\gamma' a')$ are obtained

$$\alpha_c = 6{\cdot}7 \times 10^{-5} \frac{1}{2a}\left(\frac{1}{\ln \gamma' a' + 0{\cdot}38}\right) \quad \text{N/m} \quad \text{.....(49)}$$

$$\alpha_i = 1{\cdot}32 \times 10^{-5} \gamma'^2 \left(2 - \frac{1}{\ln \gamma' a' + 0{\cdot}38}\right) \quad \text{N/m} \quad \text{.....(50)}$$

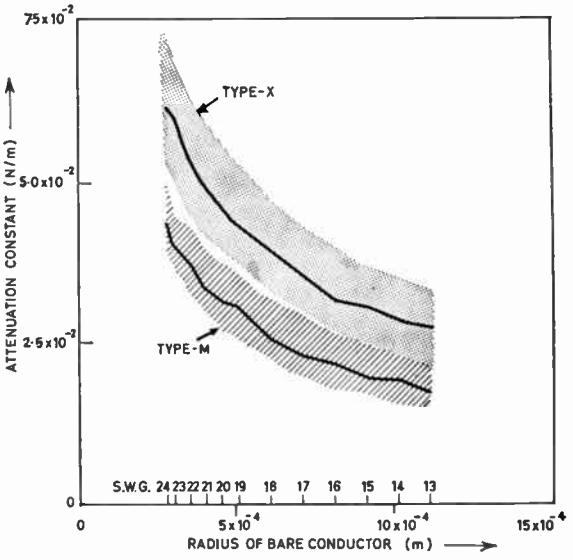


Fig. 9. Theoretical attenuation constants of 'Paramex' enamelled copper wires as a function of radius of bare conductor.

Solid-curves indicate the attenuation constants corresponding to the average dimensions and shaded regions indicate the range of possible variation due to variation in wire dimensions within manufacturer's tolerances.

$$G(\gamma' a') = 705 a'^2 \ln(a'/a) \quad \text{.....(51)}$$

It is found that the dimensional variations of these wires within the tolerances specified by the manufacturer give rise to appreciable variation in the attenuation constants. Hence, the attenuation constants for conditions of maximum and minimum attenuation and also for average dimensions are

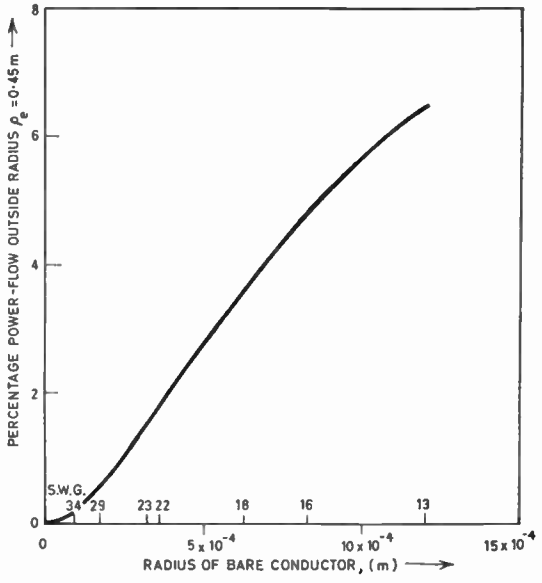


Fig. 10. Power-flow outside a radius $\rho_e = 0{\cdot}45$ metre as a function of wire-radius for bare copper wires.

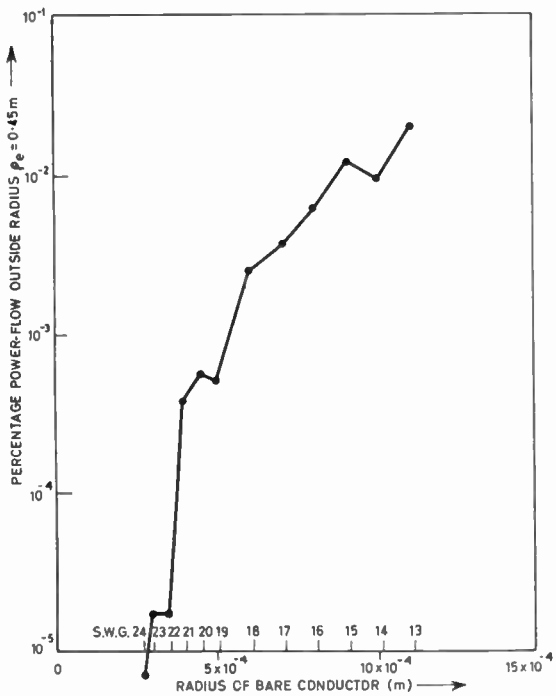


Fig. 11. Power-flow outside a radius $\rho_e = 0{\cdot}45$ metre as a function of wire-radius for enamelled copper wires ('Paramex', Type-M).

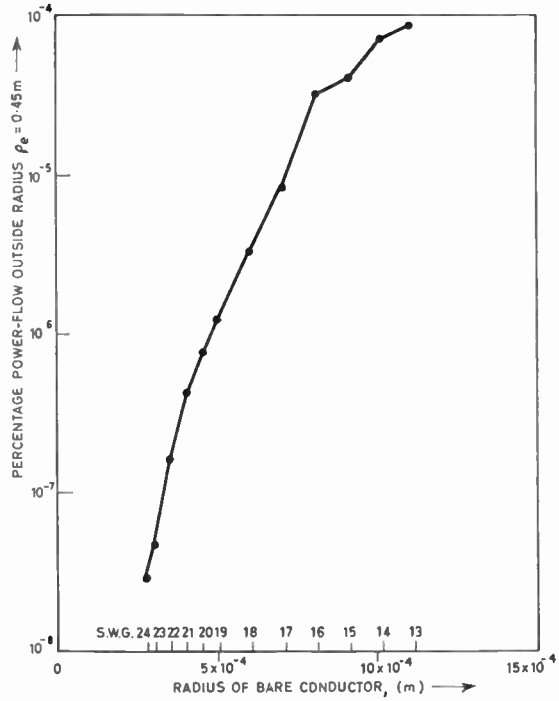


Fig. 12. Power-flow outside a radius $\rho_e = 0{\cdot}45$ metre as a function of wire-radius for enamelled copper wires ('Paramex', Type-X).

calculated for each wire. For any particular gauge of wire, the attenuation constant is maximum when the diameter of bare conductor is at its lower limit within the tolerance with the coating thickness at the upper limit, whereas the attenuation constant is minimum when the diameter of bare conductor is at the upper limit with the coating thickness at the lower limit. The results of the calculations are presented graphically (see Fig. 9). The curves corresponding to the average dimensions are shown by solid lines. The range of possible variation of the attenuation constant due to dimensional variation within the tolerance is shown by dotted and hatched regions.

6.5. *Power-flow: Resonator with Sommerfeld Line*

The percentage power-flow expressed in eqn. (42) outside a radius $\rho_e = 0{\cdot}45$ metre corresponding to the radius of the end-plates has been calculated for lines of radius varying from $0{\cdot}5 \times 10^{-4}$ metre to 12×10^{-4} metre. The results are presented graphically in Fig. 10.

6.6. *Power-flow: Resonator with Goubau Line*

The results of calculations of power-flow outside a radius $\rho_e = 0{\cdot}45$ metre for Type-M and Type-X Paramex wires are shown graphically in Figs. 11 and 12.

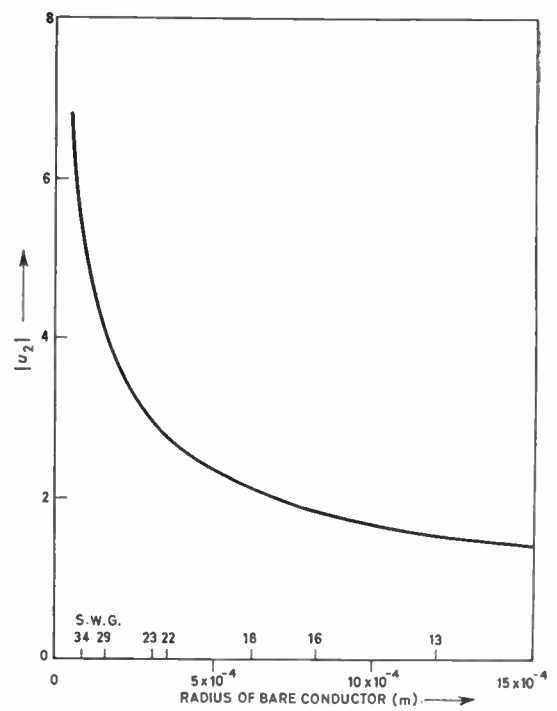


Fig. 13. Magnitude of the radial propagation constant of bare copper wire as a function of radius.

6.7. *Radial Variation of the Magnetic Field*

The variation of the magnetic field is governed by $H_1^{(1)}(ju_2\rho)$ in the case of Sommerfeld lines and by $H_1^{(1)}(j\gamma'\rho)$ in the case of Goubau lines. The variation of $|u_2|$ with respect to the radius of bare conductor is represented graphically in Fig. 13. The Hankel functions have been calculated with the aid of Bessel function tables. Some of the results of calculations are given in Figs. 21 and 22 for the sake of comparison with experimental results.

7. Measurement of Q

The experimental set-up (Figs. 14 and 15) is used to display on the oscilloscope screen the resonance curves which are photographed, magnified and analysed for computing Q. The two frequency meters are adjusted such that the pips lie on the skirts of the resonance curve, one above and the other below the resonance frequency f_0 (see Fig. 16). The two frequencies f_1 and f_2 are noted and the resonance curve is photographed. The sweep voltage is switched off and the input to the resonator is reduced by 3 dB by using the precision variable attenuator. The resulting trace, which is a vertical line with an amplitude 3 dB less than that of the resonance curve is recorded on the same film by exposing it a second time.

Assuming the frequency to be a linear function of the reflector voltage between f_1 and f_2, it is evident that the half-power bandwidth is proportional to l_1. The half-power bandwidth is

$$\Delta f_0 = (f_1 \sim f_2)\frac{l_1}{l_2} \qquad \text{.....(52)}$$

Hence,

$$Q = \frac{f_0}{f_1 \sim f_2}\frac{l_2}{l_1} \qquad \text{.....(53)}$$

8. Effect of Tilt of the End-plate

The shape of the resonance curve and the Q of the resonator depend on the planeness and parallelism of the end-plates. To test the sensitivity of the resonator to departure from parallelism of the end-plates, the bottom plate was adjusted to tilt in steps of 2 minutes from -1 degree to $+1$ degree with respect to the reference position in which the resonance curve was sharp and symmetrical. At each setting of the tilt of the end-plate, the corresponding resonance curve was photographed, magnified and analysed for Q (see Fig. 17 and Fig. 18). The variation of Q as a function of the tilt of the end-plate for different lengths of the resonator is represented graphically in in Fig. 19.

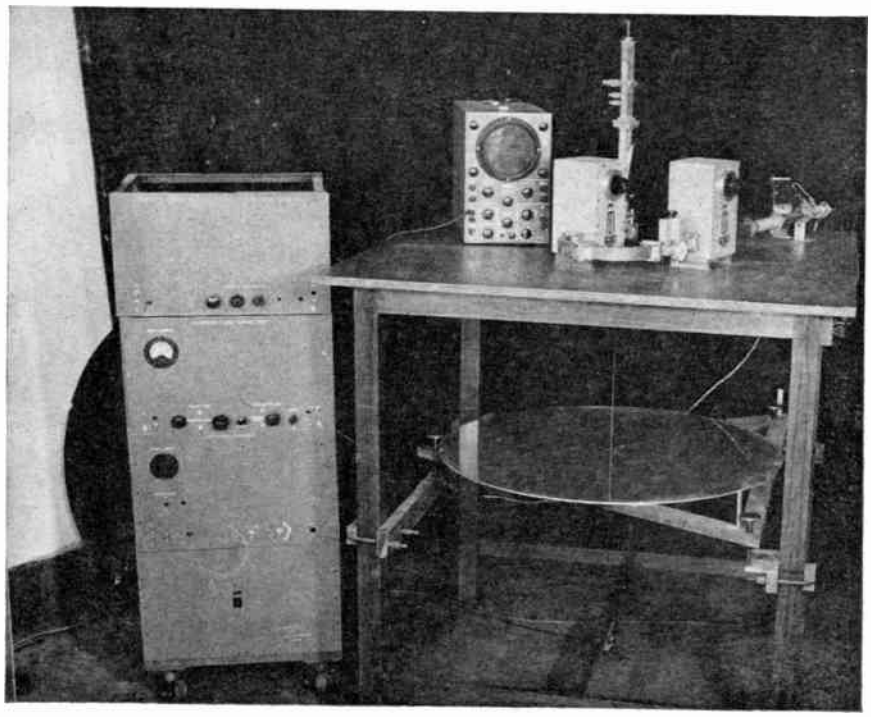


Fig. 14. Experimental set-up.

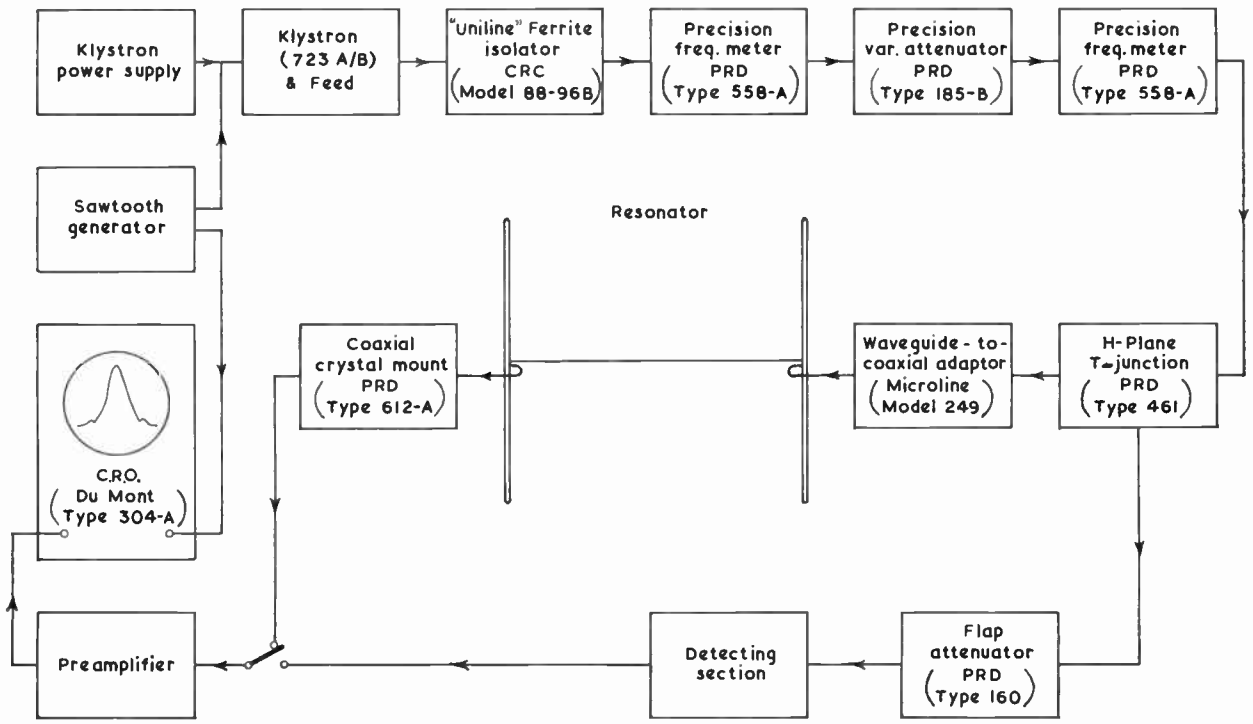


Fig. 15. Block diagram of the experimental set-up used for displaying the resonance curve.

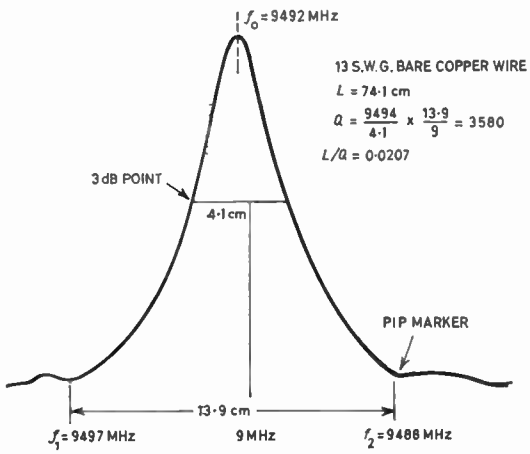


Fig. 16. Enlarged resonance curve showing the method of calculation of Q.

9. Effect of Rotation of the End-plate with the Loop

The shape of the resonance curve is modified if there is direct coupling between the exciting loop and the output loop. The degree of direct coupling between the two loops is changed if the orientation of one loop relative to the other is altered. In order to check whether there is any direct coupling between the loops, the bottom plate with the output loop was rotated about its own axis in steps of 15 degrees to cover a range of a little over 360 degrees. At each step the corresponding resonance curve was analysed for Q. The result of analysis when the Sommerfeld line is 24 s.w.g. bare copper wire and $L = 0{\cdot}498$ m is represented graphically in Fig. 20. The curve exhibits periodic variation having a period of about 40 degrees. These periodic oscillations can be taken as an indication of the presence of higher-order modes.

10. Measurement of Radial Variation of Magnetic Field

The variation of the magnetic field ($H_{\phi 2}$) in the radial direction is determined with the help of a small magnetic probe which is inserted into the resonator through a long, narrow, rectangular slot cut radially in the movable end-plate (see Fig. 15). The loop is mounted on a small carriage running on rails fixed to the outer surface of the end-plate. The output of the probe is passed through a crystal tuner to the oscilloscope (see Fig. 14) with the time-base switched off. For different positions of the vertical trace on the oscilloscope the output is maintained constant by means of a precision variable attenuator in the input circuit of the resonator. The results of measurement in the case of Sommerfeld and Goubau lines are

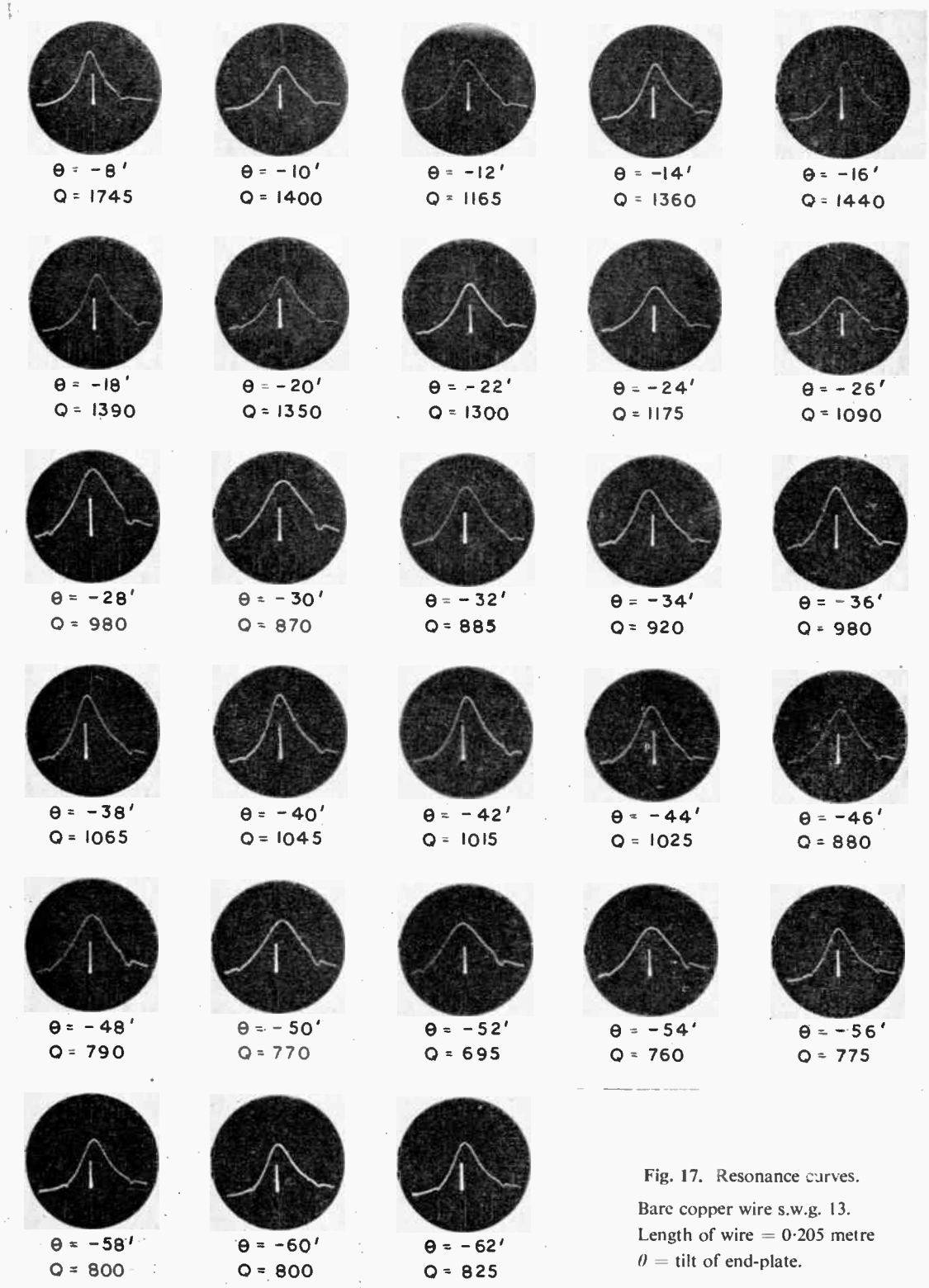


Fig. 17. Resonance curves.

Bare copper wire s.w.g. 13.
Length of wire = 0·205 metre
θ = tilt of end-plate.

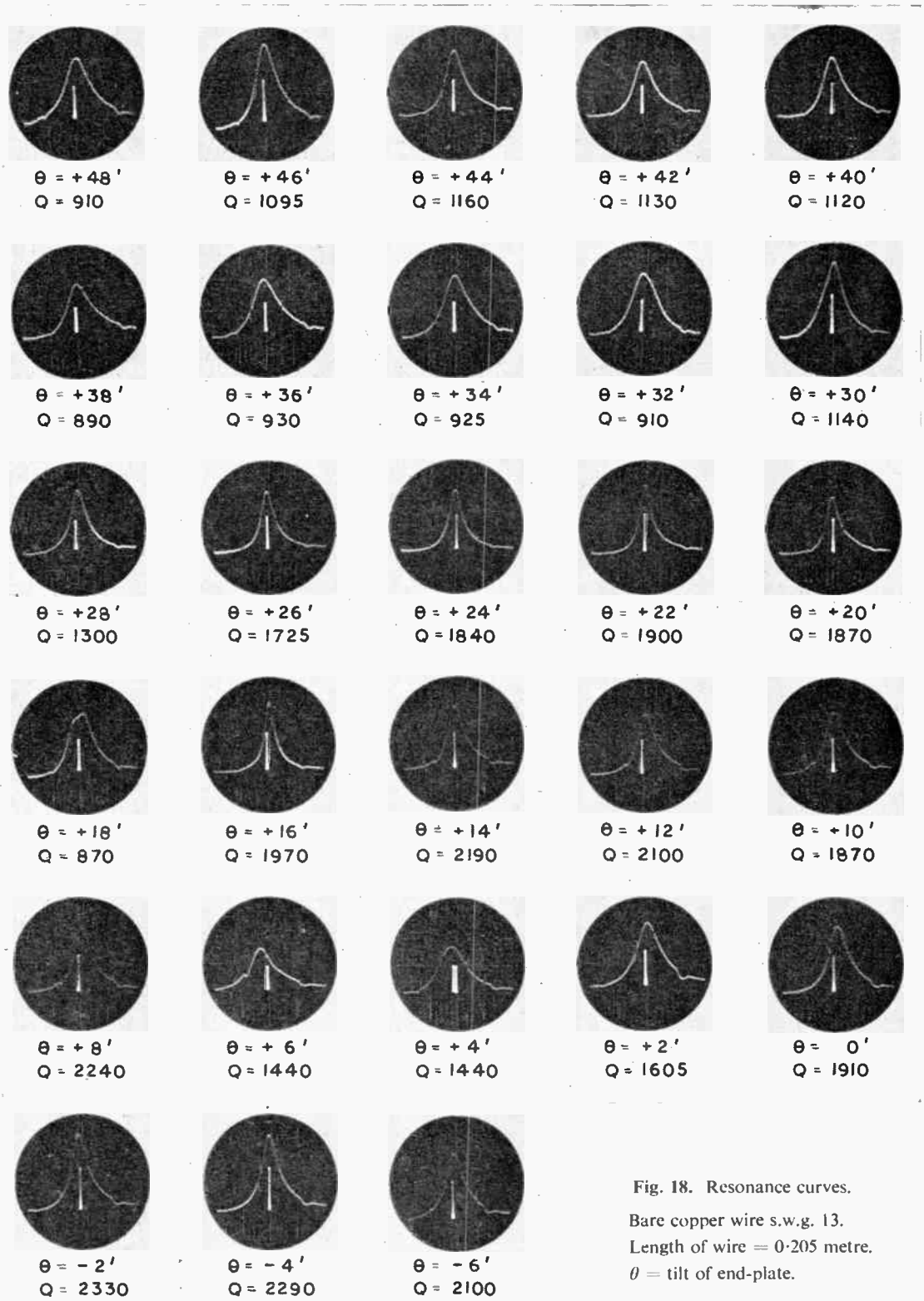


Fig. 18. Resonance curves.

Bare copper wire s.w.g. 13.
Length of wire = 0·205 metre.
θ = tilt of end-plate.

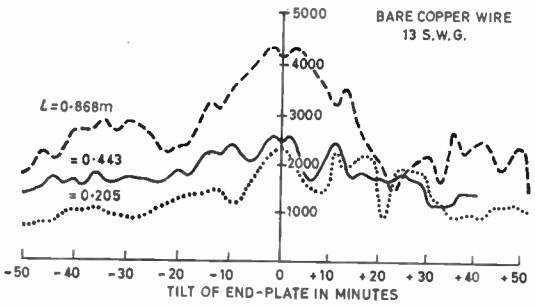


Fig. 19. Variation of Q with tilt of end-plate.

compared with the theoretical values discussed in Section 6.6 (see Figs. 21 and 22). It can be seen that there are superimposed periodic oscillations on the measured field decay curves. These oscillations are considered to be due to the presence of higher-order modes generated by the discontinuity presented to the surface-wave field at the edges of the end-plates.

11. Measurement of Attenuation Constant

The measurement of the attenuation constant of Sommerfeld and Goubau lines is based on the relation (3) which can be rewritten as:

$$\frac{L}{Q}=\frac{\alpha\lambda_g L}{\pi}+\frac{2R_m\lambda_g}{\pi Z_0} \qquad \text{.....(54)}$$

The plot of L/Q against L (see Figs. 23 and 24) is a straight line the slope of which gives the value of $\alpha\lambda_g/\pi$ and the intercept of which on the L/Q axis gives the value of $2R_m\lambda_g/\pi Z_0$. The value of λ_g is determined from a plot of the length of the resonator versus the number of half-wavelengths corresponding to it (see Fig. 25). The number of half-wavelengths is determined by moving a thin glass rod held perpendicular to the surface-wave line and observing the sharp reduction in amplitude of the oscilloscope trace when the rod passes through an antinode of the electric field. The observed values of the attenuation constants are compared with the theoretical values in the case of Sommerfeld as well as Goubau lines (see Figs. 26 and 27).

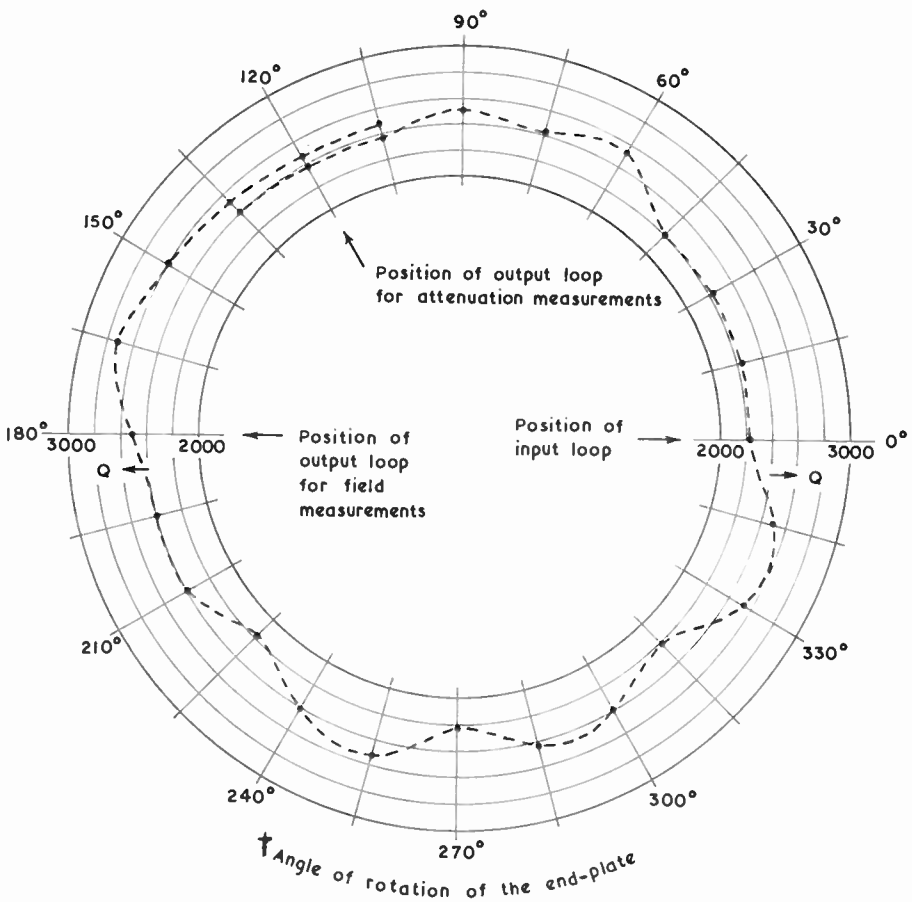


Fig. 20. Variation of Q with rotation of end-plate with the loop.

† The angle is measured from the plane of the exciting loop to that of the output loop looking from the input end towards the output end of the resonator.

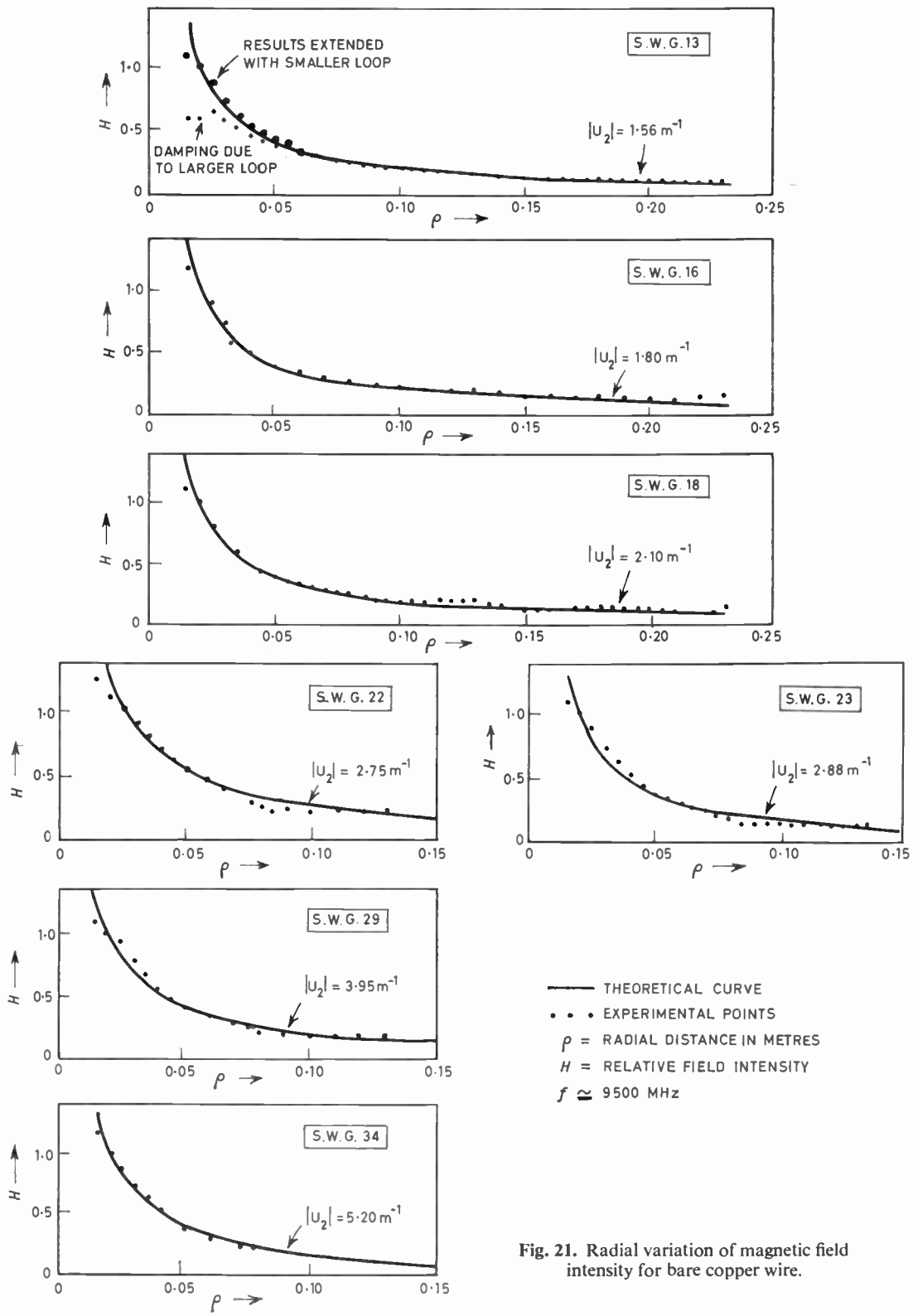


Fig. 21. Radial variation of magnetic field intensity for bare copper wire.

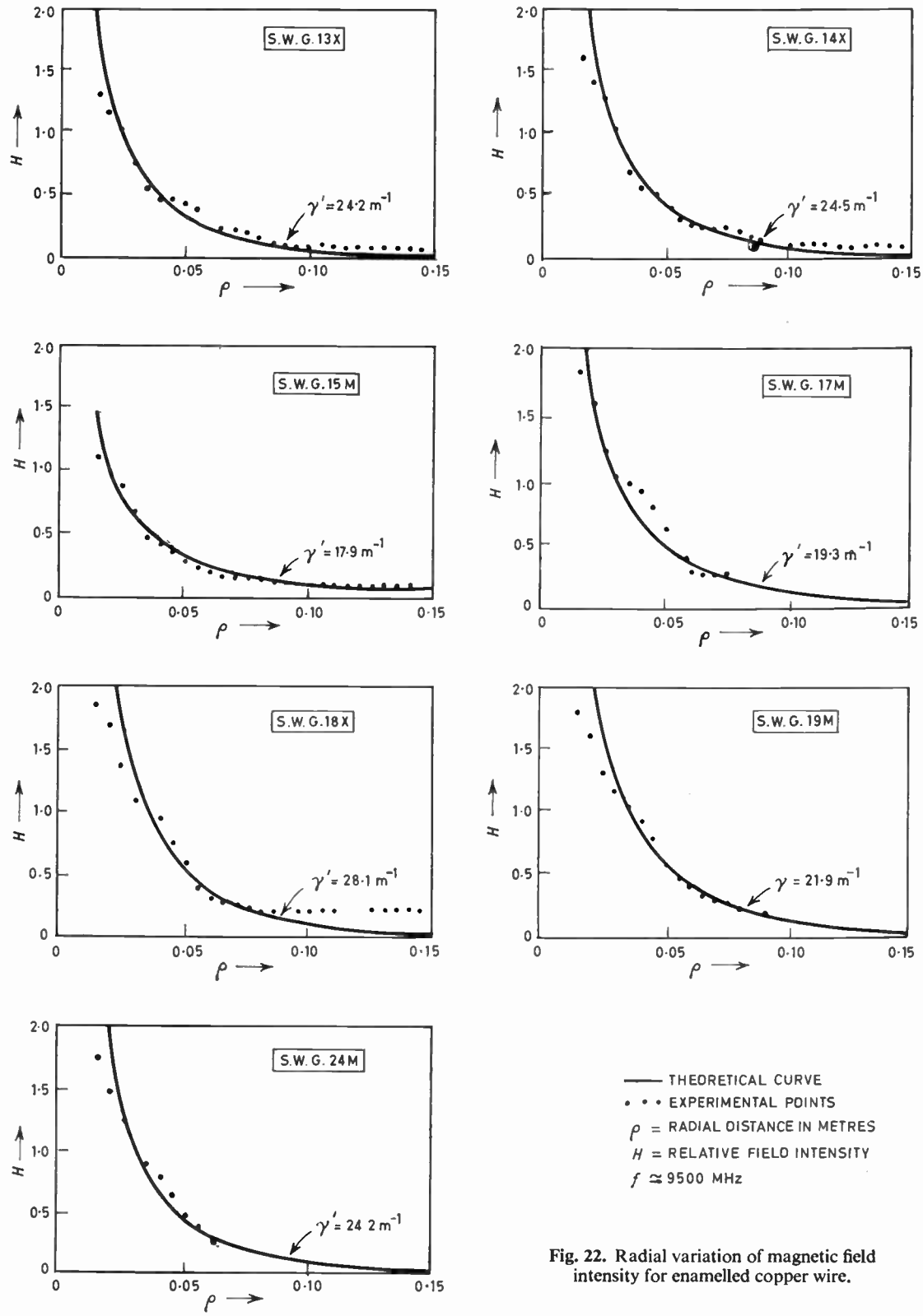


Fig. 22. Radial variation of magnetic field intensity for enamelled copper wire.

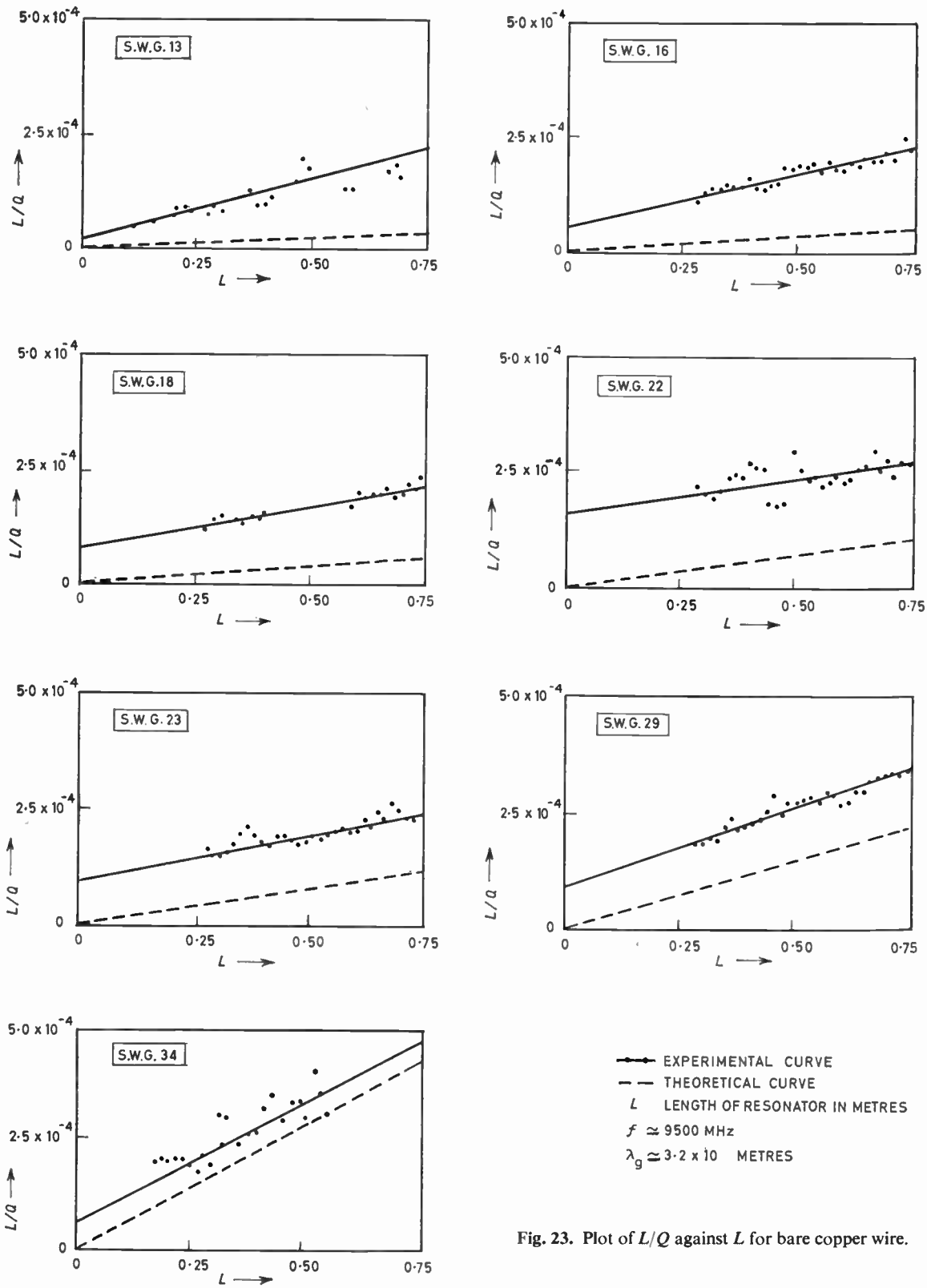


Fig. 23. Plot of L/Q against L for bare copper wire.

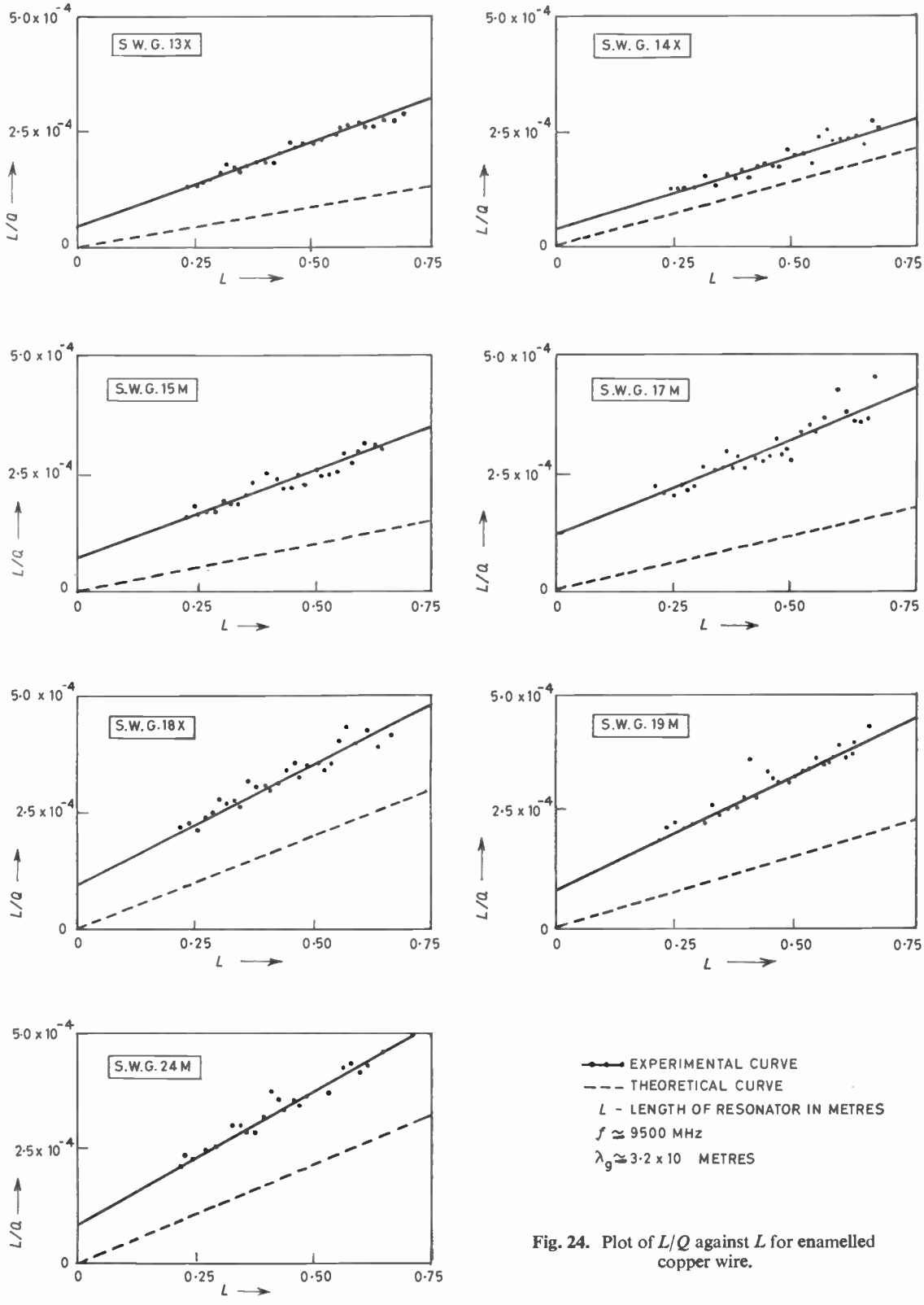


Fig. 24. Plot of L/Q against L for enamelled copper wire.

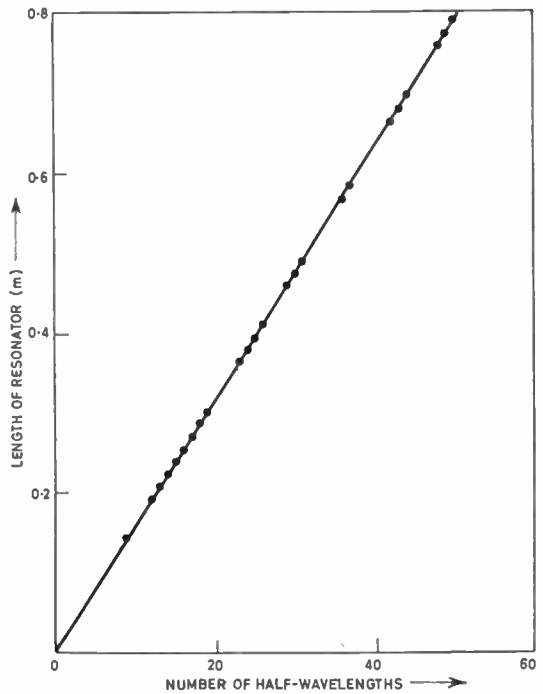


Fig. 25. Length of the resonator as a function of the number of half-wavelengths.

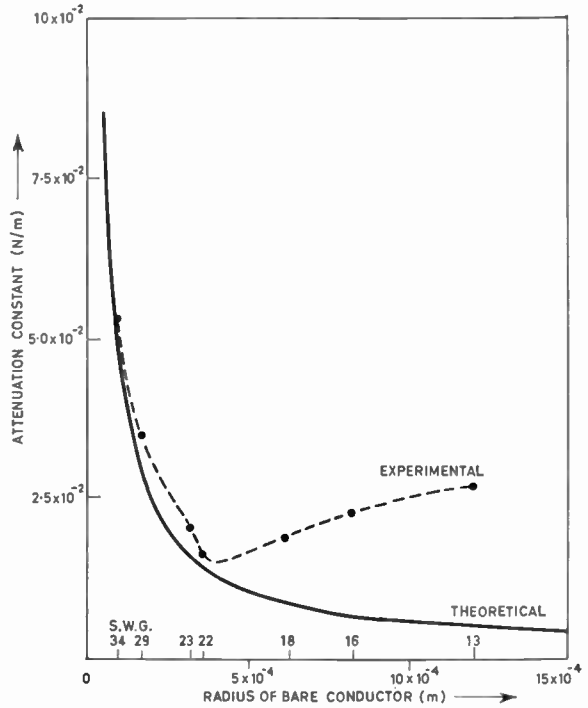


Fig. 26. Variation of attenuation constant with radius for bare copper wires.

12. Evaluation of Radiation Loss

In order to determine the contribution of radiation loss to the measured values of α, an effective value σ_{eff} for the surface-wave line is calculated and attenuation constants of wires having conductivity σ_{eff} are calculated. A sample calculation is given in Table 3.

Table 3

Calculation of the attenuation constant of wire having a radius $0{\cdot}5 \times 10^{-4}$ metre and conductivity $0{\cdot}1 \times 10^{7}\ \Omega^{-1}\ m^{-1}$

Gauge of wire	α (N/m)	α_{rel}	α/α_{rel} (N/m)	Average α/α_{rel} (N/m)
34	$5{\cdot}31 \times 10^{-2}$	0·595	$8{\cdot}9 \times 10^{-2}$	
29	$3{\cdot}46 \times 10^{-2}$	0·340	$10{\cdot}2 \times 10^{-2}$	$9{\cdot}9 \times 10^{-2}$
23	$2{\cdot}00 \times 10^{-2}$	0·185	$10{\cdot}8 \times 10^{-2}$	
22	$1{\cdot}60 \times 10^{-2}$	0·165	$9{\cdot}7 \times 10^{-2}$	

Attenuation constant of wire having radius $6{\cdot}5 \times 10^{-4}$ metre $= 9{\cdot}9 \times 10^{-2}$ N/m

Attenuation constant of the above wire relative to the attenuation constant of wire of radius $0{\cdot}5 \times 10^{-4}$ metre and $\sigma = 0{\cdot}1 \times 10^{7}\ \Omega^{-1}\ m^{-1}$ $= 0{\cdot}1285$

Conductivity σ_{eff} corresponding to the value of 0·1285 of relative attenuation constant $= 4{\cdot}3 \times 10^{7}\ \Omega^{-1}\ m^{-1}$

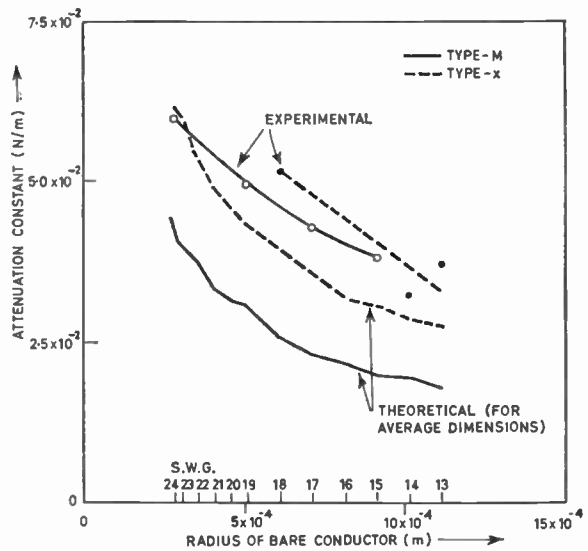


Fig. 27. Variation of attenuation constant with radius of bare conductor for 'Paramex' enamelled copper wire.

The attenuation constants of wires having conductivity σ_{eff} is calculated and compared with experimental values (see Fig. 28). The difference between the experimental and theoretical curves may be considered to be the loss due to radiation.

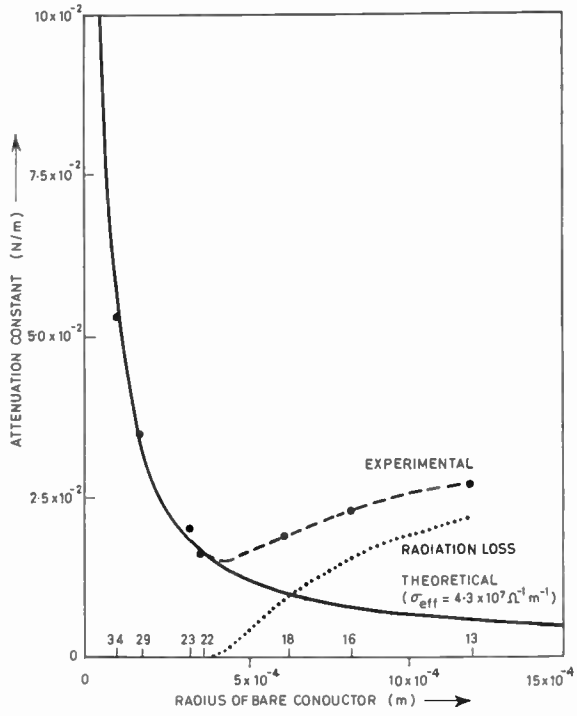


Fig. 28. Evaluation of radiation loss in the case of bare copper wires.

13. Proper Size of the End-plates

In order that the end-plates may be effective as terminations, the loss due to radiation should be negligibly small. It is observed (see Fig. 28) that for wires of diameter less than that of 22 s.w.g. wire, the radiation loss is insignificant. For 22 s.w.g. wire, the power-flow outside a radius 0·45 metre is 1·75%. Hence it can be concluded that the performance of the end-plate will be satisfactory when the power-flow outside a radius corresponding to that of the end-plates is less than 1·75%.

14. Concluding Remarks

As a result of the investigations, the following contributions are regarded as interesting:

(i) method of measuring Q,

(ii) evaluation of radiation loss and an estimation of the minimum size of the end-plates,

(iii) calculation of the percentage power-flow outside the resonator,

(iv) calculation of the effective conductivity of the Sommerfeld surface-wave line.

The nature of the variation of the magnetic field in the radial direction and the loss measurements on the surface-wave line confirm the physical realizability of Sommerfeld wave.

15. References

1. Sommerfeld, A., 'Fortpflanzung elektrodynamischer Wellen an einem zylindrischen Leiter', *Ann. Physik Chemie, Lpzg.*, **67**, pp. 233–90, 1899.
2. Sommerfeld, A., 'Uber die Ausbreitung der Wellen in der drahtlosen Telegraphie', *Ann. Physik, Lpzg.*, **28**, pp. 665–737, 1909.
3. Sommerfeld, A., 'Uber die Ausbreitung der Wellen in der drahtlosen Telegraphie', *Ann. Physik, Lpzg.*, **81**, pp. 1135–53, 1926.
4. Zenneck, J., 'Uber die Fortplanzung ebener elektromagnetischer Wellen langs einer ebenen Leiterfläche und ihre Beziehung zur drahtlosen Telegraphie', *Ann. Physik, Lpzg.*, **81**, pp. 1135–53, 1907.
5. Bouwkamp, C. J., 'On Sommerfeld's surface wave', *Phys. Rev.*, **80**, p. 294, 1950.
6. Barlow, H. M., 'Surface waves', *Proc. Inst. Radio Engrs*, **46**, pp. 1413–17, 1958.
7. Barlow, H. M., 'The power radiated by a surface wave circulating around a cylindrical surface', *Proc. Instn Elect. Engrs*, **106**, Part B, pp. 179–85, 1959.
8. Barlow, H. M., 'Surface wave supported by cylindrical surfaces', *Trans. I.R.E. on Antennas and Propagation*, **AP-7**, pp. S147–S53, Special Supplement, December 1959.
9. Barlow, H. M., and Cullen, A. L., 'Surface waves', *Proc. I.E.E.*, **100**, Part III, pp. 329–47, 1953.
10. Barlow, H. M., and Karbowiak, A. E., 'An experimental investigation of the characteristics of cylindrical surface waves', *Proc. I.E.E.*, **100**, Part III, pp. 321–28, 1953.
11. Barlow, H. M., and Karbowiak, A. E., 'An experimental investigation of the properties of corrugated cylindrical surface waveguides', *Proc. I.E.E.*, **101**, Part III, pp. 182–88, 1954.
12. Barlow, H. M., and Karbowiak, A. E., 'An experimental investigation of axial cylindrical surface waves supported by capacitive surfaces', *Proc. I.E.E.*, **102**, Part B, pp. 313–22, 1955.
13. Barlow, H. M., and Brown, J., 'Radio Surface Waves', pp. 13–15 (Clarendon Press, Oxford, 1962).
14. Cullen, A. L., 'The excitation of plane surface waves', *Proc. I.E.E.*, **101**, Part IV, pp. 225–34, 1954.
15. Cullen, A. L., 'A note on the excitation of surface waves', *Proc. I.E.E.*, **104**, Part C, pp. 472–74, 1957.
16. Cullen, A. L., 'Surface wave resonance effect in a reactive cylindrical structure excited by an axial line source', *J. Res. Nat. Bur. Stds*, **64D**, pp. 13–19, 1960.
17. Attwood, S. S., 'Surface wave propagation over a coated plane conductor', *J. Appl. Phys.*, **22**, pp. 504–9, 1951.
18. Goubau, G., 'Surface waves and their application to transmission lines', *J. Appl. Phys.*, **21**, pp. 1119–28, 1950.
19. Goubau, G., 'Single conductor surface wave transmission lines', *Proc. I.R.E.*, **39**, pp. 619–24, 1951.
20. Goubau, G., 'On the excitation of surface waves', *Proc. I.R.E.*, **40**, pp. 865–868, 1952.
21. Goubau, G., 'Designing surface wave transmission lines', *Electronics*, **27**, pp. 180–84, April 1954.
22. Goubau, G., 'Open wire lines', *Trans. I.R.E. on Microwave Theory and Techniques*, **MTT-4**, pp. 197–200, 1956.
23. Goubau, G., 'Some characteristics of surface wave transmission lines for long distance transmission', *Proc. I.E.E.*, **106**, Part B, Supplement, pp. 166–68, 1959.

24. Goubau, G., 'Waves on interfaces', *Trans. I.R.E. on Antennas and Propagation*, AP-7, pp. S140–S46, Special Supplement, December 1959.
25. Wait, J. R., 'Excitation of surface waves on conducting stratified, dielectric clad and corrugated surfaces', *J. Res. Nat. Bur. Stds*, **59**, pp. 365–77, 1957.
26. Wait, J. R., 'On the excitation of electromagnetic surface waves on a curved surface', *Trans. I.R.E.*, **AP-8**, pp. 445–48, 1960.
27. Wait, J. R., 'Electromagnetic Surface Waves: Advances in Radio Research', Vol. 4, pp. 157–217 (Academic Press, London, 1964).
28. Chatterjee, S. K., and Madhavan, P., 'Propagation of microwaves on a single wire, Part I', *J. Indian Inst. Sci.*, **37**, pp. 200–23, 1955.
29. Chatterjee, S. K., and Chatterjee, R., 'Propagation of microwaves on a single conductor embedded in three coaxial dielectrics, Part I', *J. Indian Inst. Sci.*, **38**, pp. 156–71, 1956.
30. Chatterjee, S. K., and Chatterjee, R., 'Propagation of microwaves on a single conductor embedded in three coaxial dielectrics, Part II', *J. Indian Inst. Sci.*, **39**, pp. 71–82, 1957.
31. Chatterjee, S. K., and Contractor, S. N., 'Propagation of microwaves on a single wire, Part II', *J. Indian Inst. Sci.*, **39**, pp. 107–122, 1957.
32. Chatterjee, S. K., 'Surface wave transmission at 3·2 cm wavelength', *J. Instn Engrs (India)*, **38**, pp. 875–82, 1958.
33. Chatterjee, S. K., and Chatterjee, R., 'Surface waveguide', *J. Inst. Telecomm. Engrs*, **4**, pp. 90–95, 1958.
34. Chandler, C. J., 'An investigation of dielectric rod as waveguide', *J. Appl. Phys.*, **20**, pp. 1188–93, 1949.
35. Scheibe, E. H., King, B. G., and Van Zeeland, D. L., 'Loss measurements of surface wave transmission lines', *J. Appl. Phys.*, **25**, pp. 790–97, 1954.
36. King, B. G., Scheibe, E. H., and Tatsuguchi, I., 'The physical realizability of the Sommerfeld wave on a cylindrical conductor', *Proc. Nat. Electronics Conf.*, **11**, pp. 949–57, 1955.
37. King, B. G., Tatsuguchi, I., Scheibe, E. H., and Goubau, G., 'Pseudo-resonance between parallel plates', presented at the Spring meeting of U.R.S.I., Washington, D.C., 1955.
38. Barlow, H. M., and Cullen, A. L., 'Microwave Measurements', pp. 87–92 (Constable, London, 1950).
39. Watson, G. N., 'A Treatise on the Theory of Bessel Functions', 2nd edn., p. 134 (Cambridge University Press, 1962).

16. Appendix 1: Evaluating the integral in eqns. (5a) and (7)

The integral involved in eqns. (5a) and (7) is

$$\int^{\rho} \mathrm{H}_1^{(1)}(\mathrm{j}u_2\rho)\mathrm{H}_1^{(2)}(-\mathrm{j}u_2^*\rho)\,\rho\,\mathrm{d}\rho \qquad \ldots\ldots(55)$$

This integral is evaluated as follows:

If C_μ and $\bar{C}_\nu$ denote any two cylinder functions of orders μ and ν respectively, and if k and l are constants, the following relations hold good.[39]

$$\int^{z} \left[(k^2-l^2)z - \frac{\mu^2-\nu^2}{z}\right] C_\mu(kz)\bar{C}_\nu(lz)\,\mathrm{d}z = z\left[C_\mu(kz)\frac{\mathrm{d}\bar{C}_\nu(lz)}{\mathrm{d}z} - \bar{C}_\nu(lz)\frac{\mathrm{d}C_\mu(kz)}{\mathrm{d}z}\right] \ldots(56)$$

If $\mu = \nu$, then eqn. (56) reduces to

$$\int^{z} C_\mu(kz)\bar{C}_\mu(lz)z\,\mathrm{d}z = \frac{z}{(k^2-l^2)}\left[C_\mu(kz)\frac{\mathrm{d}\bar{C}_\mu(lz)}{\mathrm{d}z} - \bar{C}_\mu(lz)\frac{\mathrm{d}C_\mu(kz)}{\mathrm{d}z}\right] \quad (57)$$

$$\frac{\mathrm{d}\bar{C}_\mu(lz)}{\mathrm{d}z} = l\left[\bar{C}_{\mu-1}(lz) - \frac{\mu}{lz}\bar{C}_\mu(lz)\right] \qquad \ldots\ldots(58)$$

$$\frac{\mathrm{d}C_\mu(kz)}{\mathrm{d}z} = k\left[C_{\mu-1}(kz) - \frac{\mu}{kz}C_\mu(kz)\right] \qquad \ldots(59)$$

Substitution of eqns. (58) and (59) in eqn. (57) yields

$$\int^{z} C_\mu(kz)\bar{C}_\mu(lz)z\,\mathrm{d}z = [z/(k^2-l^2)][lC_\mu(kz)\bar{C}_{\mu-1}(lz) - kC_{\mu-1}(kz)\bar{C}_\mu(lz)] \qquad \ldots\ldots(60)$$

Substitution of $K = \mathrm{j}u_2$, $l = -\mathrm{j}u^*$ and $z = \rho$

$$C_\mu(kz) = \mathrm{H}_1^{(1)}(\mathrm{j}u_2\rho)$$

$$\bar{C}_\mu(lz) = \mathrm{H}_1^{(2)}(-\mathrm{j}u_2^*\rho)$$

in eqn. (60) gives

$$\int^{\rho} \mathrm{H}_1^{(1)}(\mathrm{j}u_2\rho)\mathrm{H}_1^{(2)}(-\mathrm{j}u_2^*\rho)\rho\,\mathrm{d}\rho = [1/(u_2^{*2}-u_2^2)][-\mathrm{j}u_2^*\rho\,\mathrm{H}_1^{(1)}(\mathrm{j}u_2\rho)\mathrm{H}_0^{(2)}(-\mathrm{j}u_2^*\rho) - -\mathrm{j}u_2\rho\,\mathrm{H}_0^{(1)}(\mathrm{j}u_2\rho)\mathrm{H}_1^{(2)}(-\mathrm{j}u_2^*\rho)]$$

17. Appendix 2: Small argument approximations for Hankel functions

$$\mathrm{H}_1^{(1)}(\mathrm{j}u_2 s) = -\frac{2}{\pi u_2 s} \qquad \ldots\ldots(61)$$

$$\mathrm{H}_1^{(2)}(-\mathrm{j}u_2^* s) = -\frac{2}{\pi u_2^* s} \qquad \ldots\ldots(62)$$

$$\mathrm{H}_0^{(1)}(\mathrm{j}u_2 s) = \mathrm{j}\frac{2}{\pi}(m+\mathrm{j}n) \qquad \ldots\ldots(63)$$

$$\mathrm{H}_0^{(2)}(-\mathrm{j}u_2^* s) = -\mathrm{j}\frac{2}{\pi}(m+\mathrm{j}n) \qquad \ldots\ldots(64)$$

where

$$m = \tfrac{1}{2}\ln[(0{\cdot}89s)^2(a_2^2+b_2^2)] \qquad \ldots\ldots(65)$$

$$n = \arctan(b_2/a_2) \qquad \ldots\ldots(66)$$

Manuscript first received by the Institution on 23rd January 1968 and in final form on 29th April 1968. (Paper No. 1205/CC18.)

Radio Engineering Overseas . . .

The following abstracts are taken from Commonwealth, European and Asian journals received by the Institution's Library. Abstracts of papers published in American journals are not included because they are available in many other publications. Members who wish to consult any of the papers quoted should apply to the Librarian giving full bibliographical details, i.e. title, author, journal and date, of the paper required. All papers are in the language of the country of origin of the journal unless otherwise stated. Translations cannot be supplied.

PRODUCTION OF FREQUENCY ANALOGUE SIGNALS

Measured intelligence can conveniently be transmitted by frequency-analogue signals, which are insensitive to interference and are readily combined with digital techniques. This requires special converters, which can take the form of oscillators (preferably harmonic) whose frequency is solely and correctly determined by the output signals of amplitude-analogue passive transducers. To simplify further processing it is desirable to have a linear relation between the frequency (or period) and a single circuit parameter, e.g. a resistance value. An essentially linear characteristic can be achieved by applying a new kind of amplitude control: When this method is applied to an R-C oscillator with two R-C chains, a change in one of the resistances, controlled by the measured quantity, initiates a change in the other resistance of such a magnitude that the two R-C chains always vary to the same extent. A large relative variation in frequency for only a small relative variation in resistance is obtained if this oscillator is connected to a specially designed self-balancing electronic compensator. A German engineer with the Philips Laboratories in Hamburg describes a device of this type intended for use with a strain gauge bridge. Measurements of departure from linearity, temperature drift and the effects of varying the supply voltage in the experimental versions of these converters indicated errors of 0·01 to 0·05%.

'A process for converting quantities to be measured into frequencies', D. Meyer, *Philips Technical Review*, **29**, No. 6, pp. 189–196, 1968.

LINEARLY TUNABLE SELECTIVE SYSTEM

The design of a selective system linearly tunable over a wide range, throughout which its Q-factor remains constant, has aroused considerable theoretical and practical interest in recent years. The systems of particular interest in this context are amplifiers and oscillators linearly tunable over the entire audio and supersonic band.

Though R-C oscillators have been in use for a long time, they have undergone remarkably few changes over the years, and are a long way from satisfying the requirements of linear relationship between the frequency and the variable resistance (or capacitance). The problem of resolving the contradictory requirements of a wide tuning range on the one hand, and a linear tuning scale and a constant Q over the range on the other has also not been successfully tackled.

In circuits based on R-C networks the tuning frequency is inevitably a non-linear function of the variable component. In similar R-L circuits a linear dependence can be obtained, but only by achieving a proportional variation of three linear variable impedances, by mounting them on a common shaft. Apart from the practical difficulties of doing this, there is a considerable drop in the circuit-Q as the tuning frequency falls.

An oscillator circuit using two Wien bridges forms the basis of a selective circuit which can be tuned linearly over a wide range. A Soviet paper describes the circuit which uses two double-T networks and finds the relation between the scale non-linearity and the width of the tuning band. Design formulae for the circuit components are also derived.

'A selective system linearly tunable over a wide range using one variable component', I. D. Vul'man, *Telecommunications and Radio Engineering* (English language edition of *Elektrosvyaz* and *Radiotekhnika*), pp. 84–90, No. 12, 1967.

IONOSPHERIC MEASUREMENTS BY VERTICAL TRIANGULATION

The measurement of the distant ionosphere by means of ground backscatter observations has been described in many papers. The main interest has been concentrated on the maximum usable frequency, on modes and on irregularity studies. The main observations have been of echo structure and of time delay, particularly at the skip distance, as a function of frequency and of azimuth. Some observations of angle of elevation have been reported, but these were not used to deduce the height of the distant ionosphere. Estimates of the equivalent height of the ionosphere are needed for the precise interpretation of the vertical geometry of skywave propagation over short ionospheric paths; when the length of the path is about 1000 km the minor inaccuracies in the verticle angle measurement tend to be submerged by the inaccuracies which arise when the equivalent height of the ionosphere is measured at one of the terminals of the transmission path instead of at the distant point (or points) of reflection along the path.

An Australian paper briefly describes a simple experiment on locating backscatter sources which was carried out at St. Kilda, South Australia, in October 1960, in which observations were made of the angles of arrival in three dimensions and of the time delay of high-frequency backscatter returns, and suggests how these results provide an estimate of the height of the distant ionosphere using the technique of vertical triangulation. The method of Appleton and Barnett using transmitters at known distances is also discussed briefly.

'Triangulation of high frequency backscatter returns', R. F. Treharne, *Proceedings of the Institution of Radio and Electronics Engineers Australia*, **29**, No. 4, pp. 109–14, April 1968.