

THE RADIO AND ELECTRONIC ENGINEER

The Journal of the British Institution of Radio Engineers

FOUNDED 1925 INCORPORATED BY ROYAL CHARTER 1961

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

VOLUME 25

JANUARY 1963

NUMBER 1

THE PACE OF RESEARCH

UNDER the terms of its Charter the Institution is constituted to advance science in the field of radio and electronics. In pursuance of this object it has provided a forum—through the *Journal*, other publications and meetings—for disclosing new information and the possible application of new scientific and technological ideas.

Can the Institution do more to advance science by stimulating the pace of research in radio and electronics? The answer was contained in the speech given to the Radio and Electronics Industry by the President of the Institution, Admiral of the Fleet the Earl Mountbatten of Burma, K.G., on 29th November 1961. The President followed up that speech by personally appointing the first Research Committee of the Institution in February 1962. The Committee has already recommended:

(1) The establishment of the first Institution Research Scholarship—appropriately named "The Mountbatten Research Studentship". This is the first contribution of the Institution toward securing improved opportunity for research scholars.

(Other research endowments are available, mainly through the Department of Scientific and Industrial Research, but the Research Committee believes that there are insufficient facilities of this nature. It is important to increase the number of research studentships, fellowships, and the endowment of more University Chairs devoted to radio and electronics in order not only to improve the pace of research but to provide adequate background to the teaching of radio and electronic science and technology.)

(2) The preparation of a survey on radio and electronics research in Great Britain. An interim report was submitted to the President on 29th October, 1962. The final report will be completed in April 1963.

(The statistical data collected as a basis for the research survey are believed to be more up-to-date and succinct than in any other report available on current radio and electronics research in Great Britain. It shows the necessity of increasing the total value of true research and the fact that the exchange of information and experience is inadequate; this in turn reveals that insufficient mechanism exists for securing co-operation of effort and cross-fertilization of ideas between fundamental scientific work and education and industry.)

It is not pertinent at this stage to anticipate the final report nor the recommendations which might be made as a result of consultation with industry. It is apparent, however, that the most important way of securing an increase in the pace of radio and electronics research is by the extension of University facilities.

The function of the Institution in such work has most clearly been emphasized by the response of Government Departments, Universities, Colleges of Advanced Technology, and Industry to the questionnaire circulated by the Institution's Research Committee. Many of the replies which the Committee received have gone beyond the detailed questionnaire and expressed views which were of sufficient importance to merit open discussion. For this reason a Symposium was held by the Institution on 7th January 1963 which in itself has made a valuable contribution to formulating the final draft of the research report. It has provided yet another example of how the Institution can be the effective means of advancing the science of radio and electronic engineering which is the first charge laid upon it as a Chartered Institution.

L. H. B.

INSTITUTION NOTICES

The Proceedings of the Brit.I.R.E.

The first issue of the Institution's *Proceedings* is being distributed this month to members in the British Isles. The contents of *The Proceedings* for January include:

Details of all Institution meetings to be held in Great Britain during February 1963;

Proposals for elections and transfers to various grades of membership;

Library acquisitions;

The first group of papers which were set in the November 1962 Graduateship Examination with examiners' comments;

Institution notices.

As membership of the Institution overseas expands, it is hoped that "Divisional Proceedings" will be started in those countries where there is a sufficient number of members to justify such a publication.

The main purpose of *The Proceedings of the Brit.I.R.E.* is to provide members in the British Isles with full information on local activities. It is, however, believed that many members overseas, especially those having close ties with the United Kingdom, would like to receive copies regularly. Arrangements can therefore be made to send copies to members outside the British Isles at a cost of 15s. per year which represents postage and packing plus a nominal charge for the issue itself. Such subscriptions will run from January to December inclusive and will comprise nine or ten issues. It is envisaged that certain issues will be combined in view of the summer recess for Institution meetings.

Payments to the Institution from Canada and the U.S.A.

Members and subscribers in Canada and the United States are asked to note the change in address of the Institution's bankers in Toronto:

Canadian Imperial Bank of Commerce, King and Victoria Branch, 21 King Street East, Toronto, Ontario, Canada.

All remittances should be clearly marked "Account of the British Institution of Radio Engineers."

Index for Volume 24 of the *Journal*

The despatch of Indexes for Volume 24 of the *Journal* (July-December 1962 inclusive) has been postponed until the February issue. Copies will be inserted into all copies sent to members and subscribers in Great Britain and overseas.

The International Telemetry Conference

An International Conference on Telemetry is to be held in London from 23rd to 26th September 1963 under the joint sponsorship of two British engineering institutions and four American professional societies. The Conference will take place at the Institution of Electrical Engineers' building and the British, European and Commonwealth contributions are being organized jointly by the I.E.E. and the Brit.I.R.E. American contributions are being sponsored by the Institute of Electrical and Electronics Engineers (formerly A.I.E.E. and I.R.E.), the American Rocket Society, the Institute of Aerospace Sciences, and the Instrument Society of America.

Outline details of the Conference were circulated with the December issue of the *Journal* and applications for registration forms should be returned to the Institution of Electrical Engineers, Savoy Place, London, W.C.2, as soon as possible. Offers of papers are invited and synopses of proposed contributions by Brit.I.R.E. members should be sent as soon as possible to the British Institution of Radio Engineers, 9 Bedford Square, London, W.C.1, for consideration by the joint Programme Committee. The last date for submission of the complete papers is 16th April.

The Conference will include an extensive programme of technical visits and social functions and details of these activities will be circulated to those requesting programme and registration information. It is envisaged that some 70 papers will be presented in the course of eight sessions.

The 1962 List of Members

All Corporate Members, Companions, Associates and Graduates have been sent free of charge their copies of the 10th issue of the *List of Members* of the Institution. Registered students, whose names are not included in the List, may obtain copies from the Institution, price 5s. each.

Correction

The following amendments should be made to the paper, "A Method of I.F. Switching for a Microwave Diversity System", which was published in the August 1962 issue of the *Journal*:

Page 177, Fig. 10: The input resistance R should be changed to R_i ,

where R_i is given by
$$\frac{R(R_1 + R_D)}{R + R_1 + R_D}$$

Page 177, Section 6.3: The first equation should read

$$V_o = V_i \frac{R_D}{\sqrt{X^2 + R_D^2}} \times \frac{r}{\sqrt{X^2 + r^2}}$$

Two Four-Quadrant Electronic Multipliers

By

A. KRAICER,
(Associate Member) †

Presented at the Symposium on "Recent Developments in Industrial Electronics" in London on 2nd-4th April 1962.

Summary: Two instruments, which provide analogue multiplication in all four quadrants, are described.

The principle used in the first multiplier is that of varying the saturation point of a ferro-magnetic circuit in proportion to one of the inputs, so as to control the mark/space ratio of a chopper, which then modulates the other input.

The second multiplier uses the principle of modulating the phase of one square wave with respect to another similar wave in proportion to one of the input signals. Zero modulation corresponds to a phase difference of $\pi/2$. This wave then chopper modulates the other input signal, and the resultant output is combined with the original square wave in a synchronous demodulator.

After suitable filtering in both cases, the desired product is obtained. The accuracy is better than $\pm 1\%$ of full-scale output.

1. Introduction

This paper will describe the development of two four-quadrant electronic multipliers, that is, multipliers which, when presented with two input voltages $\pm X$ and $\pm Y$, will produce an output proportional to their product, XY , with the correct sign attached.

The first of these units was developed as a multiplying element to be incorporated in a respiratory analyser. The second was a subsequent design in which the use of a d.c. amplifier is avoided.

The input voltages to the multipliers are within the limits of ± 5 V and ± 4 V, respectively, for the X and Y signals; the output voltage is equal to 0.1 times the product of the input voltages.

There are several possible methods which may be adopted for achieving multiplication.

2. Multiplication Methods

2.1. The Use of Mechanically-driven Elements

The limitations here are that it is difficult to obtain fast response, so that mechanical methods are confined to very low frequencies. They also tend to be bulky.

2.2. The "Quarter-squares" Method

This is based on the fact that the product

$$XY = [(X+Y)^2 - (X-Y)^2]/4.$$

Thus, in its simplest version, the sum of the signals would be applied to one square-law device, and the difference of the signals applied to another square-law device. The difference of the outputs of the two square-law circuits is then proportional to the product XY .

† J. Langham Thompson Ltd., Watford, Hertfordshire.

Another approach to the "quarter squares" method is the addition of a high-frequency triangular wave, symmetrical about earth, with an amplitude

$$> (|X| + |Y|),$$

to the sum and difference signals of X and Y , which are then fed to two operational amplifiers. The modulus of one amplifier's output is made positive, while that of the other is made negative; the product is then obtained from the mean value of the composite signal when they are added.

This can be regarded as a time-division method of obtaining the square-law response, since the signal simultaneously controls the base and the height of a triangle, resulting in the area increasing as the square of the signal.

The latter two methods involve the difference of relatively large quantities, and the possibility of error is great if operational amplifiers are not used.

2.3. The Hall Effect

The Hall effect can be used to multiply two quantities together. One of the main drawbacks is the low output voltage of a Hall element. As a result of this, a stable, high gain, d.c. amplifier, with an output centred on earth potential, would be required to follow the multiplier.

2.4. Time Division Methods

Time division methods can be used to provide one of the forms of pulse modulation: e.g., pulse width modulation (basically, modulation of the mark/space ratio); pulse position modulation (phase modulation); modulation of the pulse repetition frequency (frequency modulation).

The use of transistors as switches makes this an attractive method.

2.5. *Assessment of Methods*

From this survey of possible methods it was considered better not to use mechanical methods because of their comparative bulk and slowness. A simple system relying on non-linear response was also rejected, since such arrangements are prone to large errors. More advanced systems based on the "quarter-squares" method tend to become strewn with operational amplifiers, and, likewise, were not followed further.

The Hall-effect multiplier was seriously considered, but it was decided that an arrangement using transistors as switches in a time-division system would be preferable, and this method has been adopted.

3. Selected Methods

Two multipliers were designed. Both employ chopper transistors to provide amplitude modulation proportional to one input signal. The modulation which is proportional to the other input signal is provided by varying the mark/space ratio in the first case, and by shifting the phase of one square wave signal with respect to another in the second case.

3.1. *Description of Multiplier using Mark/Space Ratio Modulation*

Figure 1 shows the arrangement in block schematic form.

The input, $\pm Y$, is applied to the balanced chopper as a push-pull signal; a paraphase stage providing polarity inversion. This is necessary in order to keep the mean level of the Y inputs to the chopper at zero, and obtain the four quadrants of output. Thus even though a large Y signal is present, there will be no output when, in the absence of modulation of the X input, the mark/space ratio is unity.

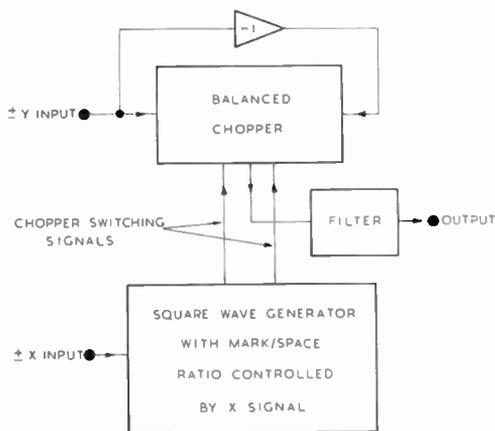


Fig. 1. General block schematic.

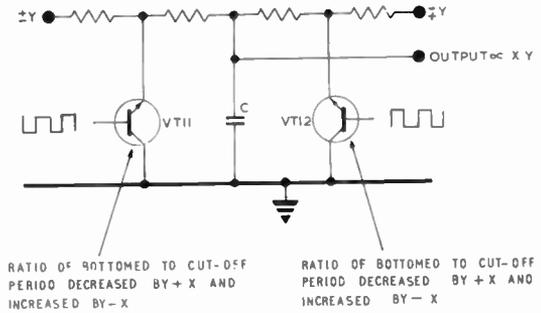


Fig. 2. Balanced chopper (basic circuit).

The filter is employed to reduce the residual output at the chopper frequency. To obtain a low value of this output, using a simple filter, it is necessary for the ratio of the chopper frequency to the highest modulation frequency to be high.

It is desirable to keep the output time-constant to a minimum in order to reduce the attenuation and phase shift at signal frequencies. This again is helped by maintaining a high ratio.

3.2. *Balanced Chopper (Basic Circuit)*

The basic circuit of the balanced chopper employed is shown in Fig. 2. The switching signal is arranged so that the associated transistor is bottomed on its positive-going excursions and cut off on its negative-going excursions. The switching signals to the two transistors are in antiphase.

When a voltage, $+ Y$, is applied to the left-hand input, a voltage, $- Y$, will appear at the right-hand input via the paraphase stage. In the absence of any X signal the current into the capacitor C will be equal and opposite on each half cycle of the switching signal—resulting in no mean potential being developed. In the presence of a positive X signal the left-hand input will contribute more to this charge, and the right-hand input will contribute less. Thus the mean charge will establish a positive voltage at the output, which will be proportional to the magnitude of Y and to the value of the mark/space ratio, i.e. to the magnitude of X . If the polarity of either input is changed, so will the polarity of the output, while if both input polarities are changed the polarity of the output will not.

3.3. *Modulator*

The generation of the chopper switching signal, with its mark/space ratio controlled by the X signal, centres round a saturable-core transformer. The principle is shown in Fig. 3. If a voltage E is applied to an inductor (as in Fig. 3 (a)) the flux in the core will rise at a linear rate proportional to $-E$ in order to satisfy the relationship $E = -N(d\phi/dt)$. After a given time, t , which will be inversely proportional to the voltage, the core will saturate.

Figure 3(b) shows the connection of such an inductor alternately across voltages e_1 and e_2 . After S1 makes, the time to saturation will be inversely proportional to e_1 , while after S2 makes, the time to saturation will be inversely proportional to e_2 . If, now, the process of the core going into saturation can be made to change the switched state, we shall obtain a wave form with a mark/space ratio of e_2/e_1 .

Figure 3(c) gives the basic circuit of such a generator. Two feedback windings are placed on the core and connected to VT13 and VT14 in such a direction as to maintain the transistor which is bottomed in that state until the core saturates, when the conducting and nonconducting roles are reversed. The voltages acting in the circuit are $V \pm X$ and $V \mp X$, where the modulus of V is always greater than that of X .

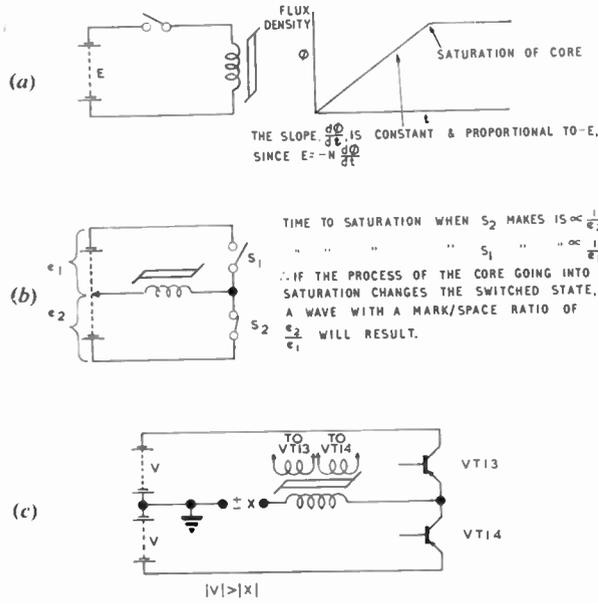


Fig. 3. Square wave generator.

3.4. Multiplier Basic Equations

The mark period is equal to $\frac{k}{V-X}$, and the space period is equal to $\frac{k}{V+X}$, where k is a constant. The mean value of the output of the multiplier will thus be equal to

$$kY \left(\frac{1}{V-X} - \frac{1}{V+X} \right) / \left(\frac{1}{V-X} + \frac{1}{V+X} \right) = \frac{kXY}{V}$$

It is interesting to note at this point that the output is proportional to $1/V$. This fact could be put to use if a divider were required. The necessity of having a push-pull signal for V could be avoided by adopting a

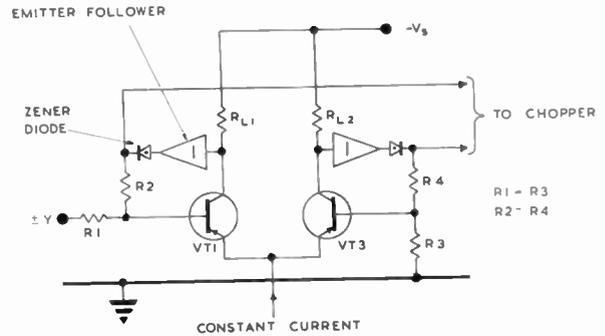


Fig. 4. Y amplifier (basic circuit).

conventional push-pull circuit. This would, however, necessitate having a push-pull X signal. If the X signal were made a constant quantity, circuit difficulties would be reduced, the result would be KY/V , where K is a new constant. The resulting circuit would give two quadrant division, and switching would be needed to give the remaining two quadrants.

Since the frequency of operation is inversely proportional to the total time of one complete cycle, the frequency equals

$$k / \left(\frac{1}{V-X} + \frac{1}{V+X} \right), \text{ or } \frac{k}{2} \left(\frac{V^2 - X^2}{V} \right).$$

Thus this is at a maximum when $X = 0$, and drops as the modulus of X increases, following a parabolic law.

3.5. Y Input Amplifier

The amplifier producing push-pull signals about earth potential from the Y input is shown in outline in Fig. 4. The arrangement is that of a long-tailed pair phase-splitter which produces a balanced output from a single-ended input. Each output is succeeded by an emitter-follower, with a Zener diode to return the mean signal potential to earth level. Voltage feedback from these points, via R1, R2 and R3, R4, linearizes the amplifier's response, and keeps the output impedances low.

4. Final Circuit

The final circuit is shown in Fig. 5. The transistor, VT2, provides a stabilized constant current for the long-tailed pair VT1, VT3. The thermistor Th1 provides temperature compensation. The mean output potential at the junction of R10, R15, is used as negative feedback to VT2 to reduce any in-phase output, while C1 decouples the base of VT2 from the chopper signal at VT11, VT12.

Transistors VT5 and VT7 provide compensated emitter loads for the emitter-followers VT4, VT6. In the absence of the cross-coupling provided by R11 and R16, and given suitable bias, the transistors VT5, VT7, would act substantially as constant current elements. The cross-couplings allow the emitter-follower tran-

sistor whose emitter voltage is the more negative, and is therefore drawing more current from the chopper network, to have a reduced supply of current from its feed transistor. Similarly the emitter-follower transistor which is drawing less current from the chopper network is provided with more current from its feed transistor; R14 assists in this action. The circuit incorporating VT5 and VT7 can be regarded as applying a negative resistance to the output. The result is that the emitter-followers have a smaller effective loading and the Zener diodes, MR3 and MR4, have a smaller variation of current through them than would otherwise be the case.

In the square-wave generator, the X input signal is fed to the emitter-follower VT9, which has as its emitter load VT10, C3 and the winding, 5-6, of T1. The output impedance of the emitter-follower is reduced to a low value at the chopper frequency by the shunting capacitor C3. This capacitor must not prevent the emitter follower from developing the required voltage by causing it to become cut-off. The feedback from the collector of VT9 to the base of VT10 reduces the effective loading on the emitter follower by causing the current drawn by VT10 to vary inversely with the load current. The emitter follower output controls the mark/space ratio. It is desirable to have a means of adjusting this ratio to unity when the X voltage is zero. This is achieved by including a transistor, VT8, which feeds a constant current into R23 in the base circuit of VT9. This current can be adjusted by RV4, in the emitter circuit of VT8, to offset the emitter/base barrier potential of the silicon transistor VT9 and, if necessary, to provide the slight adjustments required to cover any inequality

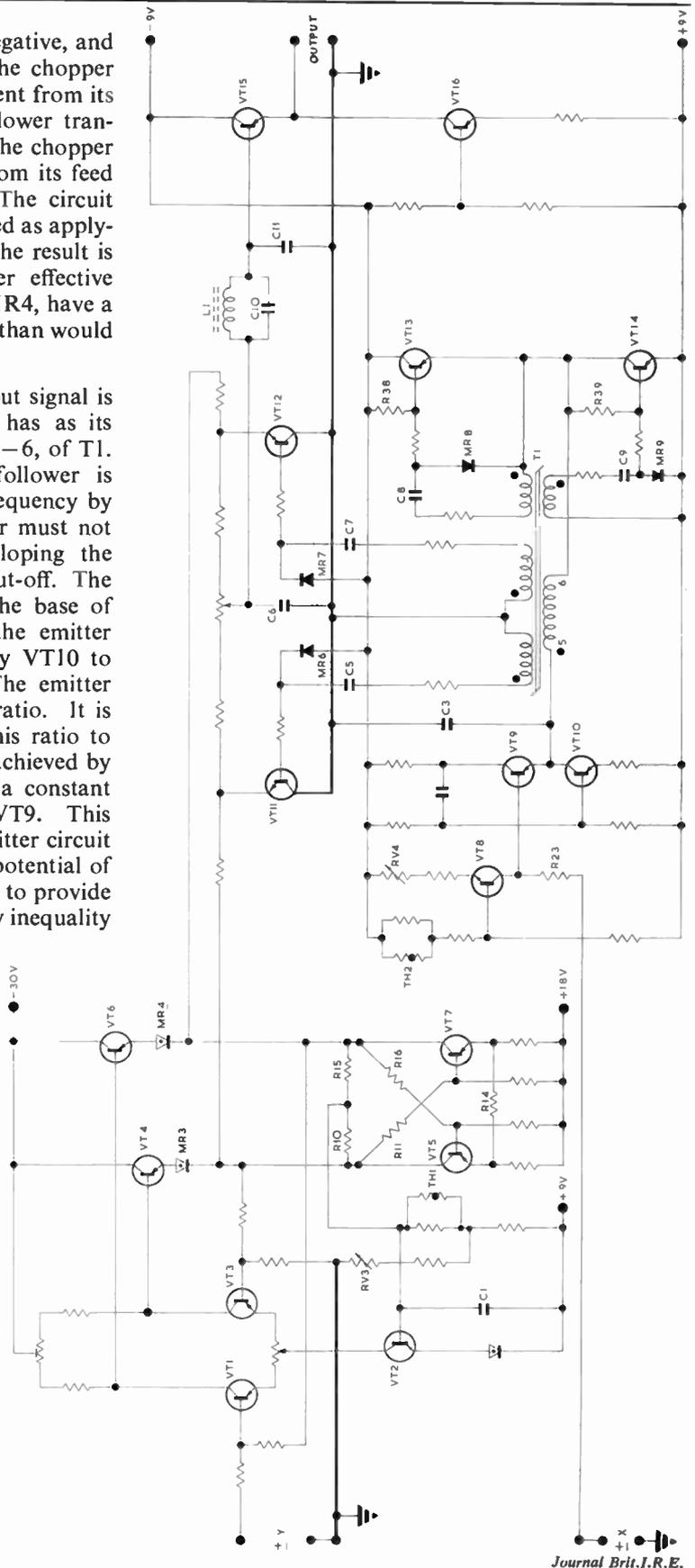


Fig. 5. Four-quadrant multiplier.

of the positive and negative supply rails to the generator. A thermistor, Th2, provides compensation against movement of the mark/space ratio with temperature. The circuit of the square wave generator centres around VT13, VT14 and T1. In order to keep the collector current requirements of VT13, VT14, to a minimum, the characteristics required of the core for T1 are that the material employed should reach its sharp saturation "knee" at a low flux density and the core volume should be low. A small number of laminations of mumetal was used in practice. Resistors R38 and R39 provide the necessary starting bias current, while the components C8, MR8 and C9, MR9 provide d.c. restoration so that the drive current available to the base circuits of the transistors will not change greatly as the mark/space ratio changes.

Similarly, the signals fed to the chopper circuit are d.c. restored to -9 V by C5, MR6 and C7, MR7. This ensures that the chopper transistors VT11, VT12 can be bottomed and cut-off by the switching signal at all values of the Y input.

C6, L1, C10, C11 form the filter. The filter is tuned so that the ripple output is not greater than 5 mV over

the full range of X input, where the switching frequency approximately covers the range of 1.5 kc/s to 2 kc/s.

The provision of an output emitter-follower VT15, with its constant current feed VT16, means that in order to have the output at earth potential, the mean potential of the output of the Y amplifier must be made negative of earth, by means of RV3, to offset the barrier potential of the emitter/base junction of VT15.

The possibility of improving the linearity of the magnetic circuit by the use of negative feedback was considered, but the phase modulation system described below was adopted in preference.

5. Phase Modulation System

It was considered that this approach would avoid the use of a d.c. single-ended to push-pull amplifier, and also the necessity for d.c. restoration circuits, and in these respects would be superior to the first system.

5.1. Block Schematic with Waveforms

Figure 6 shows the overall block schematic with the pertinent waveforms.

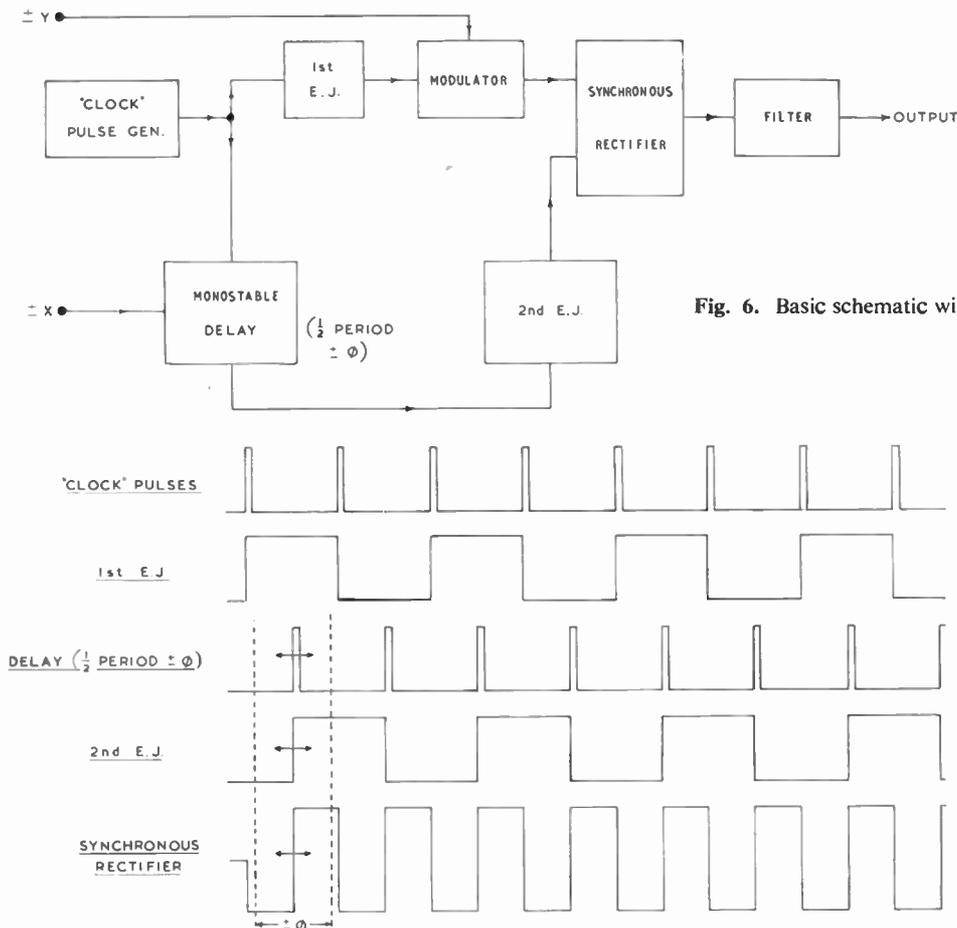


Fig. 6. Basic schematic with waveforms.

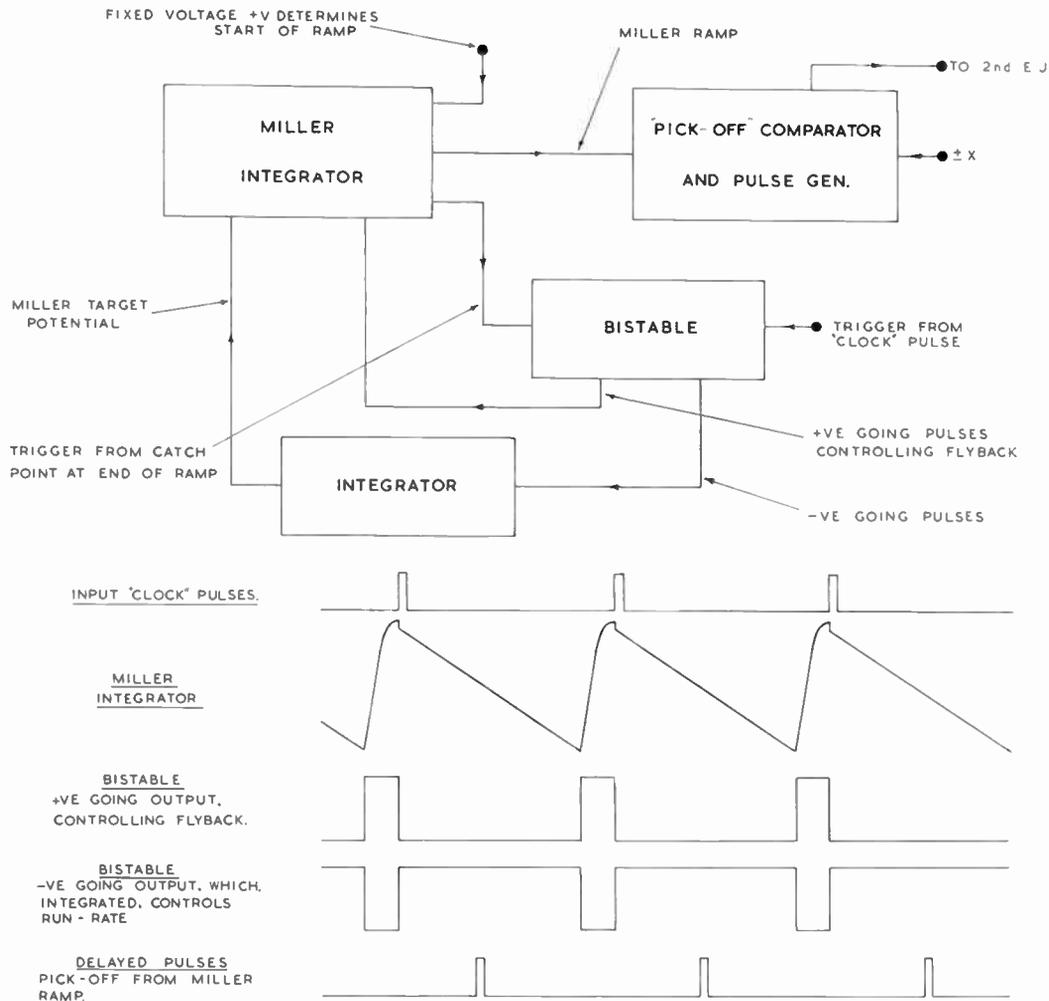


Fig. 7. Monostable delay.

The voltage, $\pm Y$, into the modulator controls the amplitude of the signal being fed by the first Eccles-Jordan circuit to the synchronous rectifier. The amplitude and polarity of the feed to the synchronous rectifier are both directly controlled by the Y signal—there being no signal passed when Y is zero.

Considering the circuit when the X input is at zero and the Y input held at a fixed positive value, we have the condition where the two Eccles-Jordan circuits, 1st and 2nd E.J., are presenting constant amplitude inputs to the synchronous rectifier, but with the phase of the output of 2nd E.J. being delayed on that from 1st E.J. by $\pi/2$ radians. This phase delay is obtained by driving the first Eccles-Jordan circuit directly from a "clock" pulse generator, while the second Eccles-Jordan circuit is driven from the output of a delay circuit, which delays the pulses by half the "clock" period. Under these conditions the mean

output from the synchronous rectifier will be zero, as is shown by the relevant waveform.

If now an X signal is applied to the delay circuit the delay period will change and the mean output from the synchronous rectifier will no longer be zero. The change in the delay, ϕ , is made directly proportional to the X input. Positive values of X reduce the delay, while negative values of X increase it. The effect is to phase modulate the signal feeding through the 2nd E.J. into the synchronous rectifier. The output of the synchronous rectifier will be directly proportional to the degree of phase modulation, i.e. to the magnitude of X , and the polarity will be dependent on whether ϕ was positive or negative, i.e. to the polarity of X .

When an X signal is applied and the Y signal is at zero, the mean output of the synchronous rectifier will also be zero.

Conditions thus exist for four-quadrant multiplication. All the elements in this block schematic are fairly standard items, with the exception of the delay block, which might be called the "hub" of the multiplier.

5.2. Delay Schematic

Figure 7 gives further details of this unit and shows the waveforms at various points.

The basic timing element is a voltage ramp, produced by a Miller integrator. The ramp is initiated by a "clock" pulse, and the potential at a point on the ramp halfway between the "clock" pulses is arranged to be earth potential. Now the X signal controls the "pick-off" point for generating the pulses which trigger the second Eccles-Jordan circuit. Thus, when the X signal is at zero, the delayed pulses will have a delay of half the clock period, as required. The application of X signals will then provide an earlier or later "pick-off".

Without compensation, a change of the "clock" period, or of the slope of the run-down, will be equivalent to a change in the X input. In order to maintain the values of the "clock" period and the ramp slope correctly related it is desirable to incorporate a feedback, or servo, system, so that if the "clock" frequency changes, or if the Miller run-rate changes, a correction will be made.

This feedback is derived from the bistable circuit shown, which controls the quiescent period of the Miller stage. The bistable is made to change state by a trigger derived from a fixed potential reached at the end of the Miller run-down. This cuts off the Miller amplifier and allows the integrating capacitor to be recharged. At the next "clock" pulse the bistable state is changed back again, and the Miller run is initiated. If now the run-rate or the "clock" frequency changes, the time which the bistable spends in the "flyback" position will change. An output from this bistable circuit is integrated and the integrated output is fed back to modify the Miller target potential, the polarity being that required to provide the desired correction.

5.3. Alternative Arrangement

A simplification of the circuit can be achieved by letting the Miller ramp generator be free-running. This can be done by changing the bistable circuit

controlling the flyback period to a monostable circuit—the internal time constant will now determine the end of the flyback period and the beginning of the run. The flyback pulse can replace the "clock" pulse for triggering the first Eccles-Jordan circuit.

The drawback here is that the ratio of the Miller run-time to the flyback time, which should be constant, may vary. With care in the design of the Miller circuit, and the monostable circuit controlling the flyback, it would appear possible still to maintain a high degree of accuracy while retaining the advantage of circuit simplification. An investigation of this arrangement is in progress.

6. Conclusion

Two four-quadrant, analogue, multipliers have been described, one in detail and one in outline. Performance figures for the first system, on the basis of the prototype, are: the accuracy is better than $\pm 1\%$ of full scale output over the ambient temperature range 10°C to 35°C (this range would appear capable of extension if the need arose); the residual output at chopper frequency is not greater than 0.125% of full scale output and the -3 dB point is not less than 10 c/s. In other applications this bandwidth could be increased to several hundred cycles per second by suitable circuit modification. Mains supply voltage variations of $\pm 10\%$ have a negligible effect on the output.

The second system exists in "breadboard" form, and to date the results obtained would indicate that a superior performance to the above can be expected. An accuracy of better than 0.2% of full-scale deflection, with the -3 dB point at 500 c/s, is typical of early results.

7. Acknowledgments

The author would like to thank the directors of J. Langham Thompson Ltd. for permission to publish this paper.

Particular acknowledgments are made to Mr. F. Cox and Mr. J. V. Sims for their assistance in the discussions and work on these multipliers, and in the preparation of the paper.

Manuscript first received by the Institution on 3rd April 1962, in revised form on 11th May and in final form on 5th July 1962 (Paper No. 778.)

© The British Institution of Radio Engineers, 1963

Automatic Landing Developments

The ability to land and take-off under "zero visibility" weather conditions is one of the important developments to which airlines can look forward in this decade.

The Technical Committee of the International Air Transport Association, which made this claim in a report submitted at the 18th Annual General Meeting of I.A.T.A. in Dublin, say that its principal objective is to assure the greatest possible regularity and economy of operation consistent with an overriding concern for safety.

Landing and take-off under "zero visibility" weather conditions is an I.A.T.A. goal which will be reached in three progressive stages.

The new range of British jet airliners is designed to carry equipment to enable them to effect one stage of the fully automatic landing which would be used in "zero visibility" conditions. Vickers *VC-10* and *Super VC-10* four-jet inter-continental airliners, will have this built-in provision for fully automatic landing equipment. The three-jet de Havilland *Trident* medium range airliner has also been designed to accommodate the automatic landing equipment pioneered and developed by Smith's Aviation Division. Like the *VC-10*, the *Trident* will first be fitted with "autoflare" equipment. This is an automatic control system to produce a "flare-out" or change in the aircraft's fore and aft attitude and its flight path during landing at the point where it levels out before the final touchdown on the runway. During this "flare-out" the pilot remains in control of the rudder and ailerons.

With the *VC-10*, all the equipment necessary to achieve automatic "flare-out" is carried in the aircraft. The basic element in this system used in the *VC-10*, the PB.20 autopilot, has been developed in the U.K. by Elliott Brothers (London) into a twin autopilot system. It is integrated with the *VC-10's* powered flying control system to provide coupled approach with automatic control of the throttles to the point of touchdown.

On the *VC-10* the two autopilot systems are completely separate with only one controlling the aircraft at any time. A failure in any of the component parts of either system automatically operates a change-over to one of the other systems. Using this "autoflare" system, the *VC-10* will make a normal approach using i.l.s. (Instrument Landing System) down to the height at which the pilot would normally take over full control of the aircraft. This is about 150–200 ft. At this point i.l.s. is automatically disconnected leaving the autopilot to control the throttles and elevators. The *VC-10* will continue to descend, maintaining a constant attitude decided by a computer, to the "flare-out" point at 50 ft above the runway. At this point the flare computer, which uses

signals fed into it by a radio altimeter, takes over control. It automatically closes the throttles and the landing is completed.

Using the "autoflare" system, the pilot must still provide directional control with the rudder from 150–200 ft to touchdown using visual references. Full automatic landing requires more computers and ground installations than are necessary for "autoflare". The *VC-10's* system, being self-contained and therefore dependent on the existing i.l.s. only during the early stages of the landing approach, will provide the basic experience for development of a fully automatic landing system.

A further development for this blind landing system is now being tested which will also control an aircraft's landing run along the runway (roll-out). Developed as a private venture by the Radio Division of Standard Telephones and Cables, a new type of Instrument Landing System Localizer aerial may become the first in the world to meet I.C.A.O. recommendations for full automatic landing including the roll-out phase.

The i.l.s. localizer aerial used in this system measures 160 ft wide and 18 ft high. The great size of the aerial is dictated by the need for high accuracy of its radio beam. Because it will be located 1600 ft from the stop end of the runway and on an extension of its centre line, this huge aerial has been designed to break upon impact should it be struck by an overshooting aircraft.

Based on a localizer aerial which is standard equipment at all major airports in the United Kingdom and many overseas airports, the new aerial emits a narrow 4 degrees wide radio beam of high stability. It is thus unaffected by weather conditions or, perhaps more significant, by aircraft flying over it which can deflect the beam.

The beam is picked up by airborne receivers in airliners equipped for automatic landings, and is used to operate a "cross pointer" indicator in the cockpit. The information from this instrument enables the pilot to steer a predetermined course to the runway threshold. This information is fed into the autopilot control system when the aircraft is being landed automatically. In certain cases the equipment could be used by the pilot to take-off in zero visibility weather conditions.

A description of the British automatic landing system (Autoland) was given in a paper by Mr. J. S. Shayler, "Radio-Guidance Elements of the B.L.E.U. Automatic Landing System for Aircraft", published in the January 1961 *Journal*; more recently Mr. F. R. Gill gave a paper on "The Problems of Blind Landing of Aircraft" before the Merseyside Section and a report on this paper appears in the January 1963 issue of the *Proceedings of the Brit.I.R.E.*

Recommended Method of Expressing Electronic Measuring Instrument Characteristics

8. RECTANGULAR-PULSE GENERATORS

Prepared by the Technical Committee of the Institution and based on a report compiled by Commander J. R. Young, R.N., M.A. (Associate Member)†

Introduction

This is the eighth set of recommendations in a series which has the objective of influencing the uniformity of presentation of information on the features, characteristics and performance of electronic measuring instruments and thus assisting in their comparative assessment and selection by potential users. The establishment of standards of performance is not an objective of these recommendations.

The characteristics dealt with in this set of recommendations refer to instruments which generate a rectangular pulse which may be varied in amplitude and duration. Certain generators are capable of producing two or more pulses, the additional pulses being subject to variable delays relative to either the main pulse or a triggering source.

Table of Contents

1. General Details	3. Triggering Facilities	6. Voltage Output
1.1 Power supply requirements	3.1 Internal	6.1 Ranges
1.2 Temperature range	3.2 External	6.2 Accuracy of calibration
1.3 Overall dimensions	3.3 Internal impedance	6.3 Stability
1.4 Weight	4. Characteristics of the Main Pulse	6.4 Attenuator
1.5 Construction and finish	4.1 Pulse duration	7. Source Impedance
1.6 Valve and/or transistor complement	4.2 Rise time	7.1 Main pulse
1.7 Accessories	4.3 Decay time	7.2 Additional pulse
1.8 Ancillary equipment	4.4 Overshoot	8. Synchronizing Output
	4.5 Maximum duty cycle	8.1 Length of synchronizing pulse
	4.6 Delay	8.2 Delay range
2. Pulse Repetition Frequency	5. Additional Pulse Characteristics	8.3 Accuracy of delay
2.1 Range	5.1 Provision of additional pulse	8.4 Re-setting accuracy
2.2 Calibration accuracy	5.2 Pulse duration	8.5 Polarity of synchronizing pulse
2.3 Re-setting accuracy	5.3 Delay	8.6 Amplitude
2.4 Incremental adjustment	5.4 Rise time	8.7 Rise time
2.5 Stability	5.5 Decay time	8.8 Source impedance
	5.6 Overshoot	

FEATURE	METHOD OF EXPRESSION	REMARKS
Part 1.—GENERAL DETAILS		
1.1 Power supply requirements	.. volts a.c./d.c. .. c/s and/or battery .. watts (voltage change .. %)	Maximum voltage variation for which the stated accuracies hold good must be given.
1.1.1 Protection	.. in mains circuit .. in h.t. circuits	Characteristics to be stated.
1.2 Temperature range	.. °C to .. °C	Maximum ambient temperature range for which the stated accuracies hold good with nominal power supplies.

† Approved by the Council for publication on 26th September 1962 (Report No. 26).

TECHNICAL COMMITTEE

FEATURE	METHOD OF EXPRESSION	REMARKS
1.3 Overall dimensions	Height .. in (.. cm) Width .. in (.. cm) Depth .. in (.. cm)	
1.4 Weight	.. lb (.. kg)	
1.5 Construction and finish		Where the instrument conforms to particular standards or specifications, these should be named.
1.6 Valve and/or transistor complement	Type numbers	State if special selection is required.
1.7 Accessories		Details of any connectors and adaptors, etc., supplied.
1.7.1 Connectors		State type.
1.8 Ancillary equipment		Where specific filter plug-in adaptors or additional attenuator and probes are recommended for use in conjunction with the instruments, their features should be expressed.

Part 2.—PULSE REPETITION FREQUENCY

2.1 Range	.. c/s to .. kc/s in .. bands	
2.2 Calibration accuracy	± .. %	This is the maximum error between the indicated and actual frequencies.
2.3 Re-setting accuracy	± .. %	This is the re-setting accuracy of the main control.
2.4 Incremental adjustment		The method used should be stated.
2.4.1 Range	± .. c/s or ± .. % of main setting	
2.4.2 Calibration accuracy	± .. %	
2.5 Stability	± .. % change without re-adjustment	Maximum change in frequency over any period of 10 min within a 7-hr period commencing 60 min after switching on. During this 10-min period the supply voltage and temperature are assumed to be sensibly constant.

Part 3.—TRIGGERING FACILITIES

3.1 Internal		
3.2 External		State type of trigger pulse, sine wave, amplitude minimum and maximum, polarity and minimum pulse duration.
3.3 Internal impedance	.. pF in series with .. kΩ	

EXPRESSING ELECTRONIC MEASURING INSTRUMENT CHARACTERISTICS

FEATURE	METHOD OF EXPRESSION	REMARKS
Part 4.—CHARACTERISTICS OF THE MAIN PULSE		
4.1 Pulse duration	.. s to .. s	Measured at 50 per cent of the peak amplitude. State whether continuously variable or adjustable in steps and specify multiplying range used if any.
4.1.2 Accuracy of pulse duration	.. %	Accuracy to be stated as a percentage of indicated pulse duration.
4.2 Rise time	.. s (10% to 90%)	Range on which rise time is measured to be stated.
4.3 Decay time	.. s (90% to 10%)	Range on which decay time is measured to be stated.
4.4 Overshoot	Not more than .. %	To be stated as percentage of nominal pulse amplitude.
4.4.1 Tilt	Not more than .. %	To be stated as percentage of nominal pulse amplitude.
4.5 Maximum duty cycle	.. % of cycle time or maximum mark to space ratio	State whether warning lamp or other form of warning indication fitted.
4.6 Delay	.. s to .. s with reference to commencement of trigger pulse	State if variable.
4.6.1 Accuracy of delay	.. %	Accuracy to be stated as a percentage of indicated delay.
4.6.2 Re-setting accuracy	.. %	Accuracy to be stated as a percentage of indicated delay.

Part 5.—ADDITIONAL PULSE CHARACTERISTICS

5.1 Is additional pulse provided?	YES/NO	
5.2 Pulse duration of additional pulse	.. s to .. s	Measured at 50 per cent of the peak amplitude. State whether continuously variable and ranges used.
5.2.1 Accuracy of pulse duration	.. %	Accuracy to be stated as a percentage of indicated pulse duration.
5.3 Delay	.. s to .. s	Delay to be quoted relative to commencement of trigger pulse and restriction on overlap time to be stated. State whether continuously variable and ranges used.
5.3.1 Accuracy of delay	.. %	Accuracy to be stated as a percentage of indicated delay.
5.4 Rise time	.. s (10% to 90%)	Range on which measured to be stated.
5.5 Decay time	.. s (90% to 10%)	Range on which measured to be stated.
5.6 Overshoot	Not more than .. %	To be stated as percentage of nominal pulse amplitude.
5.6.1 Tilt	Not more than .. %	To be stated as percentage of nominal pulse amplitude.

FEATURE	METHOD OF EXPRESSION	REMARKS
Part 6.—VOLTAGE OUTPUT		
6.1 Ranges	.. V to .. V	State positive or negative. Give lowest scale reading that is within stated accuracy for each range.
6.1.1 Voltage ranges of additional pulse	.. V to .. V	Restriction regarding additional pulse with respect to main pulse (if any) to be stated.
6.2 Accuracy of calibration	.. % of scale reading	Maximum error to be given for each range.
6.2.1 Calibration reference	Internal or external and accuracy	Give type and method of setting.
6.3 Stability	.. % change without re-adjustment	Maximum change over any period of 10 min within a 7-hr period commencing 30 min after switching on.
6.4 Attenuator		Give details of attenuator type and ranges.
6.4.1 Attenuator accuracy	.. % of scale reading	Maximum error to be given for each range.
6.4.2 Attenuator resetting accuracy	.. % of scale reading	Maximum error to be given for each range.
Part 7.—SOURCE IMPEDANCE		
7.1 Main pulse source impedance	.. Ω on .. range	Impedance on each range to be given if different and if separately available. State resistive and reactive components.
7.2 Additional pulse source impedance	.. Ω on .. range	
Part 8.—SYNCHRONIZING OUTPUT		
8.1 Length of synchronizing pulse	.. s	
8.2 Delay range	.. s to .. s	Delay to be stated relative to commencing of trigger pulse and if continuous or stepped control.
8.3 Accuracy of delay	.. %	Accuracy to be stated as percentage of indicated delay.
8.4 Re-setting accuracy	.. %	Accuracy to be stated as percentage of indicated delay.
8.5 Polarity of synchronizing pulse		
8.6 Amplitude	.. volts	Peak volts to be stated into given load.
8.7 Rise time	Less than .. s	
8.8 Source impedance	.. Ω	

The other Recommendations in this Series are:

- | | |
|---|--|
| 1. "Amplitude-modulated or frequency-modulated signal generators" (January 1958). | 4. "Valve voltmeters" (April 1960). |
| 2. "Cathode-ray oscilloscopes" (January 1959). | 5. "A.c. bridges" (August 1960). |
| 3. "Low-frequency generators" (March 1960). | 6. "Stabilized power supplies" (February 1961). |
| | 7. "Wave and distortion analysers" (October 1962). |

The dates given are the issues of the *Journal* in which the particular recommendation appears; reprints of individual items may be obtained from the Institution, price 1s. 6d. each. Complete sets of the eight reprints are available, price 10s. 6d. per set.

Long Range V.H.F. Air/Ground Communications

By

D. B. CLEMOW, B.Sc.(Eng.)†

AND

E. H. BRUCE-CLAYTON‡

Presented at a meeting of the Radar and Navigational Aids Group in London on 11th April 1962.

Summary: In recent years among airlines and civil aviation administrations interest has developed in the use of v.h.f. radiotelephony for air ground communications at ranges much in excess of those conventionally regarded as the maximum obtainable. The paper reviews the reasons for this development, describes a series of tests conducted from an experimental station in the Persian Gulf and analyses the results obtained. The possibilities of developments to achieve further increases in range are discussed.

1. Introduction

In civil aviation, the need for long range air/ground communications derives from the necessity for aircraft to maintain contact with the Air Traffic Control Centre throughout the latter's area of responsibility. This may require ranges of up to 1000 nautical miles or more, often extending over oceans or undeveloped territories, which preclude the use of intermediate ground stations connected back to the Centre by landlines or radio links. Except in the comparatively few areas of the world where it has been economic and practicable to adopt the latter solution, the normal medium for long range air/ground communication in civil aviation is high-frequency radiotelephony, using conventional a.m. d.s.b. modulation and frequencies generally between 3 and 13 Mc/s.

The development of air traffic, and the ground control services, has imposed an increasing load upon the air/ground channels, which has emphasized the shortcomings of h.f. radio-telephony. Mutual interference both between ground stations and aircraft has worsened as more users share the available frequencies, and this has been aggravated by unauthorized users—a symptom of deteriorating international discipline in the h.f. spectrum. The normal hazards of h.f. communications—such as rapidly varying propagation conditions, high atmospheric noise levels and selective fading—are of course still present, and during the next few years will indeed become more evident with the decline of the sunspot number.

At a typical h.f. r.t. ground station four or five frequencies are monitored and depending on the time of day, distance from the station and sunspot cycle epoch so the aircraft chooses a suitable frequency upon which to call. At distances of about 300–400 nautical

miles it is necessary to use the lower h.f. bands which inevitably are subject to high noise levels. The consequence is that communication is difficult and sometimes impossible.

Aircrews, who generally now have to operate h.f. r.t. without the aid of a specialist radio officer, are becoming exasperated with the inefficiency and shortcomings of h.f. r.t. in various overseas regions. A solution must be found, and in recent years attention has been directed to the possibility of using v.h.f. r.t. to ranges considerably beyond that conventionally accepted as the maximum, with a view to easing the load on the h.f. r.t. or perhaps on some routes replacing it altogether.

The use of the aeronautical v.h.f. band (118–132 Mc/s) has a number of advantages over h.f.; atmospheric noise is low; propagation is much less dependent upon frequency, season, or time of day; due to the absence of an ionospheric reflected signal the management of frequency allocation is much more effective; and high-gain directional aerials are of acceptable dimensions and cost.

For practical planning purposes in civil aviation (where 50-watt ground transmitters and a unity gain aerial mounted 60 ft high are typical) it has been assumed that the air/ground v.h.f. range was confined to a radio horizon calculated from a standard lapse rate of refractive index with height. Of course it was known that considerable variations from the standard atmosphere occurred—tending, in the warmer climates, to increase range.

Also, it had become increasingly evident from work in other fields that the use of greater gain and system sensitivity would enable a signal to be detected beyond this conventional horizon. But in practice the familiar formula

$$R = 1.22\sqrt{h}$$

† International Aeradio Ltd., Engineering Division, Southall, Middlesex.

‡ International Aeradio Ltd., 40 Park Street, London, W.1.

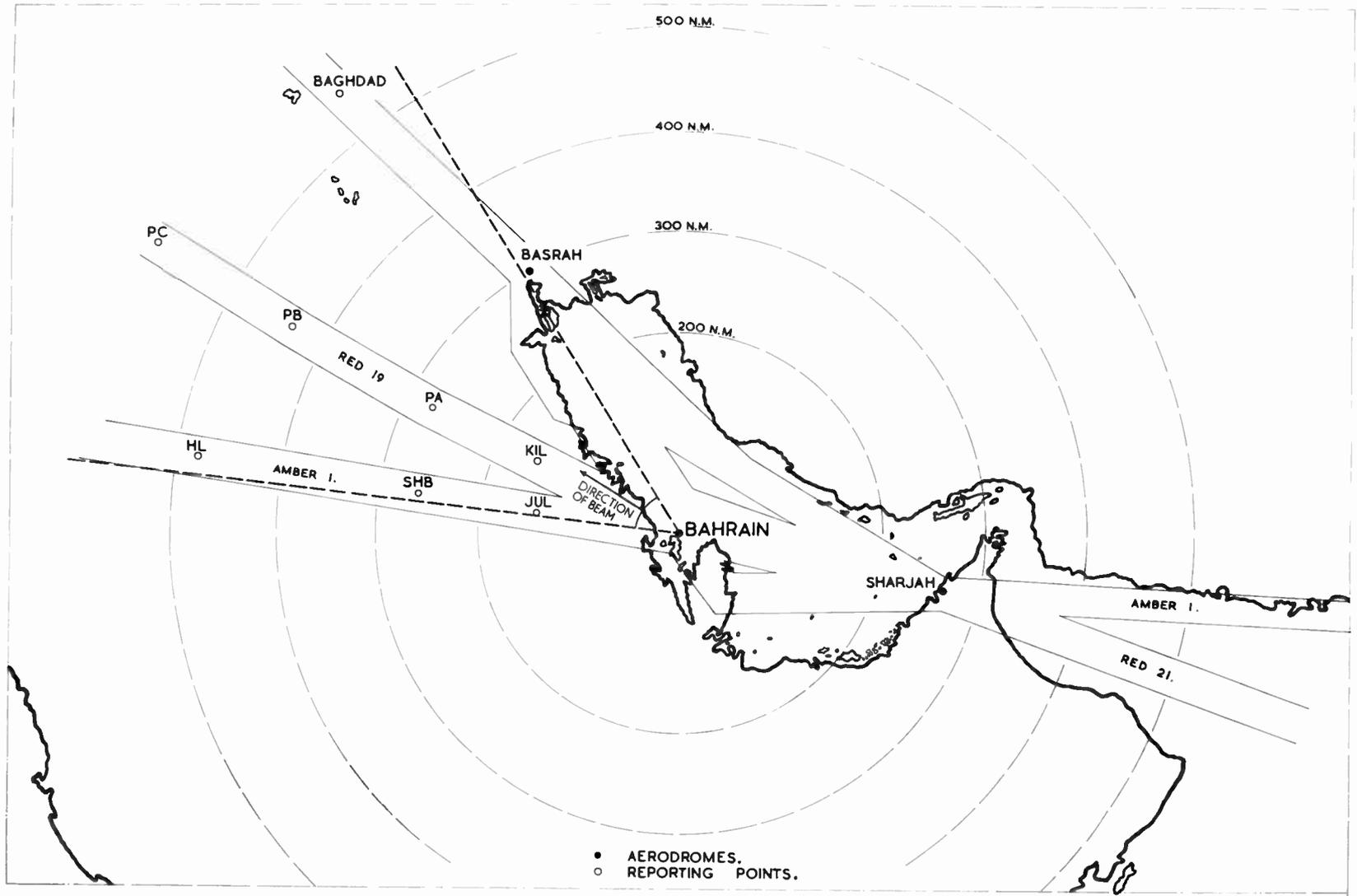


Fig. 1. Relationship of beam to Airways in Bahrain area.

where R = range in nautical miles and h = height in feet,

(giving for example a maximum range of 244 nautical miles at 40 000 ft), conditioned people's thinking, and it is only in the last few years that civil aviation has begun to look beyond this horizon.

If the possibility of greater v.h.f. range was to be exploited for practical operational application it was necessary to establish what ranges might be obtained with given conditions and equipment. To this end, American, British and French authorities and airlines have conducted various trials during the last four years—the results of these trials will be summarized later. The planner must also be given another piece of information, and one that has not always been given sufficient weight in the published results of trials. This is the percentage of probability that can be placed upon the range predictions. Ranges of, say, a thousand miles achieved on a few occasions are technically most interesting, but cannot form the basis for practical planning. It is thought that if authorities are to be induced to install long range v.h.f. stations and to change their operational practices they must receive an assurance that over a period of say a year, a given range will be achieved at least by 90% of the aircraft, or by the same aircraft nine times out of ten, at the first attempt. The theoretically desirable goal of 100% probability is not a realistic and economic possibility in practical operational conditions—nor is it essential, because ground station coverage planning would always allow a range margin.

2. The Bahrain Trials

In March 1956 the authors' company started a series of tests in the United Kingdom using transmitters with powers in excess of 50 watts and directional high gain aerials, but the station was not ideally sited and for operational reasons it was not possible to amass the amount of data necessary to predict the range probability. Hence, two years ago, it was decided to establish a station in a more suitable location.

It was noticed that other studies of this kind had been on over-sea paths and generally the stations were sited on high ground. Many air routes are of course over land, and high ground does not always exist, or its use may be costly or inconvenient. So a station was erected at Bahrain Airport in the Persian Gulf, where the site is practically at sea level and the aerial was beamed on the overland route to the North West.

Figure 1 shows a map of the area, with the airways and air traffic control reporting positions marked. The station is sited only 60 ft from the seashore, 4 ft above mean sea level and there are virtually no obstructions in the forward direction.

Along airways Amber 1 and Red 19 the path is over sea for the first 20 nautical miles (n.m.) and thence over Saudi Arabia where the land is essentially arid and flat with a rise to the interior giving heights of 900 to 1500 ft at 350 n.m. Along the airway to the North the path is over sea for the first 250 n.m. and then it is over arid land with no substantial height. To the rear the path is over the Island of Muharraq and whilst this has a maximum elevation of only about 25 ft there are local obstructions such as low buildings and palm groves. Beyond this the path is over sea except between 260 and 315 n.m. where the airways cross the Oman peninsula and the land rises to a height of between 2000 and 3500 ft.

The aerial array, which is common to the transmitter and receiver, comprises eight 6-element vertically polarized Yagi aerials vertically stacked and correctly phased on a 100 ft guyed lattice mast, the height of the centre of the array being 60 ft above the ground. The power gain of the array is 18 dB referred to a half-wave dipole and the horizontal beam width is ± 26 deg to the half power points. A vertically polarized half-wave dipole over a ground plane is fitted to the top of the tower. The operator can switch this to the receiver instead of the Yagi array so enabling a comparison of signal strengths to be made.

The transmitter has a carrier output power of 1 kW with high level modulation. A vogad/clipper unit is fitted to ensure that the modulation depth is kept as high as possible consistent with modulation peaks not over-modulating the transmitter and so tripping the h.t. power supply. The receiver has a noise factor of about 6 dB and a sensitivity of 1 microvolt p.d. for a signal/noise ratio of 13 dB. The muting circuit is noise compensated and is set to operate on the weakest possible signal. It is difficult to be precise about the signal level which will open the muting but in later calculations a figure of 0.5 μ V p.d. has been used which it is certain must be pessimistic. The transmitter and

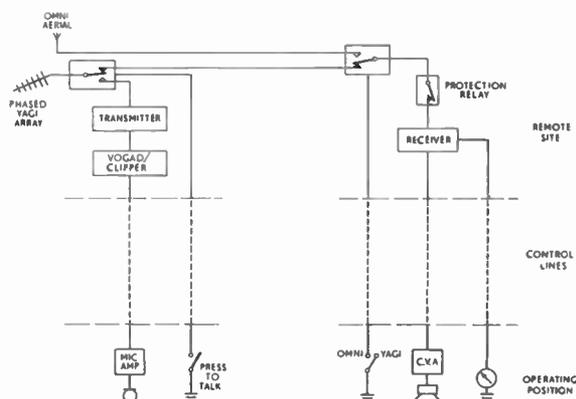


Fig. 2. Schematic diagram of long-range v.h.f. ground installation at Bahrain.

receiver are operated remotely over underground telephone lines, the received signal being amplified in a constant volume amplifier. The a.g.c. voltage of the receiver is presented to the operator on a meter calibrated in 5 dB segments and the operator logs received signal strengths. The installation is depicted in Fig. 2.

The frequency used for the tests was 126.7 Mc/s, this being an operational area control channel for the Bahrain Centre, though a special operator operated the channel for the tests. The aircraft involved were all regular airline services of various companies, and apart from their willing co-operation, no special arrangements existed. This method of conducting the trials had some disadvantages, for example, the airborne equipment characteristics could vary from aircraft to aircraft; because of other duties, or distractions, they could not always call as often as one would have liked, and when the call was not at a reporting point, the position had to be estimated. However, there were substantial advantages too; the trials reflected true operational conditions with a wide cross-section of aircraft, and enabled a very large sample of results to be accumulated over a period of several months.

The trials began in June 1961, and are still running, but this paper presents the results up to and including February 1962. However, during the period June to September no signal strength meter was fitted and estimates of the monitoring operator were relied on. A definite numerical value cannot be placed upon the operator's assessment, but despite this it is felt that they should not be excluded altogether as they cover the summer months and so give an indication of the difference between the two seasons.

3. Characteristics of the Aircraft Involved

Figure 3 illustrates the numbers of aircraft included in the analysis, and their height. Flights subsequent to the fitting of the meter—October to February—are shown separately. It will be seen that about 65% of the aircraft were between 30 and 35 000 ft and only 10% less than 25 000 ft. There are few aircraft flying

in this region at low altitudes. Flights in the main beam are shown separately from those to the rear of the aerial. The sample for the forward direction when the signal strength meter was fitted totals 400 aircraft, on each of which from 2 to 8 measurements of signal strength were obtained during the course of each flight.

The majority of the aircraft flying through the Bahrain Flight Information Region are commercial jet transports such as the DeHavilland *Comet*, the Boeing 707 and the Douglas DC8. In these aircraft a transmitter with a power output of 25 W is commonly used. The ultimate sensitivity of receivers is of the order of 1 or 2 μ V p.d., with one airline having a low noise pre-amplifier with a noise factor of 5.5 dB, but they are all fitted with a muting or squelch circuit which effectively controls the sensitivity. The level at which the muting operates is, in some aircraft, preset on the ground during servicing and in others is under the Captain's control. In the former case, the level varies from Operator to Operator, the lowest known to the authors being 1.5 μ V and the highest 5 μ V p.d. In the latter case the range over which the Captain has control is normally about 1 to 4 μ V p.d.

On all three types of aircraft aerials protruding from the fuselage are used, in some cases above, in other cases below, or a mixture of both. Short feeder lengths having a loss of about 1 dB are common. The radiation patterns of these aerials are nominally circular, but in practice they are distorted, and there can be a difference in gain of about 3 dB from front to rear.

A few piston engine and turbo-prop aircraft are flying in the area. The radio equipment they carry is similar except that in some aircraft the transmitter has a power output of 10 W and the receiver is in general set to open at a rather higher signal level due to the greater noise level in these aircraft.

4. Conduct of the Trials

The method of working was as follows. Approaching aircraft were asked, on h.f. r.t., to begin calling Bahrain as far away as operational circumstances permitted, and to continue to call at intervals after establishing v.h.f. contact. Departing aircraft were asked to call at intervals until contact was lost. The signal strength, aircraft position, height, and heading were logged for each call. Calls well within the conventional radio horizon were not included in the statistics.

The whole of the analysis is based upon measurements of the signal strength in the air to ground direction; from both theory and practice it was known that the range in the opposite direction—from ground to air—was always better.

Since the characteristics of the airborne transmitting installation are all very similar, the only variables are

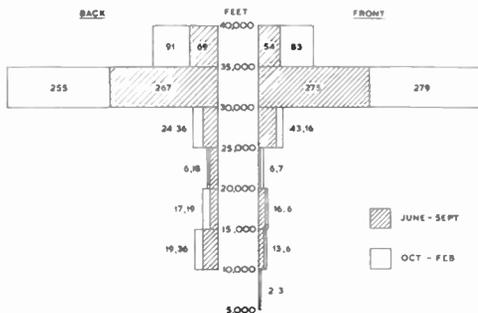


Fig. 3. Number of aircraft included in the analysis.

the state of serviceability of the airborne transmitter and the radiation pattern of its aerial. It was concluded that in practice the difference between aircraft rarely exceeded 5 dB.

5. Analysis of Airways in the Beam

During the period June to September 1961 the signal strength meter was not fitted and therefore the operator's assessment of the signal strength was recorded. Due to the good a.g.c. action of the receiver and the constant volume amplifier, however, it was difficult for him to detect any difference in the signal except when an aircraft was at extreme range. It was realized that inherently this method must give pessimistic maximum ranges because the chance of every aircraft making its last call at the maximum possible range is small.

For the period October 1961 to February 1962 an attempt to overcome this deficiency was made. It was possible to do this as the signal strength at the most distant point was known and from that position it could be calculated how much further the aircraft could have travelled before the signal had fallen below $0.5 \mu\text{V}$ at the ground receiver.

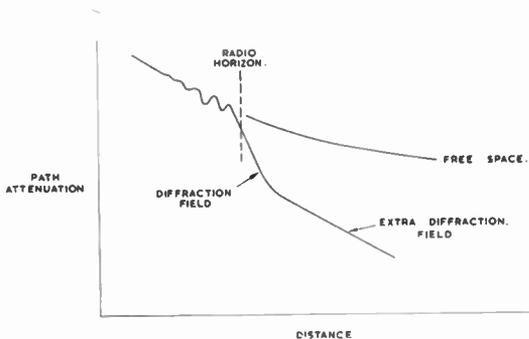


Fig. 4. Typical path attenuation/distance curve for air/ground path.

As a result of the analysis of many ground-to-ground non-optical v.h.f. circuits new theories have been developed in the past decade. For air-to-ground circuits, however, little analytical work of a similar nature has been carried out but from the information which has been published the indications are that for a short distance beyond the radio horizon the path attenuation increases rapidly until some level below the free space signal is reached and then the attenuation rate changes to a lower value. Figure 4 illustrates this.

Throughout this paper the radio horizon has been calculated on the basis of the earth's effective radius being four-thirds of the actual radius.

The diffraction field has been the subject of much investigation and C.C.I.R. under Resolution 22 calculated the signal strength in this region. Various theories for the more distant field have been suggested includ-

ing ducting, tropospheric scattering, partial refractions in an atmosphere which has a refractive index graduation and specular reflections from an elevated layer or layers.

According to the C.C.I.R. curves the path attenuation in the diffraction zone increases at the rate of about 0.75 dB/n.m. For the extra-diffraction zone however due to the variations of the theories there is no agreement for the attenuation rate and figures varying from 0.4 to 0.02 dB/n.m. have been quoted.

An endeavour was made to check this pattern of attenuation from the results. Figure 5 shows five typical flights, with signal strengths plotted against distance beyond the horizon. The steepness of the slope indicates the rate of attenuation being experienced. The *Comet* at the top has a slope representing an attenuation rate of 0.07 dB per mile . The *707* on the next graph shows a rate of 0.05 dB for about a hundred miles then increasing to 0.66 dB per mile . The next *707* slope is about 0.13 dB , and the next one 0.1 dB . The *Comet* at the bottom is 0.34 dB .

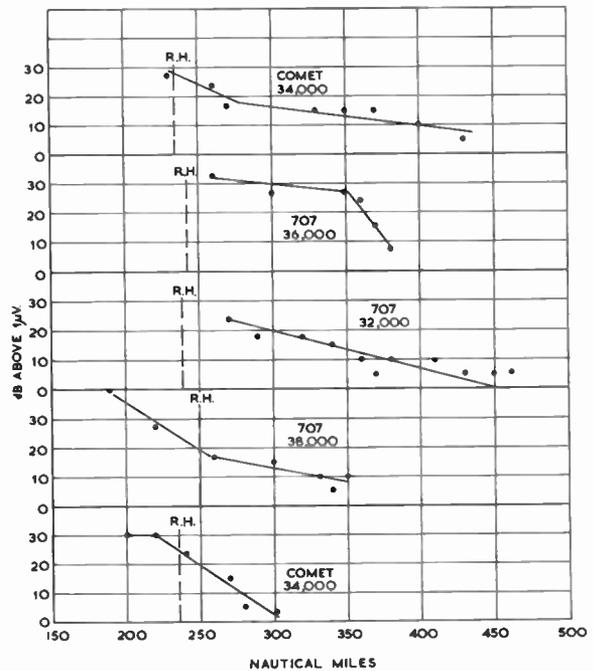


Fig. 5. Signal strength at ground receiver for five typical flights in the beam.

Figure 6 shows a similar representative series. The *Comet* at the top has an attenuation rate of 0.7 dB per mile . The *Caravelle* was a curious case—the signal strength actually increased out to about 150 miles beyond the radio horizon before it began to fall. The next three flights all have slopes of about 0.1 dB per mile . Both these figures seem to support the theory of a lower rate of attenuation well beyond the horizon than in the diffraction zone just beyond the horizon.

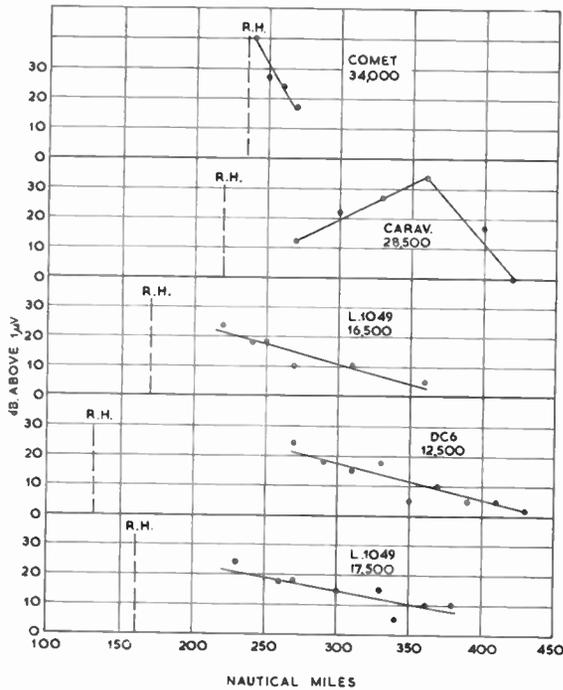


Fig. 6. Signal strength at ground receiver for five typical flights in the beam.

The signal strength/distance curves for aircraft in the beam were plotted and all those for which there were sufficient readings to define a curve were analysed and the slope of the curve estimated. It was found that it was difficult to assess this for several reasons: the possibility of a ± 2.5 dB inaccuracy in the signal strength reading; a possible inaccuracy in the distance; and finally the signal fluctuates in level, and as the aircraft transmission is short, it cannot be certain that an average value has been obtained.

It was not possible to assess any curves close to the radio horizon where the attenuation rate is of the order of 0.75 dB/n.m. because for a drop in level of 30 dB at a speed of about 480 knots the time interval would be about 5 minutes. During such a period there were never sufficient readings to define a curve and as it happens the logs included comparatively few readings just after the radio horizon where this rate occurs.

Altogether there were about forty flights with sufficient signal strength measurements to define accurately a rate of attenuation. These are shown in Fig. 7. Each dot represents the rate of attenuation experienced by an aircraft at a certain distance beyond the radio horizon. The distance plotted is the mid-point of each flight examined. It can be seen that there is a spread in the results from about 0.06 to 0.2 dB/n.m. for distances of 100 to 200 n.m. beyond the radio horizon and about 0.15 to 0.25 dB/n.m. for distances of 60 to 100 n.m. beyond the radio horizon.

From the information derived it was concluded that the rate of attenuation in the region from 50 miles to 200 miles beyond the horizon averages 0.2 dB/n.m. or less.

For those aircraft who did not call at maximum range it was calculated how much further the aircraft could have travelled before the signal would have been lost at the ground receiver, using two attenuation rates, namely 0.75 dB/n.m., the classical theory rate, and 0.2 dB/n.m., a rate which has been derived above.

In order to give a mental picture of the kind of ranges that were obtained, Fig. 8 displays the maximum ranges obtained in the month of October 1961 along airway Red 19. The curve plots height against distance and is prepared on "four-thirds earth radius" graph paper. The conventional radio horizon allowing for the 60 ft height of the aerial has been inserted and so has a line showing the position of an aircraft radiating 25 W to produce 0.5 μ V at the ground receiver as calculated from the C.C.I.R. curves. An attenuation rate of 0.2 dB/n.m. to correct for true maximum range has been used where it was necessary to do so. This diagram convincingly demonstrates how far beyond the horizon satisfactory communication can in fact be sustained with individual aircraft. During that month there were about 20 which reached 450 n.m., most of them flying at over 30 000 ft but a few as low as 12 000 and 18 000 ft. But, as stated earlier, this kind of diagram can be misleading because it fails to show the percentage probability of a certain range being achieved, and of course it only shows one month's results.

It was not practical to analyse on the basis of an infinite number of heights and distances from Bahrain so height intervals of 5000 ft and distance intervals of 10 n.m. were assumed. For each height interval the number of aircraft having a maximum range within each 10 n.m. interval was set down and from these figures a probability of an aircraft attaining a particular range was calculated. It was now possible to

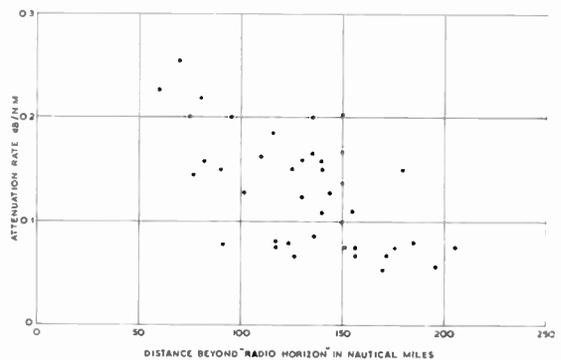


Fig. 7. Attenuation rate at various distances beyond the radio horizon.

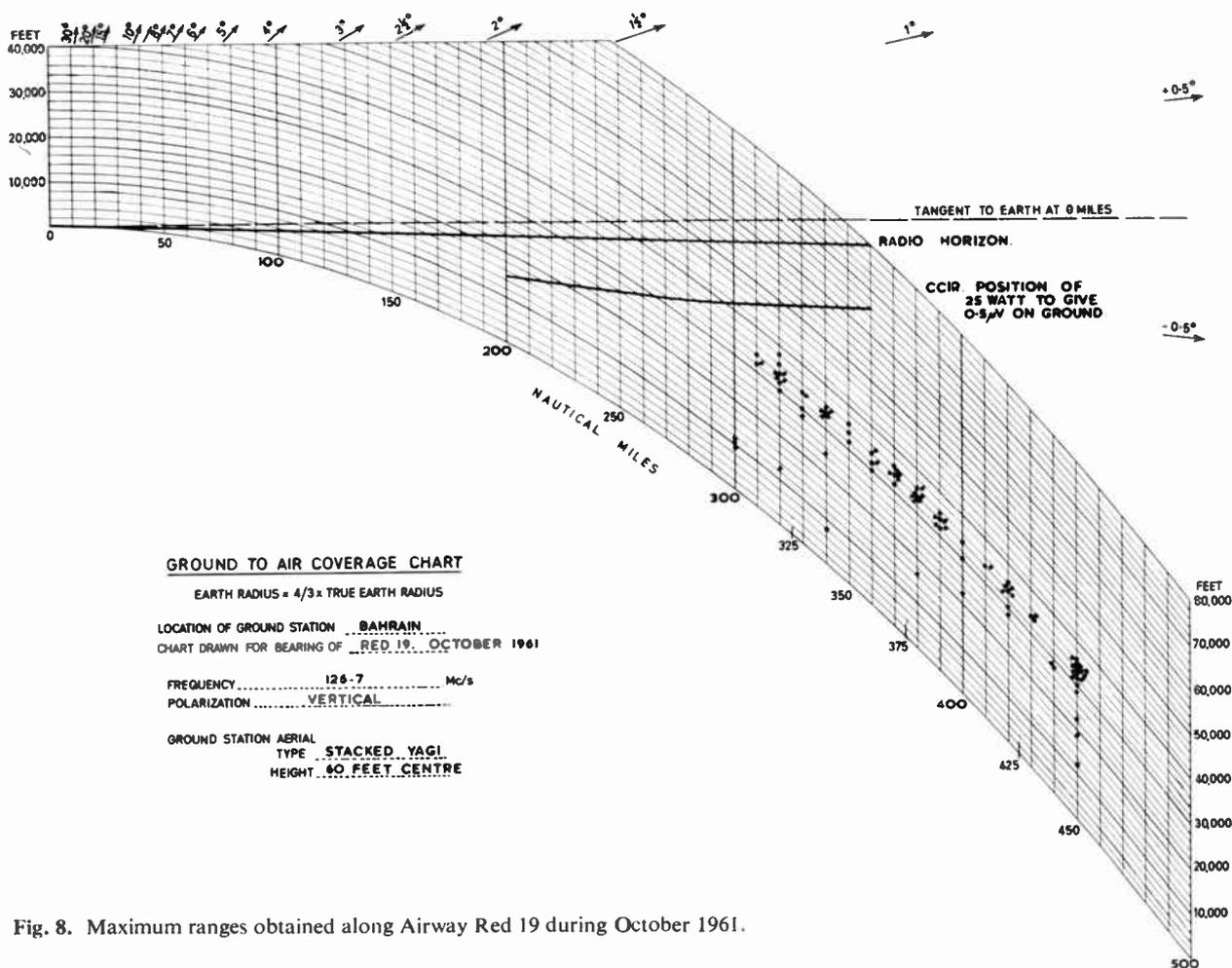


Fig. 8. Maximum ranges obtained along Airway Red 19 during October 1961.

present the main conclusions of these trials in graphical form.

Figure 9 shows for airway Red 19 the probability of an aircraft obtaining any range during the summer and winter months. During the winter months the 90% certainty range (for attenuation rate of 0.2 dB/n.m.) is 325 n.m.

Figure 10 shows similar information for airway

Amber 1. In the winter months the 90% certainty range for the same attenuation rate is 315 n.m.

There is good correlation between the results for the two airways. For the winter months the ranges calculated for an attenuation rate of 0.75 dB/n.m. are the same to within a mile or two and the 0.2 dB/n.m. curves differ by a few miles only for probabilities in excess of 80%, Red 19 being the greater. For the

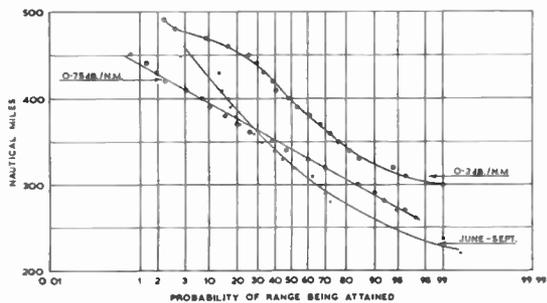


Fig. 9. Probability of an aircraft attaining any range along Airway Red 19.

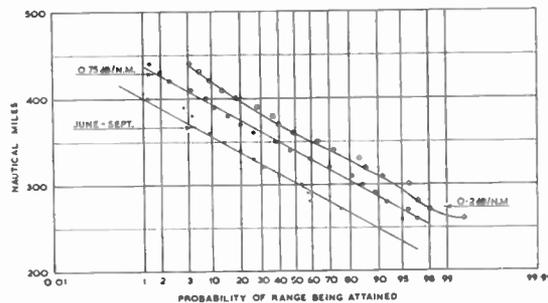


Fig. 10. Probability of an aircraft attaining any range along Airway Amber 1.

summer months for probabilities greater than 60% the ranges for the two airways are similar but for probabilities less than this the range is greater along Red 19. An examination of the individual months within the two seasons was made but essentially they were similar to the curves shown.

Due to the two different methods of computation which necessarily had to be employed there is some doubt about what ranges would have been obtained in the summer months. For these two overland airways it would appear that there is no great difference between the summer and winter conditions except that along airway Red 19 there is a greater probability of attaining ranges in excess of 360 n.m. and more especially so in the summer. This may possibly be due to the fact that airway Red 19 is in the centre of the beam whereas airway Amber 1 is for some distance within the edge of the beam where the system gain is 3 dB less.

The probabilities quoted have included aircraft of all heights but as 93% of the aircraft were within the height interval of 25 000 to 45 000 ft, the probabilities to a few per cent apply to aircraft at any height within this interval.

If an aircraft at a height of 32 500 ft is considered as being typical then the radio horizon distance is 230 n.m. and along airway Red 19 for a 90% certainty at an attenuation rate of 0.2 dB/n.m. the radio horizon was exceeded by 41%.

Table 1

Ranges obtained with aircraft flying at relatively low heights on the Red 19 Airway

July to September 1961			October 1961 to February 1962		
Height	Range	Percentage Beyond Radio Horizon	Height	Range	Percentage Beyond Radio Horizon
14 500	165	64	10 500	280	110
8 500	170	42	5 500	300	200
10 000	245	88	10 500	300	126
10 500	250	88	11 500	300	117
14 500	265	71	9 500	320	150
13 500	275	83	11 500	320	131
12 500	290	102	9 500	380	196
13 500	300	100	11 500	380	175
12 500	315	120	12 500	450	215
11 500	315	128			
14 500	360	132			
13 500	385	156			
12 500	470	230			

For aircraft below about 20 000 ft the probability curves do not apply. The number of aircraft included in the logs is so few that the only conclusion which can be drawn is that there is a trend to show that if the maximum range is related to the radio horizon then relatively low flying aircraft achieve much greater ranges. Table 1 shows, for the two periods July to September 1961 and October 1961 to February 1962, the aircraft which were logged below 15 000 ft. For each aircraft the range which it achieved as a percentage beyond the radio horizon has been calculated.

There is an airway to the north of Bahrain to Baghdad along which comparatively few aircraft fly. This, coupled with the fact that the Bahrain Flight Information Region extends for about 120 n.m. only in this direction means that few reports are obtainable on this airway. There were too few aircraft in the logs to draw any firm conclusions but compared to airway Red 19 the high flying aircraft seem to have a tendency to greater ranges whilst the low flying aircraft had similar ranges. Most of the flights were in July–September period and no conclusions as to the difference between winter and summer can be drawn. It should be noted that in this direction, unlike the other airways in the beam, the path is over sea for the first 250 n.m.

The airline schedules are such that the majority of the aircraft transit Bahrain during the hours of darkness and consequently there were insufficient day flights to draw any conclusions. There did not appear to be any significant difference between ranges which were obtained at day and night.

6. Analysis of Airways to the Rear

Figures 11 (a) and (b) show the variation of signal strength with distance to the rear for 10 typical flights. It is immediately apparent that these are quite different to those for airways in the beam and the fall in signal strength follows a much less regular pattern. Generally there is a fall-off in signal strength to an extremely low value or zero in the region of 250–280 n.m. followed by an enhancement to a higher but still low value and then a falling off at a very slow rate. Signals may persist for up to ranges of 400 to 500 n.m. It is interesting to note that the path is over land between 260 and 315 n.m.

Due to the irregularity of the signal strength and the apparent slow rate of attenuation it was felt that any assessment of attenuation rate would be inaccurate. For airways in the beam drops in signal strength of 20 dB occurred in distances of the order of 100 n.m. and therefore any small inaccuracies in the signal level or in distance would not greatly affect the attenuation rate. To the rear however signal strength variations of only 5 or 10 dB were obtained for distances of 100 or even 200 n.m. and therefore due to the inaccuracies in the measurements wide variations of

attenuation rates could be obtained. It was concluded that an attenuation rate and hence a maximum range could not be predicted as was done for signals in the beam.

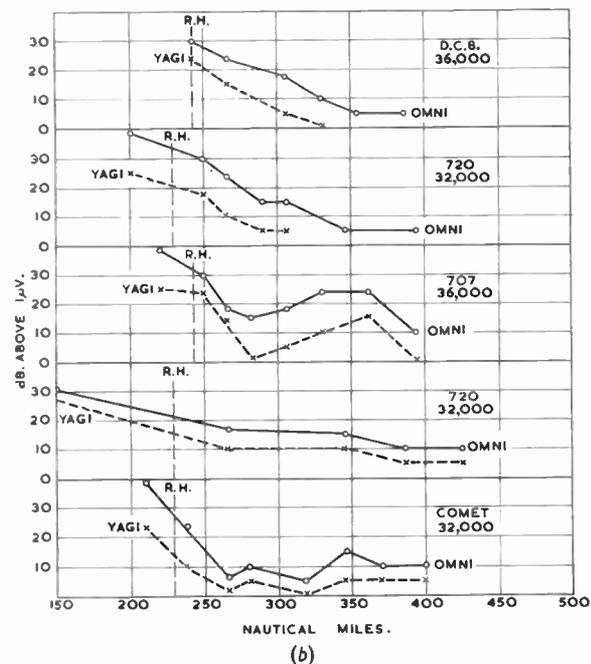
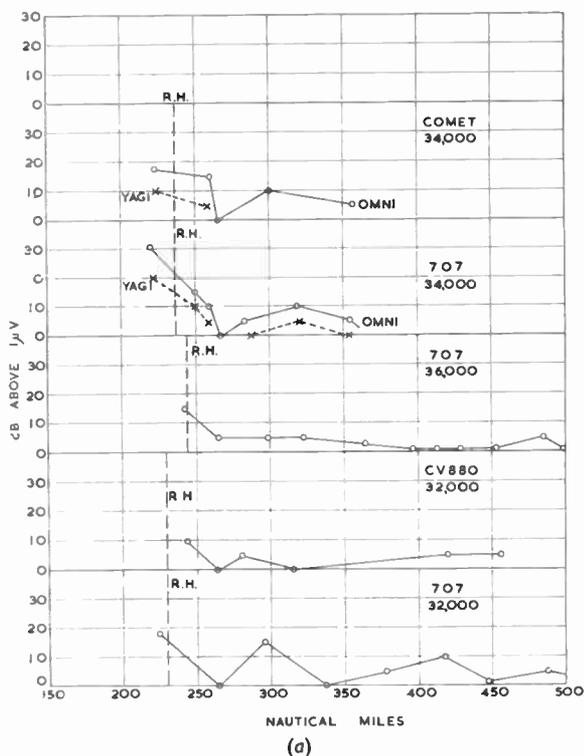


Fig. 11. Signal strength at ground receiver for five typical flights to the rear.

Because of this difficulty, for the period July 1961 to February 1962 for both airways to the rear, points from 265 to 450 n.m. were taken and at each of these the percentage of aircraft that actually attained them was calculated. This method must give pessimistic results because the last call was rarely at the maximum possible range.

Figure 12 shows the analysis for airways to the rear. The first thing that is immediately apparent is that there is a difference between the summer months of July to October and the winter months of November to February. In the summer an aircraft had about 85% to 90% chance of attaining a range of 270 n.m. but in the winter this drops to 40% to 60%.

It is interesting to compare these figures with those for the summer months of Red 19 and Amber 1 in the beam as these were analysed in a similar manner (see Table 2).

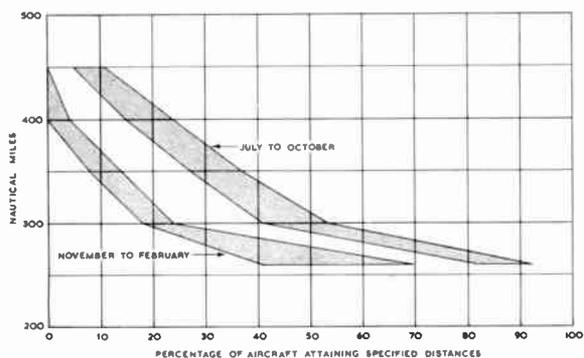


Fig. 12. Percentage of aircraft attaining specified ranges to the rear.

Considering that the gain to the rear is 21 dB less than that to the front it is surprising to see that during the summer months the probability of achieving ranges of up to 400 miles is the same to within a few per cent along Red 19 and greater than along Amber 1. During the winter months to the rear the probability is always less than along Red 19 and less than Amber 1 for 270 and 300 n.m. but about the same for 350 and 400 n.m.

Table 2

	Summer Probability		Winter Probability
	Red 19	Amber 1	Rear
270 n.m.	85	80	85
300	65	50	50
350	35	13	35
400	15	1	20

It is interesting to compare results obtained at equidistant points in the forward and rear directions. For example, at 270 n.m. at 30000 ft—just beyond the radio horizon—81% of the aircraft in the winter months gave a signal exceeding 5 μ V at the receiver. At the equivalent point to the rear only 25% of the aircraft exceeded this value.

It has been stated before that over the sea path to the rear the characteristics of the signal are that it falls to a low level about the radio horizon and then decays very slowly, instead of falling off quite rapidly as it does over land. Further where the path crossed land the signal has a definite tendency to fall to an extremely low value and then recover to a higher but still low value.

An effort was made to estimate the ranges which were obtained on the omnidirectional aerial to the rear but the variation was large from a few miles to well over 100 n.m. It was concluded that no useful information could be obtained from the figures as the operator cannot transmit on this aerial and the aircraft, once it went beyond the Yagi array range, may not have tried to keep in communication.

The probability percentages quoted were derived from analysis of aircraft of all heights, but there were few below 25 000 ft. Below this height the trend seems to be that if their maximum range is related to their radio horizon then relatively they did better than the aircraft at greater altitudes.

7. Correlation with Meteorological Conditions

The effect of meteorological conditions on the ranges obtained were naturally of interest, especially in view of other studies which have sought to establish a correlation between meteorological parameters and both ground/ground and air/ground ranges.

Four days were picked when long ranges were being experienced and three days of short ranges and from the radiosonde ascent data from Bahrain the potential refractive modulus K was plotted.

For the present purpose it is sufficient to regard K as a parameter derived from the true refractive index n , but which no longer varies appreciably with height in a well mixed atmosphere. A plot of K as a function of height will therefore emphasize local differences in the thermodynamic properties and hence the refractive index of the atmosphere. Local deviations from the normal refractive index gradient will appear as a sloping instead of a vertical line.

Bahrain radiosonde station is unfortunately probably not wholly representative of the conditions existing along the flight path, as it is a few miles off the coast of a large land mass. However it was found that on the four days when ranges of the order of 350 to 450 n.m. were obtained the K plot showed consistently greater slopes than on the three days when ranges of the order of 275 to 325 n.m. were common.

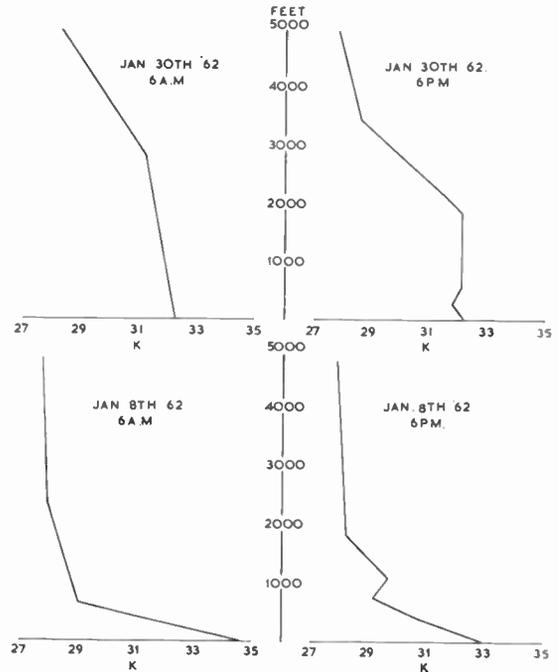


Fig. 13. K plots for a typical day of short ranges (above) and a day of long ranges (below).

For comparison purposes K plots for two days are shown on Fig. 13 the upper curves being a typical short range day and the lower ones a long range day. It will be noticed that on the days when long ranges were being obtained the big change in K was occurring at low altitudes, of the order of 1000 ft whilst the aircraft were at about 30 000 ft. On the basis of seven days, rigid conclusions cannot be drawn but the tendency is interesting.

Vergara, Levatich and Carroll† analysed the P.A.A. San Francisco flights in conjunction with the meteorological data available using a different analytical method and they concluded that there was some correlation between the meteorological conditions and the ranges.

8. Other Trials

A brief summary of other work which has been carried out in this field indicates that similar though often longer ranges have been obtained.

Long range v.h.f. trials were carried out from Shannon in South West Ireland by Pan American World Airways.‡ The installation was identical to that at Bahrain but the site elevation was 880 ft a.m.s.l.,

† W. C. Vergara, J. L. Levatich and T. J. Carroll, "V.H.F. Air-Ground Propagation Far Beyond the Horizon and a New Meteorological Parameter", Bendix Corporation.

‡ "V.H.F. Tropospheric Scatter Techniques for the Aeromobile Communications Service", Interim Report No. 2, Pan American World Airways Inc., 1961.

3 miles inland, with a clear view westwards to the Atlantic. Tests were carried out in 1958 and a report stated that range was insignificantly affected by height in respect of aircraft operating between 14 000 ft and 37 000 ft. Some aircraft were contacted out to 350 n.m., others reached 600 n.m. but the majority reached the vicinity of 450 n.m.

From a report issued in 1960 of tests carried out at the same station in 1959, 95 flights have been analysed and the probability curve shown in Fig. 14 constructed. It is interesting to note that if all aircraft are taken into account the 90% certainty range is 312 n.m. and the 50% range is 390 n.m., with some aircraft achieving a range of 500 n.m. or more. If aircraft above 25 000 ft only are considered then to a few per cent the ranges are unaffected; aircraft below 25 000 ft however did not ever exceed a range of 400 miles. These trials were carried out with a muted ground receiver and the ranges are given on the basis of a last call but whether this represented the maximum range obtainable is not known.

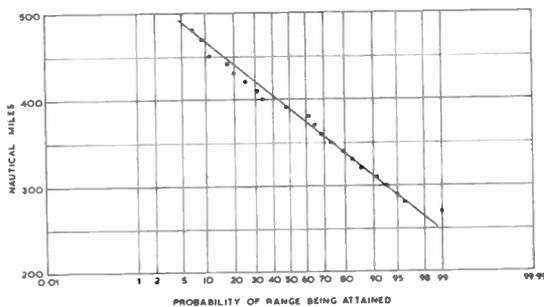


Fig. 14. Probability of an aircraft attaining any range from P.A.A. station at Shannon.

From a site at San Francisco 1450 ft a.m.s.l. with a clear view to the west, P.A.A. carried out some tests with 40 aircraft flying to Hawaii during 1959 and 1960.† The installation was similar to that at Shannon but the receiver was preceded by a low noise pre-amplifier and the aerial in this case had a gain of 15.5 dB. The maximum ranges obtained varied considerably, one aircraft reaching 385 n.m., seven between 400 and 500 n.m., eighteen between 500 and 1000 n.m., eleven between 1000 and 2000 n.m. and three over 2000 n.m. The aircraft heights varied between 10 000 ft and 35 000 ft and there appeared to be no immediate correlation between height and range except that ranges in excess of 1200 n.m. were achieved only by aircraft flying higher than 25 000 ft. It is important to note that the ground receiver was unmuted in these tests and signal levels down to 0.15 µV were recorded. An analysis of the flights carried out by Vergara, Levatich and Carrol concluded that in the neighbourhood of the radio horizon the attenua-

tion rate was 0.75 dB/n.m. or less and this persisted until the signal level was somewhere between 20 and 40 dB below the free space value. For the short flights the attenuation rate thereafter was about 0.15 dB/n.m. and for the long flights about 0.02 to 0.04 dB/n.m.

A further station has recently been established at Beirut by P.A.A. It is located at a height of 6230 ft a.m.s.l. 25 miles east of Beirut, so that for an aircraft at 30 000 ft the radio horizon is 310 n.m. Two aerials, back to back, each with a gain of 15.5 dB, are aligned Eastwards and Westwards. The radio equipment is again similar to the San Francisco equipment. No details of the performance of this installation have been released but it is understood that ranges of between 380 and 500 n.m. have been obtained with jet aircraft.

The Ministry of Aviation have conducted a series of trials over the past few years from Mangersta in the Isle of Lewis at a site 230 ft a.m.s.l. A 1 kW transmitter, a receiver with a noise factor of about 6 dB and an aerial having a gain of 12 dB were used. It is understood that with jet aircraft ranges of 350 to 400 n.m. were being obtained but no figure for the probability of these ranges being obtained is available.

Starkey‡ analysed several flights made at 10 000 ft in England where transmission from a 120 kW e.r.p. transmitter on 91.3 Mc/s at a height of 400 ft were measured at distances out to 200 n.m. beyond the radio horizon. It was found that the attenuation rate near the radio horizon was, in general, slightly less than that calculated for the diffraction field and that, when the signal had fallen to about 40 or 50 dB below free space, the attenuation rate changed to a rate of the order of 0.2 dB/n.m. It was shown that the results, in this particular case, could be explained by the simple hypothesis of specular reflection from temperature inversion at the tropopause.

9. System Sensitivity

The system sensitivity of the air-to-ground path of the Bahrain installation is 174 dB made up as follows:

Aircraft transmitter	14 dBW
Aircraft feeder loss	-1 dB
Aircraft aerial	0 dB
Ground aerial	18 dB
Ground feeder loss	-1 dB
Ground receiver (0.5 µV p.d.)	144 dBW
		174 dB

The comparative figure for the ground-to-air path depends on the muting setting of the aircraft receiver.

‡ B. J. Starkey, "Some aircraft measurements of beyond-the-horizon-propagation phenomena at 91.3 Mc/s", *Proc. Instn Elect. Engrs*, 103, B No. 12, pp. 761-3, November 1956.

† See footnote on page 26.

If this is assumed to be $2 \mu\text{V}$, the figure is 178 dB, made up as follows:

Ground transmitter	30 dBW
Ground feeder loss	-1 dB
Ground aerial	18 dB
Aircraft aerial	0 dB
Aircraft feeder loss	-1 dB
Aircraft receiver	132 dBW
					178 dBW

In order to determine how these results correlated with theory and with other work, the C.C.I.R. curve for the air to ground path with the system sensitivity assumed above is shown in Fig. 15. Through the 325 n.m. (the range obtained with a 90% probability along airway Red 1 in the beam) and the 0 dB point is drawn a line with a slope of 0.2 dB/n.m. which is the attenuation rate determined from the results. This intersects the C.C.I.R. curve about 35 dB below the free space field. A more likely curve is shown dotted, which intersects the C.C.I.R. curve 30 dB below the free space field.

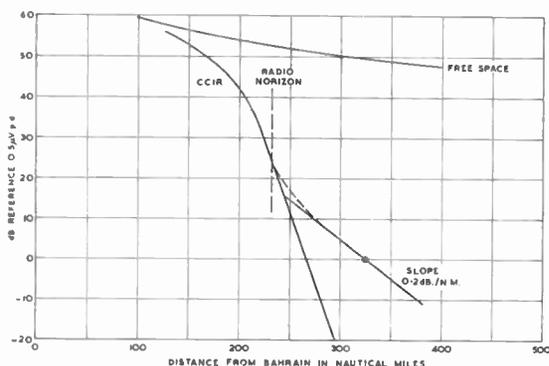


Fig. 15. Signal strength at ground receiver versus distance curve, derived from results.

It appears therefore that these results can be explained to a fair degree by a combination of strict diffraction theory and figures which have obtained during these and other similar trials.

Due to the fact that the aircraft has to report back on h.f. and in general assumes that the contact is lost when the ground station can no longer hear the aircraft and it gets no reply to a call, there are few reports of the difference between the air/ground and ground/air paths. The 34 reports of such cases vary in distance from 112 n.m. to 15 n.m., all showing a greater ground-to-air range. The large spread can be attributed to a variety of reasons and of course the difference is critically dependent on the muting setting of the airborne receiver but it is important to note the large difference in range which can be encountered.

With the present installation at Bahrain there is a deficiency in the air-to-ground path of about 4 dB, this could be overcome by the use of a receiver with a noise factor of the order of 3 dB and a better muting circuit. It is felt that there is little point in using a receiver with a better noise factor than 3 dB as from the available information it appears that the average noise due to cosmic radiation at the frequencies of the aeronautical v.h.f. band is equivalent to an aerial noise factor of 2 dB. With a 3 dB receiver the overall noise factor will be about 5 dB compared with about 7 dB for a 6 dB receiver. Although the receiver had a noise-compensated muting circuit it is known that this is not the best that can be achieved and if a codan unit were employed then the combination of this and the better receiver would equalize the gain of the two paths. The 90% probability range overland would then be of the order of 345 n.m. This is still not sufficient for many air routes, therefore the means of achieving greater range has to be considered.

10. Possibilities of Further Range Improvement

Any aerial gain that can be achieved affects both paths and is therefore advantageous. The value of aerial gain that can be achieved on the ground is determined by economics. If it is assumed that the highest mast that can be used is 200 ft then if stacked Yagi aerials are to be used the possible gain will depend on the horizontal beamwidth required. If the configuration of the airways is such that 75 deg or 50 deg is required, then the maximum gain possible with the above parameters is 21 dB, but if a horizontal beamwidth of 25 deg is suitable the limit is 24 dB. Arrays using other types of aerial are possible; corner reflectors and full wave dipoles in front of a reflecting screen are two such examples. There is also the familiar parabolic dish, but one 60 ft in diameter has a gain of only 24 dB at this frequency. A gain of 21 dB does not involve too high an economic penalty and is likely to be within the average budget.

It is known that airline operators are averse to using any directional aerial because this involves the Captain in yet another task of selection. There is also a problem when the aircraft is not flying directly to or from the station, as happens when an aircraft crosses the edge of a Flight Information Region. Despite this however it is suggested that in the future aerials with a gain of the order of 6 dB to 10 dB fore and aft may be used.

With the present airborne equipment the muting setting has to be made less sensitive than it could be, as otherwise there is a tendency for it to open when the aircraft is flying over densely populated and hence noisy areas. It could however be improved and it is known that one manufacturer, at least, has a circuit similar to the codan which opens only when a readable

signal/noise ratio is obtained. The performance would then be limited by aircraft noise, which for a modern jet is generally less than $1 \mu\text{V}$.

The present 25 W output power of the airborne transmitter could be increased and it is suggested that a 250 W transmitter is practical. A prototype equipment of this power has already been flown in an aircraft, as far as is known quite successfully.

On the ground a 10 kW transmitter would be practical though use of this power would involve careful engineering and if the station was located close to other aeronautical band v.h.f. installations interference would result which might prove costly to overcome.

If all these gains were incorporated the total system gain would become 207 dB for the ground-to-air path and 209 dB for the air-to-ground path made up as follows:

Air-to-Ground Path

Aircraft transmitter . . .	24 dBW
Aircraft feeder loss . . .	-1 dB
Aircraft aerial	10 dB
Ground aerial	21 dB
Ground feeder loss	-1 dB
Ground receiver, say	156 dBW
	209 dB

Ground-to-Air Path

Ground transmitter	40 dBW
Ground feeder loss	-1 dB
Ground aerial	21 dB
Aircraft aerial	10 dB
Aircraft feeder loss	-1 dB
Aircraft receiver	138 dBW
	207 dB

Based on the Bahrain results for jet aircraft flying overland a station at sea level with the increased sensitivities should have a range, for a 90% probability, of the order of 495 n.m.

The measures necessary to achieve such ranges involve some considerable cost both on the ground and in the aircraft. However, if it were possible to relinquish h.f. circuits on many sectors the cost may be justified in view of the increased reliability and a "first call" type of operation which is achieved on v.h.f. A special v.h.f. airborne installation capable of use on a comparatively few channels especially allocated for this service is probably the best solution, as this would lead to a cheaper installation. Initially this would have to be carried in addition to the existing h.f. installation but perhaps in the future this latter would be unnecessary.

Other approaches are possible to achieve further gain which have not been included in the above figures. These include the location of the ground station on high ground, speech processing, diversity reception and a narrowing of the bandwidth.

Ground station height increases the range as the radio horizon distance is increased, and it has been suggested that further benefits accrue as there is a better "take-off" for the transmission and the lower lobe of the aerial pattern is depressed. It is difficult to put a value on speech processing but experiments have indicated that 3 or 6 dB apparent gain is possible. Before any diversity gain can be assessed further tests must be carried out to determine the correlation of the signals on two displaced aeriels. On ground-to-ground circuits up to 6 dB is possible by this technique and it would seem likely that this order of gain might apply to the air-to-ground path. On the ground-to-air path the possibility of using diversity reception is remote especially if directional aeriels are used on the aircraft as the fitting of two such aerial arrays would be a formidable task. Present receivers have a bandwidth of the order of 30 kc/s which is dictated by the stability of the equipments. Techniques for improving the latter are common knowledge and if advantage was taken of these and the bandwidth reduced to 7.5 kc/s a system gain of 6 dB in each direction would be possible.

11. Conclusions

The work at Bahrain was not conducted under rigorously controlled experimental conditions but it does represent a very large and typical sample taken under practical operational conditions. Specific general conclusions cannot therefore be drawn but the following points have emerged.

Further confirmation has been provided that v.h.f. air-to-ground communication well beyond the conventional radio horizon is a practical proposition using existing airborne equipment and ground stations of reasonable cost.

In the climate of the Persian Gulf for an overland path for a high degree of probability there is no evidence of any difference in ranges obtainable between the winter and summer but for an oversea path there is a tendency to greater ranges in the summer.

The range to be expected depends on geographical factors and in particular overland ranges are likely to be less than oversea.

Aircraft height, at least within a wide band of altitudes, has much less effect on range than it does within the radio horizon.

There was confirmation that far beyond the radio horizon the attenuation rate is considerably lower than classical theory suggests.

The desirability of presenting results of trials on a range probability basis is shown. Until the theory of long range operation has been further developed such a presentation is essential if planning authorities are to be assisted.

The influence of meteorological conditions has been detected but no correlation was possible with the data at our disposal.

12. Acknowledgments

We wish to record our appreciation to all the air-crews whose co-operation was so essential to our work,

to the staff in Bahrain, and to our colleagues who assisted us, particularly those who did much of the laborious analysis calculations. We are also indebted to the Ministry of Aviation and the Royal Air Force for facilities they kindly extended to us in Bahrain. We thank them all for having helped us to take this look beyond the horizon.

Manuscript first received by the Institution on 6th April 1962 and in final form on 23rd May 1962 (Paper No. 779/RNA17).

© The British Institution of Radio Engineers, 1963

DISCUSSION

Under the Chairmanship of Dr. J. W. R. Griffiths (Associate Member)

Mr. W. E. Brunt: I congratulate the authors on an excellent paper; in my opinion, it is a competent account of a project which was itself well done. And not only have they provided a great deal of quantitative information but they have displayed it to good effect. I think they have made a significant contribution to the subject.

This does not necessarily mean, of course, that I agree with all of the conclusions drawn. On the whole I am in agreement, but there are some points which are debatable in particular where they pass from describing what has been achieved, and forecast what might next be done.

I understand the authors to say that further extensions of range are merely a matter of more circuit gain, irrespective of the manner in which this gain is obtained, and that the improvement in range to be expected may be readily calculated.

Our experience in B.O.A.C. leads us to believe that matters are not quite so simple; for example, the authors propose to gain 10 dB by increasing the sensitivity of airborne receivers. Now this is not difficult to do; but it might be quite difficult to live with. These receivers are equipped with muting, commonly set at $3 \mu\text{V}$; there is a good operational reason for this and one which tends to become more rather than less pressing.

Next, the authors have said that Fig. 9 is probably the most significant of their diagrams and I think they are right. Among the several things it shows is the way in which range varies with attenuation factor. As I understand it, this diagram does not bear out the authors' forecast. Perhaps they could explain where I have gone wrong.

I am particularly pleased to see the association of range and probability, for I have found this to be the most satisfactory way of bringing the subject down to earth. It has a most salutary effect upon over-optimistic claims and a wonderfully unifying effect upon the results obtained in similar programmes. Take, for example, Figs. 9 and 14; here are plotted the results of trials carried out in quite different circumstances, and yet the 95% ranges are almost identical.

Furthermore, I have found that once the operational range of a v.h.f. station is defined as "that maximum distance at which the probability of successful two-way contact is not less than 95%", then the Bell propagation curves are a reliable guide to performance, at least as far as the near diffraction zone.

It is true, of course, that the Bell curves give no hint of the so-called "plateau" in the far diffraction zone, for which the authors have provided still further evidence. There have been several attempts to explain this plateau in terms of some new mechanism of propagation, but I believe that what is slowly emerging is that what is involved is simply the old mechanism working in ways not yet perfectly understood.

Much of the mystery which surrounds this phenomena has been removed, for me at least, by a recent paper by Bean† where he says that the refractive index of the standard atmosphere varies with height not linearly as had been supposed, but exponentially. Consequently, the magnitude of diffraction to which a ray is subject varies with the vertical angle of radiation; from which follows a natural tendency to focusing in the far diffraction zone, with "ducting" as a special case of this common phenomena.

Mr. M. T. O'Dwyer: In 1957 in conjunction with an American airline we commenced a series of tests along similar lines with one of our 1 kW transmitters located at Shannon Airport. Over a period of several months a large number of trials were conducted, with very diverse results. No firm correlation could be established between limit range and aircraft altitude, heading or type and it was suspected that variations in the airborne receiver sensitivities were obscuring the issue.

A special calibrated receiver was therefore made up and flown many times on the Atlantic route. A pattern emerged which enabled mean path attenuation to be plotted against distance, and it is interesting to observe that the figures obtained are close to those derived by the

† B. R. Bean, "The radio refractive index of air", *Proc. Inst. Radio Engrs*, 49, pp. 260-73, March 1962

authors of the present paper, i.e. about 1 dB per n.m. in the "diffraction" region and about 0.14 dB per n.m. in the "scatter" region.

As a result of these tests we installed for the airline a permanent station at Ballybunion, Co. Kerry, about 40 miles S.W. of Shannon Airport. A prototype of the Bahrain aerial array was erected on a site 880 ft a.m.s.l. and the station was controlled by radio-link from Shannon. This station is still in routine use and the results obtained have been reported on by the airline.

An identical installation was made for another American airline at Carrickerry and tests have indicated that the ranges obtained are practically identical.

The authors of this paper should be congratulated on their analysis of the results at Bahrain and on their presentation.

The authors (in reply): We would like to thank Mr. Brunt for his comments and compliments. We agree that further range extension, derived from increased system gain, will demand airborne equipment improvements among other things. Perhaps over-optimistically, we do not share Mr. Brunt's hesitation about the practicability of such improvements. For example, we foresee that aircraft will be equipped with a separate long range v.h.f. installation, which would be used only at ranges and heights remote from external noise sources, thus permitting the use of a more sensitive receiver. Evidence of the low rate of attenuation in the extra-diffraction zone is accumulating—Mr. O'Dwyer has mentioned other examples this evening—and we feel that the necessary equipment improvements to take advantage of this fact reasonably can be expected.

Mr. Brunt had some difficulty over our forecast of further range improvement. We think this is explained by the fact that in the lecture presentation we referred to an "approximate" system gain of 200 dB, i.e. an increase of 22 dB over the present figure. In fact the precise figure on which we based our forecast was 208 dB, an increase of 30 dB from the present 178 dB. At an attenuation rate of 0.2 dB per mile, this would yield the extra 150 miles that we forecast.

Lt.-Col. D. R. W. Thomas (Member): Can the authors say if they intend in their future experiments to investigate quantitatively the field strength distribution over a number of paths of a given distance. Work has been done on estimating the field strength distribution at say 20 km from a 20-watt transmitter over different land paths for a given type of terrain. It would be a very useful contribution to propagation knowledge if similar information could be obtained from these very valuable experiments.

The authors (in reply): We did make some effort to measure field strengths at varying distances, as suggested by Colonel Thomas. The table above shows the percentage of aircraft whose signal strength reached the values shown, measured at Bahrain when the aircraft was over one of three different reporting positions.

Mr. M. Settelen: Has any thought been given by the authors to the use of frequency modulation in such a system? The B.B.C. v.h.f. broadcasting transmitters have a power of 10 kW and from a range point of view could

Reporting Position	PA	SHB	Sharjah
Distance from Bahrain 270 n.m.	265 n.m.	266 n.m.	
0	7%	12%	45%
5 μ V	13%	7%	30%
10 μ V	17%	12%	17%
15 μ V	15%	24%	4.5%
20 μ V	20%	14%	2.5%
25 μ V or more	28%	31%	1%

provide interesting information if a suitable monitoring receiver was carried in an aircraft.

Little work has been done over the past several years in the improvement of v.h.f. aircraft aeriels and it could well be that suitable development effort could produce a directional aerial with automatic alignment on a ground transmitter.

The authors (in reply): We note Mr. Settelen's confirmation that little attention has been paid to improvement in the airborne aerial; we think that the airlines should consider encouraging work in this direction. Regarding the use of f.m., we think this would be a disadvantage at the longer ranges because the signal/noise ratio deteriorates so rapidly once the signal falls below the limiter level.

Brigadier E. J. H. Moppett: Commenting first on Mr. Settelen's point about the use of frequency modulation, this is not acceptable for aviation communications as the capture effects prevent the hearing of weak distress signals through stronger signals.

I note that the authors made no mention of the mechanism of propagation which would account for their observations. It may not be germane to their argument, but this is most important in relation to aerial siting. In the past scatter propagation has been suggested as contributing to long range v.h.f. transmission, but this idea is now not so often supported. The indications supporting the occurrence of scatter propagation have been its independence of aircraft height within wide limits and "drop-outs" of less than one second duration. On the evidence given in the authors' paper the first point is not conclusive. However, did they note any "drop-outs" during the trials and have they any ideas generally on the mechanism of the observed propagation?

The authors (in reply): Brigadier Moppett asked if we noticed any signal "drop-outs". Certainly there were none of sufficient duration or frequency of occurrence to warrant any record of them in the trials, but once, at extreme range, one of the authors noticed a cyclic variation of signal strength between zero and readable, lasting for about a minute. As Brigadier Moppett observes, we have not attempted to explain the propagation mechanism and we think it would be imprudent to do so on the basis of the data obtained to date in these trials.

Mr. Caradoc Williams: The authors have discussed in their paper the wave propagation conditions and their

observations relating to a sea level receiving site. For v.h.f. propagation beyond the horizon, one is primarily interested in the radiation properties of the ground station aerial at very low angles of elevation and under these conditions it might be expected that the nature of the ground and its reflection properties are important at very great distances from the aerial and less so at distances close to the aerial. Therefore, whether the stations are sited near the shore or well inland at elevated sites might appear to be relatively unimportant, provided they are not screened by topographical features. Perhaps the authors would care to make some observations concerning the heights of aerial sites and their preferred locations for beyond the horizon communications.

The authors (*in reply*): Regarding Mr. Caradoc Williams' observations on ground-station aerial-siting, our experience is of course limited to Bahrain, but we agree that an unobstructed horizon is a great advantage and is probably more important than natural height. We also attach importance to an electrically quiet site.

Mr. J. C. Farmer: There is little doubt that taken over several years and including periods of minimum sunspot activity and high magnetic intensity, the availability of h.f. radio-telephony ground-air circuits will be well below 90 per cent.

The importance of the experiments described in this paper, which help to confirm those carried out in other parts of the world, is that they indicate a very worth-

while extension of the normal range for v.h.f. ground-air communications by about 100 nautical miles at a reasonable cost.

I am inclined to agree with Mr. Brunt that the difficulties of obtaining a high probability of contact over greater distances are considerable, and I would ask the authors whether they consider that parametric amplifiers would assist in this direction.

I should also point out a possible penalty of these experiments. If it is clearly shown that consistent increases of range on v.h.f. above those originally assumed for purposes of frequency planning are likely, we may expect eventually greater interference problems and thus more restriction on frequency repetition.

The authors (*in reply*): In theory, parametric amplifiers would appear to have some application in this context, as suggested by Mr. Farmer, but we doubt whether the gain would be economically justified. A receiver noise factor of less than 3 dB is not warranted, due to extra-terrestrial noise; reducing it to say 2 dB by the use of a parametric amplifier would reduce the overall noise by only 1 dB.† Mr. Farmer's warning about the effect on v.h.f. frequency allocation is valid, though as we remarked, the problem is essentially simpler than with the present H.F. system.

† F. Horner, "Extra-terrestrial radio noise as a source of interference in the frequency range 30-1000 Mc/s", *Proc. Instn Elect. Engrs*, 107, Part B, No. 34, pp. 373-6, July 1962 (I.E.E. Paper No. 3267E).

The Closed-Coupler Reciprocity Calibration of Probe Microphones

By

E. ASPINALL, B.Sc. †

Presented at a joint meeting of the Acoustics Group of the Institute of Physics and the Physical Society and the Electro-Acoustics Group of the Institution in London on 17th January 1962.

Summary: Although the closed-coupler reciprocity method of pressure calibration of probe microphones offers, in principle, considerable practical advantages over the current method of using a Rayleigh disk in a resonant tube it is shown that only by careful choice of size of coupler can accurate calibrations by the reciprocity method be achieved, and even these calibrations are restricted to the frequency range below about 1 kc/s.

List of Symbols

I	current
T	identification letter for a transducer
E	electro-motive force
M	pressure sensitivity of a microphone, in terms of the ratio of e.m.f. generated by the microphone to the sound pressure applied to the microphone
Z_A	acoustical impedance (of the cavity or coupler)
Z_{12}	the ratio of the e.m.f. produced by a probe microphone "1" to the current through a second probe microphone "2" when the two probe microphones are connected acoustically by a coupler of known acoustical impedance Z_A (the probe microphone "2" being used as a generator of sound in the coupler and the probe microphone "1" being used as a detector of the sound produced in the coupler by probe microphone "2")
λ	wavelength (of sound)
D	diameter of spherical coupler
V	volume of spherical coupler
ρ	density of air
c	velocity of sound, in air
ω	pulsatance
f	frequency
$j = \sqrt{-1}$	

1. Introduction

It has been the practice for many years in the Research Branch of the British Post Office to make use of the technique of a Rayleigh disk suspended at the centre of a resonant tube¹ for determining the pressure calibration of probe microphones.^{2, 3, 4} This practice

† Post Office Research Station, Dollis Hill, London, N.W.2.

has proved to be well-founded and a high degree of accuracy has been established for frequencies in the range used in telephony and possibly also up to 7–8 kc/s, but by comparison with the closed-coupler reciprocity method^{5, 6} of pressure calibration there exists a major disadvantage in operating the Rayleigh disk method, in that the tube length must be altered to suit the frequency of calibration, or vice versa. This can lead to a long and tedious procedure if a continuous calibration is required over a considerable frequency range, especially when the probe microphone being calibrated has a large number of peaks and troughs in its acoustical characteristics.

Another disadvantage of the Rayleigh disk method is that special equipment is required, particularly the disk itself and its extremely sensitive and delicate suspension, whereas the closed-coupler reciprocity method requires, in principle, only a special but simple small hard-walled coupler. (Both methods require commonplace electrical equipment such as oscillators, amplifiers, attenuators, and meters, all suitable for the audio-frequency range.) This disadvantage of the Rayleigh disk method is an important one at a time when users of acoustical equipment wish to check, even if not actually calibrate, their own acoustical measuring instruments, particularly probe microphones.

2. Recapitulation on the Reciprocity Method

During the 1940s reciprocity methods were developed in America, particularly in relation to the calibration of condenser microphones, to such an extent that in 1949 the American Standards Association issued a standard⁶ based on the reciprocity method and since then this standard has been used as a basis for calibrating condenser microphones in many parts of the world. Reciprocity principles and their application to closed-coupler techniques, such

as are used in the American standard, are described fully in publications on acoustics (such as references 4 and 5, the latter containing a large bibliography on the subject) and it will be sufficient in this paper merely to state the basic principles, the relevant formulae and the conditions that are to be met in applying the method to the calibration of probe microphones.

If two electroacoustic transducers are coupled together by a cavity of small acoustical impedance, and if a current (I_1) through one transducer (T1) used as a generator, or source of sound in the cavity causes an e.m.f. (E_2) to be produced in the second transducer (T2) used as a microphone, or detector of sound, in the cavity then

$$M_1 M_2 = \frac{E_2}{I_1} \cdot \frac{1}{Z_A}$$

$$= \frac{Z_{21}}{Z_A}$$

If now the second transducer (T2) is replaced by a third transducer (T3) and the two transducers T1 and T3 are used as generator and microphone respectively in the manner in which T1 and T2 were used, then

$$M_1 M_3 = \frac{Z_{31}}{Z_A}$$

And if lastly the first transducer T1 is now replaced by the second transducer T2 and the two transducers T2 and T3 are used as generator and microphone respectively in the manner in which T1 and T2 were used, then

$$M_2 M_3 = \frac{Z_{32}}{Z_A}$$

The above three equations are sufficient to determine individually M_1 , M_2 and M_3 and it is readily seen that

$$M_1 = \sqrt{\frac{Z_{21} Z_{31}}{Z_{32}}} \cdot \frac{1}{Z_A}$$

with corresponding expressions for M_2 and M_3 . Apart from mentioning that these formulae apply for linear reversible transducers, which is the case for the sort of probe microphone in use in the Post Office, there are two other conditions to be met both of which relate to the cavity or coupler. First, the acoustical impedance of the coupler must be known and must be small compared with the acoustical impedance at the probe tip of each of the microphones being used (unless these impedances are known, in which case they can be taken into account in the microphone sensitivity determination). Second, the coupler must behave acoustically as a simple capacitance element, or, put another way, for a given sound source the level of the sound pressure throughout the volume of the coupler must be uniform.

3. Assessment of Coupler Requirements

3.1. Shape

A cylindrical shape of coupler had been found apposite in America for use with condenser microphones, in which case two microphones form the two ends of the cylinder. Results of work such as that of Mawardi⁷ on the pressure distribution within a hard-walled cylinder when there is a small sound source in the cylinder wall indicated that a cylindrical shape would be likely to have pressure, and particle velocity, distributions which depart markedly from uniformity for the size of cylinder which would be necessary for the frequency range of interest in calibrating probe microphones. The spherical shape appeared more attractive than any other shape, in that it was easy to specify and make, it had no discontinuities such as the cylinder possessed at its ends, and analytical or numerical solution of the sound field equations with spherical symmetry seemed to be more likely than with other shapes. (It may be noted here that the spherical shape also allows the three probe microphones inherent in this use of the reciprocity method to be placed with the tips of the probes on a great circle of the sphere and this would enable a complete experimental routine to be completed without disturbing the cavity and the three microphones; potentially, this offers a great practical advantage.)

3.2. Size

In order to assess a suitable value for the size of the spherical coupler, an attempt was first made to calculate the pressure distribution and the particle velocity distribution set up within a spherical cavity by a small source of sound on the surface of the cavity. An analytical solution was obtained to the field equations in the form of an infinite series each term of which contained Legendre polynomials. These polynomials have only been tabulated, other than for a few special values of the argument, for orders up to the first thirty-two and only by undertaking an inordinate amount of computation would this theoretical approach have been of help. Accordingly, a two-fold experimental approach was adopted, one investigation dealing with pressure distribution and the other dealing with acoustical impedance determinations.

The results of the investigation relating to pressure distribution which is described in Appendix 1 show that when the ratio of the wavelength (λ) of the sound to the diameter (D) of the sphere is less than about 40 there is a significant change in sound pressure from one point in the cavity to another, and as one would expect the smaller the ratio the more severe the change.

The results of the acoustical impedance measurements which are described in Appendix 2 show that

when the ratio λ/D is less than about 20 the impedance of a spherical cavity departs materially from being inversely proportional to frequency, i.e. from behaving as a capacitance element.

Thus, to satisfy the second of the two conditions stated at the end of Section 2, the indications are that on the one hand the ratio λ/D should not be less than about 40 and on the other hand the ratio λ/D should not be less than about 20. There is, however, no criterion by which to determine precisely what the minimum value of λ/D should be for application to this reciprocity method and it is sufficient to take $\lambda/D = 30$ as a guide to the value below which it is very likely that calibration results will be poor in accuracy if not even worthless, and above which there is an increasing possibility of accurate calibrations being achieved.

Dealing now with the first of the two conditions at the end of Section 2, it was noted that the precise value of the acoustical impedance of an individual probe microphone is not usually known and in general the impedance cannot be measured readily. This impedance is, however, several thousand acoustical ohms or more throughout the frequency range of interest and so the impedance of the cavity must be small compared with several thousand acoustical ohms. As a guide, about seven hundred ohms may be taken as the upper limit of the impedance of the cavity, and as this impedance is given by the expression

$$Z_A = \frac{\rho c^2}{j\omega V} = \frac{3\rho c^2}{j\pi^2 f D^3}$$

this places a lower limit on the value of D which is inversely proportional to the cube root of frequency.

3.3. Results of Assessments

The combination of the effects of the two conditions given at the end of Section 2 enables a range of values

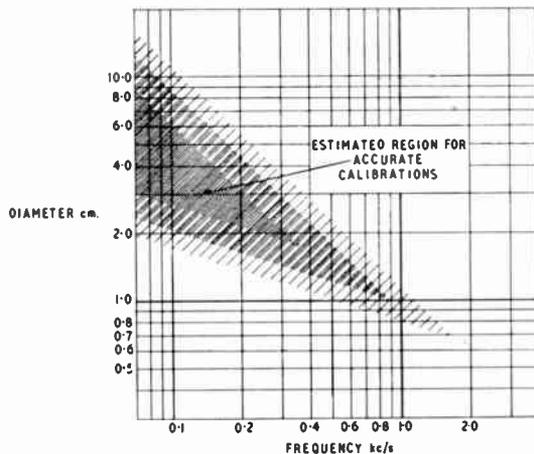
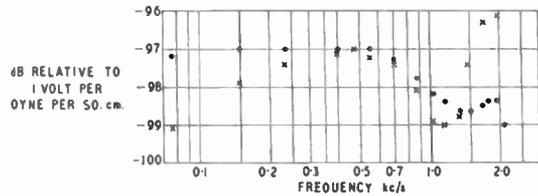
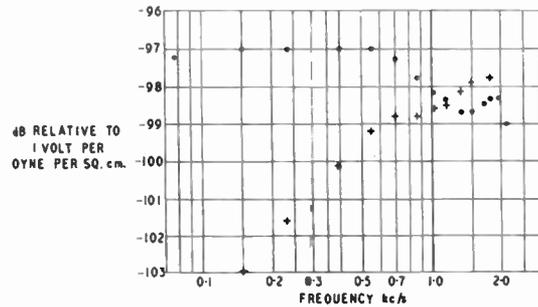


Fig. 1. Relationship between diameter of cavity and frequency for accurate reciprocity calibrations.



(a) Using spherical cavity diameter 1 cm (for reciprocity calibration).



(b) Using spherical cavity diameter 0.5 cm (for reciprocity calibration).

Fig. 2. Sensitivity frequency characteristics of probe microphones. Typical comparisons of calibrations by Rayleigh disk method (*) and reciprocity method (x or +).

of diameter for the sphere to be determined throughout the frequency range of interest and the permissible ranges are illustrated on Fig. 1.

It appears from Fig. 1 that it is extremely unlikely that a single spherical cavity can be found to cater for even a substantial part of the whole frequency range, that it is extremely unlikely that there is any volume at all to cater for the frequencies higher than a very few kc/s, but that there is a possibility of obtaining a cavity for frequencies up to 1 or perhaps even 2 kc/s. Notwithstanding this bleak prospect, two cavities of diameters 1 cm and 0.5 cm respectively were adopted, the larger cavity being used to attempt reciprocity calibrations, and the smaller cavity being intended to test the validity of the above assessments on size restrictions.

4. Results of Attempts at Calibration

Typical of the results obtained by carrying out the reciprocity technique on several probe microphones are those given on Fig. 2 which show in absolute terms not only the pressure calibration by the reciprocity method but also for comparison the pressure calibration of the same probe microphone by the well-established Rayleigh disk method. It will be seen from Fig. 2 (a) that using a spherical cavity of diameter

1 cm the reciprocity method yields results whose accuracy is not tolerable when the frequency of calibration exceeds about 1.5 kc/s or is lower than about 200 c/s. The failure at high-frequency is due to the cavity failing to behave as a simple acoustical capacitance, and the failure at the low-frequency is due to the cavity impedance becoming too high. These results are in conformity with the statements made in Section 3.3 when dealing with the assessments of cavity size.

The results shown on Fig. 2 (b) obtained using a spherical cavity of diameter 0.5 cm emphasize the failure at low frequencies and again confirm the statements made in Section 3.3.

5. Conclusions

The investigation shows that the closed-coupler reciprocity method, attractive though it is, has a very restricted field of application for the pressure calibration of probe microphones commonly used in the Post Office and indeed in other acoustical laboratories in Great Britain; there is no chance of achieving accurate calibrations by the reciprocity method for frequencies above about 1 kc/s, and only by careful choice of coupler can accurate calibration be achieved below about 1 kc/s.

6. Acknowledgment

Acknowledgment is made to the Engineer-in-Chief of the Post Office for permission to make use of information contained in this paper, and to the author's colleagues at the P.O. Research Station who have collaborated in the investigation.

7. References

1. The International Telephone Consultative Committee XVIIth Plenary Assembly, October 1954. Annexes to Volume IV ("Green Book"), Annex 5 pp. 47-52 "Theory of calibration of microphones with a stationary wave tube and a Rayleigh disk". (Published by the International Telecommunication Union, Geneva, 1957.)
2. R. B. Archbold, "An experimental probe microphone for the measurement of sound pressures". *P.O. Elect. Engrs J.*, 45, Part 4, p. 145, January 1953.
3. W. West, "Probe microphones as laboratory standards", *Acustica*, 4, No. 1, p. 131, 1954.
4. A. N. Rapsey, "Some investigations relating to the calibration of probe microphones", *J. Brit.I.R.E.*, 25, No. 1, p. 38, January 1963.
5. L. L. Beranek, "Acoustic Measurements", p. 123 *et seq.* (John Wiley, New York, 1949.)
6. American Standards Association, "Method for the Pressure Calibration of Laboratory Standard Pressure Microphones". Z.24.4-1949.
7. Osman K. Mawardi, "Measurement of acoustic impedance", *J. Acoust. Soc. Amer.*, 21, p. 86, March 1949.
8. E. W. Ayers, E. Aspinall and J. Y. Morton, "An impedance measuring set for electrical, acoustical and mechanical impedances", *Acustica*, 6, p. 11, 1956.

8. Appendix 1

Sound Pressure Distribution in a Spherical Cavity (due to a small source on the surface of the cavity)

Since the fundamental field equations show that sound pressure distribution is a function of the ratio of the wavelength of the sound to the diameter of the cavity rather than a function of the absolute magnitude of either, it was possible to use a large sphere,

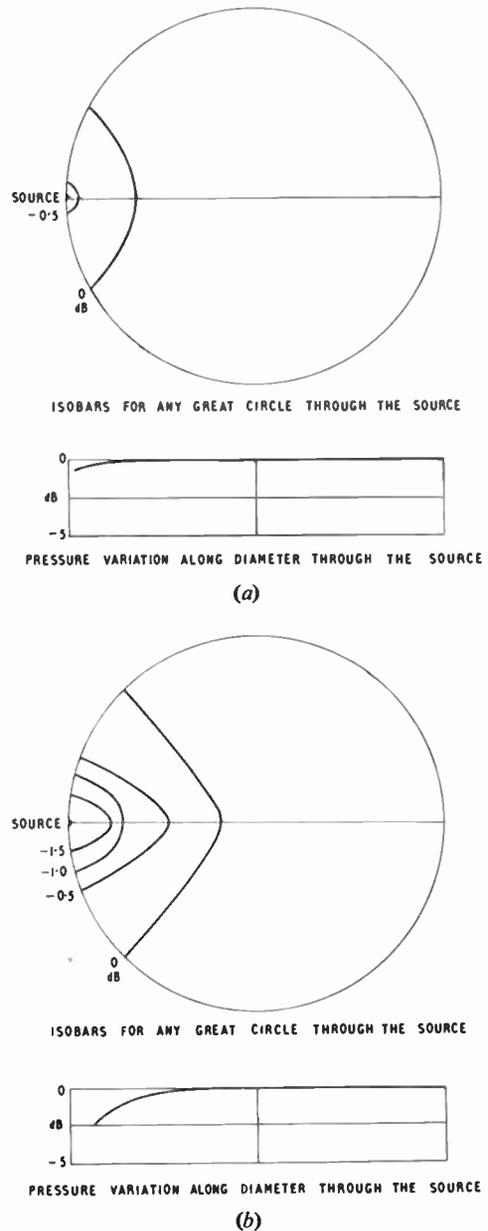


Fig. 3. Sound pressure distribution inside a spherical cavity set up by a small source on the surface. Ratio of wavelength (λ) of sound to diameter (D) of sphere: (a) $\lambda/D = 40$; (b) $\lambda/D = 20$; (c) $\lambda/D = 10$.

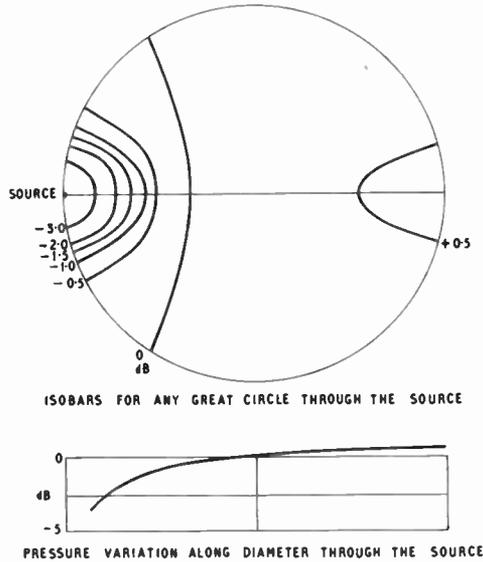


Fig. 3. (c)

of diameter about 6.4 cm and correspondingly larger wavelengths, and to make use for the most part of existing measuring and indicating equipment.

A small sound source was provided on the surface of the sphere (by using a probe microphone as a sound generator). Arrangements were made for measuring the sound pressure generated by this source at different positions, relative to the source, within the sphere and also on the surface of the sphere. The instrument used for this was the type of probe microphone the calibration of which is the main subject of this paper. Although the probe is of small size and the acoustical input impedance is several thousand acoustical ohms, it was realized that if the pressure distribution in the sphere departed from uniformity it might be possible for the impedance at the area in the sphere at which the probe was to be placed to be of the same order or even larger than that of the probe microphone. This would happen if over the small area concerned the sound pressure were high and the particle velocity were low. The measurements were therefore repeated using a special probe microphone whose acoustical input impedance was made several times that of the normal one used. The results of the measurements using this special probe microphone were very nearly the same as those obtained with the normal probe microphone, and thereby confirmed that the pressure

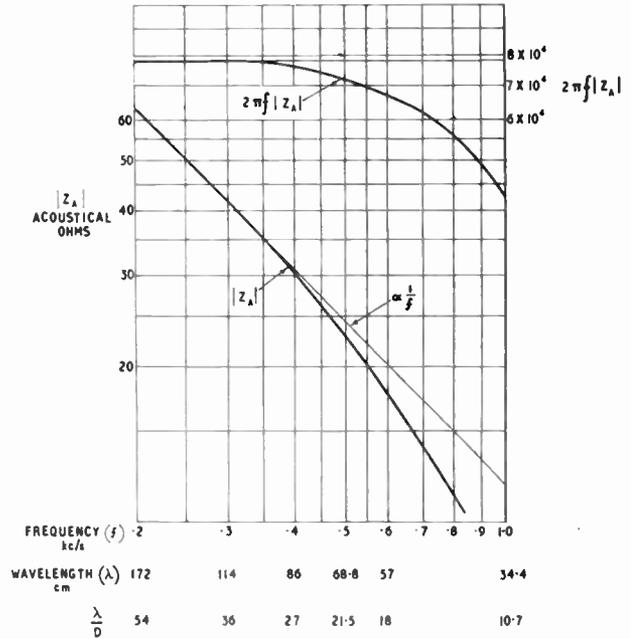


Fig. 4. Acoustical impedance of a spherical cavity of diameter (D) 3.2 cm.

distributions shown on Fig. 3 are very nearly correct. For smaller values of the ratio λ/D the pressure distribution becomes even more non-uniform than as shown on Fig. 3 (c); for instance, when $\lambda/D = 2.5$ the change in pressure round the surface of the sphere is as much as 20 dB.

9. Appendix 2

Acoustical Impedance of a Spherical Cavity

In order once again to make use of equipment⁸ proven by other acoustical impedance measurements to be satisfactory, a rigid-walled sphere of internal diameter 3.2 cm was made and its acoustical impedance at a small area on the surface of the sphere was measured. The results are given on Fig. 4. (It may be noted in passing that the acoustical impedance of the sphere has zero reactance when the ratio of wavelength to diameter is about 7.)

Manuscript received by the Institution on 7th March 1962.

(Cont. No. 57/EAS)

© The British Institution of Radio Engineers, 1963

Some Investigations Relating to the Calibration of Probe Microphones

By

A. N. RAPSEY, M.Sc.†

Presented at a joint meeting of the Acoustics Group of the Institute of Physics and the Physical Society and the Electro-Acoustics Group of the Institution in London on 17th January 1962.

Summary: It has been the practice in the Post Office Research Station to use the technique of a Rayleigh disk in a resonant tube for obtaining pressure calibrations of microphones, particularly probe microphones. The accuracy of such calibrations has long been known from fundamental considerations to be suspect above about 4 kc/s, for the tube of diameter 4.5 cm which has been found convenient for general use so far. This paper describes and gives the results of an attempt to check the accuracy of this limit.

1. Introduction

Lord Rayleigh first observed that a light disk suspended in a "cylindrical sounding resonator" tended to set itself across the tube, that is at right angles to the direction of the alternating air current, and further that the extent of this turning was proportional to the "energy" of the motion.

Figure 1 shows the apparatus in which this principle can be used to calibrate probe microphones. The light disk is suspended from a fine glass or fused silica fibre so that it is at the centre of a resonant tube. The source of sound is at one end of the tube and the probe microphone is mounted with the end of its probe flush with the inside surface of the other end. The tube and the suspension column must be well sealed to prevent leakages which would alter the pressure in the tube. An optical system is used to measure the deflexion of the disk.

The relationship between the torque on the disk and the particle velocity of the sound wave was developed by W. Koenig‡ in 1891 and is

$$L_0 = S\phi = \frac{1}{6}\rho d^3 \bar{V}^2 \sin 2\theta$$

where L_0 is the torque on the disk,

ϕ is the angle of rotation of the disk from its initial position, caused by the sound wave,

ρ is the density of the medium in the tube,

d is the diameter of the disk,

\bar{V} is the r.m.s. particle velocity of the sound wave,

and θ is the angle between the normal to the disk and the direction of propagation of the sound wave, i.e. the axis of the tube. In practice θ is arranged to be 45 deg and thus $\sin 2\theta = 1$.

From this expression the particle velocity can be obtained from the angle of rotation of the disk.

Now for a stationary wave tube at odd multiples of half-wave resonance, there is a particle velocity maximum at the centre of the tube and a pressure maximum at the ends of the tube. Under these conditions the pressure at the ends of the tube

$$P = \rho c \bar{V}$$

where c is the velocity of sound in the medium and \bar{V} is the particle velocity at the centre of the tube.

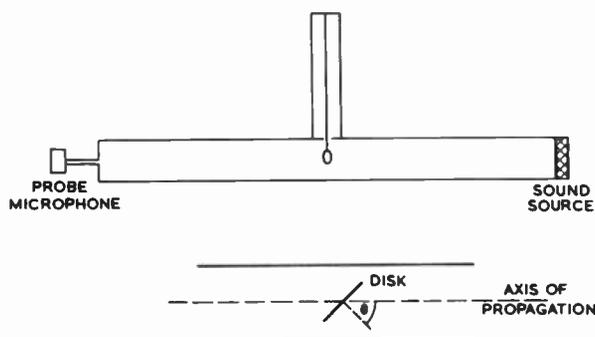


Fig. 1. Principle of Rayleigh disk apparatus.

Thus the pressure at the end of the tube and hence at the probe of the microphone can be calculated from the rotation of the Rayleigh disk. The output from the microphone is measured electrically and the microphone sensitivity calculated. This process gives values of the microphone sensitivity at a series of spot frequencies determined by the length of the tube and the velocity of sound in the medium. Values at different frequencies are sometimes obtained by varying the length of the tube.

† Post Office Research Station, Dollis Hill, London, N.W.2.

‡ W. Koenig, *Ann. der Phys.*, 43, p. 43, 1891.

The Koenig expression was developed with the following assumptions:

- (a) the fluid medium was incompressible,
- (b) the disk was an infinitely thin ellipsoid,
- (c) the disk is rigid and rigidly held,
- (d) the sound scattered by the disk can be neglected,
- (e) there are no forces due to heat conduction or viscosity,
- (f) there is no vortex motion around the disk.

Most of these conditions clearly do not exist in practice and correction factors have been developed by various workers to account for actual conditions. These corrections were very thoroughly discussed by R. A. Scott.†

2. Existing Apparatus

The Rayleigh disk apparatus in use at the Post Office Research Station has a tube 96.5 cm long and 4.5 cm in diameter, with a mica Rayleigh disk 0.68 cm in diameter suspended at its centre by a glass fibre. The disk has a small plane mirror stuck to its surface. Figure 2 (a) shows how a beam of light from a projector enters the tube through a glass window at A, is reflected from a fixed mirror B, mounted inside the tube, and emerges from another glass window at D.

Two criteria may determine the upper limiting frequency at which this apparatus can be used to give accurate results. The first is given by Rayleigh's criterion for the propagation of plane waves in a tube, which states that such waves will exist provided that

$$\frac{\text{diameter of the tube}}{\text{wavelength of the sound}} < 0.59$$

This gives a limiting frequency of 4.5 kc/s.

The second criterion is the establishment of the first transverse mode of resonance which should affect the sound pressure at the microphone. This could occur when the diameter of the tube equals half a wavelength. In this case this occurs at 3.8 kc/s.

Thus, although the apparatus has been used up to 7 kc/s, the results have been generally held to be suspect above about 4 kc/s.

3. New Apparatus

It was decided to test this assumption by constructing a new apparatus on physical similarity principles with a scale factor of 1 : 3. Thus a tube of 1.5 cm diameter was made. This should have made the limiting frequencies 13.5 kc/s on Rayleigh's criterion and 11.4 kc/s due to transverse resonance.

The size of the disk was scaled down by the same factor, making its diameter 0.23 cm. After careful

† R. A. Scott, *Proc. Roy. Soc.*, 183A, p. 296, 1945.

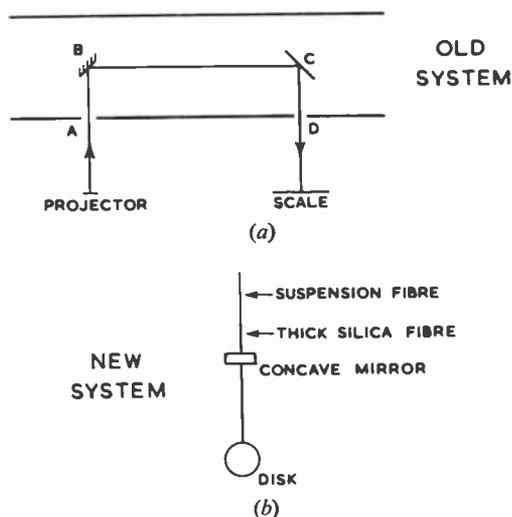


Fig. 2. Optical systems used with Rayleigh disk apparatus.

consideration of the corrections already mentioned, it was decided to make it 0.0025 cm thick out of platinum, to provide sufficient rigidity and inertia.

It was no longer possible to attach a mirror directly to so small and thin a disk without seriously upsetting its performance, and it proved to be very difficult to make the disk flat and smooth enough to act as its own mirror. Thus it was decided to try to improve the optical system as shown in Fig. 2 (b). The new Rayleigh disk was fastened to the lower end of a piece of fused silica fibre about 6 cm long, which was thick compared to the suspension fibre. A small concave mirror was placed near the upper end of this thicker fibre so that it hung near the base of the suspension column. A window cut in this column allowed light from a projector to be reflected from this mirror on to a scale. A second mirror was mounted back to back with the first to produce a more evenly balanced system which hung vertically. These arrangements produced a bright, sharp image on the screen and rendered the fixed mirror inside the tube superfluous.

4. Results

The sensitivities of a number of probe microphones were measured on the old wider and new narrower tubes in quick succession. A typical result is shown in Fig. 3. Below 4 kc/s the sensitivities measured with the new tube are generally slightly higher than those measured with the old tube. Above 4 kc/s there is an interesting state of affairs which can perhaps be interpreted as follows.

With the old tube of 4.5 cm diameter transverse modes of resonance should occur at 3.8, 7.6 and 11.4 kc/s. There appears to be no effect at the lowest of these frequencies, perhaps because it is below the frequency given by Rayleigh's criterion for plane wave

propagation in the tube, so that there may be no transverse component of the sound waves to produce a resonant condition. However, there are notable effects at, or very near to, the other two frequencies.

It is interesting to note that these effects will be most noticeable with probe microphones, for larger microphones will integrate the response over their sensitive areas. In the extreme case of a microphone whose area covers the whole area of the end of the tube, the effects of transverse resonance will probably not be detected.

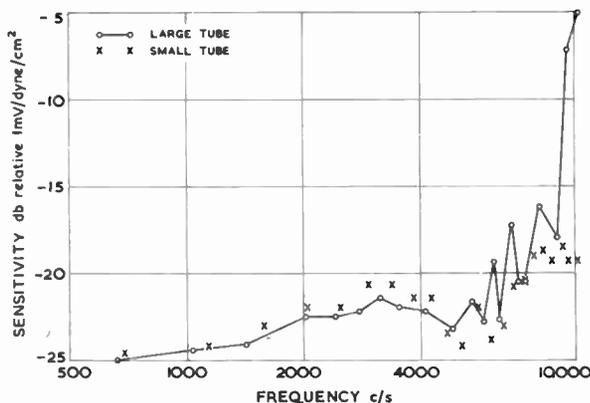


Fig. 3. Typical microphone calibrations.

Figure 3 shows no values below 700 c/s because at some of these frequencies disk deflexions were observed which were in the reverse direction, i.e. the disk turned towards a position in line with the axis of the tube. This effect was first reported by Skinner†; however, his apparatus was quite different and he attributed his results to causes peculiar to that apparatus. Figure 4 shows a smooth curve joining the observed series of maximum and minimum deflexions. Investigations showed that these “negative” deflexions were caused by resonance phenomena in the suspension column of the apparatus. Measurements made with the tip of a probe microphone in the hole through which the suspension fibre enters the tube, showed that the whole suspension column acted as a Helmholtz resonator and that the “negative” deflexions corresponded to its resonant frequencies. Experiments with the disk alone and with the mirrors alone, showed that both were affected. Slight departures from the strictly vertical hanging position of the disk and

† C. H. Skinner, *Phys. Rev.*, 27, p. 346, 1926.

mirror assembly caused this air movement to produce a turning moment.

This moment decreased rapidly with increasing frequency and above 700 c/s was not detectable at the sound pressures used in the main tube, i.e. 50 dyne/cm².

The expression for the fundamental resonant frequency of a Helmholtz resonator is

$$F = \frac{C}{2\pi} \sqrt{\frac{A}{lv}}$$

where A is the area of the neck of the resonator,

l is the length of the neck of the resonator,

and v is the volume of the resonator.

By a suitable choice of these parameters the resonant frequencies could be moved below the range considered for microphone calibrations.

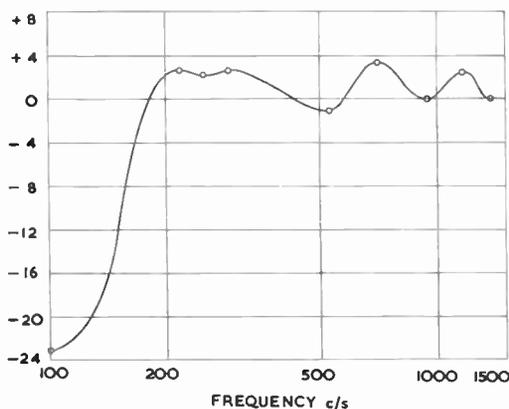


Fig. 4. Disk deflexions at low frequencies.

5. Conclusion

These results appear to show that the limit of confidence in the results from the older, larger tube is approximately correct. Once the Rayleigh criterion for plane wave propagation in the tube has been exceeded, transverse resonances can be set up leading to inaccurate results.

6. Acknowledgment

Acknowledgment is made to the Engineer-in-Chief of the Post Office for permission to make use of the information contained in this paper.

*Manuscript received by the Institution on 7th March 1962.
(Cont. No. 58/EA6)*

© The British Institution of Radio Engineers, 1963

The Use of Transfer Matrices in the Analysis of Conditions of Oscillation

By

K. G. NICHOLS, M.Sc.
(Associate Member)†

Summary: The condition of oscillation of a feedback oscillator in terms of its open loop transfer matrix is deduced. The method of determining the open loop transfer matrix for a feedback oscillator is demonstrated by considering illustrative examples, and the frequencies of oscillation and the starting conditions of oscillation are deduced for some of these examples.

1. Introduction

A variety of methods for obtaining the conditions of oscillation of feedback oscillators are available. In such analyses it is usual to assume that the operation of the active devices is restricted to substantially linear regions of their characteristics. The active devices may then be represented by linear incremental equations, or linear equivalent circuits, expressed in terms of constant parameters which are functions of the operating point of the device.

Triode valves, pentode valves with fixed screen and suppressor to cathode voltages, and transistors are three-pole active devices which in oscillator circuits are often treated as linear devices for the purpose of determining oscillation conditions. Such three-poles may be represented by 2×2 matrices, in particular by transfer matrices,^{2,3} as can also the normal passive circuit elements.^{1,2,3}

In what follows the condition of oscillation of a feedback oscillator is first deduced in terms of an overall, open loop, transfer matrix, and then the methods of determining the overall transfer matrix from the transfer matrices of the individual circuit elements is demonstrated, for some common oscillator circuits.

2. Conditions of Oscillation in terms of the Open Loop Transfer Parameters

In Fig. 1 the incremental equivalent circuit of a feedback oscillator is shown. This comprises a four-pole network which has the whole of the output signal fed back as the input signal to the four-pole. The transfer matrix^{1,2,3} of the four-pole network is taken as

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \dots\dots(1)$$

That is the input voltage and current are related to the output voltage and current by (assuming for the present that the feedback loops are broken)

† Department of Electronics, University of Southampton.

$$\begin{pmatrix} v_1 \\ i_1 \end{pmatrix} = A \begin{pmatrix} v_2 \\ i_2 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} v_2 \\ i_2 \end{pmatrix} \dots\dots(2)$$

It is now required to find conditions on the elements of the transfer matrix A such that when the feedback loops in Fig. 1 are closed the system oscillates. When the loops are closed, necessarily

$$\begin{pmatrix} v_2 \\ i_2 \end{pmatrix} = \begin{pmatrix} v_1 \\ i_1 \end{pmatrix} \dots\dots(3)$$

so that on substituting eqn. (3) into eqn. (2),

$$\begin{pmatrix} v_1 \\ i_1 \end{pmatrix} = A \begin{pmatrix} v_1 \\ i_1 \end{pmatrix}$$

that is $(A - I) \begin{pmatrix} v_1 \\ i_1 \end{pmatrix} = 0 \dots\dots(4)$

where I is the 2×2 unit matrix,

or $\begin{pmatrix} a_{11} - 1 & a_{12} \\ a_{21} & a_{22} - 1 \end{pmatrix} \begin{pmatrix} v_1 \\ i_1 \end{pmatrix} = 0 \dots\dots(5)$

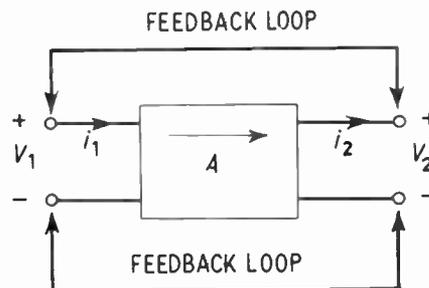


Fig. 1. Equivalent circuit of the feedback oscillator.

These last equations have non-trivial solutions for v_1 and i_1 (that is oscillatory signal) if and only if

$$\begin{vmatrix} a_{11} - 1 & a_{12} \\ a_{21} & a_{22} - 1 \end{vmatrix} = 0 \dots\dots(6)$$

that is, if and only if

$$a + 1 = a_{11} + a_{22} \dots\dots(7)$$

where $a = a_{11} \cdot a_{22} - a_{12} \cdot a_{21} \dots\dots(8)$

Equation (7) gives the conditions of oscillation of the system.

If the four-pole with transfer matrix A is unilateral, the determinant of the matrix is zero,³ that is

$$a = 0 \quad \dots\dots(9)$$

and so the condition of oscillation given by eqn. (7) reduces to

$$1 = a_{11} + a_{22} \quad \dots\dots(10)$$

It follows that an investigation of the conditions of oscillation of the oscillator reduces to a determination of the transfer parameters a_{11} and a_{22} , of the open loop system, in terms of the complex frequency s . Equation (7) or (10), whichever is appropriate, is a polynomial equation in s and if the equation has any root with positive real part oscillation commences and grows exponentially in amplitude until non-linearity caused by saturation of a valve or transistor limits the amplitude of oscillation.

In most cases it is required that the circuit shall just oscillate. This means that the root with greatest real part shall have this real part just equal to zero, that is the root is of the form $j\omega$, where ω is the real frequency. The determination of the starting conditions of oscillation thus amounts to computing a_{11} and a_{22} in terms of the real frequency ω and then separating out the real and imaginary parts of eqn. (7) or (10), whichever is appropriate. The real part will then give the starting condition of oscillation and the imaginary part the frequency of oscillation. This technique is illustrated in the examples which follow.

3. Tuned Anode Oscillator

A basic circuit configuration of a tuned anode oscillator is shown in Fig. 2. Figure 3 shows the equivalent circuit of the oscillator, the feedback loop being completed by the connection of terminal Y to terminal X.

The four-pole between the input and output terminals can be broken down into a series-cascaded arrangement of basic four-poles each having a simple transfer matrix.

The overall transfer matrix is thus,^{1,2,3}

$$A = -\frac{1}{\mu} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & r_a \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ j\omega C & 1 \end{pmatrix} \begin{pmatrix} 1 & R+j\omega L \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & j\omega M \\ -\frac{1}{j\omega M} & 0 \end{pmatrix} \begin{pmatrix} 1 & R'+j\omega L' \\ 0 & 1 \end{pmatrix} \dots\dots(11)$$

each matrix in this product arising from an elementary circuit element.

For example the factor

$$A_1 = -\frac{1}{\mu} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \dots\dots(12)$$

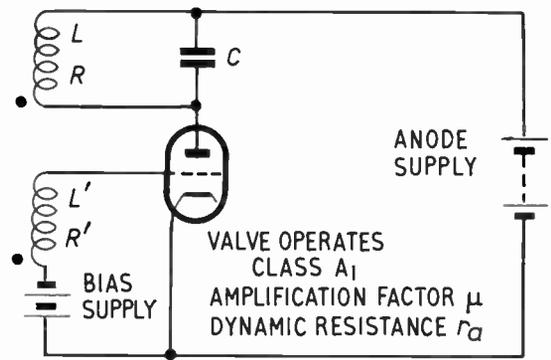


Fig. 2. A basic form of tuned anode oscillator.

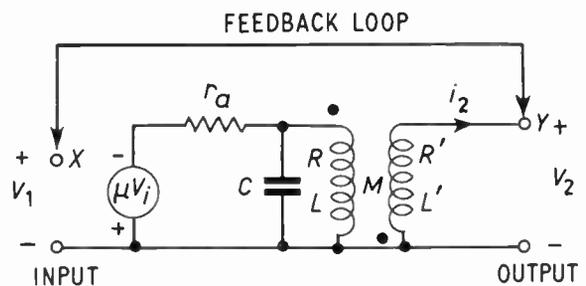


Fig. 3. Linear class A1 equivalent circuit of a tuned anode oscillator.

is the transfer matrix of the voltage controlled generator μv_1 ,³ and the factor

$$A_5 = \begin{pmatrix} 0 & j\omega M \\ -\frac{1}{j\omega M} & 0 \end{pmatrix} \quad \dots\dots(13)$$

is the transfer matrix of an ideal mutual inductance M .³ The primary and secondary inductances and resistances are represented by the factors

$$A_4 = \begin{pmatrix} 1 & R+j\omega L \\ 0 & 1 \end{pmatrix} \quad \dots\dots(14)$$

and

$$A_6 = \begin{pmatrix} 1 & R'+j\omega L' \\ 0 & 1 \end{pmatrix} \quad \dots\dots(15)$$

respectively.

Multiplying out, step by step, the matrix factors of eqn. (11) gives the transfer matrix

$$A = -\frac{1}{\mu} \begin{pmatrix} -\frac{1}{j\omega M} [r_a + R - \omega^2 L C r_a + j\omega(L + C R r_a)] & , & a_{12} \\ a_{21} & , & 0 \end{pmatrix} \dots\dots(16)$$

Note that a_{12} and a_{21} have not been computed as they are not required for eqn. (7) or (10) ($a = 0$, as follows) and that also $a_{22} = 0$, which is a consequence of the zero a_{21} and a_{22} elements in the first factor of the right-hand side of eqn. (11). Another immediate consequence of eqn. (11) is that $a = |A| = 0$; this follows since the determinant of the first factor on the right-hand side of eqn. (11) is zero. That is, it requires only one unilateral element in a series-cascaded set to make the whole network unilateral.³ Since $a = 0$, eqn. (10) may be used to obtain the conditions of oscillation as follows:

$$-\frac{j}{\mu\omega M} [r_a + R - \omega^2 L C r_a + j\omega(L + C R r_a)] = 1$$

or $(L + C R r_a) - j \cdot \frac{r_a}{\omega} \left(1 + \frac{R}{r_a} - \omega^2 L C\right) = \mu M$ (17)

Equating real and imaginary parts gives

$$M = \frac{C R}{g_m} + \frac{L}{\mu} \dots\dots(18)$$

and $\omega = \frac{1}{\sqrt{L C}} \left(1 + \frac{R}{r_a}\right)^{\frac{1}{2}} \simeq \frac{1}{\sqrt{L C}} \dots\dots(19)$

as the oscillation starting condition and oscillation frequency respectively.

4. Tuned Grid Oscillator

This differs from the case of the tuned anode oscillator considered in Section 3 in that the anode load inductance L is now untuned whereas the grid load inductance L' is now tuned. This means that the capacitor C now shunts L' and not L .

The matrix factor

$$A_3 = \begin{pmatrix} 1 & , & 0 \\ j\omega C & , & 1 \end{pmatrix} \dots\dots(20)$$

on the right-hand side of eqn. (11) will now appear at the end of the product.

Since matrix multiplication is non-commutative, this alteration of the order of the factors in eqn. (11) will lead to a different open loop transfer matrix for the oscillator. The further development of this example follows closely that of the tuned anode oscillator of Section 3 and so will not be carried out.

5. Transistor Ladder Oscillator

A transistor oscillator, using a symmetrical ladder network to provide the necessary phase reversal, will now be considered. The basic configuration of a 3-stage ladder network oscillator is shown in Fig. 4.

The incremental equivalent circuit of the ladder oscillator of Fig. 4 is shown in Fig. 5.

In this simple version of the oscillator the first resistor R of the ladder is also the d.c. load for the collector of the transistor. The emitter decoupling capacitor C_e causes the transistor to operate in common emitter configuration at the oscillation frequency (otherwise the transfer parameters of the transistor stage can be expressed in terms of the common emitter transfer parameters of the transistor and in terms of C_e and R_e).³

The open loop transfer matrix can again be determined as the product of the transfer matrices of the series-cascaded elements. To this end the transfer matrix of one stage of the ladder is computed as

$$\begin{pmatrix} 1 & , & 0 \\ \frac{1}{R} & , & 1 \end{pmatrix} \begin{pmatrix} 1 & , & \frac{1}{j\omega C} \\ 0 & , & 1 \end{pmatrix} = \begin{pmatrix} 1 & , & \frac{1}{j\omega C} \\ \frac{1}{R} & , & 1 + \frac{1}{j\omega C R} \end{pmatrix} \dots\dots(21)$$

The transfer matrix of the 3-stage ladder is then

$$\begin{pmatrix} 1 & , & \frac{1}{j\omega C} \\ \frac{1}{R} & , & 1 + \frac{1}{j\omega C R} \end{pmatrix}^3 = \begin{pmatrix} 1 + \frac{3}{j\omega C R} + \frac{1}{(j\omega C R)^2} & , & \frac{3}{j\omega C} + \frac{4}{(j\omega C)^2 R} + \frac{1}{(j\omega C)^3 R^2} \\ \frac{3}{R} + \frac{4}{j\omega C R^2} + \frac{1}{(j\omega C)^2 R^3} & , & 1 + \frac{6}{j\omega C R} + \frac{5}{(j\omega C R)^2} + \frac{1}{(j\omega C R)^3} \end{pmatrix}$$

$$= \begin{pmatrix} 1 - j3x - x^2 & , & -R x(j3 + 4x - jx^2) \\ \frac{1}{R}(3 - j4x - x^2) & , & 1 - j6x - 5x^2 + jx^3 \end{pmatrix} \dots\dots(22)$$

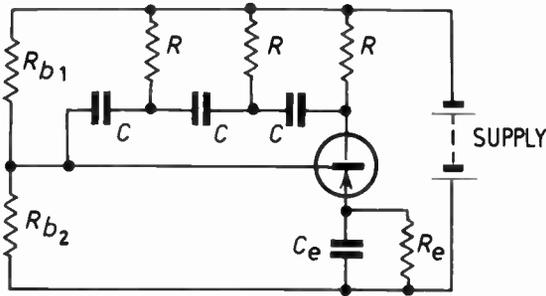


Fig. 4. Basic configuration of a three-stage ladder oscillator.

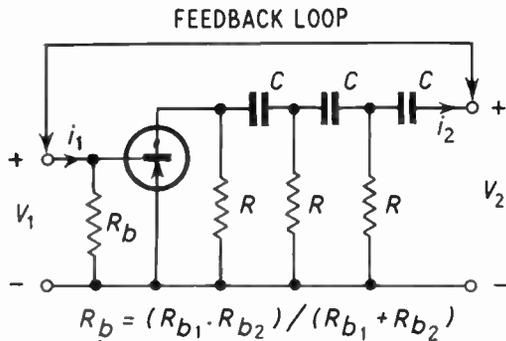


Fig. 5. Equivalent circuit of three-stage ladder oscillator.

where
$$x = \frac{1}{\omega CR} \quad \dots\dots(23)$$

Taking the transfer matrix of the transistor as

$$A_t = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} \quad \dots\dots(24)$$

for the whole system the open loop transfer matrix is

$$A = \begin{pmatrix} 1 & 0 \\ \frac{1}{R_b} & 1 \end{pmatrix} \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} \times \begin{pmatrix} 1-j3x-x^2 & -Rx(j3+4x-jx^2) \\ \frac{1}{R}(3-j4x-x^2) & 1-j6x-5x^2+jx^3 \end{pmatrix} \quad \dots\dots(25)$$

The determinant of this transfer matrix is given by

$$a = |A| = \begin{vmatrix} 1 & 0 & \alpha_{11} & \alpha_{12} & 1 & 0 & 1 & \frac{1}{j\omega C} \\ \frac{1}{R_b} & 1 & \alpha_{21} & \alpha_{22} & \frac{1}{R} & 1 & 0 & 1 \end{vmatrix}^3 = \begin{vmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{vmatrix} = \alpha_{11}\alpha_{22} - \alpha_{12}\alpha_{21} = \alpha \quad (\text{say}) \quad \dots\dots(26)$$

since the determinant of the transfer matrix of a passive network is unity.^{3, 4}

The matrix multiplication on the right-hand side of eqn. (25) gives the elements a_{11} and a_{22} of A as

$$a_{11} = \alpha_{11}(1-j.3x-x^2) + \frac{\alpha_{12}}{R}(3-j.4x-x^2) \quad \dots\dots(27)$$

and

$$a_{22} = -\left(\frac{\alpha_{11}}{R_b} + \alpha_{21}\right)Rx(j.3+4x-jx^2) + \left(\frac{\alpha_{12}}{R_b} + \alpha_{22}\right)(1-j.6x-5x^2+jx^3) \quad \dots\dots(28)$$

The conditions of oscillation can now be deduced by substituting eqns. (26), (27) and (28) into eqn. (7) namely

$$1 + \alpha = a_{11} + a_{22} \quad \dots\dots(29)$$

and then separating real and imaginary parts.

The details of this procedure depend largely on the exact form in which the transfer matrix of the transistor is given. For the sake of example it will be assumed that the frequency of oscillation is low enough for the low frequency, common emitter, hybrid parameters of the transistor to be appropriate.

It is readily shown³ that the expression for the transfer matrix in terms of the hybrid parameters is

$$\begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} = -\frac{1}{h_{21}} \begin{pmatrix} h & h_{11} \\ h_{22} & 1 \end{pmatrix} \quad \dots\dots(30)$$

where
$$h = h_{11}h_{22} - h_{12}h_{21} \quad \dots\dots(31)$$

and so
$$\alpha = \frac{h_{12}}{h_{21}} \quad \dots\dots(32)$$

Equations (27) and (28) now become respectively,

$$a_{11} = -\frac{1}{h_{21}} \left[h(1-j3x-x^2) + \frac{h_{11}}{R}(3-j4x-x^2) \right] \quad \dots\dots(33)$$

and

$$a_{22} = \frac{1}{h_{21}} \left[\left(\frac{h}{R_b} + h_{22}\right)Rx(j3+4x-jx^2) - \left(\frac{h_{11}}{R_b} + 1\right)(1-j6x-5x^2+jx^3) \right] \quad \dots\dots(34)$$

Substituting eqns. (32), (33) and (34) in eqn. (29) gives

$$h_{21} + h_{12} = -c_0 + c_2x^2 + jx(c_1 - c_3x^2) \quad \dots\dots(35)$$

where

$$c_0 = 1 + \left(\frac{3}{R} + \frac{1}{R_b}\right)h_{11} + h \quad \dots\dots(36)$$

$$c_1 = 6 + \left(\frac{4}{R} + \frac{6}{R_b}\right)h_{11} + 3Rh_{22} + 3\left(1 + \frac{R}{R_b}\right)h \quad \dots\dots(37)$$

$$c_2 = 5 + \left(\frac{1}{R} + \frac{5}{R_b}\right)h_{11} + 4Rh_{22} + \left(1 + \frac{4R}{R_b}\right)h \quad \dots\dots(38)$$

$$c_3 = 1 + \frac{1}{R_b}h_{11} + Rh_{22} + \frac{1}{R_b}h \quad \dots\dots(39)$$

Separating the real and imaginary parts of eqn. (35) gives

$$x^2 = \frac{c_1}{c_3}, \quad \text{or} \quad \omega = \frac{1}{CR} \sqrt{\frac{c_3}{c_1}} \quad \dots\dots(40)$$

as the frequency of oscillation, and

$$h_{21} + h_{12} = c_2x^2 - c_0 = \frac{c_2c_1 - c_3c_0}{c_3} \quad \dots\dots(41)$$

as the starting condition of oscillation.

In view of the orders of magnitude of the low frequency, common emitter, hybrid parameters of a transistor, and of the usual range of values of R and R_b , eqns. (36) to (41) can often be appreciably simplified. For example the low frequency h_{12} parameter is invariably negligible compared with h_{21} and so may be omitted from eqn. (41). (This amounts to assuming that the transistor is unilateral and that $\alpha = 0$.) Again in the expressions for c_0 , c_1 , c_2 , and c_3 , that is in eqns. (36), (37), (38) and (39) respectively, the leading terms are predominant giving as a first approximation from eqn. (40)

$$\omega = \frac{1}{CR} \cdot \frac{1}{\sqrt{6}} \quad \dots\dots(42)$$

and from eqn. (41)

$$h_{21} = 29 \quad \dots\dots(43)$$

frequency counterpart, namely

$$h_{21}^f = \frac{h_{21}}{1 + j\left(\frac{\omega}{\omega_0}\right)} \quad \dots\dots(44)$$

where ω_0 is the common emitter 3-dB cut-off frequency.

Low frequency values might still be used for the remaining parameters. Strictly the determinant h should be modified as well, but as terms involving h in eqns. (36), (37), (38) and (39) are invariably neglected, this refinement may be omitted.

Equation (35) now becomes

$$h_{21} = \left[1 + j\left(\frac{\omega}{\omega_0}\right)\right] [(c_2x^2 - c_0) + jx(c_1 - c_3x^2)] \quad \dots\dots(45)$$

Separating real and imaginary parts of eqn. (45) and using obvious approximations, leads to not-too-complicated expressions for the oscillation frequency and starting condition.

6. Colpitts Oscillator

As a further example the Colpitts oscillator depicted in Fig. 6 will be considered. The incremental equivalent circuit of the oscillator of Fig. 6 is usually drawn as shown in Fig. 7.

The equivalent circuit of Fig. 7 may be redrawn as shown in Fig. 8 which is more convenient for the application of the technique developed in this paper.

From Fig. 8 the open loop transfer matrix of the system is seen to be

$$A = \begin{pmatrix} 1 & 0 \\ \frac{1}{R_b} & 1 \end{pmatrix} \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \frac{1}{j\omega L'} & 1 \end{pmatrix} \begin{pmatrix} 1 & \frac{1}{j\omega C'} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ j\omega C_2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & R + j\omega L \\ j\omega C_1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \dots\dots(46)$$

Equation (42) is the well-known expression for the 180 deg phase shift frequency of a 3-stage ladder network, and eqn. (43) expresses the fact that the current gain of the transistor should equal the attenuation of the network at its 180 deg phase shift frequency.

A better approximation would include the terms involving h_{11} , and possibly those involving h_{22} , in the expressions for the coefficients c_0 , c_1 , c_2 , and c_3 . At low frequencies the terms in h are negligible.

At higher frequencies the hybrid parameters have imaginary parts and the separation of the real and imaginary parts of eqn. (29) is more involved. As a first step in this direction the short circuit current gain parameter h_{21} could be replaced by its high

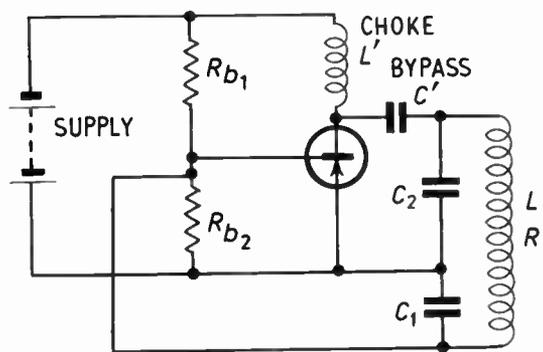


Fig. 6. A transistorized Colpitts oscillator.

where $\begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix}$ is the transfer matrix of the transistor in common emitter configuration.

From eqn. (46) it follows that

$$a = |A| = \begin{vmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{vmatrix} = \alpha \quad (\text{say}) \quad \dots\dots(47)$$

as explained in Section 5.

For simplicity, R_b will be assumed to have a negligible shunting effect across the input of the transistor, L' a negligible shunting effect across the output of the transistor, at the oscillation frequency, and C' a negligible reactance compared with the π network following C' , at the oscillation frequency. These assumptions correspond to taking R_b , L' and C' equal to infinity so that the factor matrices on the right-hand side of eqn. (46) corresponding to these elements, become unit matrices.

Equation (46) then reduces to

$$A = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ j\omega C_2 & 1 \end{pmatrix} \begin{pmatrix} 1 & R+j\omega L \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ j\omega C_1 & 1 \end{pmatrix} \\ = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} \begin{pmatrix} (1-\omega^2 LC_1)+j\omega C_1 R & R+j\omega L \\ -\omega^2 C_1 C_2 R+j\omega(C_1+C_2-\omega^2 LC_1 C_2) & (1-\omega^2 LC_2)+j\omega C_2 R \end{pmatrix} \quad \dots\dots(48)$$

These assumptions have been made as simplifications for the sake of example and do not detract from the power of the method in any way. (Note that in the ladder oscillator example of Section 5 R_b was kept finite.)

If now it is supposed that the oscillator frequency is low enough for the transistor to be represented by its low frequency hybrid parameters, then, as in Section 5,

$$\begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} = -\frac{1}{h_{21}} \begin{pmatrix} h & h_{11} \\ h_{22} & 1 \end{pmatrix} \quad \dots\dots(30)$$

where $h = h_{11} h_{22} - h_{12} h_{21} \quad \dots\dots(31)$

and $\alpha = \frac{h_{12}}{h_{21}} \quad \dots\dots(32)$

From eqns. (48) and (30)

$$a_{11} = -\frac{1}{h_{21}} \{h[(1-\omega^2 LC_1)+j\omega C_1 R] + h_{11}[-\omega^2 C_1 C_2 R+j\omega(C_1+C_2-\omega^2 LC_1 C_2)]\} \quad \dots\dots(49)$$

and

$$a_{22} = -\frac{1}{h_{21}} \{h_{22}(R+j\omega L) + (1-\omega^2 LC_2)+j\omega C_2 R\} \quad \dots\dots(50)$$

Substituting eqns. (32), (49) and (50) into the oscillation condition eqn. (7) gives

$$-(h_{21}+h_{12}) = [1+h_{22}R+h-\omega^2(LC_2+h_{11}C_1C_2R+hLC_1)] + j\omega[C_2R+h_{11}(C_1+C_2)+h_{22}L+hC_1R-\omega^2h_{11}LC_1C_2] \quad \dots\dots(51)$$

Equating real and imaginary parts of eqn. (51), the imaginary part gives the oscillation frequency as

$$\omega = \left\{ \frac{C_1+C_2}{LC_1C_2} + \frac{R}{h_{11}} \cdot \frac{1}{LC_1} + \frac{h_{22}}{h_{11}} \cdot \frac{1}{C_1C_2} + \frac{hR}{h_{11}} \cdot \frac{1}{LC_2} \right\}^{\frac{1}{2}} \quad \dots\dots(52)$$

The order of magnitude of the low-frequency parameters of the transistor is such that

$$\omega = \left\{ \frac{C_1+C_2}{LC_1C_2} + \frac{R}{h_{11}} \cdot \frac{1}{LC_1} \right\}^{\frac{1}{2}} \quad \dots\dots(53)$$

with close approximation, and if R is appreciably less than h_{11} , not an obvious condition, then

$$\omega = \sqrt{\frac{C_1+C_2}{LC_1C_2}} \quad \dots\dots(54)$$

which is the well-known expression for the oscillation frequency of a Colpitts oscillator.

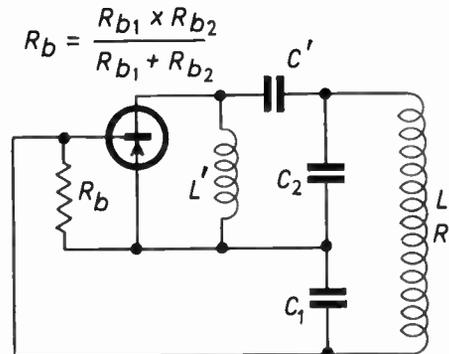


Fig. 7. Equivalent circuit of oscillator of Fig. 6.

The real part of eqn. (51) gives, as the starting condition of oscillation,

$$h_{21}+h_{12} = \omega^2(LC_2+h_{11}C_1C_2R+hLC_1) - (1+h_{22}R+h) \quad \dots\dots(55)$$

where ω^2 is given by eqn. (52).

Bearing in mind the order of magnitude of the hybrid parameters and using the eqn. (54) value of ω^2 , eqn. (55) reduces approximately to

$$h_{21} = \frac{C_2}{C_1} + h_{11} \left(\frac{C_1+C_2}{L} \right) R \quad \dots\dots(56)$$

7. Hartley Oscillator

As a concluding example the basic Hartley oscillator configuration will be considered. This oscillator differs from the Colpitts oscillator considered in Section 6 only in the form of the π network of the last three elements of Fig. 8. For a Hartley oscillator this π network must be replaced by the network of Fig. 9.

This network is made up of two 3-pole networks in parallel-parallel connection. The transfer matrices of the two 3-poles are

$$A = \begin{pmatrix} 1 & R_1 + j\omega L_1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & j\omega M \\ -\frac{1}{j\omega M} & 0 \end{pmatrix} \begin{pmatrix} 1 & R_2 + j\omega L_2 \\ 0 & 1 \end{pmatrix}$$

$$= -\frac{1}{j\omega M} \begin{pmatrix} R_1 + j\omega L_1 & (R_1 + j\omega L_1)(R_2 + j\omega L_2) + \omega^2 M^2 \\ 1 & R_2 + j\omega L_2 \end{pmatrix} \dots\dots(57)$$

and

$$B = \begin{pmatrix} 1 & \frac{1}{j\omega C} \\ 0 & 1 \end{pmatrix} \dots\dots(58)$$

respectively.

Now it can be shown³ that the transfer matrix of two such networks in parallel-parallel connection is

$$C = \frac{1}{a_{12} + b_{12}} \begin{pmatrix} a_{12}b_{11} + a_{11}b_{12} & a_{12}b_{12} \\ (a_{12} + b_{12})(a_{21} + b_{21}) - (a_{11} - b_{11})(a_{22} - b_{22}) & a_{22}b_{12} + a_{12}b_{22} \end{pmatrix} \dots\dots(59)$$

where

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \dots\dots(60)$$

and

$$B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \dots\dots(61)$$

are the transfer matrices of the two networks respectively.

If the matrices of eqns. (57) and (58) are used as the matrices A and B , respectively, the resultant transfer matrix given by eqn. (59) is rather involved, although it is still possible to manipulate it. In order to avoid an unduly lengthy matrix and yet still preserve the essential features of the analysis, the matrix of eqn. (57) will be simplified by taking $R_1 = R_2 = 0$ and by neglecting $\omega^2 M^2$ in comparison with $\omega^2 L_1 L_2$.

Equation (57) then reduces to

$$A = -\frac{1}{j\omega M} \begin{pmatrix} j\omega L_1 & -\omega^2 L_1 L_2 \\ 1 & j\omega L_2 \end{pmatrix} \dots\dots(62)$$

Equations (58), (62) and (59) now give for the transfer matrix of the network of Fig. 9.

$$C = -\frac{1}{M + \omega^2 L_1 L_2 C} \begin{pmatrix} L_1(1 - \omega^2 L_2 C) & j\omega L_1 L_2 \\ \frac{1 - \omega^2(L_1 + L_2 + M)C}{j\omega} & L_2(1 - \omega^2 L_1 C) \end{pmatrix} \dots\dots(63)$$

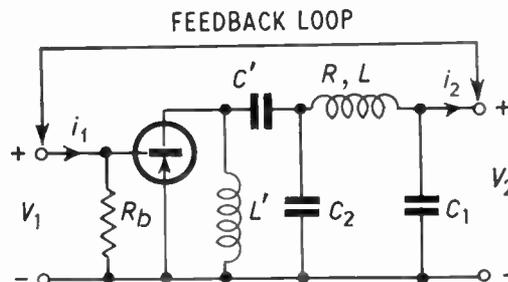


Fig. 8. Oscillator of Fig. 7 redrawn as three-pole with feedback loop.

The matrix on the right-hand side of this last equation is to take the place of the second factor on the right-hand side of eqn. (48) in the analysis of the Colpitts oscillator. Then in a way similar to that developed

for the Colpitts oscillator, the conditions of oscillation of the Hartley oscillator are given by, in terms of the low frequency hybrid parameters of the transistor,

$$(h_{21} + h_{12})(M + \omega^2 L_1 L_2 C)$$

$$= hL_1(1 - \omega^2 L_2 C) + h_{11} \left(\frac{1 - \omega^2(L_1 + L_2 + M)C}{j\omega} \right)$$

$$+ h_{22} \cdot j\omega L_1 L_2 + L_2(1 - \omega^2 L_1 C) \dots\dots(64)$$

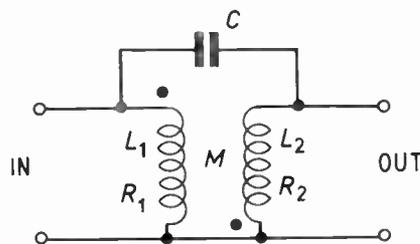


Fig. 9. Feedback network of a Hartley oscillator.

Separating real and imaginary parts of eqn. (64), the imaginary part determines the frequency of oscillation as

$$\omega = \left[(L_1 + L_2 + M)C \left(1 + \frac{h_{22}L_1L_2}{h_{11}(L_1 + L_2 + M)} \right) \right]^{-\frac{1}{2}} \dots\dots(65)$$

or
$$\omega \simeq \frac{1}{\sqrt{(L_1 + L_2 + M)C}} \dots\dots(66)$$

The starting condition of oscillation is given by the real part of eqn. (63), namely

$$(h_{21} + h_{12})(M + \omega^2 L_1 L_2 C) = hL_1(1 - \omega^2 L_2 C) + L_2(1 - \omega^2 L_1 C) \dots\dots(67)$$

Bearing in mind the orders of magnitude of the low frequency parameters of the transistor and using the value of ω^2 given by eqn. (66), eqn. (67) reduces to approximately

$$h_{21} = \frac{L_2(L_2 + M)}{M(L_1 + L_2 + M) + L_1 L_2} \dots\dots(68)$$

If $M = 0$, eqn. (68) is

$$h_{21} = \frac{L_2}{L_1} \dots\dots(69)$$

analogous to the Colpitts oscillator case, eqn. (56) (with $R = 0$).

8. Conclusions

A systematic technique for the determination of the oscillation frequency and starting condition of feed-

back oscillators has been presented. Using mainly the simple laws of matrix multiplication for series-cascaded circuit elements, the open loop transfer matrix of the system is established, the oscillation conditions being simply related to this matrix. The technique allows the possibility of analysing the effects of components and residuals, normally neglected, on the frequency and starting condition of oscillation.

After submitting this paper it came to the author's notice that a substantially similar method of analysis had been presented by J. H. Brodie of the University of Tasmania.⁵

9. References

1. L. A. Pipes, "Applied Mathematics for Engineers and Physicists", Chap. X. (McGraw-Hill, New York, 1946.)
2. J. S. Bell and K. Brewster, "An introduction to matrices and their use in transistor circuits", *Electronic Engineering*, **31**, p. 98, February 1959.
3. K. G. Nichols, "A matrix representation of linear amplifiers", *J. Brit.I.R.E.*, **21**, No. 6, p. 517, June 1961.
4. L. A. Pipes, "The matrix theory of four-terminal networks", *Phil. Mag.*, **30**, p. 370, 1940.
5. J. H. Brodie, "Matrix analysis of oscillators", (letter), *Trans. Inst. Radio Engrs (Circuit Theory)*, CT-7, pp. 69-70, March 1960.

Manuscript received by the Institution on 11th August 1961 (Paper No. 780).

British Electronic Telephone Exchange

On 12th December 1962, the Postmaster General, the Rt. Hon. Reginald Bevins, M.P., opened the first British electronic telephone exchange to go into public service.

The exchange, Highgate Wood, London, has 800 lines, and operates on the principle of time division multiplex using pulse amplitude modulation which enables 100 speech channels to be carried simultaneously on a single highway. Setting-up and register functions are also time-shared. This is believed to be the first exchange employing this combination of multiplexing and modulation techniques to be used for public traffic anywhere in the world, and it is certainly the first t.d.m. type of electronic exchange to go into public service in Europe.

Highgate Wood is in the nature of an advanced field experiment, following the success of a smaller exploratory installation at the G.P.O. Research Department at Dollis Hill, London, in 1959.† Three further electronic exchanges of more sophisticated type are now being built. Two of these, Pembury and Goring-on-Thames, will replace manual exchanges and the third, at Leighton Buzzard, will replace an early type of automatic exchange. Two of these exchanges are based on slightly different applications of the time division pulse amplitude modulation principle. The third, that at Leighton Buzzard, is based on space division switching. They should come into service in about two years time, and will be essentially trials of potential production systems, and, unlike Highgate Wood, will not have conventional stand-by electro-mechanical automatic exchanges.

The design and construction of the new electronic exchanges is the result of an extensive programme of co-operative research and development between the Post Office and the five principal British manufacturers of exchange equipment. The co-ordinating body in this work has been the Joint Electronic Research Committee, set up in 1956, on which are represented the Post Office and the five firms—Associated Electrical Industries, Automatic Telephones and Electric Co., Ericsson Telephones, the General Electric Company and Standard Telephones and Cables.

The first exchange, though thoroughly operational, is not entirely typical of the systems that are to follow it. The design of Highgate Wood was completed before semiconductor devices for all the required functions were generally available or sufficiently well proved. Consequently, though diode switching is used throughout the time-sharing circuits and transistors are used exclusively in subscribers' line units, the magnetic drum circuits and the register apparatus, several thousand thermionic valves are incorporated

into other sections of the equipment—for example, pulse amplifiers driving coaxial cables. All valves and transistors are British, including specially made valves for the pulse output amplifiers.

The new Mark II exchanges now being built will be transistorized, more compact and will present no special ventilating problems. Power requirements will be considerably reduced and the reliability factor can be expected to increase.

The time division multiplex system on which Highgate Wood is based enables highways and switching paths through the exchange to carry 100 simultaneous channels in the form of 1- μ s amplitude-modulated pulses at a pulse repetition frequency of 10 kc/s. Overall frequency control is maintained by a 1-Mc/s quartz crystal.

Incoming lines and junctions are connected to the exchange highways and to the various time-shared functions by coincidence gates which in turn are controlled by 100- μ s magnetostrictive delay line circulating stores. Each connection uses a pair of multiplex channels, spaced 50 μ s apart, to overcome sidetone which would be troublesome if the same channel pulse were used for within-group connections.

The use of t.d.m. techniques is of course only within the exchange complex itself. Conventional amplitude modulation at voice frequencies either direct or on multiplexed carriers is employed to subscribers and other exchanges. It is, however, probable that some more sophisticated system of higher transmission efficiency, such as pulse code modulation, may eventually be introduced between exchanges at least and these exchanges will have to be electronically switched.

A focal point of the Highgate Wood system is a magnetic drum "memory" store which constitutes the central information library of the exchange. On the drum, all the permanent factors relating to each line are stored in the form of a 28-bit word containing data as to the type of line, class of service and directory address (the relationship between the subscriber's telephone number and the physical location of the corresponding exchange equipment).

Additionally, a 7-bit word is temporarily written into the drum to provide state-of-line information—busy or parked. The drum also carries out the important function of recording subscribers' meter pulses. Other drum tracks carry translator information and provide the digitally-coded instructions required by the setting-up equipment for the routing of calls.

The drum makes one revolution every 28 ms and

† "An experimental electronic telephone exchange", *J. Brit. I.R.E.*, 19, No. 11, page 726, November 1959.

each subscriber's line is scanned once every 224 ms to coincide with a test pulse of 1 μ s, applied through the line connecting gates to each free line in turn to recognize whether it is looped (handset lifted) or not. Junctions from outside exchanges are scanned at a faster rate because, although a delay of approximately a quarter of a second can be accepted in applying dialling tone to a subscriber's line (it is physically impossible to lift a handset to the ear in that period), it is not permissible to delay the receipt of information from an incoming junction. For this reason, junctions are scanned once in every revolution of the drum—every 28 ms.

When a line—or junction—is detected as calling, an indication is given to a channel selector together with the instruction that a connection with the register equipment is required. The fact that the calling line is now busy is recorded in the state-of-line section of the memory drum.

A much-simplified picture of the handling of a call originating on a Highgate Wood exchange line is as follows:

As soon as the test pulse identifies the fact that a line requires attention, it causes a signal to be passed to the setting-up apparatus. This causes "busy" to be written in the state-of-line sector of the magnetic drum track relating to the calling line, and the setting-up equipment takes from the drum the permanent information—directory address, etc.—of the line, which it stores. This information is passed to the register as soon as a free time-slot has been allocated, and the identity of this time-slot is written into the line connection stores (also 100- μ s delay lines) which inhibits seizure of the line by any other call. A tone store is advised of the identity of the line calling and the antiphase pulse train controlling the outward speech highway (50 μ s away from the inward highway) passes dialling tone to the subscriber.

This train of events, which is only in broad outline, has possibly occupied a few thousandths of a second from the moment the test signal scanned the subscriber's line and detected that a call was required.

Dialling information from the subscriber is passed through the multiplex to the register equipment which counts and stores the make-and-break of the dial signals in modified binary code in the circulating delay lines. With a delay period of 900 μ s these lines can store nine decimal digits for each time-slot of the 100 channels.

The first three digits of the dialled information are fed back via the translator to the memory drum, ten tracks of which consist of permanent translations of the various exchange destinations and services (HIG for Highgate Wood; INF for supervisor). The incoming binary-coded signals are compared with the magnet-

ically-recorded tracks and when coincidence is achieved, the drum reads out the equipment address of the outgoing junction or line group and any additional routing instructions that are needed to steer the call to its destination. This information returns to the register and is again stored. In addition, in every case the wanted connection is referred to the temporary recordings on the drum. If the line is already engaged, instructions are sent back to the supervisory equipment to generate "busy" tone. If the line is available, the required line information is staticized to connect the called line to the time-division multiplex highways and the "busy" signal is written on to the drum. The final steps are to set up equipment which selects the necessary free channels to establish the connection, apply ringing tone to the called line and to record metering information.

An important aspect of electronic exchanges is that they must be capable of uninterrupted service for 24-hours a day, and one of the design parameters is that they should be capable of operation for reasonably long periods—week-ends, for example—without maintenance. Because much of the equipment controlling the traffic is common to a large number of lines, failure cannot be tolerated. For this reason, a large amount of the exchange apparatus consists of routining equipment which constantly monitors the various sections to detect fault conditions. Pulse patterns are set up at the various gates and compared with the resulting output pulses to indicate faults. Where equipment is duplicated (in some cases it is triplicated) the routining equipment regularly switches in each section in turn, at a periodicity of some 30 seconds, so that every piece of redundant equipment is in regular periodic use until fault conditions are detected.

Faults are pin-pointed by substitution analysis and when the offending equipment is identified, its location is indicated by lamp signals. Throughout the system, maintenance is simplified by enabling any faulty piece of equipment to be withdrawn and replaced as a unit, obviating on-the-spot repairs. Designs for future exchanges include the facility of automatic fault indication. At Highgate Wood, some 25 per cent of the electronic equipment is related to routiners used in fault analysis.

Summarizing the advantages to be gained by the gradual changeover to electronic exchanges the Post Office expect: the size and power requirements will be smaller and hence the cost of buildings and other services will be less; the speed of connection will be reduced, on a London-Glasgow call for instance from 6-8 seconds to about 1 second; servicing and maintenance time and costs will be reduced, due to increased reliability expected from semiconductor switching compared with electromechanical.

The Portrayal of Body Shape by a Sonar or Radar System

By

A. FREEDMAN, Ph.D. †

Presented at the Symposium on Sonar Systems in Birmingham on 9th–11th July 1962.

Summary: When acoustic waves are incident upon a rigid body or electromagnetic waves are incident upon a perfectly conducting body, the amplitude of the scattered radiation usually varies in a very complicated manner with direction of incidence, with direction of observation and with frequency. Even if consideration is restricted to back-scattering, the resulting scattering directivity curves appear to offer no clue to the shape of the scattering body. Yet the visual identification of objects by scattered radiation is an everyday experience. This paper endeavours to clarify the mechanism whereby the eye perceives shape and to see how this is applicable to methods of display in sonar and radar equipment. The analysis, which is qualitative, is based upon a recent theory of the mechanism of scattering of short wavelength acoustic radiation by a rigid body immersed in a fluid medium, and its electromagnetic counterpart.

1. Introduction

When acoustic or electromagnetic radiation is incident upon an arbitrary body, even under the simplest boundary conditions the amplitude of the scattered radiation usually varies in a very complicated manner with direction of incidence, with direction of observation and with frequency. Even if consideration is restricted to back-scattering, the resulting scattering directivity curves appear to offer no clue to the shape of the scattering body. Yet every day we are identifying objects visually by scattered radiation. This paper will endeavour to clarify the mechanism whereby the eye perceives shape and to see how the understanding gained may be applied to methods of display in equipment such as sonar or radar.

Under normal conditions, vision is essentially a phenomenon in which the wavelengths are small in comparison with the dimension of the objects being viewed. Use will therefore be made, in a purely qualitative way, of a relatively recent theory¹ explaining the mechanism of direct scattering of acoustic waves by a rigid body or of electromagnetic waves by a perfectly conducting body when the general body dimensions and radii of curvature are large in terms of wavelengths. An introduction to that mechanism was given in a paper delivered at an earlier symposium of the Acoustics Group of the Physical Society,² and the mathematical details and examples of application for acoustic back-scattering have appeared in two recent papers.^{3, 4} But, as the approach is unlikely to be generally familiar a brief description of the scattering mechanism will now be given.

† Admiralty Underwater Weapons Establishment, Portland, Dorset.

2. Description of the Scattering Mechanism

For convenience, the description will be in terms of the acoustic case, although it applies to the corresponding electromagnetic case also.

On the top right-hand side of Fig. 1, which is a purely qualitative illustration, is shown a relatively smooth, generally convex, rigid body. On the top left are the two transducers which are distant from the scattering body and will, in the first instance, be assumed non-directional. The receiver is sited at an arbitrary point outside the geometrical shadow region (but not very close to the shadow boundary). At Fig. 1(b) is shown a plot of $A(R)$, the projection towards the transmitter of all those parts of the scattering body which are visible from both transducers and are within total travel path R . $A(R)$ increases with R until all the directly visible part of the scattering body is encompassed. The shadow regions provide no contribution to the scattering process being described.

A plot of $dA(R)/dR$, shown at Fig. 1(c), exhibits two discontinuities, $D(A, 1, 1)$ and $D(A, 3, 1)$, corresponding to sudden changes of slope in the curve of $A(R)$ versus R .† The total paths R_1 and R_3 at which these occur are associated, respectively, with the nearest point of the body and with some assumed prominent feature which has been indicated schematically in Fig. 1(a). Associated with another lesser feature, a change of slope of $dA(R)/dR$ has been shown at R_2 , together with a further change of slope at the limit, R_4 , of the visible part of the body. The curve of $d^2A(R)/dR^2$

† The nomenclature $D(A, g, n)$ has been adopted to denote the value of a discontinuity at total path R_g in the n th derivative, with respect to R , of $A(R)$.

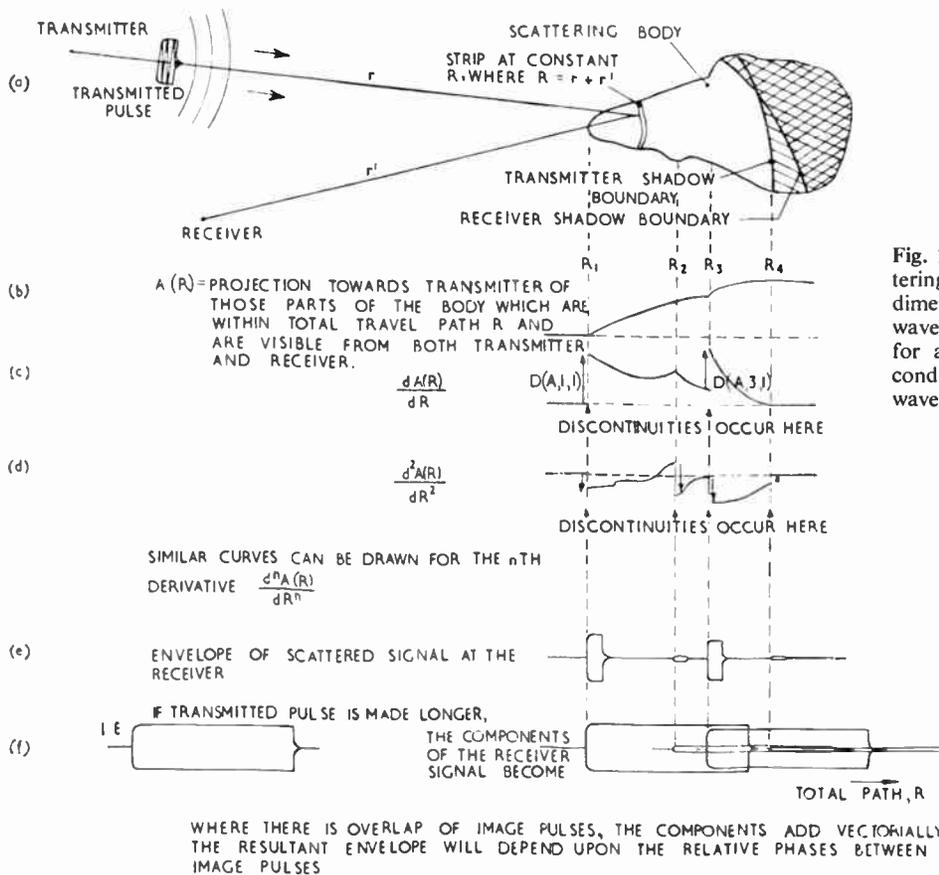


Fig. 1. Principles of direct scattering for bodies whose range and dimensions are large in terms of wavelengths. Body assumed rigid, for acoustic waves or perfectly conducting, for electromagnetic waves.

thus exhibits four discontinuities. Similar curves could be drawn for the n th derivative of $A(R)$.

The interpretation of the previously mentioned theory^{1,3} is that if a pulse, such as that shown in Fig. 1(a) is transmitted towards the object, each discontinuity in $d^n A(R)/dR^n$ ($n = 0, 1, 2, \dots$) engenders a component towards the scattered signal, which component, on the assumption of certain bandwidth limitations, has the same envelope as the transmission pulse. The resultant of all scattering components created at the same total path by discontinuities in various derivatives of $A(R)$, which resultant is the quantity that is physically observable, also has an envelope identical with that of the transmitted pulse and is therefore termed an "image pulse".

The magnitude of each scattering component is proportional to the size of its generating discontinuity, and also generally decreases fairly rapidly with increase in the order, n , of derivative in which the discontinuity occurs. Thus, for a given image pulse, usually only the first one or two orders of derivative in which discontinuities occur need to be considered. The magnitude of each component is also proportional to (frequency)¹⁻ⁿ. The phase within each component

depends upon the total path associated with its generating discontinuity.

For the transmitted pulse and the scattering body shown in Fig. 1(a), the envelope of the scattered signal would be approximately as shown in Fig. 1(e). The image pulses at R_1 and R_3 have been drawn proportionally to $D(A, 1, 1)$ and $D(A, 3, 1)$, respectively, it being assumed that the contributions due to discontinuities in $d^2 A(R)/dR^2$ are small compared with those in $dA(R)/dR$. The amplitudes of the image pulses at R_2 and R_4 have been drawn proportional to the relevant discontinuities in $d^2 A(R)/dR^2$, but on a smaller scale than the image pulses at R_1 and R_3 .

If the transmission pulse is lengthened as in Fig. 1(f), the image pulses will overlap. Where there is overlap, the components add vectorially and the resultant envelope will depend upon the relative phases between image pulses.

All the foregoing description may be extended quite simply to the case where the transducer directivities are not uniform over the scattering body. The contribution of each element of body surface to the total projected area must be weighted by the transmitter and receiver directivities appropriate to the position of that surface element. The projected area, $A(R)$, is replaced

by the directivity-weighted projected area, $A_w(R)$, the principles remaining otherwise the same. The shape of the curve of $A_w(R)$ versus R will differ from that of $A(R)$ versus R , but the discontinuities will occur at the same values of R , the transducer directivities being themselves continuous. The magnitudes of the discontinuities will however be altered.

The loci of constant R are spheroidal surfaces having the transmitter and receiver as foci. Thus if the position of one of the transducers, say the transmitter, is changed while the other stays fixed, the positions of the image-pulse-generating discontinuities on the surface of the scattering body will also shift.

3. The Echo from a Cylinder

We turn next to an example for the situation where transmitter and receiver are coincident. In this situation we may use the range parameter, r , instead of total path, R . $A(r)$ then replaces $A(R)$.

As the scattering body, we choose a smooth, right circular cylinder. Whereas an exact quantitative treatment presents mathematical difficulties,⁴ the use of simplifying assumptions regarding the projection of the cylinder provides a qualitative picture of the echo structure. The fact that the levels thereby obtained for the individual image pulses are quantitatively incorrect is not very important in the present context, as the main purpose here is to illustrate a mechanism.

Assuming that the incident wavefront can be considered plane over the extent of the cylinder, then with the direction of incidence other than parallel to or normal to the cylinder axis, the parts of the body

which are directly irradiated consist of the nearest end-disc and of one half of the curved surface. (Fig. 2(a).)

We note that $dA(r)/dr$ is proportional to the projected area of that part of the surface lying within an elementary range interval dr . Thus, for the irradiated end-disc, the shape of $dA(r)/dr$ over the range extent of the disc is a half ellipse, but this will be approximated by the shape shown in Fig. 2(b), which is not vertical at its near and far limits.

Strips between ranges r and $r+dr$ on the irradiated part of the cylindrical surface increase from zero length at the nearest range, r_1 , of the cylinder to a maximum at range r_2 . (Fig. 2(a).) Between ranges r_2 and r_4 , the length of strip remains constant, and between r_4 and r_5 it drops down towards zero again. The mode of increase between r_1 and r_2 and that of decrease between r_4 and r_5 are functions of aspect angle, but for simplicity a linear increase and a linear decrease will be assumed. The curve of Fig. 2(c) results.

Addition of this curve to that for the end-disc gives the total $dA(r)/dr$. (Fig. 2(d).) $d^2A(r)/dr^2$, illustrated in Fig. 2(e), exhibits five discontinuities, whose relative magnitudes are shown in Fig. 2(f).

Ignoring the contribution of discontinuities in higher order derivatives of $A(r)$, the echo structure for non-directional transducers and a short transmission pulse will be of the type illustrated in Fig. 2(g). Figure 2(h) shows the echo components resulting from an arbitrary lengthening of the transmission pulse. There is now overlap of image pulses, and the resulting envelopes will depend on their relative phases. Envelope (or semi-envelope) curves for two arbitrarily

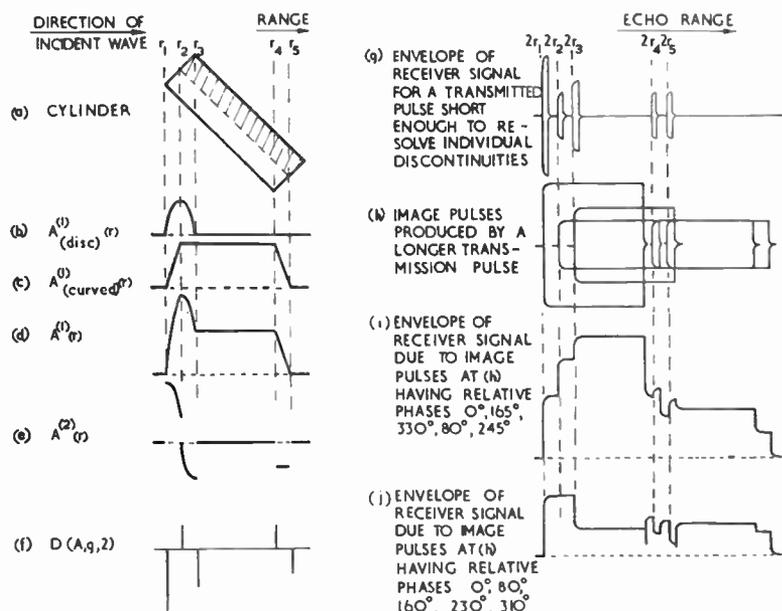


Fig. 2. Echo from right circular cylinder.

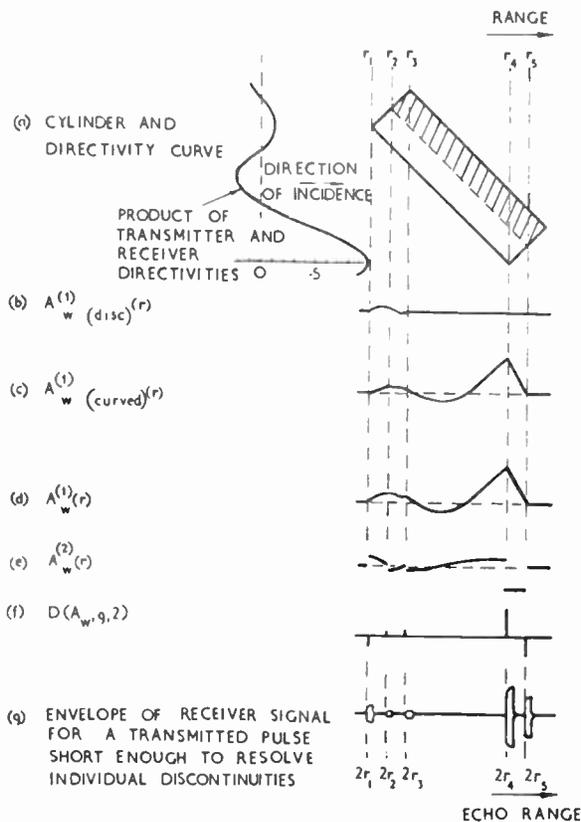


Fig. 3. Echo from right circular cylinder when the transducers are directional.

chosen combinations of relative phase are presented in Fig. 2(i) and (j).

We now examine the effect of transducer directivity on the echo. Assume that the transducers are non-directional in elevation, i.e. normal to the plane of Fig. 2(a), but that in azimuth the product of transmitter and receiver directivities (which we shall term the "product directivity curve" or the "product beam") is as shown in Fig 3(a), in which the beam peak is arbitrarily directed towards the far end of the cylinder. (Only about half of the symmetrical directivity curve is illustrated.)

A computation in which each curve is weighted approximately by the corresponding directivity value modifies the former curves of Fig. 2(b) and (c) to the shapes shown in Fig. 3(b) and (c). Then proceeding as before, one arrives at the echo structure of Fig. 3(g), in which the relative amplitudes of the image pulses differ markedly from those shown in Fig. 2(g). Not only are these amplitudes a function of the shape of the product directivity curve, but they also depend upon the beam centre bearing relative to the cylinder.

The dependence on the latter quantity will now be examined. It can be shown that each of the five discontinuities for the cylinder is, in effect, concen-

trated at a single point.¹ As the product beam is moved over the body, then the response due to such a discontinuity will vary as the value of the product beam on the latter's bearing, and it will have its peak level (equal to the non-directivity-weighted level) when the peak of the product beam is directed towards the discontinuity. Thus, for each discontinuity, a plot of its directivity-weighted magnitude versus centre bearing of product beam will have the shape of the product directivity curve mirrored about the beam axis.

For any given centre bearing of product beam, we therefore know both the distribution of ranges and of directivity-weighted magnitudes of the discontinuities. For a stipulated envelope of transmission pulse one can therefore construct curves of echo envelope (or semi-envelope) versus range for fixed centre bearings. Sets of such curves are shown in Figs. 4 and 5 for the same two transmission pulses as were depicted in the previous two figures. The positions of the zero lines have been spaced out proportionally to the centre bearing of the product beam. For the long pulse case, the same relative phases of image pulses have been assumed as previously.

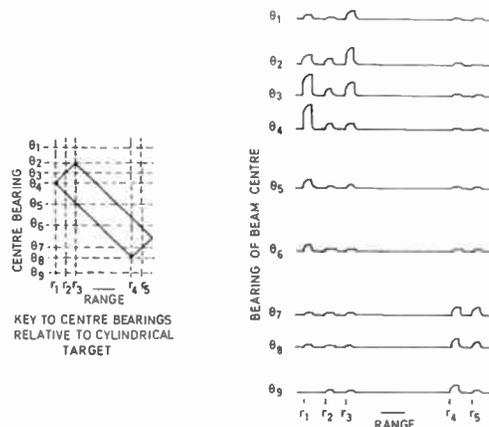


Fig. 4. Envelopes of receiver signal for short pulse case. Position of zero lines spaced proportionally to centre bearing of product beam.

4. The Echo Field

Let us assume a situation in which, while all the rest of the geometry remains static, the transmitter-receiver combination can rotate about a given axis. For example, let the elevation remain fixed and let the transducers rotate in azimuth. At any one bearing, information received at any given instant is nominally associated with a particular range, a particular azimuth angle, and the fixed elevation angle. The above angles are, of course, those corresponding to the beam centre and the above range corresponds to there and back travel of the front part of the transmitted pulse. However, from the foregoing treatment

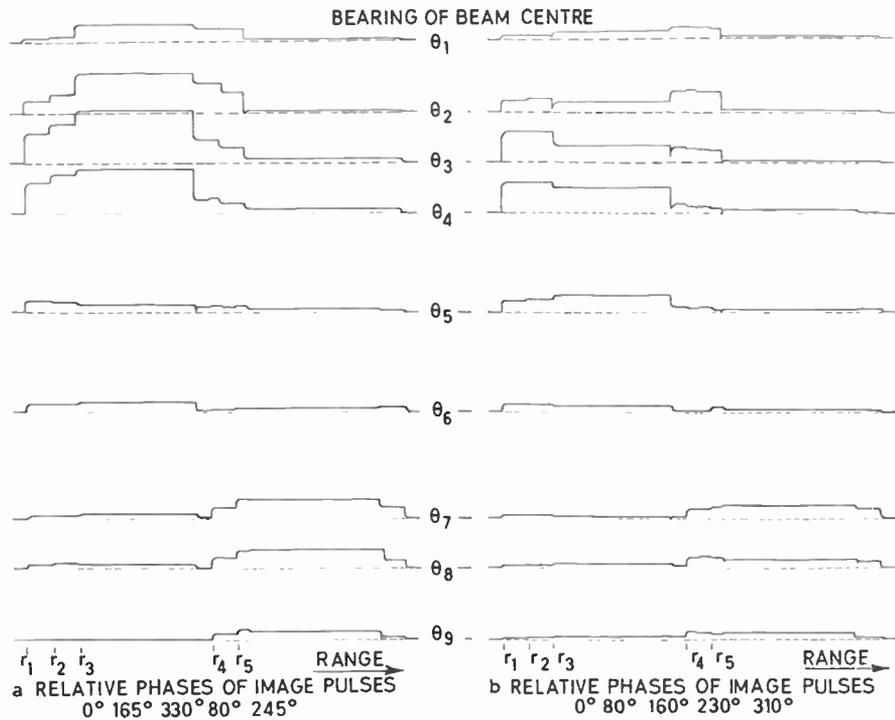


Fig. 5. Envelopes of receiver signal for long pulse case. Position of zero lines spaced proportionally to centre bearing of product beam.

it is known that the information contributing to any instantaneous value of the echo is collected over a finite volume in a manner governed by the transducer product beam and the transmission pulse. It is nevertheless convenient to associate the instantaneous echo value with a given point in space which corresponds to the centre bearing and elevation of the product beam and to the front of the transmitted wave packet. Each point in space will then have a certain echo level associated with it, and the distribution of such values will be termed the "echo field". The echo field is thus the field distribution one would obtain by plotting, at each point in the field, the corresponding receiver voltage amplitude produced when the product beam is directed on to that point. It is what one would "see" if one could observe in all directions at once while yet retaining resolution for each particular direction. The echo field is, of course, a function of the transmission pulse and of the product beam.

In any actual system, if the product beam varies in shape as it points in different directions, then allowance can be made for this in drawing the echo field. For the present, it will be assumed that the product beam remains constant for all field points.

Where, as in our example, the product directivity is uniform in elevation, the distribution of the echo field in all planes of constant elevation will be the same and all the information concerning the field will be given

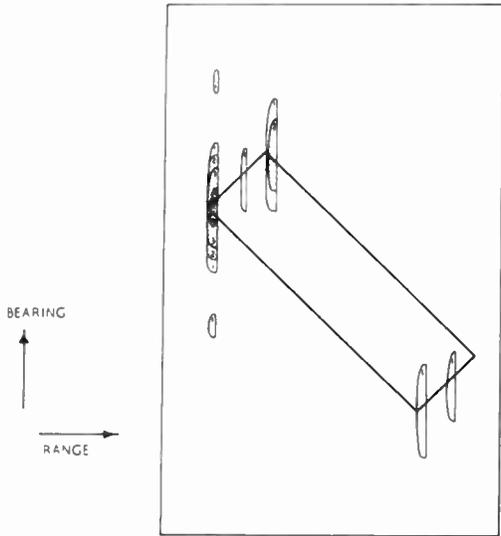
by the latter's distribution versus range and bearing.

From sets of curves such as those of Figs. 4 and 5, it is a simple matter to obtain contour curves of the echo field by joining together points of equal receiver response. Figure 6 shows echo field maps for the cylinder derived in this manner from Figs. 4 and 5; additionally the echo field for an intermediate transmission length is depicted. For convenience, the very low levels have been omitted from the charts. It is evident that the echo field for the short transmission case shows a much more direct relationship to the shape of the scattering body than does that for the long transmission case.

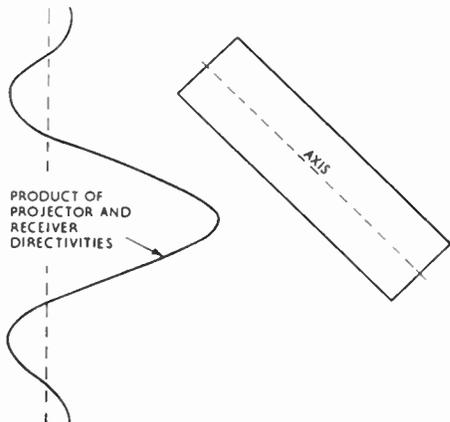
Subject to the assumptions of our analysis, such maps as those shown in Fig. 6 (and those corresponding to other angles of elevation, when the product beam is directional in the vertical plane) convey all the information that it is possible to obtain with the particular geometrical situation of the scattering body and the given transducers. The aim of any system of display can but be to illustrate as much as possible of the useful part of this information. Any limitations that exist in the information available in the echo field will also be present in the final display.

In practice, it is not generally possible to display the whole of the echo field, although this is feasible in theory. A sampling technique, in the form of scanning, has usually to be adopted.

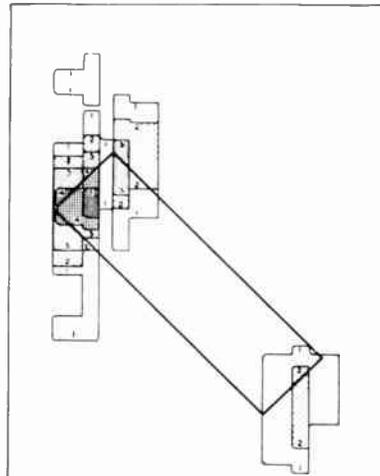
RELATIVE POSITIONS,
IN RANGE OF THE
IMAGE PULSES



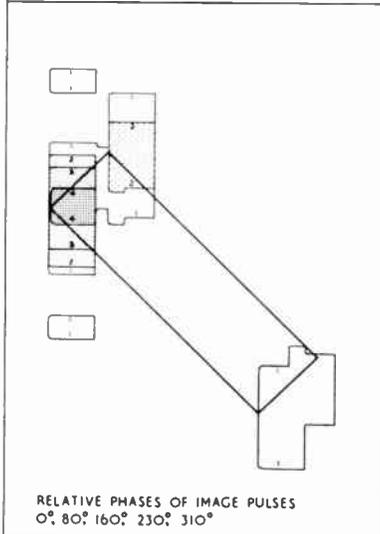
b. ECHO FIELD FOR SHORT TRANSMISSION PULSE.



a. CYLINDRICAL TARGET AND THE PRODUCT BEAM PRODUCING THE ECHO FIELD.

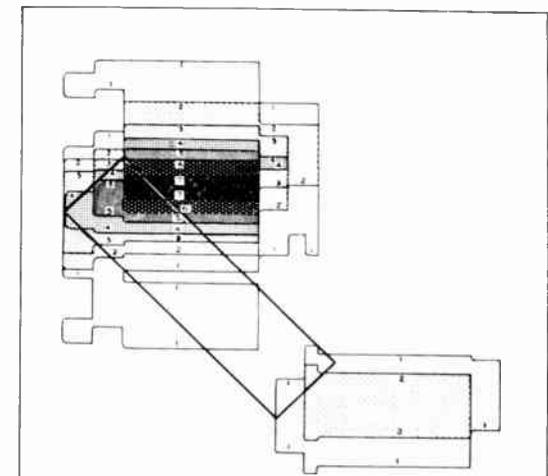


RELATIVE PHASES OF IMAGE PULSES
 $0^\circ, 165^\circ, 330^\circ, 80^\circ, 245^\circ$

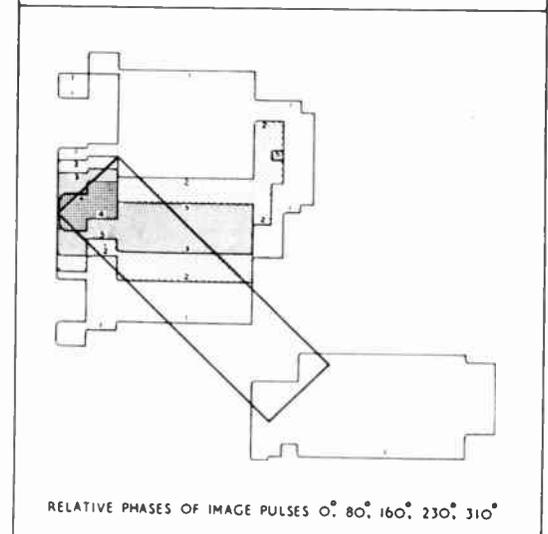


RELATIVE PHASES OF IMAGE PULSES
 $0^\circ, 80^\circ, 160^\circ, 230^\circ, 310^\circ$

c. ECHO FIELD FOR INTERMEDIATE TRANSMISSION PULSE.



RELATIVE PHASES OF IMAGE PULSES $0^\circ, 165^\circ, 330^\circ, 80^\circ, 245^\circ$



RELATIVE PHASES OF IMAGE PULSES $0^\circ, 80^\circ, 160^\circ, 230^\circ, 310^\circ$

d. ECHO FIELD FOR LONG TRANSMISSION PULSE.

Fig. 6. The echo field.

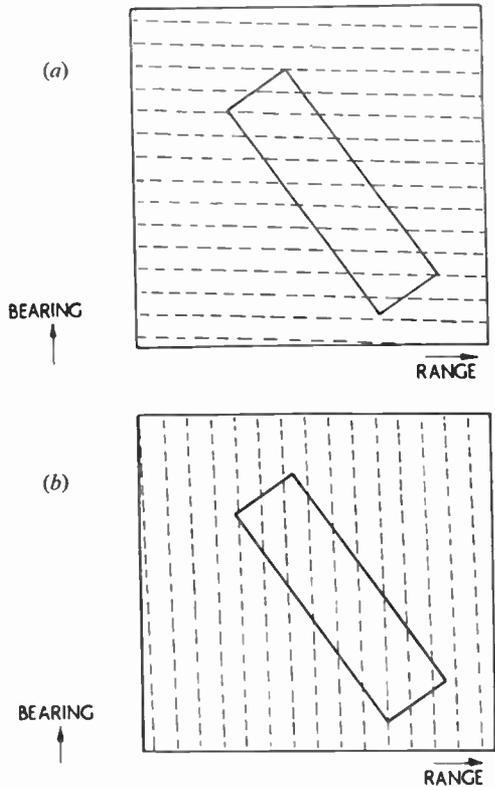


Fig. 7. Sampling loci within the echo field.

- (a) Scan line grid using one scan line per pulse repetition interval.
- (b) Scan line grid compatible with complete scanning within one pulse repetition interval.

5. Sampling the Echo Field

One form of scanning, commonly used in radar, consists of slowly rotating the transducers at constant angular velocity in azimuth, while transmission and echo reception are repeated at a fairly rapid rate. Providing the range extent of the region of interest is much smaller than its mean range, and assuming a Cartesian range-bearing representation, the loci of the parts of the echo field being sampled correspond to a grid of equispaced, parallel lines inclined at a small angle to a line of constant bearing. (Fig. 7(a).) The scanning spot of the brightness-modulated cathode ray tube in the display system is made to follow a corresponding grid pattern. (The conventional sonar presentation on a chemical recorder corresponds to a series of pictures in which there is only one such scan line (at constant bearing) per picture.)

On the above scan system, only one scan line is obtained per transmission, and so the time to build up a picture is equal to the pulse repetition period multiplied by the number of scan lines in the picture. In practical equipments, this method is usually fast

enough when using electro-magnetic waves, owing to the latter's high phase velocity, but tends to be too slow when using acoustic waves. To overcome this drawback, Tucker, Welsby and Kendell⁵ have shown how a lateral scanning grid can be achieved within the period of a single pulse repetition interval.

On a Cartesian range-bearing presentation, the loci of the echo field regions sampled by the latter method correspond to a grid of equispaced parallel lines inclined at a small angle to a line of constant range (Fig. 7(b)), and the cathode-ray tube is made to follow such a scan.

Positioning a set of scan lines of this type across a picture of the echo field (such as one of the pictures of Fig. 6), the variation of echo level which would be present along each line can be deduced. Alternatively, as a more convenient procedure, the sets of scan lines can be superposed on a diagram of the kind shown in Figs. 4 and 5. The echo level can be read off at each intersection of a scan line with one of the given lines of constant bearing, and hence the variation of the level along the scan lines may be deduced. Such distributions, obtained with the help of Fig. 4, are shown in Fig. 8 for a scan line interval equalling the transmission length. The convention adopted in that figure is that thickness of trace represents echo field level (or trace brightness, assuming linearity). Unless the range of the scattering body is precisely known, the relative positions, in range, of the scattering body and the scanning grid cannot be precisely defined to

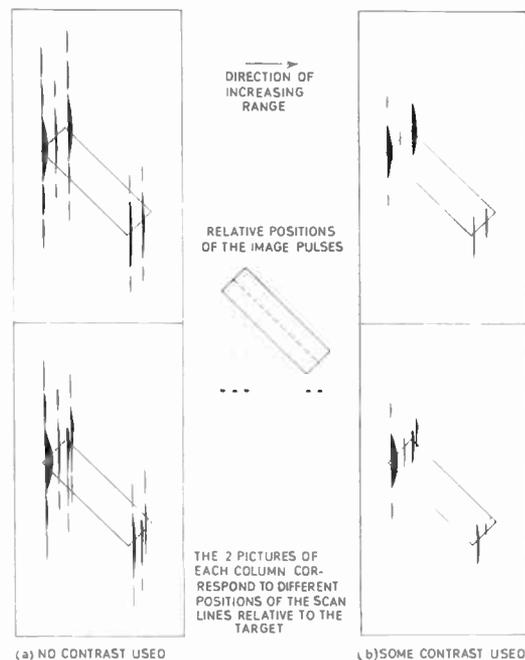


Fig. 8. B scan display for short transmission pulse, using one scan line per pulse length.

within one scan line interval. Because of this indeterminism, different echo field samples may be obtained, as is indicated by the two diagrams of Fig. 8(a). Figure 8(b) illustrates the effect of removing the lower levels, i.e. of "applying some contrast".

The quality of the sampling in reproducing the features of the echo field will be a function of the relativities of the scan line separation, the transmission length and the body size. Bandwidth considerations will naturally play a dominant part. Some of the criteria are discussed in Reference 5, although not from the viewpoint of portrayal of body shape. The problems are very similar to the corresponding ones of television presentation.

6. The Use of a Simplified Area of Resolution

By considering in detail the effect of various degrees of overlap of image pulses, both in range and in angular extent, interesting information can be obtained regarding the structure of the echo field.¹ However, instead of concerning ourselves with such detailed structure, it will suffice for our purpose to determine approximately the regions of the field providing an echo response. To do this, a grossly simplified model of the transmitted wavepacket will be utilized. We shall assume the product beam to be non-directional in elevation, whilst its directivity in bearing will be assumed to have unit value over an angular aperture, β , and to be zero outside this. (Fig. 9(a).) (The beamwidth, β , of this hypothetical beam may, arbitrarily, be deemed to be equal to the beamwidth to the 3 dB points of an actual product beam, although the exact

selection of such beam limits is of secondary importance in the present thread of argument.) The envelope of the transmission pulse will, furthermore, be assumed rectangular and of duration τ . (Fig. 9(b), where c represents the wave velocity.) The horizontal cross section of the wave packet relative to the position of a point, P, in the echo field is shown in Fig. 9(c). Owing to the "there and back" travel, the region from which simultaneous echo contributions may be obtained due to the wave packet at that position is an area of half the range extent of the wave packet, as illustrated in Fig. 9(d). Thus, for any given point of discontinuity, Q, on a scattering body, the beam could be pointed up to an angle of $\pm\beta$ from the bearing of the discontinuity, and the latter would still make a contribution to the echo, providing the field point, P, were at a range anywhere between the range, r , of Q and $r + (c\tau/2)$. Thus, in the echo field, a discontinuity at a point will produce a contribution to the echo over the rectangular region shown in Fig. 9(e) (on the Cartesian system which has been used). This region will be termed the "area of resolution". The concept of this area of resolution is an approximation, and it obscures certain facts¹; nevertheless, it provides a useful means of assessing how diffuse will be the echo field representation of an object under different conditions of transmission length and product beamwidth.

Figure 10 shows the area of resolution associated with each of the five discontinuities of the cylinder for each of the three previously used transmission lengths and for three different widths of product beam. For each of the nine situations portrayed, the shaded areas (which may include regions of overlap) represent the region in which the echo field will show a response.

A better idea of how transmission length and product beamwidth affect the diffusion of a picture in the echo field can be gained by going from a smooth body which has only a few discontinuities to a body where discontinuities may be considered to be distributed at many points over the surface. For the several different sizes of area of resolution portrayed at the top of Fig. 11, Fig. 11(a) shows the possible region of diffusion within the echo field for an irregular, approximately cylindrical body on which discontinuities may be distributed anywhere over the surface, while Fig. 11(b) illustrates the same thing for an irregular body of arbitrary shape. The trends are fairly obvious.

The simplified model used above tells us little about the distribution of levels within the region of echo response, other than that where many rectangular areas of resolution overlap, the levels are likely to be high; it does, however, provide a useful, albeit approximate, idea of the extent of the region of echo response, although even this can be misleading on occasion.

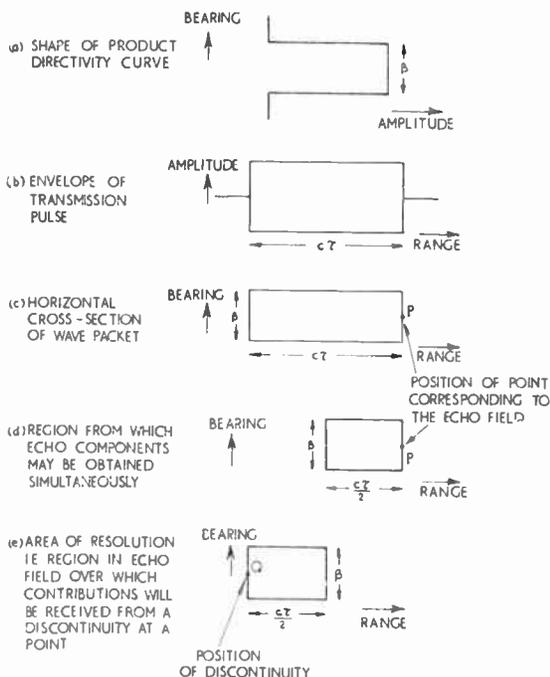


Fig. 9. Simplified area of resolution.

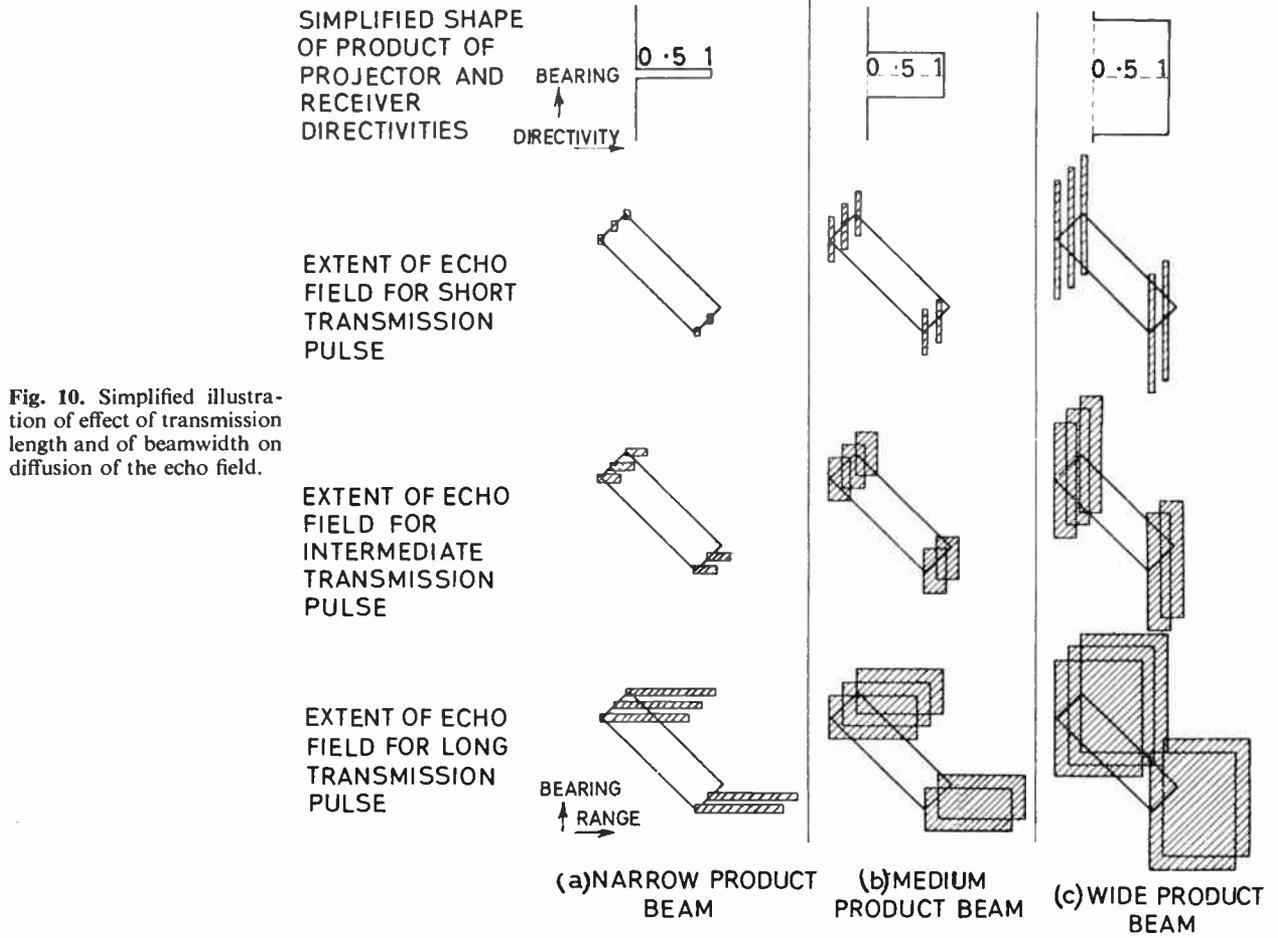


Fig. 10. Simplified illustration of effect of transmission length and of beamwidth on diffusion of the echo field.

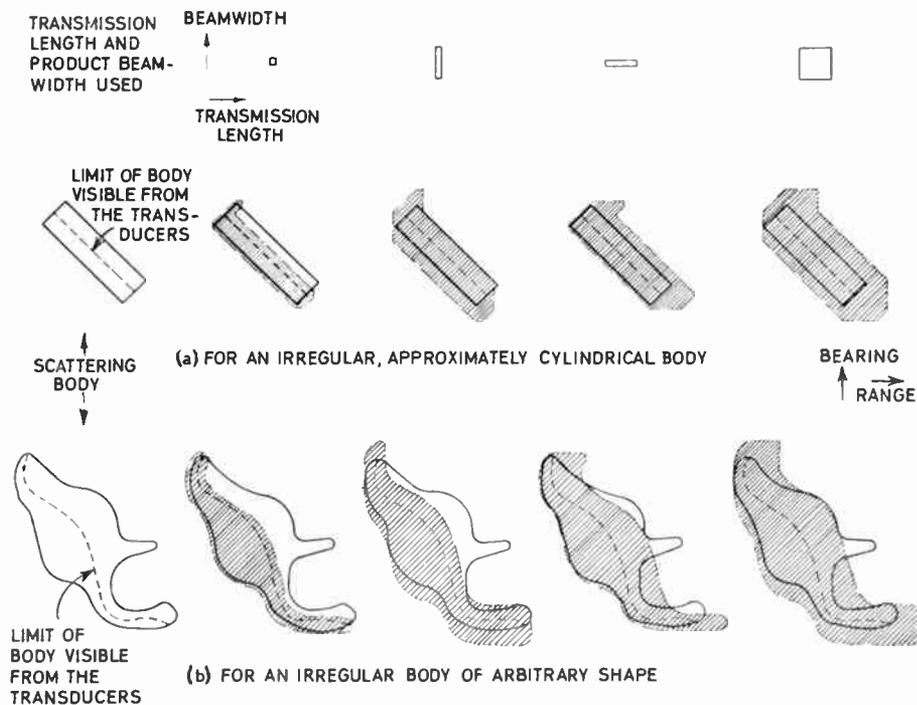


Fig. 11. Simplified illustration of effect of transmission length and of beamwidth on the total possible region of diffusion in the echo field.

7. The Criteria for Resolution within the Echo Field

On the assumption of a two-dimensional, brightness-modulated display showing the variation of the echo field (or of that part of it that is sampled) versus bearing and range, we may now state the criteria for any desired degree of resolution of an isolated scattering body within the echo field.

From consideration of Fig. 11(b), it is evident that if the shape of the visible part of a body is to be reproduced with reasonable faithfulness in the echo field picture, it is necessary for the transmission length to be shorter than the shortest range extent of the visible part of the body at any one bearing (i.e. not the total range extent of the visible part of the body) and for the width of the product beam to be less than the smallest bearing extent of the visible part of the body at any one range (i.e. not the total bearing extent of the body).

The above generalizations are of necessity rather sweeping and may conveniently be relaxed where consideration of individual cases indicates that resolution of some part of the outline of the body, for which the range extent or the bearing extent happens to be small locally, is not important for recognition of the object.

If the aspect at which a particular body will be viewed cannot be predicted, then it is necessary to apply the above criteria to the smallest range and bearing extents which could occur for any aspect.

When the aim is not solely to recognize the general shape of the body, but to have a higher degree of resolution, in order to distinguish individual features on its surface, then the same general considerations apply to any one given area of surface as applied previously to the body as a whole. Resolution of an individual feature on the surface will again depend upon the relativity of transmission length and width of product beam to the relevant region of the surface, but there is the additional complication that the effects of adjacent surface regions may overlap into the area under investigation and make it more difficult to disentangle the latter's geometry. This is rather akin to having several discrete scattering bodies in relatively close proximity. Thus, it seems likely that, compared with the criteria for determining the outline of a body, the criteria for distinguishing features on the latter's surface will be somewhat stricter than is indicated by a mere scaling down of transmission length and width of product beam in proportion to the dimensions of the surface details involved.

8. Three-dimensional Echo Field Distributions and their Representation

We turn from consideration of the case where the product beam is non-directional in elevation to the

more general case where it is directional both in elevation and bearing.

The concept of the simplified area of resolution may straightforwardly be extended by assuming that, in elevation, the directivity of the product beam has unit value over an angular aperture, γ , and is zero outside it. This leads to the simplified volume of resolution shown in Fig. 12. For any given centre bearing and at any given instant, this represents the spatial region which can contribute components to the total echo by virtue of any echo-creating discontinuities which happen to be within that region. The same types of limitation which apply to the simplified area of resolution will obviously also apply to the simplified volume of resolution.

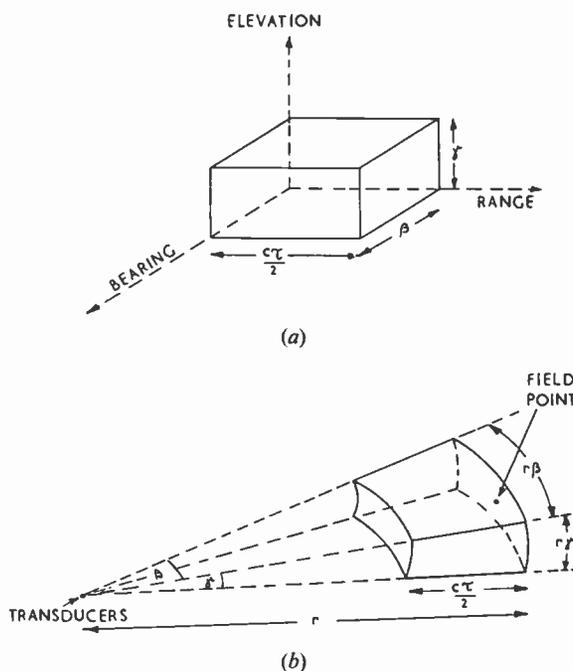


Fig. 12. Simplified volume of resolution. (a) In Cartesian co-ordinates. (b) In pictorial view.

The criteria of resolution described in Section 7 for the two-dimensional case still apply, with the obvious extensions to allow for resolution in the third dimension.

Since the echo field is now a variable in three dimensions, a choice of parameters is available for its portrayal in the form of a two-dimensional representation. Either a range parameter and an angular parameter (range-bearing or range-elevation presentation), or two angular parameters (bearing-elevation presentation), may be selected. The relationship between the two types of presentation is not quite the same as between a plan and an elevation view of an object, for these presuppose viewing the latter from

two directions at right angles to one another, in which case different parts of the object are seen on each occasion. Range-bearing and bearing-elevation presentations of the echo field both relate to the same part of the scattering body and no information is obtained of that part of it which is within the geometrical shadow region from the one viewing point.

To produce, say, a range-bearing presentation, one could keep the elevation fixed and scan in one of the manners previously described. At any given point on the resultant picture, all available contributions will have been received from all elevations within the simplified volume of resolution (or, more precisely, the output consists of a directivity-weighted integration, in elevation, over the product beam). To produce a bearing-elevation presentation, a television type scan would be appropriate. At any given point on the resultant picture all available contributions will have been received from all ranges within the volume of resolution, and this, it must be remembered, is changing its position with time. Pictures from gradually increasing distances will be presented, unless the irradiation is of the c.w. type, in which case the picture represents the sum of the effects of all relevant discontinuities up to infinite range.

The selection of the type of presentation is usually governed by the degree of resolution, in each of the co-ordinate parameters, afforded by a given system. In radar and underwater echo location systems the range resolution is generally good and the angular resolution relatively poor; the presentation is therefore mostly of the range-bearing (or range-elevation) type. In optical systems, the irradiation is mostly of the c.w. type and there is no range resolution, but the angular resolution is normally good; the optical displays to which we are accustomed are therefore of the bearing-elevation type.

9. The Relationship between an Optical Receiver and a Receiver and Display System Utilizing a Scanning Technique

Optical displays, be they presented on a photograph or viewed directly by the eye, utilize lens systems. It is therefore relevant to see how the use of such a system is related to the assumptions made, so far, about the mechanism of obtaining information from the echo field, namely to sampling in the form of scanning.

A point source in the object plane of a lens system does not simply give rise to a corresponding image point in the image plane but, because of the diffraction pattern associated with the lens aperture, it occasions a distributed image consisting of a bright central spot surrounded by fainter rings. Similarly, illumination at a point in the image plane does not stem solely from the corresponding point in the

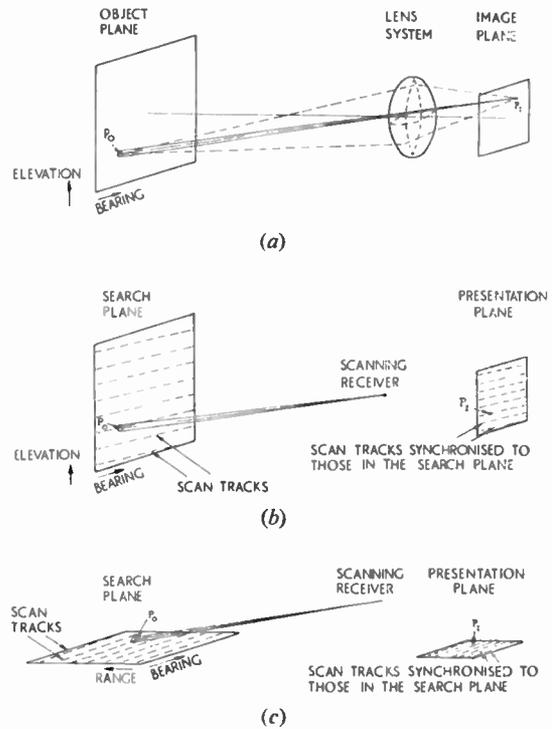


Fig. 13. Comparison between a presentation using a lens system and presentations using scanning techniques.

- (a) Schematic of lens system.
- (b) Schematic of scanning system for bearing-elevation presentation.
- (c) Schematic of scanning system for range-bearing presentation.

object plane but has been collected from all the sources within a volume determined correspondingly by diffraction. Thus, with each element of surface in the image plane there is associated a beam pattern whose beamwidth, in an average optical system, is very small in comparison with beamwidths normally encountered in radar and underwater echo location equipments. (Fig. 13(a).)

Assuming a photographic plate to be sited in the image plane, each grain of photographic material on the plate can be considered to be one of a distribution of receivers, each such elementary receiver having small beamwidth, both in bearing and elevation, and each pointing in the direction of the corresponding point in the object plane. If, instead of a camera, the optical system is deemed to be the eye, the same type of description applies, the grains of photographic material being replaced by the rods and cones on the retina of the eye. If the beamwidth of the projector supplying the irradiation is much greater than that associated with each elementary receiver, the directivity of each product beam will be, effectively, the same as that of the elementary receiver.

Whereas, when using a scanning system, one product beam was pointed successively at different directions in the search field whilst the resulting signals were displayed in the corresponding positions on a cathode ray tube (Fig. 13(b) and (c)), a lens system is equivalent to having a distribution of product beams, pointing all the time in the various directions in the search field, each elementary receiver causing its resultant signal to be displayed in the appropriate position on the display. Apart from the fact that the one system samples the echo field temporally whilst the other maintains surveillance for all directions continuously, the mechanisms of the two systems are essentially the same. Thus, apart from the material of Section 5, what has been said in previous Sections about the display of the echo field holds for a lens system as well as for a scanning system. It will therefore hold for an optical display, providing there is compliance with the previously assumed condition of irradiation from the direction of the receiver.

It is also evident that the distribution which we have termed the echo field is the same thing as the image field of optics for the particular situation where the optical illumination comes solely from the direction of the observer.

The type of picture produced by such unidirectional irradiation and the influence of multidirectional irradiation will be examined in the next Section.

Before proceeding to this examination, we note that if the lens system of Fig. 13(a) represents a television camera with its photo-electric screen in the image plane, then Fig. 13(b) represents the type of system whereby the image plane distribution is transferred to the screen of the cathode-ray tube. The search plane in this case is the image plane of the camera.

It is possible that the retina of the eye is scanned in some such manner.

The use of a lens system, coupled with image plane scanning of a suitable transducing surface, is obviously applicable to acoustic⁶ and to radar equipments.

10. Types of Picture produced by Unidirectional and Multidirectional Irradiation

The type of picture seen when a smooth convex body is illuminated by light from a single source near the observer is not the same as the normal view of the body. Shape is not seen, and the picture consists only of so-called "glints" or "highlights", which correspond to the directions in which high levels occur within the echo field. Pictures showing the effect can be seen in Reference 7 and it is also commonly observed when an aircraft in flight is caught in a searchlight beam. A similar type of picture is obtained when an aircraft in flight "reflects glints" due to sunlight falling on it.

This latter effect calls for a slight extension of the echo field concept, for the direction of observation no longer coincides with the direction of irradiation. We have seen, in Section 2, that the scattering mechanism is the same for all directions outside the geometrical shadow region; alteration of the direction of observation, outside the shadow region, merely alters the positions and magnitudes of the discontinuities. Thus, a system possessing good angular resolution will present highlights whose position on the scattering body will vary as the bistatic angle is altered.

If a body is irradiated by more than one projector (from different directions), the contributions of the several sources will be superposed (due allowance being made for phase). The highlights due to the several sources will be seen and, if there are sufficient sources, the shape of the body will become discernible. This effect can sometimes be observed when an aircraft in flight is illuminated by several searchlights.

That it is possible to see the different appearances of an aircraft in flight, when subjected to unidirectional and to multidirectional searchlight illumination, is due to the fact that relatively little scattering is produced by the body's surroundings, so that irradiation comes, effectively, only from the directions of the searchlights used.

In general, when we view an object, it is irradiated from all directions by light which has been scattered off surrounding bodies. This is equivalent to having a distribution of sources around the object. For a given direction of observation, the highlights from the various sources merge to give a complete portrayal of the visible part of the surface.

The various sources will be at arbitrary distances from the object being illuminated, and the time delays associated with each incident beam will therefore also be arbitrary. However, this introduces no confusion in the picture, as the presentation, by not taking account of range, does not depend upon the time of travel of the radiation. Also, as there are many sources and as the radiation from each is essentially of a c.w. nature, the phases of the various incoming waves do not matter, for the superposition of the components will tend, by probability laws, to produce a relatively constant result. So far, we have been implicitly analysing the situation for single frequency irradiation. The frequency spread present when using white light will increase the statistical tendency to a constant picture depending little on the phases of the contributing waves.

If illumination from one source is appreciably higher than from the others, the highlights due to that particular source will tend to stand out and behave in the same manner as when the illumination is only of the unidirectional type. Familiar examples of this

occur when an electric lamp or the sun glint off a smooth curved surface.

When radiation is incident from a single direction upon a body whose surface is rough, multiple scattering between different parts of the surface produces an effect similar to irradiation from many directions, and a picture intermediate in type between the highlight picture and the full optical picture will be obtained. Many good examples of this can be seen in photographs of the moon taken from the earth, the uni-directional irradiation stemming, of course, from the sun. A considerable amount of detail can be observed in many lunar photographs.

Even with multi-directional irradiation, optical views will tend to appear flat, i.e. two-dimensional, when observed monocularly. As is well known, it is the superposition of two slightly different views in binocular vision which gives a sensation of depth of view.

11. Characteristics of Possible Types of Presentation

Basically, there is no reason why optical views should be restricted to a bearing-elevation presentation any more than there is why radar or underwater echo location pictures should be restricted to a range-bearing presentation.

The types of picture which can be obtained from a generally convex body, in a uniform free field, are grouped in Table 1. This assumes that sufficient resolution is available to prevent undue spreading and that radiation does not penetrate into the body with subsequent re-radiation.

For a body with large concave features, the extra time taken by rays in traversing multiple scattering paths will be greater than for the small concavities associated with rough bodies. The distortion in range-bearing presentations will be greater than for rough,

convex bodies and may make such a presentation unrecognizable. Similar effects will occur when there is more than one scattering body in the field or when there is some penetration of a body, with subsequent re-radiation after scattering at internal surfaces.

12. Conclusions

A mechanism of scattering has been described which is applicable to both acoustic and electromagnetic radiation when the wavelength is small compared with the dimensions of the scattering body. The directly scattered radiation at a field point other than in (or very close to) the geometrical shadow region is composed of a number of discrete signals. The envelope of each of these is, subject to certain bandwidth limitations, a copy of that of the transmission signal. These components (which have been termed "image pulses") are generated by certain discontinuities associated with the projection of the scattering body in the direction of the transmitter. When the image pulses overlap in time, the resultant scattered signal consists of their vectorial addition, and if the transmitted signal is long enough to cause a region of steady-state overlap, the latter region corresponds to the c.w. level. Thus, for a transmission of sufficient length, a non-directional receiver will measure a scattering level corresponding to a point on one of the usual distribution-in-angle scattering curves of the body.

If the receiver is replaced by the eye, then instead of just measuring the total scattering level, the eye acts, in effect, as a directional receiver which scans the scattered contributions from each part of the body (i.e. the highlights) and displays them in the correct relative positions on the retina. Normally the irradiation on a body is from many directions, and superposition of the various sets of highlights formed gives a complete picture, showing the shape of the scattering body.

Table 1
Types of picture obtainable from a convex, opaque body

Type of irradiation	Type of surface	Type of picture	
		Bearing-elevation presentation	Range-bearing presentation
Single source	Smooth	Highlight type of picture	Highlight type of picture
	Rough	Approaching a normal picture	Approaching a normal picture, but some distortion due to extra time taken by rays in traversing multiple scattering paths
Distributed sources	Smooth or Rough	Normal picture	No proper picture, owing to different travel distances from the different sources

The above principles have been applied to forecast the types of picture it should be possible to produce by a sonar or radar system for smooth and for rough bodies, when the irradiation is either produced by a single source or by distributed sources, and when the presentation is either of the bearing-elevation or of the range-bearing kind.

13. Acknowledgment

This paper is based on material forming part of the author's Ph.D. Thesis.¹ It is published by permission of the Admiralty.

14. References

1. A. Freedman, "The Formation of Acoustic Echoes in Fluids", Ph.D. thesis, London University, 1961.
2. A. Freedman, "A new interpretation of acoustic echo formation". Paper read at "Underwater Acoustics" Symposium of the Acoustics Group of the Physical Society, London, 1960.

3. A. Freedman, "A mechanism of acoustic echo formation", *Acustica*, **12**, No. 1, p. 10, 1962.
4. A. Freedman, "The high frequency echo structure of some simple body shapes", *Acustica*, **12**, No. 2, p. 61, 1962.
5. D. G. Tucker, V. G. Welsby and R. Kendell, "Electronic sector scanning", *J. Brit.I.R.E.*, **18**, p. 465, 1958.
6. L. D. Rozenberg, "Survey of methods used for the visualization of ultrasonic fields", *Soviet Physics—Acoustics*, **1**, No. 2, p. 105, 1955.
7. Research Analysis Group, Committee on Undersea Warfare, National Research Council, "Physics of Sound in the Sea". Part III. "Reflection of sound from submarines and surface vessels", pp. 380, 382–385.

Manuscript received by the Institution on 8th June 1962 (Paper No. 781/SS28).

© The British Institution of Radio Engineers, 1963

POINTS FROM THE DISCUSSION

Mr. M. J. Daintith†: Why is a body insonified from all directions more readily detected than one illuminated from one direction only?

The Author (in reply): I have not stated that a body will necessarily be easier to detect if irradiated from all directions rather than from one only. What I have been dealing with, in essence, is the form of presentation under ideal, free-field conditions. I believe I have shown that, under such conditions (and assuming an appropriate † Admiralty Underwater Weapons Establishment.

type of display), when a body is insonified from one direction only, the picture presented will consist solely of highlights, while when it is insonified from many directions body shape will be seen.

Ease of detection is a further, distinct problem, yet to be tackled, in which the background of the signal (noise, reverberation, the presence of other objects) must be taken into account. It does, however, seem likely that the possibility of recognizing an object's shape will, in general, increase the probability of detection.

Coherent and Non-Coherent Demodulation of Envelope-Modulated Radio Signals

with particular reference to the Enhanced-Carrier System

By

P. V. INDIRESAN, Ph.D.

(Associate Member) †

AND

Professor

D. G. TUCKER, D.Sc. (Member) ‡

Summary: The ordinary "linear detector" and the coherent demodulator are briefly discussed as the two extreme cases of demodulators for envelope-modulated signals, since in the former the rectifier (or modulator) is switched by the complete input signal, while in the latter a pure local reference tone is used. The synchrondyne and enhanced-carrier systems are regarded as intermediate cases in which the switching is effected by an impure carrier derived from the incoming signal. The enhanced-carrier system, using filters to extract the carrier from the incoming signal, appears to have received little attention in the literature; and the distortion caused by tuning errors in the enhanced-carrier path and the interference due to unwanted input signals are here particularly considered.

1. Introduction

Receivers for envelope-modulated signals usually rely on the so-called linear detector for demodulation. Such a detector, in which the output voltage follows the envelope of the received signal, inevitably gives an output from unwanted signals applied to it, and selectivity therefore has to be obtained by filtration before the detector. If the filter were ideal, i.e. with a constant small loss throughout the pass-band and infinitely rapid cut-off at the edges of it, then high selectivity would be compatible with absence of frequency-response distortion of the wanted signal. But in practice such a result is not attainable, and good audio frequency-response generally has to be sacrificed if reasonable selectivity is to be obtained. In theory this difficulty may be mitigated by using coherent demodulation in which the signal channel is sufficiently broad to avoid distortion and selectivity is obtained by a low-pass filter *following* the demodulator.

In principle the linear detector does not differ from the coherent demodulator as much as may appear at first sight. It is readily shown¹ that a linear detector may be regarded as a self-switched modulator, so that when a single envelope-modulated signal is being received it is only a special or extreme case of a coherent demodulator where the signal and the switching voltages are identical. At the other extreme is the ideal coherent demodulator in which the switching voltage is totally independent of the input signal and is obtained directly from the carrier source. The properties of both the linear detector and the coherent demodulator are well-known.

† Electrical Engineering Department, University of Roorkee, India; formerly at the University of Birmingham.

‡ Electrical Engineering Department, University of Birmingham.

In between these two extremes lie the enhanced-carrier system and the synchrondyne. Possibly the latter could be regarded for some purposes as a special case of the former, but for the present purpose it is more convenient to regard the enhanced-carrier system as one in which the enhanced-carrier path is provided with filters to remove the sidebands and interference as far as possible, while the synchrondyne uses a synchronized oscillator for this purpose. Figure 1 shows the schematic arrangements considered. The synchrondyne and its advantages and disadvantages are relatively well-known, but the enhanced-carrier system appears to be little known; in particular, the effects of mistuning in the carrier-selection path, and of interference in this path, appear not to have been discussed in the literature. This therefore will be the main contribution of the present paper.

2. Summary of Properties of Linear Detector and Coherent Demodulator, including the Synchrondyne

The envelope-modulated input signal (wanted) will be taken as

$$e_1 = a(1 + m_1 \cos p_1 t) \cdot \cos \omega_1 t \quad \dots\dots(1)$$

and the interfering signal as

$$e_2 = ax(1 + m_2 \cos p_2 t) \cdot \cos \omega_2 t \quad \dots\dots(2)$$

Then in the linear detector (Fig. 1(a)), assuming $x \ll 1$, the audio output of p_1 and its harmonics can be shown² to be approximately

$$a \left[m_1 \left(1 - \frac{x^2}{4} \right) \cos p_1 t - \frac{x^2}{4} m_1^2 \cos 2p_1 t \right] \dots\dots(3)$$

while that of p_2 and its harmonics will be approximately

$$a \left[\frac{x^2}{2} m_2 \cos p_2 t + \frac{x^2}{4} m_2^2 \cos 2p_2 t \right] \dots\dots(4)$$

apart from a constant numerical factor. It is clear that the relatively small interfering signal has only a second-order effect on the wanted output; but whereas in the absence of the stronger signal the output of p_2 would be proportional to x , in the presence of the strong wanted signal it is proportional to $x^2/2$. Thus the demodulated output of the weaker signal is depressed by the presence of the stronger signal to a fraction $x/2$ of the amplitude it would have if it were the only signal present. This effect is called "detector discrimination". It means that if the selective circuits prior to detection reduce the unwanted signal to $1/n$ of the wanted signal, in the output of the detector it will actually be reduced to $1/2n^2$ of the wanted output. This is physically reasonable, since in the self-switched modulator conception of the linear detector, both signal and switching voltage benefit from the pre-detector selectivity.

The coherent demodulator³ is different in that it is a modulator whose switching function is generated by a pure local carrier, free of sidebands and interference (see Fig. 1(b)). This local carrier must be of the same frequency as that of the wanted signal, and should preferably be exactly in phase with it. If, in fact, it has a phase angle θ_1 relative to the signal, so that the switching function has a fundamental term $h_1 \cos(\omega_1 t + \theta_1)$, then the output (after h.f. components are removed by filtration) from an input $e_1 + e_2$ (given by eqns (1) and (2)) is

$$\frac{1}{2}ah_1 \left[m \cos \theta_1 \cos p_1 t + x \cos [(\omega_1 - \omega_2)t + \theta_1] + \frac{x}{2} m_2 \{ \cos [(\omega_1 - \omega_2 + p_2)t + \theta_1] + \cos [(\omega_1 - \omega_2 - p_2)t + \theta_1] \} \right] \dots (5)$$

The output amplitude of signal p_1 is proportional to $\cos \theta_1$, but there is no output of p_2 and the interference is a modulation on carrier frequency $\omega_1 - \omega_2$. If this is above the required audio range, it can be removed by filtration *after* the demodulator. But the effect of the filter is, of course, only linear; it is not enhanced by detector discrimination as in the linear detector.

The fact that a linear detector gives an output of unwanted signal p_2 means that its selectivity must be provided *before* the detector, i.e. at a high frequency where adequate discrimination can be provided usually only at the expense of a poor frequency-response of the audio output. The coherent demodulator is free of this difficulty.

The synchrodyne³ is, in principle, a coherent demodulator where the local carrier is obtained from an oscillator synchronized to the carrier of the wanted signal. In the simplest form, the synchronization is effected by injecting some of the incoming signal into

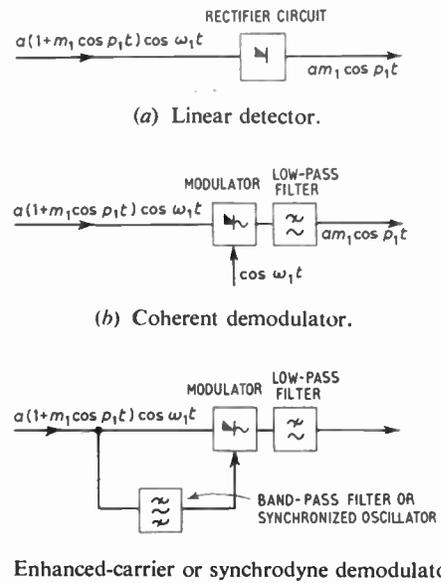


Fig. 1. The demodulator circuits considered.

the oscillator circuit, which is tuned very nearly to the correct frequency. Although this ensures that the dominant output of the oscillator is at the correct frequency, yet it also allows a small amount of the interfering signal into the switching signal. The discrimination of the synchronized oscillator against interfering signals is a non-linear phenomenon, and some discrimination is provided against a small interfering signal even when no frequency-discrimination can be effective⁴; in general, the discrimination is good, but cannot make the interfering effect always negligible. Thus some demodulated output of the interfering modulation frequency p_2 must occur at the output of the demodulator. The analysis of the effect of this interference is closely related to that of the enhanced-carrier system discussed in Appendix 2.

Another effect in the synchrodyne is the production of non-linear distortion in the audio output due to phase-modulation of the oscillator by the modulation on the synchronizing signal.⁵ This occurs whenever the natural frequency of the oscillator is not exactly equal to the wanted carrier frequency. It can be largely eliminated by automatic phase control, a technique which is now becoming very well-known,⁵ and which may be applied with or without direct injection of the synchronizing signal into the oscillator. It has been shown by Costas that automatic phase control can be effective even when no carrier is present.⁹ It would also appear to be possible to eliminate the non-linear distortion by a feedback arrangement whereby the modulation is fed back in anti-phase from the output of the demodulator to the oscillator circuit.

3. The Enhanced-Carrier System

The earliest recorded attempt to improve the detection of the amplitude modulated signals by increasing the magnitude of the carrier is probably the patent taken out by Robinson⁶ in 1922. In his arrangement the carrier component is filtered out of the input signal by a narrow band filter and then added, after amplification, to the signal, which had passed undistorted through a wideband channel. The resultant carrier-reinforced voltage was detected in the usual manner. The increased carrier level makes detector discrimination more effective and thus helps to reduce interference. A later account of the subject is given by Crosby.¹⁰ However, the narrow band channel output will not contain the carrier only but will have a little of the sideband components as well. The higher modulation frequencies are much more attenuated than the lower ones, resulting in a non-uniform audio frequency response. Under ideal conditions this distortion may, however, be suppressed by using a switching type modulator (as in Fig. 1(c)) instead of a linear detector, because the amplitude variations in the switching voltage will not appear in the detector output. This is the circuit which will be considered in the remainder of this section. Exactly the same effect is obtained by the use of a limiter in the carrier path of the original arrangement, as proposed by Crosby.

3.1. Non-linear Distortion due to Mistuning

When the narrow-band circuit is mistuned, the two sidebands of the input signal are amplified unequally and subjected to unsymmetrical phase-shifts, so that the resultant voltage applied to the switching terminals of the modulator is no longer purely amplitude-modulated, but is phase-modulated as well. Thus the

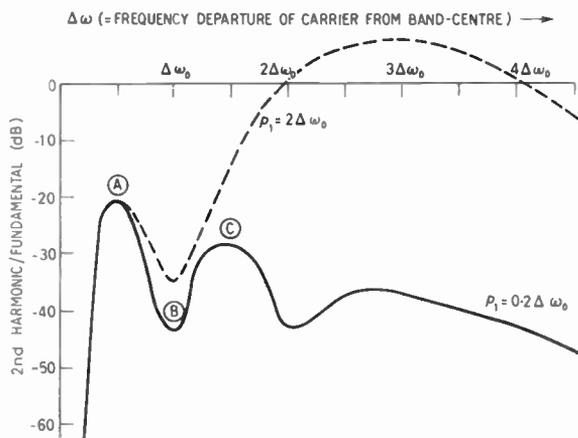


Fig. 2. Enhanced-carrier demodulator with mistuning in carrier path: typical approximate graphs of second harmonic/fundamental ratio against amount of mistuning. $\Delta\omega_0$ is half the 6 dB bandwidth. The selective circuit chosen for the carrier path is two stages of critically-coupled double-tuned circuits. Depth of modulation (m_1) = 0.3.

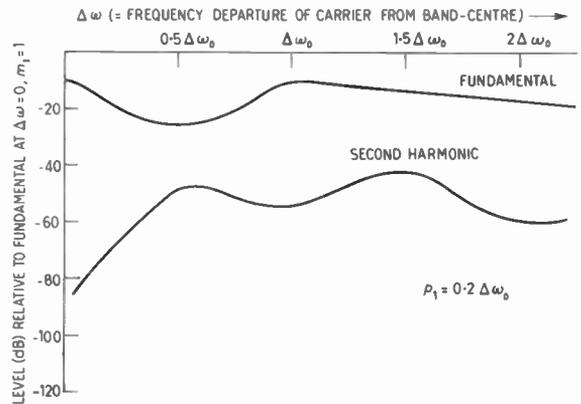


Fig. 3. Actual levels of fundamental and second-harmonic for conditions of Fig. 2, with modulating frequency (p_1) = $0.2\Delta\omega_0$.

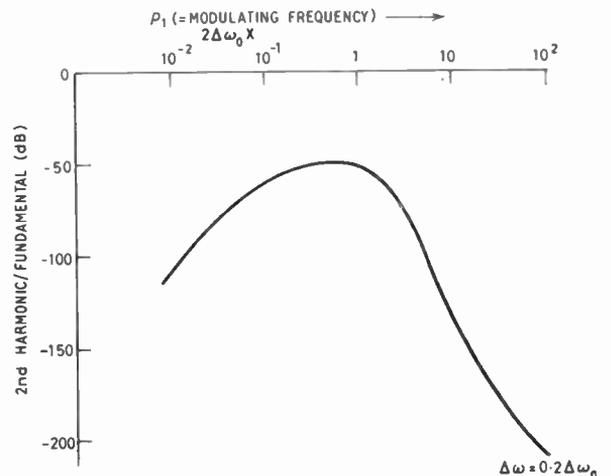


Fig. 4. Second harmonic/fundamental ratio as a function of modulating frequency (p_1) for a mistuning ($\Delta\omega$) = $0.2\Delta\omega_0$. ($m_1 = 0.3$).

switching function of the modulator is affected, and non-linear demodulation results. The magnitude of the effect depends on the circuit characteristics involved, the amount of mistuning and the frequency of the modulating signal. The basic analysis is given in Appendix 1, with the assumption of a small depth of modulation and other restrictions which are specified. It will be apparent from the formulae obtained that numerical calculations are rather tedious (even with an electronic computer, great care has to be taken to get phase angles into the right range) and that the results are rather critically dependent on the actual phase-shifts produced by the tuned circuits. For the latter reason it is also difficult to obtain numerical results in a general form. But a good idea of some of the features which are likely to be important in practice may be obtained from Figs. 2, 3 and 4, which show (in approximate form) typical behaviour when the narrow-

band selective circuits comprise two stages of critically-coupled double-tuned circuits. These figures have been drawn after studying results obtained from a computer for various values of the parameters.

For the circuit considered, the amplitude response was taken as

$$G(\omega) = 1 / \left[1 + \left(\frac{\Delta\omega}{\Delta\omega_0} \right)^4 \right]$$

and the phase response as

$$\theta = 2 \tan^{-1} \frac{1}{\sqrt{2}} \left[\frac{\Delta\omega}{\Delta\omega_0} - \frac{\Delta\omega_0}{\Delta\omega} \right]$$

where $\Delta\omega_0$ is half the 6 dB bandwidth, and $\Delta\omega$ is the difference between the centre frequency of the circuit and the frequency being considered. The phase response is plotted in Fig. 5.

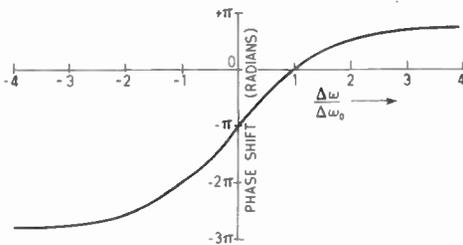


Fig. 5. Phase response of two stages of critically-coupled double tuned circuits.

Referring now to Fig. 2, it will be seen that the quantity graphed is the level of the second harmonic relative to the fundamental output signal at the same amount of mistuning. The abscissa shows the amount of mistuning, i.e. the amount by which the carrier frequency differs from the centre frequency of the tuned circuits. Naturally, when the mistuning is zero, there is no harmonic generated. As the mistuning is increased, the harmonic level rises (the actual levels depending on the modulation frequency and on the depth of modulation) and the fundamental level falls, until when $\theta = \pi/2$ rad the fundamental reaches its minimum and causes the peak in the curve marked A. After this, the fundamental rises again and the second harmonic falls (as shown for $p_1 = 0.2\Delta\omega_0$ in Fig. 3) until at the mistuning which makes $\theta \approx \pi$ rad there is the minimum harmonic ratio marked B in Fig. 2. After this, the shape of the curve depends markedly on the modulation frequency. If the latter is larger than the half-bandwidth ($\Delta\omega_0$), the condition is eventually reached where one sideband lies at the centre of the tuned band, while the carrier is greatly attenuated. At this point the harmonic level may well exceed that of the fundamental, as shown in the upper curve of Fig. 2. If the modulation frequency is smaller than $\Delta\omega_0$, the distortion tends to improve after the peak C which does not reach above peak A. This

discussion assumes, of course, that at all points sufficient voltage is maintained to control the modulator.

It is clear that, in practice, the amount of mistuning should not be allowed to exceed a fairly small fraction of $\Delta\omega_0$. Figure 4 shows how the second-harmonic/fundamental ratio varies with modulating frequency at a mistuning of $0.2\Delta\omega_0$. It is evident that this amount of mistuning may represent a reasonable limit in many practical cases. It is not surprising that the maximum distortion occurs for a modulating frequency of the order of the half-bandwidth, since this condition gives the maximum asymmetry of the sidebands obtainable without appreciable amplitude discrimination.

3.2. The Effect of an Interfering Signal

When (even in the absence of mistuning) there is an interfering signal present in the input, the switching voltage suffers a change in amplitude and in phase, and in general there will be an a.f. output in the detector due to the interfering signal. This will be mainly of two types, (a) a beat whistle equal to the frequency difference between the wanted signal and the interfering signal, and (b) a component due to the modulation in the interfering signal. In broadcast practice the beat frequency component will lie outside the desired audio band and hence may be removed by a low-pass filter. The modulation component will however lie within the required audio band and cannot be eliminated. It is of interest, therefore, to find out how much of such interference will be present. A theoretical calculation is given in Appendix 2, but unfortunately the results are again rather complicated, although one case has been worked out for comparison with measurements.

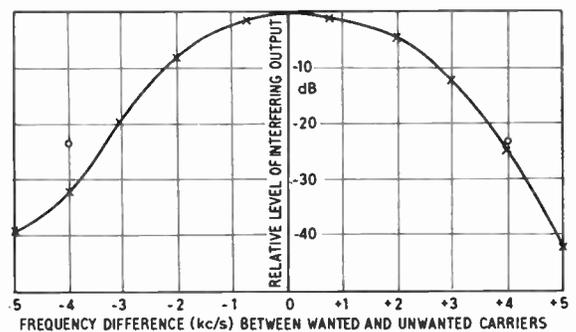


Fig. 6. Enhanced-carrier demodulator: discrimination against interfering signals. Graph shows the level (dB) of interfering audio output relative to wanted output, versus the frequency difference (kc/s) between wanted and unwanted carriers.

Crosses: experimental points.

Circles: theoretical result.

N.B. The linking of experimental points by the simple curve shown gives only an approximation to the true shape.

An experimental measurement of the magnitude of interference using a two-stage critically-coupled double-tuned circuit with $\Delta\omega_0 = 2$ kc/s was carried out as follows: two signals, one modulated at 400 c/s and the other at 1000 c/s were applied to the detector; the modulation depth was 0.3 in both. The former was exactly tuned to the centre frequency of the system, whereas the latter was detuned from resonant frequency by varying amounts. When both were tuned to the resonant frequency the outputs of the 400 and the 1000 c/s components were adjusted to be equal to each other. The reduction in the 1000 c/s component for various amounts of frequency separation was measured by a wave analyser. The results have been shown in Fig. 6. For an interfering signal 4 kc/s away from the wanted one, the relative level of 1000 c/s in the output was calculated by the formulae of Appendix B to be -23.2 dB, while the measured results are -24 dB and -32 dB on the two sides of resonance respectively. Allowing for the experimental difficulties of exact lining-up, this agreement is probably acceptable. The calculations assumed $m_1 = m_2 = 0.3$ and $a_1 = a_2$. It can be seen that the curve of suppression of interference follows very roughly the response of the carrier-enhancing filter.

4. Post-detector Equalization in the Linear Detector

It is clear that the true coherent demodulator is not practicable for radio receivers, and that the quasi-coherent (i.e. intermediate) systems such as the enhanced-carrier and synchrodyne demodulators have considerable circuit complexity and a somewhat imperfect performance. Thus it is still attractive to use a linear detector and make an attempt to overcome its disadvantage of being unable to give a good selectivity without sacrificing audio frequency-response. One simple way of doing this is to allow the audio frequency-response to become poor as the result of the pre-detector filters, and then to restore it by means of an equalizer. This method was described⁷ in 1931 under the name of "Autotone", but seems to have been associated with the improvement of "straight" receivers and so has been unused since the superheterodyne system was introduced. It is, of course, applicable to the latter. Its most serious disadvantage is probably that it worsens the ratio of harmonics to fundamental if non-linear distortion is present, and the signal/interference ratio will be worsened if there is interference producing high audio frequencies.

5. Conclusions

It has been shown that both coherent and non-coherent demodulation systems can be arranged to give good high frequency response without sacrificing selectivity. The non-coherent detector system is much

simpler and can be introduced into the receivers already in use. On the other hand the coherent detector system has the advantage that it gives faithful response right up to 100% modulation. It has also been shown that the conception of coherent demodulation giving perfect rejection of interference signals is not valid except when the switching voltage is obtained directly from the carrier source. When it is derived from the input signal, the switching voltage is no longer truly free from interference components and the discrimination against unwanted signals actually tends to be smaller (for a given selective circuit) than that in non-coherent detectors. In general, whatever selectivity is lost in the broadband signal channel must be restored in the narrowband channel. Hence the narrowband circuit of a coherent detector needs to be much more sharply tuned than a linear detector circuit, and distortion will arise if the tuned response is not accurately symmetrical about the carrier frequency.

6. Acknowledgments

The authors are indebted to Dr. J. W. R. Griffiths and Mr. E. D. R. Shearman for much fruitful discussion and advice during this work, and to Mr. C. R. Fry for assistance in computing.

7. References

1. D. G. Tucker, "Linear rectifiers and limiters", *Wireless Engineer*, 29, p. 128, 1952.
2. D. G. Tucker, "Modulators and Frequency-Changers". (Macdonald, London, 1953.)
3. D. G. Tucker, "The history of the synchrodyne and homodyne", *J. Brit.I.R.E.*, 14, p. 143, April 1954. (This paper gives a very full bibliography.)
4. D. G. Tucker and G. G. Jamieson, "Discrimination of a synchronized oscillator against interfering tones and noise", *Proc. Instn Elect. Engrs*, 103, Part C, p. 129, 1956. (I.E.E. Monograph No. 146R, August 1955.)
5. D. G. Tucker and R. A. Seymour, "The synchrodyne as a precision demodulator", *Wireless Engineer*, 27, p. 227, 1950.
6. E. Y. Robinson, British Patent No. 201,591; 1922.
7. F. L. Devereux and H. F. Smith, "Autotone", *Wireless World*, 30, p. 186, 1931.
8. K. R. Sturley, "Radio Receiver Design", p. 455, Vol. 1. (Chapman and Hall, London, 1953.)
9. J. P. Costas, "Synchronous communications", *Proc. Inst. Radio Engrs*, 44, p. 1713, 1956.
10. M. G. Crosby, "Exalted-carrier amplitude and phase-modulation reception", *Proc. Inst. Radio Engrs*, 33, p. 581, 1945.

8. Appendix 1

Non-linear Distortion due to Mistuning in the Enhanced-Carrier System

Let the tuned-circuit system of the enhanced-carrier (i.e. narrow-band) path, at a given amount of mistuning, produce a phase-shift of θ_1 at the carrier fre-

quency ω_1 , and $(\theta_1 + \alpha_1)$ and $(\theta_1 - \beta_1)$ at the sideband frequencies $(\omega_1 + p_1)$ and $(\omega_1 - p_1)$ respectively.

Let the amplitude responses at the sideband frequencies be x' and y' respectively, relative to the carrier.

The input voltage is

$$e_1 = a_1(1 + m_1 \cos p_1 t) \cdot \cos \omega_1 t \quad \dots\dots(6)$$

and thus the voltage applied to the switching terminals of the modulator is

$$e_0 = a_1 \cos(\omega_1 t + \theta_1) + a_1 x_1 \cos(\omega_1 t + \theta_1 + p_1 t + \alpha_1) + a_1 y_1 \cos(\omega_1 t + \theta_1 - p_1 t - \beta_1) \dots\dots(7)$$

where $x_1 = m_1 x' / 2$ and $y_1 = m_1 y' / 2$.

This can be re-arranged as

$$e_0 = a_1 \cos(\omega_1 t + \theta_1) [1 + x_1 \cos(p_1 t + \alpha_1) + y_1 \cos(p_1 t + \beta_1)] - a_1 \sin(\omega_1 t + \theta_1) \cdot b_1 \sin(p_1 t + \phi_1) \quad \dots\dots(8)$$

where $b_1^2 = x_1^2 + y_1^2 - 2x_1 y_1 \cos(\alpha_1 - \beta_1) \quad \dots\dots(9)$

and $\phi_1 = \tan^{-1} \left[\frac{x_1 \sin \alpha_1 - y_1 \sin \beta_1}{x_1 \cos \alpha_1 - y_1 \cos \beta_1} \right] \dots\dots(10)$

Now assume that $m_1 \ll 1$. Then in eqn (8), the first term will be the larger, since x_1 and y_1 are much smaller than unity (except possibly at very large amounts of mistuning), and it contains an amplitude modulation which will not affect the switching function of the modulator.

The second term, which is in quadrature with the first, represents therefore a phase-modulation effect of maximum deviation approximately

$$b_1 [1 - c_1 \cos(p_1 t + \psi_1)] \quad \dots\dots(11)$$

and of frequency p_1 . Here

$$c_1^2 = x_1^2 + y_1^2 + 2x_1 y_1 \cos(\alpha_1 - \beta_1) \quad \dots\dots(12)$$

and $\psi_1 = \tan^{-1} \left[\frac{x_1 \sin \alpha_1 + y_1 \sin \beta_1}{x_1 \cos \alpha_1 + y_1 \cos \beta_1} \right] \dots\dots(13)$

The switching function is thus phase-modulated, and has a fundamental component proportional to

$$\cos \{ \omega_1 t + \theta_1 - b_1 [1 - c_1 \cos(p_1 t + \psi_1)] \sin(p_1 t + \phi_1) \} \quad \dots\dots(14)$$

and since we assume $b_1 \ll 1$, this is approximately

$$\{ 1 - \frac{1}{2} b_1^2 [1 - c_1 \cos(p_1 t + \psi_1)]^2 \} \cos(\omega_1 t + \theta_1) + b_1 [1 - c_1 \cos(p_1 t + \psi_1)] \sin(\omega_1 t + \theta_1) \cdot \sin(p_1 t + \phi_1) + \frac{1}{2} b_1^2 [1 - c_1 \cos(p_1 t + \psi_1)]^2 \cos(\omega_1 t + \theta_1) \cdot \cos(2p_1 t + 2\phi_1) \dots\dots(15)$$

The relevant output of the demodulator comprises the modulation-frequency components in the product of (6) and (15), and these are readily seen to include:

Fundamental

$$a_1 [m_1 \cos \theta_1 \cos p_1 t + b_1 \sin \theta_1 \sin(p_1 t + \phi_1)] \quad (16)$$

where terms in m_1 , b_1 , and/or c_1 of third and higher orders are ignored since all three quantities are small.

The modulus of this output component is easily shown to be

$$a_1 [m_1^2 \cos^2 \theta_1 + b_1^2 \sin^2 \theta_1 + m_1 b_1 \sin 2\theta_1 \sin \phi_1]^{\frac{1}{2}} \quad (17)$$

Second harmonic

$$\frac{1}{2} a_1 b_1 [m_1 \sin \theta_1 \sin(2p_1 t + \phi_1) - c_1 \sin \theta_1 \sin(2p_1 t + \phi_1 + \psi_1) + \frac{1}{2} b_1 \cos \theta_1 \cos(2p_1 t + 2\phi_1)] \dots\dots(18)$$

where terms in m_1 , b_1 and/or c_1 of fourth and higher orders are ignored.

The modulus of this output component may be shown to be

$$\frac{1}{2} a_1 b_1 \{ (m_1^2 + c_1^2 - 2m_1 c_1 \cos \psi_1) \sin^2 \theta_1 + \frac{1}{4} b_1^2 \cos^2 \theta_1 + \frac{1}{2} b_1 \sin 2\theta_1 [c_1 \sin(\phi_1 - \psi_1) - m_1 \sin \phi_1] \}^{\frac{1}{2}} \dots\dots(19)$$

Since b_1 and c_1 both include a factor m_1 , it is clear that the fundamental output is proportional to m_1 and the second harmonic output to m_1^2 .

9. Appendix 2

Interference in an Enhanced-Carrier System

Symbols will be used as in Appendix 1, with subscript "2" when referring to the interfering signal. The input signal is

$$a_1(1 + m_1 \cos p_1 t) \cdot \cos \omega_1 t \quad \dots\dots(20)$$

and the interfering signal is

$$a_2(1 + m_2 \cos p_2 t) \cdot \cos \omega_2 t \quad \dots\dots(21)$$

For the time being, consider m_1 and m_2 to be made zero, and take the response of the tuned system in the enhanced-carrier path to be such that the interfering signal is reduced $1/r$ times relative to the wanted signal. (N.B. $r < 1$). Assume the circuit is correctly tuned to the wanted carrier, i.e. $\theta_1 = 0$. Then the switching function of the modulator can be shown² to be proportional to

$$J_0(k) \cos \omega_1 t + J_1(k) \cos(\omega_2 t + \theta_2) + \dots \dots(22)$$

where the remaining terms have no influence on the present calculation. Here $k = a_2 r / a_1$.

Now $k \ll 1$, and so we may write

$$J_0(k) \simeq 1 - \frac{k^2}{4} \quad \text{and} \quad J_1(k) \simeq \frac{k}{2} \quad \dots\dots(23)$$

It is important in this case to retain the second term in the expansion of $J_0(k)$. Thus the switching function is proportional to

$$\left(1 - \frac{k^2}{4} \right) \cos \omega_1 t + \frac{k}{2} \cos(\omega_2 t + \theta_2) + \dots\dots(24)$$

The output of the modulator will be given by the product of (20) + (21) and (24), with $m_1 = m_2 = 0$.

Now let m_2 be finite, but small compared to unity. Due to the asymmetry of the tuning circuits with respect to ω_2 , the interfering signal becomes phase as well as amplitude-modulated in the enhanced-carrier path, as discussed for mistuning in Appendix 1. Both kinds of modulation affect the switching function; the amplitude modulation causes k to vary at the frequency p_2 so that k is replaced by

$$k[1 + c_2 \cos(p_2 t + \psi_2)] \quad \dots\dots(25)$$

where $c_2^2 = x_2^2 + y_2^2 + 2x_2 y_2 \cos(\alpha_2 - \beta_2) \quad \dots\dots(26)$

and $\psi_2 = \tan^{-1} \left[\frac{x_2 \sin \alpha_2 + y_2 \sin \beta_2}{x_2 \cos \alpha_2 + y_2 \cos \beta_2} \right] \quad \dots\dots(27)$

The phase-modulation adds a term (to a first approximation†) to θ_2 thus:

$$-b_2 \sin(p_2 t + \phi_2) \quad \dots\dots(28)$$

where it is assumed that $b_2 \ll 1$; thus the complete switching function is now proportional to

$$\{1 - \frac{1}{4}k^2[1 + c_2 \cos(p_2 t + \psi_2)]^2\} \cos \omega_1 t + \frac{1}{2}k[1 + c_2 \cos(p_2 t + \psi_2)] \cos[\omega_2 t + \theta_2 - b_2 \sin(p_2 t + \phi_2)] \dots\dots(29)$$

† According to the analysis of Appendix 1, this phase modulation should also have a factor $[1 - c_2 \cos(p_2 t + \psi_2)]$, but it can readily be shown that this leads only to terms of negligible magnitude provided it may also be assumed that $c_2 \ll 1$.

The modulator output is given by the product of this and (21)+(22), with m_1 still zero. We are concerned with the output component at frequency p_2 .

Now if $b_2 \ll 1$, we may write

$$\begin{aligned} \cos[\omega_2 t + \theta_2 - b_2 \sin(p_2 t + \phi_2)] \\ \simeq \cos(\omega_2 t + \theta_2) + b_2 \sin(\omega_2 t + \theta_2) \cdot \sin(p_2 t + \phi_2) \end{aligned} \quad \dots\dots(30)$$

and thus the output component at frequency p_2 is proportional to

$$\begin{aligned} \frac{1}{4}[kc_2(-a_1 k + a_2 \cos \theta_2) \cos(p_2 t + \psi_2) + \\ + ka_2 m_2 \cos \theta_2 \cos p_2 t - \\ - ka_2 b_2 \sin \theta_2 \sin(p_2 t + \phi_2)] \quad \dots\dots(31) \end{aligned}$$

Had the wanted signal been modulated ($m_1 \neq 0$) and the unwanted signal been absent, the output of frequency p_1 would have been proportional to

$$\frac{1}{2}a_1 m_1 \cos p_1 t \quad \dots\dots(32)$$

From (31) and (32) the interference-suppression ratio can be determined for any given conditions, and the effectiveness of the carrier-enhancing filter determined. Unfortunately the calculation is tedious.

Manuscript first received by the Institution on 1st November 1961 and in final form on 1st March 1962. (Paper No. 782).

© The British Institution of Radio Engineers, 1963

of current interest . . .

Electronic Standardization by the I.E.C.

The complexities of modern electronic components, particularly semiconductors, valves and the rapidly developing microstructures, are being simplified by the International Electrotechnical Commission. To a considerable degree standards developed in the United Kingdom by VASCA (The Electronic Valve and Semiconductor Manufacturers' Association) and BVA (The British Radio Valve Manufacturers' Association) are being accepted.

Mr. P. A. Fleming, technical secretary of VASCA, has reported that considerable agreement was reached at the recent meetings of the I.E.C. in the standardization of technical language, comparison data, measurements and physical dimensions. The standardization of technical data quoted by manufacturers of semiconductors is looked upon as a major achievement.

Simplification of technical terms has already been agreed on. For instance, one of the technical committees has already standardized the term thyristor as the international description for a controlled rectifier.

Particularly important, Mr. Fleming states, is the work now being done on standardization of the methods of measurement. Future international agreement will centre on the developing field of microelectronics—extremely high component-density semiconductor structures.

New Method for Plating Thin Wire

The problems concerned in the plating of thin wire to an even thickness, by maintaining the uniformity in the composition of the electrolyte, can be overcome as a result of a new process developed by the British company, Ultra Electronics.

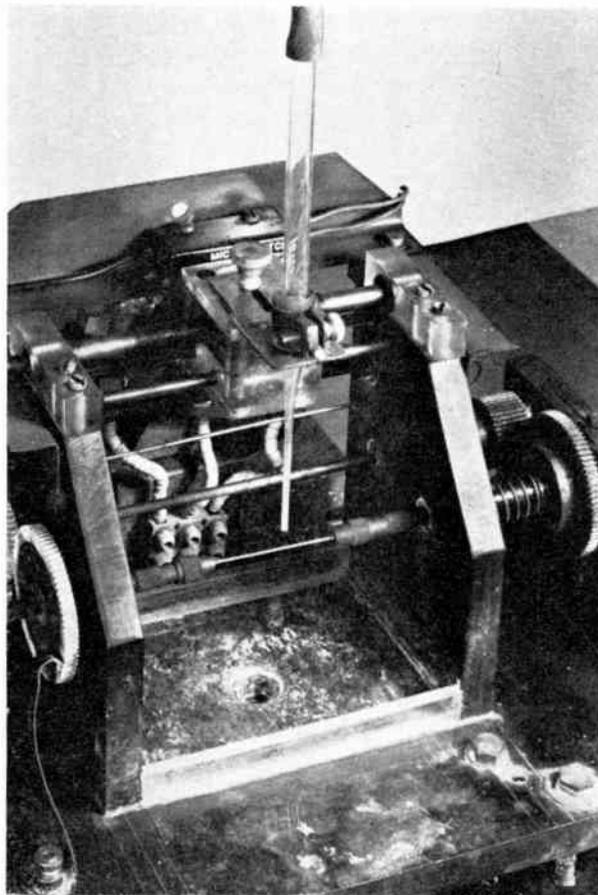
Wire with a diameter no greater than a thousandth of an inch can be plated by using a small lathe with driven pin chucks at each end, thus maintaining it in a state of tension. While it is being rotated, the wire is sprayed with an electrolyte, from a nozzle or jet, kept at a suitable electric potential which is different from that of the wire. The rotating wire ensures that all of the wire gets equal plating. In alloy plating, the solution tends to change its composition around the anode and the cathodes. With the new method of spraying on to the wire all the time, this eliminates the possibility of deterioration.

The jet is made of glass, with the anode inside. The electrolyte is run or pumped from a reservoir through the moving jet on to the rotating wire, with a voltage applied between the jet anode and the wire. By this means, fast plating can be achieved. And because the electrolyte is continuously replaced, the thickness and composition of the alloy coating is kept constant at all times.

The technique was first envisaged for electrolyte machining of wires to accurate dimensions, and the process subsequently developed for the plating of wires with magnetic alloys for use in computer devices. At present, lengths of wire up to nine inches can be dealt with in this way, but it would be possible to deal with greater lengths.

New Laboratory Site for National Institute for Research in Nuclear Science

The National Institute for Research in Nuclear Science have applied for planning permission to build a new Laboratory at Daresbury in North Cheshire. The Laboratory will house a 4000 meV synchrotron, the authorization for which was announced by the Minister for Science in July. The Laboratory will be used as a research centre by university physicists, particularly by those from the universities of Glasgow, Liverpool, and Manchester. The N.I.R.N.S. has operated the Rutherford Laboratory at Harwell (adjoining the Atomic Energy Research Establishment) for the past three years.



The new method of plating thin wire developed by the Ultra Electronics Materials Research Laboratory.

Cathode-Ray Tube Displays for Fish Detection on Trawlers

By

P. R. HOPKIN,
(Associate Member)†

Presented at the Symposium on Sonar Systems in Birmingham on 9th–11th July 1962.

Summary: Current echo-sounder recorder and cathode-ray tube display techniques are reviewed and their shortcomings stated. An improved cathode-ray tube display is described in which the seabed echo always occurs in the same position on the screen irrespective of changes in range due to the contour of the seabed or vertical motion of the trawler. A magnetic recording drum with out-of-contact heads stores the echoes from the range adjacent to the seabed and enables them to be displayed repeatedly during each sounding period.

1. Introduction

Echo-sounders were originally fitted on trawlers for navigational purposes about 1930. When instruments employing ultrasonic frequencies and recording displays were introduced in 1933, reports of echoes from fish shoals indicated that echo-sounders of this type could be used for fish detection. In 1935, Sund¹ published a recording of cod made on board a Norwegian research vessel at Lofoten, and since then echo-sounders have been used increasingly in fishery research.^{2, 3}

Trawler fish detection was improved with the introduction of cathode-ray tubes to display echoes from a limited range above the seabed. Experience with this type of display emphasized the difficulty of detecting fish when adjacent to the seabed and has led to detailed investigation of the problem.

2. Requirements

The trawler skipper is chiefly interested in the volume swept by the trawl-mouth and, as the headline height of the standard small Granton trawl used from the Humber ports is about 8 feet at the centre of the mouth, the last fathom or so above the seabed shown on the echo-sounder is very important. He needs to know the depth of water, as this determines the amount of warp paid out, and, in certain circumstances, it is advantageous to follow a seabed contour line by steering a course to keep the depth constant. The nature of the seabed is also important, particularly on fishing grounds scattered with rocks and other hazards to the trawl. It is useful to know the type of fish, whether it is of sufficient size to be of commercial value, and whether distributed in shoals or spread out over the seabed.

For British distant-water trawlers the maximum requirement is to detect a small fish such as a codling

† Kelvin Hughes Division, S. Smith & Sons (England) Ltd., Ilford, Essex.

(30 cm length) in depths down to 280 fathoms in all weather conditions in which fishing can be carried on. In particular, the display system should be capable of showing fish when spread out close to the seabed.

3. Current Equipment

3.1 Recorder Displays

The construction of recorders used in echo-sounding has been described^{4, 5} and their properties investigated.^{6, 7} The transmission pulse rate, and hence the scale of the chart, is determined by the need to obtain unambiguous echoes in the maximum required depth. In a typical rotary-arm recorder the stylus arm revolves at 100 rev/min giving a transmission interval of 0.6 seconds and a maximum unambiguous range of 240 fathoms. This is achieved by a "phasing" scale which displays 60 fathoms in 5½ in. Consequently the range of chief interest, i.e. the headline height above the seabed, occupies a narrow band of 0.12 in. Timing marks are recorded on the chart, usually at one minute intervals, so that the occurrence of fish markings can be related to the time, and hence position in which they appeared. Figure 1 shows a length of chart made by a recording echo-sounder of this type in which the contour of the seabed, and echoes from single fish and small shoals are seen.

While this form of display is efficient when detecting fish in midwater, it is not satisfactory when fishing with a bottom trawl. In certain weather conditions, the trawler is rising and falling two fathoms and the serration of the seabed profile thus produced masks fish echoes close to the seabed. The trace-to-trace correlation, which is an important feature of the recorder display,⁷ also deteriorates under these circumstances.

The recording paper has a limited density range and unless the sensitivity control of the echo-sounder is set critically, the fish echoes will be presented at satura-

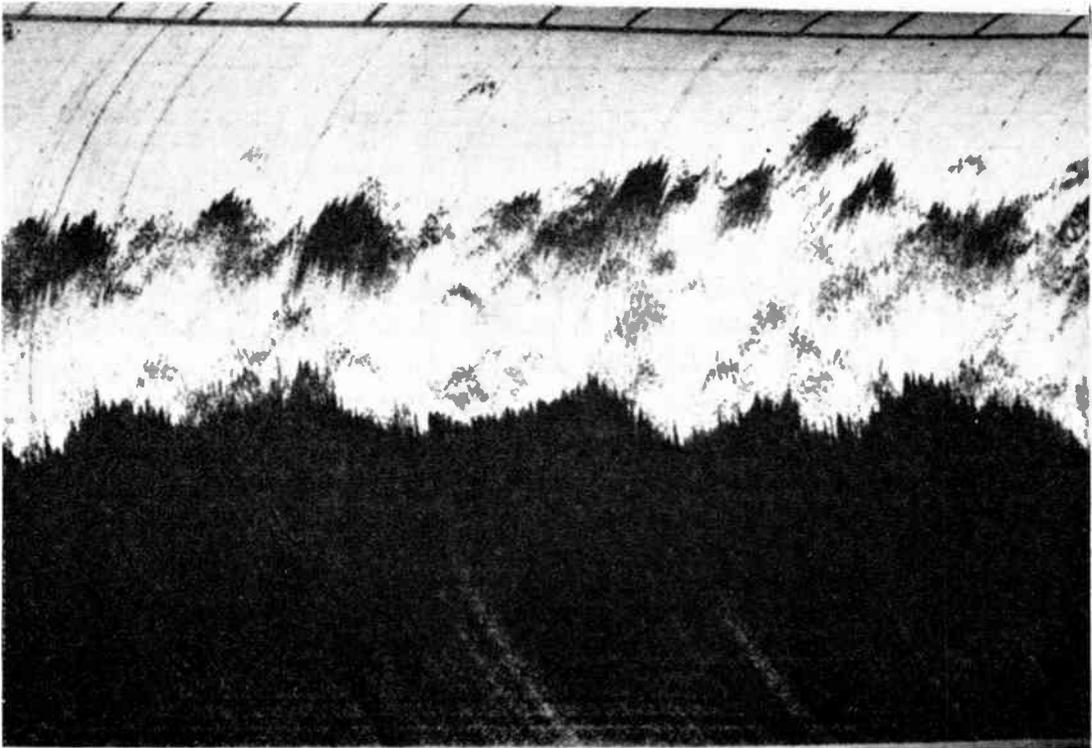


Fig. 1. Recorder chart showing echoes from cod. (S.T. *Lancella*; depth 110 fathoms, Barents Sea.)

tion density and those which adjoin the seabed will be indistinguishable from it. In recent years the introduction of amplitude discriminating techniques such as “white-line”⁸ and “grey-black”⁹ recording has improved the performance of recorders for trawler fish detection echo-sounders.

3.2. Cathode-ray Tube Displays

A cathode-ray tube has the advantage that any partial range may readily be expanded to fill the cathode-ray tube screen without increasing the transmission pulse rate.

The usual arrangement utilizes an unrectified “A-scan” with a vertical time-base displaying a range of 5 to 15 fathoms on a screen of about 5 in. diameter. This is shown in Fig. 2 which is a 10 fathom screen range display showing the seabed echo and echoes from a small shoal of fish. The start of the time-base is delayed with respect to the transmission pulse so that the display may be adjusted to include the seabed echo and the range just above it. This delay is produced by means of contacts which are “phased” relative to the transmission contacts¹⁰ or by using an electronic delayed pulse generator circuit which is triggered by the transmission pulse.

A cathode-ray tube display employing an electronic delayed pulse generator is illustrated in schematic form in Fig. 3. In this case one electro-acoustic transducer is used for both transmission and reception so

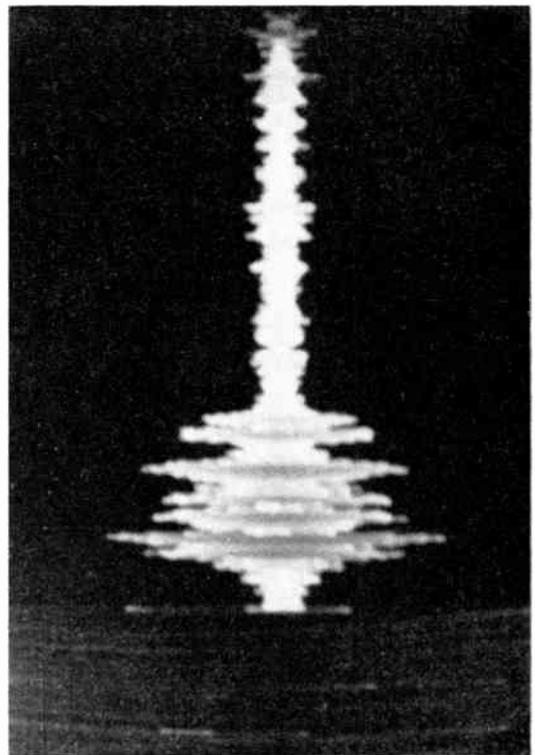


Fig. 2. Cathode-ray tube display showing echoes from a small fish shoal and the seabed. (S.T. *Primella*; depth 40 fathoms, North Sea.)

that at the transducer terminals the transmitted pulse is very large compared with the received echoes. A suitable pulse is formed from the transmission to trigger the delayed pulse generator which, in turn, triggers the time-base. As the amount of delay and the time-base range are adjusted by means of manual controls, any partial range may be displayed with varying degrees of expansion. To avoid a stationary spot on the cathode-ray tube, the beam is cut off except during the time-base stroke. The bright-up pulse may be derived conveniently from the time-base circuit.

A high-gain amplifier amplifies the echo signals before they are applied to the horizontal deflection plates of the cathode-ray tube. The gain is manually controlled and is usually adjusted so that random noise picked up by the transducer produces a small deflection of the trace.

Various methods are used to find and also indicate the depth of water. In instruments which use the

cathode-ray tube as the only means of display, a time-base triggered at the instant of the transmission pulse covers the total range.^{10, 11} The delay control for the expanded range is set either by the use of a calibrated scale¹⁰ or an electronically generated strobe.¹¹ In instruments which employ a recorder operating simultaneously and where the triggering contacts for the time-base are provided on the recorder mechanism, the expanded range is indicated on the recorder scale by a pointer or cursor. Where the recorder and cathode-ray tube displays are housed in separate cases, it is more convenient to use an electronic delayed pulse generator circuit with the manual control on the cathode-ray tube unit. In this case, marker pulses generated at the start and finish of the time-base are fed to the recorder and produce lines on the recorder chart indicating the displayed range.¹²

In addition to the improvement in the scale of the display, the cathode-ray tube has the great merit of show-

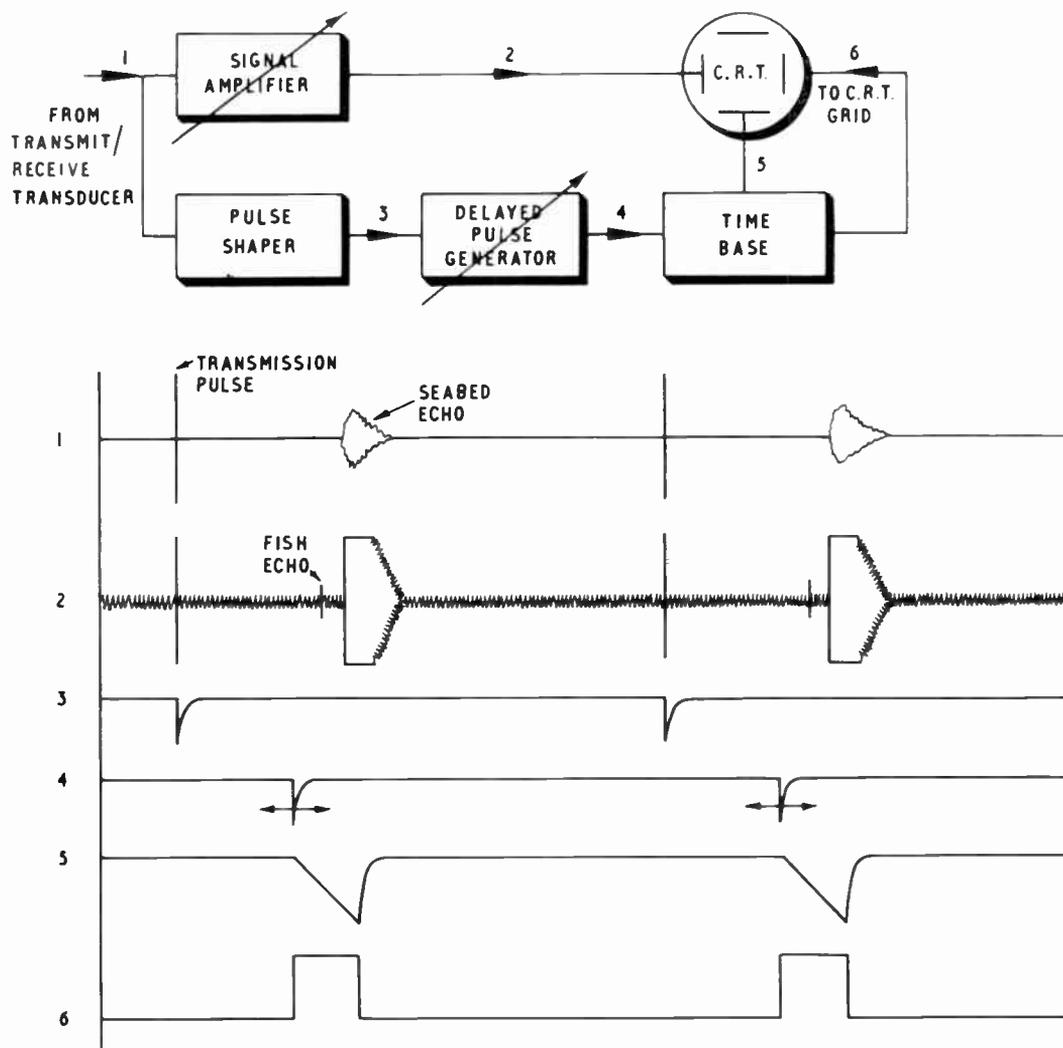


Fig. 3. Schematic and waveforms of c.r.t. display using electronic delayed pulse generator.

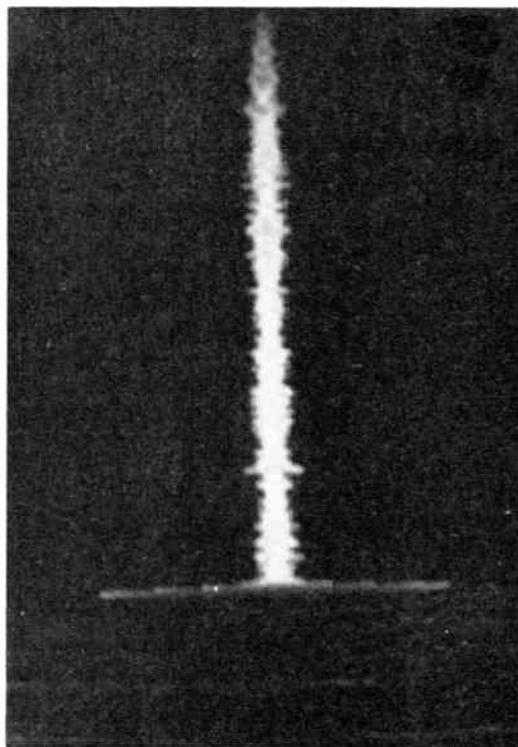
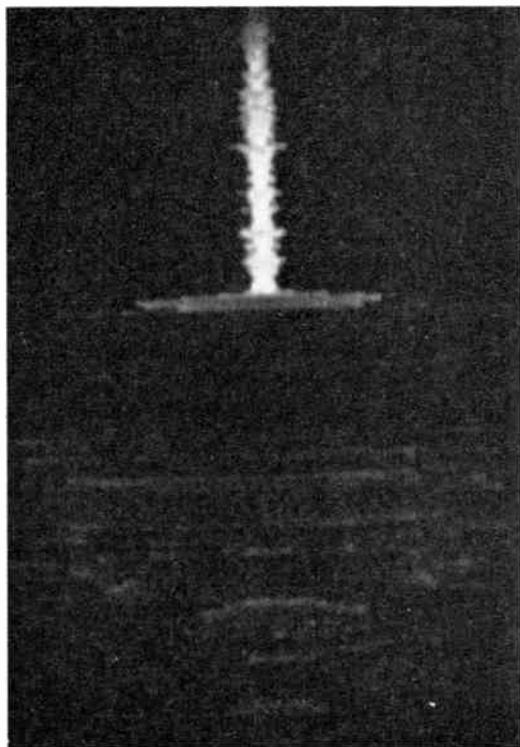


Fig. 4. Fish echoes merging into the seabed echo. (S.T. *Primella*; depth 60 fathoms, Barents Sea.)

ing fish echoes even when they merge with the leading edge of the seabed echo. Photographs of such echoes are shown in Fig. 4. This advantage is due to the spot velocity being much lower when tracing the fish echo than when tracing the seabed echo, making the former appear bright and easily distinguishable.

Another important feature of cathode-ray tube displays is that the amplitude and character of the fish echoes are faithfully represented and this assists in the determination of size and type of fish.^{13, 14}

In rough weather, when the vertical motion of the trawler is of the order of 12 ft, the displacement of the seabed echo which occurs from one trace to the next is a large fraction of the displayed range. Also, if the depth of water is varying, frequent adjustment of the delay control is necessary to follow this change. Therefore, it is only possible to use a displayed range of less than 10 fathoms in favourable conditions although, ideally, a range of 5 fathoms or less is required.

Since the displayed range is a small fraction of the maximum range allowed by the transmission pulse rate, the signals are displayed as a flash on an otherwise blank screen. While this is adequate for observation of the echo amplitude and position to be made and compared with subsequent traces, such a display is tiring to watch for long periods.

A tube with a long persistence screen is, unfortunately, of no assistance as the brightness decays rapidly at first and persists at a comparatively low level. This results in one trace being painted over another, and as the echoes do not occur in the same relative positions because of the vertical motion of the trawler, a confused display results. The major disadvantage of cathode-ray tube displays is that, unlike recorders, they must be watched continuously if full benefit is to be derived.

While the introduction of cathode-ray tube displays has greatly improved fish detection on trawlers, they have their own limitations and it is understandable that instruments which use both recorder and cathode-ray tube displays are the most usual form of fish detection equipment on British distant-water trawlers.

4. Seabed-locked Displays

The cathode-ray tube display is greatly improved by arranging that the seabed echo appears on the trace at the same place irrespective of the vertical movement of the trawler or the change in seabed contour. This allows short time-base ranges to be used and provides automatic following of the seabed contour. A tube with a long persistence screen could be used if required.

Three methods have been proposed to achieve this.

4.1 Echo Signal Delay

The fish echoes received just before the seabed echo may be displayed on a time-base which is itself triggered by the seabed echo if a form of signal delay is used.¹⁵ The time-base is made longer than the delay time, and the displayed position of the seabed echo is determined by the delay time and the time-base velocity. Since these are constant, the seabed echo appears at the same place on successive traces irrespective of its actual time of arrival, and fish echoes are displayed on a time-base which is locked to the seabed echo.

The signal delay may be obtained by an electrical or acoustical delay line, or by magnetic recording in which the difference in time between recording and replaying constitutes the delay.

4.2. Transmission Pulse Delay

This method involves transmitting a pulse at one frequency followed after a short delay by a second pulse at another frequency.¹⁶ The seabed echo from the first pulse is used to trigger the time-base and the echoes from the second pulse are amplified by a frequency-selective amplifier for display.

The time-base duration is slightly greater than the interval between the two transmitted pulses so that the seabed and adjacent fish echoes are displayed. The seabed echo occurs at the same place on each trace providing the vertical movement of the trawler is not significant during the intervals of transmitting the pulses on each frequency and reception of the seabed echoes on each frequency. Since the transmission delay is short this restriction is easily met.

4.3. Frequency Modulation

An interesting method has been proposed for obtaining a seabed-locked output with a frequency-modulated echo-sounder.¹⁷ The echoes appear as semi-continuous tones having frequencies which are a function of the range. If a narrow transducer beam-angle is used, the seabed echo will give a substantially pure tone of much greater amplitude than the fish echoes. When applied to a rectifier, new tones will be produced which are difference frequencies between fish and seabed echoes and are therefore a function of height above the seabed. The output of the rectifier is filtered to remove the input frequencies and leave only those from echoes occurring in the required range above the seabed.

5. Development of a Seabed-locked Cathode-ray Tube Display

In developing an instrument of this type for commercial trawlers, all the useful features of existing instruments must be retained, especially the detection of fish echoes which merge with the seabed echo and the accurate representation of fish amplitudes.

The signal delay method using magnetic recording meets these requirements, and being a signal storage and not merely a delay system, it has the important additional feature that the signals may be repeated any number of times.¹⁵

The magnetic recording medium is a continuous band on the periphery of a drum and it is rotated at such a speed that it completes a revolution in a time not much greater than the longest time-base used. The magnetic recording system is switched to the record condition in anticipation of the arrival of

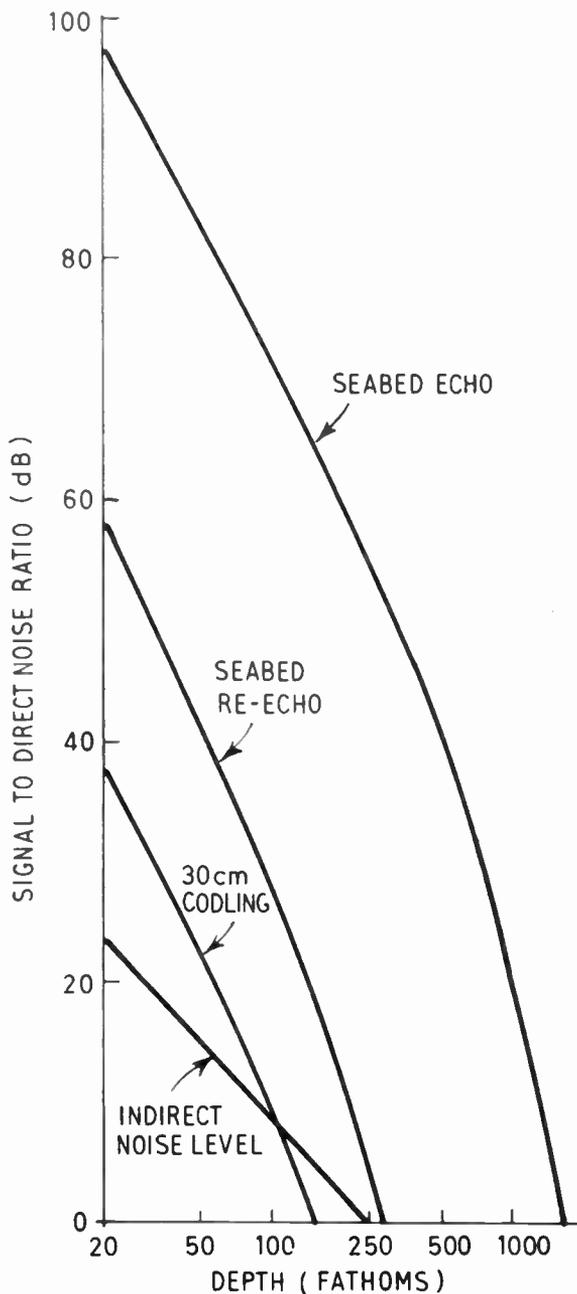


Fig. 5. Relative signal levels.

echoes occurring in the range of interest, and switched back to the replay condition soon after the arrival of the seabed echo. Thus recorded echoes repeat with every revolution of the drum and this continues until echoes from the subsequent transmission arrive and are recorded.

The presentation is now greatly improved as the screen appears blank only during the short time (about 50 milliseconds) when new signals are being recorded on the drum. The time-base frequency is 50 per second and this gives an apparently continuous display of signals which greatly enhances the brilliance of the trace.

5.1 Selection of the Seabed Echo

It is essential to the operation of a seabed-locked display that the seabed echo is reliably selected from all the other signals appearing at the transducer.

It can be seen from Fig. 5 that the amplitudes of fish echoes, seabed re-echoes (i.e. echoes of the seabed which are reflected by the surface of the sea and again by the seabed), and random noise, are smaller than the seabed echoes over the range of depths (20 to 280 fathoms) in which the equipment is required to operate. The random noise is generated acoustically by the trawler (e.g. by cavitation occurring at the propeller) and, due to the directivity pattern of the transducer, the level received indirectly by reflection from the seabed exceeds the direct path noise in depths of less than 250 fathoms. The transmission pulse, especially in an equipment which uses one transducer for both transmission and reception, produces a break-through signal which is greatly in excess of the seabed echo.

The selection system is illustrated in Fig. 6. At the input, the transmission pulse break-through is very

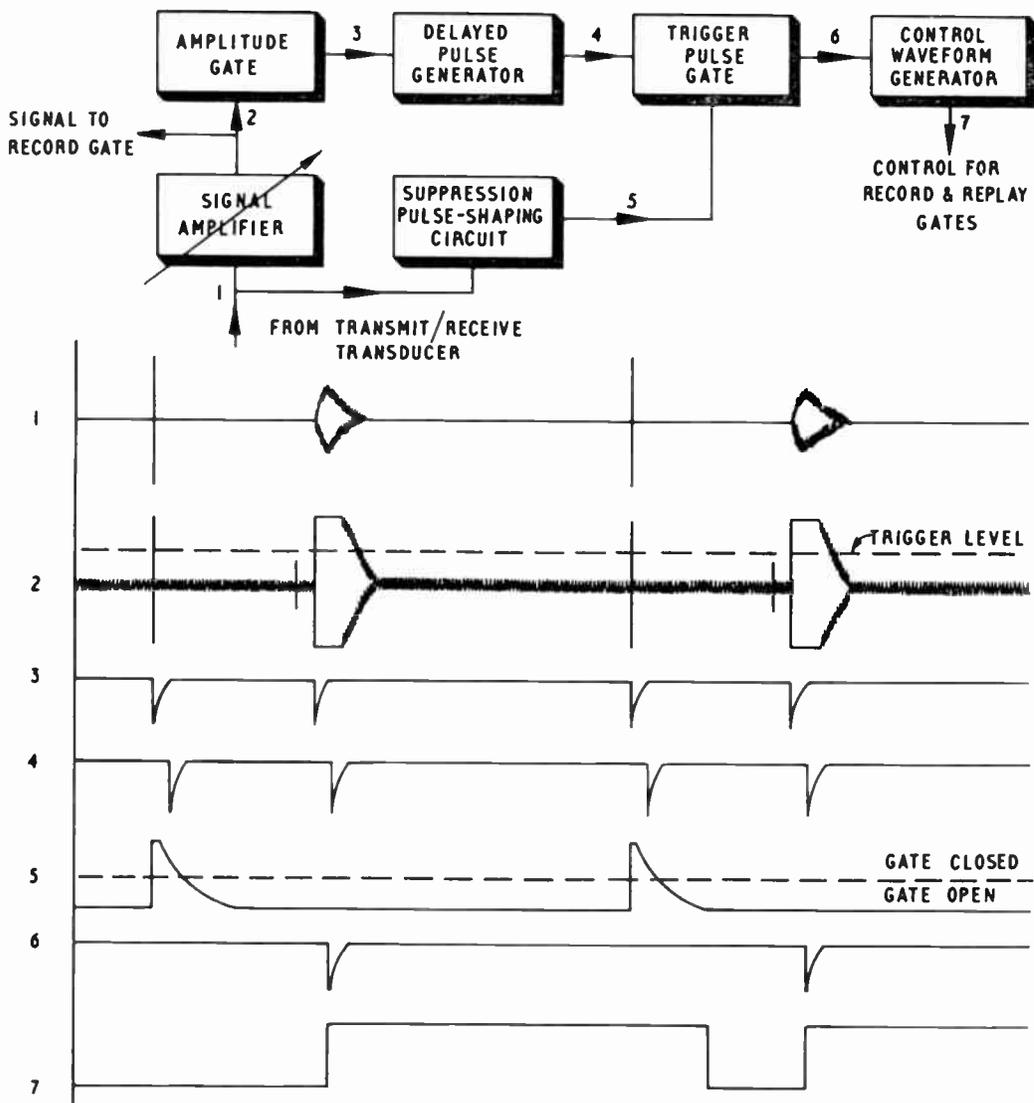


Fig. 6. Schematic and waveforms of seabed echo selection system.

large compared with the seabed echo, but at the output of the signal amplifier both signals are of the same amplitude due to saturation. Signals of smaller amplitude than the seabed echoes are eliminated by an amplitude gate. The delayed pulse generator is therefore triggered twice during each sounding period and generates two pulses delayed by about 4 ms.

The delayed pulse produced by the transmission break-through is eliminated by a gating circuit which is operated by a waveform developed from the large transmission pulse to the transducer. The gate is held closed for about 50 ms so that strong fish echoes, plankton echoes and reverberations from aerated water which may trigger the delayed pulse generator within the first 20 fathoms are also eliminated. The remaining delayed pulse produced by the seabed echo is used to trigger a waveform generator, the output of which is a pulse with a duration of about 50 ms less than the interval between transmission pulses. (In a typical system with a transmission period of 600 ms the pulse duration is 550 ms.) Once triggered, the pulse generator circuit is insensitive to further triggering occurring during the pulse. Therefore signals of sufficient amplitude to trigger the delayed pulse generator, such as echoes due to large shoals and bursts of noise which occur during the pulse, do not affect the operation of the system. The waveform thus produced is used to control all functions of the instrument.

The interval between the end of one pulse and the beginning of the next will be varied by the vertical motion of the trawler and the rate of change of depth. In practice this is not greater than 10 ms so that the interval is never less than 40 ms. The system will remain locked to the seabed echo in a wide range of depths and conditions, providing the manual gain control of the signal amplifier is adjusted to keep the displayed fish signals within the cathode-ray tube screen diameter. Even if the transducer is completely blanketed by aerated water so that no seabed echoes are received, the system will lock again when the aeration has cleared.

5.2 Magnetic Record/Replay System

The control waveform, produced as described in the previous section, is used to change the magnetic recording system from record to replay at the correct time in the sounding period. This has to be accomplished in less time than the magnetic recording drum takes to make one revolution. Therefore an entirely electronic switching system is employed in conjunction with a single head for both recording and replaying.¹⁸

5.2.1 Magnetic recording electronics

The magnetic record/replay system is illustrated in Fig. 7. The falling wavefront which terminates the

control pulse closes the replay gate, opens the record gate and switches on the bias and erase oscillator. The output of the record amplifier is added to the high frequency bias waveform and fed to the record/replay head. Previously recorded signals are erased by the erase head. This is situated so that the recording medium passes it just before reaching the record/replay head. Simultaneous recording and erasing takes place until the commencement of the next control pulse. The rising wavefront closes the record gate, switches off the bias and erase oscillator and opens the replay gate. The control pulse commences 4 ms after the seabed echo arrives, as explained in Section 5.1, so that a portion of the seabed echo is stored on the drum together with signals arriving in the previous 12 ms. The record/replay head picks up the recorded signals which are then amplified before passing to the replay gate.

At the output of the replay gate, the recorded signals are repeated every revolution of the drum until the end of the control pulse switches the system back to the record condition. For clarity, only four repetitions of the stored signals are shown in Fig. 7, but in a practical case, the transmission pulse period is 0.6 seconds, the drum revolution period is 20 ms, and there are approximately twenty-seven repetitions.

5.2.2 Magnetic recording drum

The magnetic recording drum was designed specifically for this instrument. It is required to rotate at 3000 rev/min, operate at the signal frequency of 30 kc/s with a bandwidth of not less than ± 2 kc/s and have a dynamic range of at least 50 dB. This dynamic range is necessary to preserve the brightness contrast between fish and seabed echoes when displayed on the cathode-ray tube, as described in Section 3.2. The amplitude modulation should not exceed 10% as it then becomes noticeable on the display.

The long life requirement of the equipment makes out-of-contact operation of the record/replay and erase heads essential. The effect of spacing a replay head from the magnetic recording medium has been examined in detail by Wallace,¹⁹ and he has shown that the loss introduced is

$$54.6 (d/\lambda) \text{ dB.}$$

where d = distance between head and medium

λ = wavelength.

Because of the magnitude of this loss and the dynamic range requirement, it is necessary to make the recorded wavelength much greater than the head-medium spacing. The minimum spacing which may be used is determined by the precision, and hence the cost, of manufacturing the drum assembly. It is also

desirable to use the gap width which gives maximum output for this condition.

The magnetic drum assembly designed to these requirements (Fig. 8) has a 3½-in. diameter drum, the head-medium spacing is 0.0015 in. and the effective head gap width 0.003 in. The frequency response due to wavelength losses is within 3 dB from 18 kc/s to 80 kc/s. The recorded wavelength of a 30 kc/s signal is 0.018 in, and the loss due to spacing is about 5 dB. Frequency-dependent losses in the record/replay head are small compared with the wavelength loss as a ferrite core and a non-conducting gap spacer are used. The erase head also has a ferrite core and is energized by oscillations of approximately 400 kc/s. The recording medium is magnetic oxide in resin, the surface being diamond turned after coating so that it is concentric with the axis of rotation. The drum is supported by two end-loaded ball bearings, mounted in a

substantial housing, and is driven by a hysteresis motor through a flexible coupling.

5.3 Cathode-ray Tube Display

The signals fed from the replay gate for display on the cathode-ray tube also control triggering of the time-base. Since the magnetic recording drum is rotating at a constant velocity, the replayed seabed echoes repeat with a constant period. Thus the seabed echo always appears at the same place on the screen irrespective of its actual time of arrival.

The system is illustrated in Fig. 9. An amplitude gate selects the replayed seabed echo and it is set to operate at a level not much greater than the largest amplitude fish echoes. This enables trigger pulses to be derived, their timing being relatively unaffected by the slope of the leading edge of the seabed echo. The delayed pulse generator triggered by these pulses enables

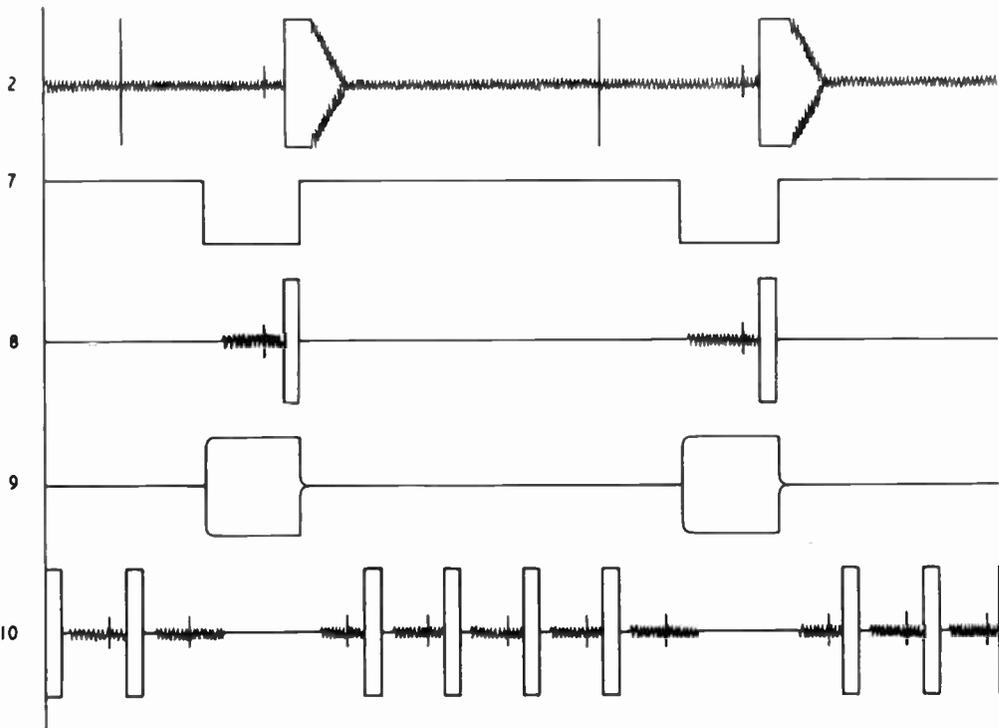
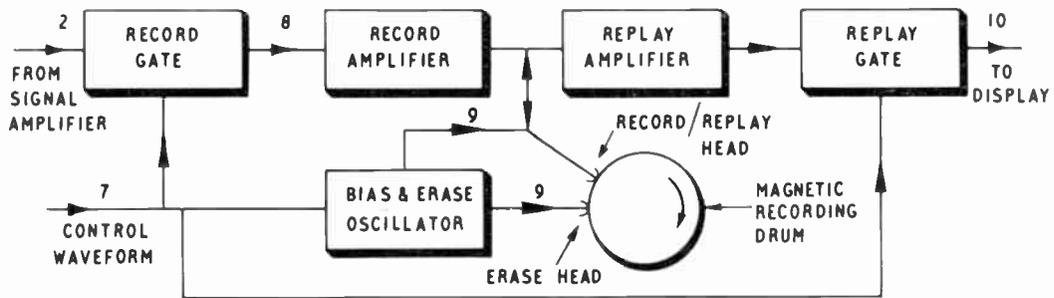


Fig. 7. Schematic and waveforms of magnetic record/replay system.

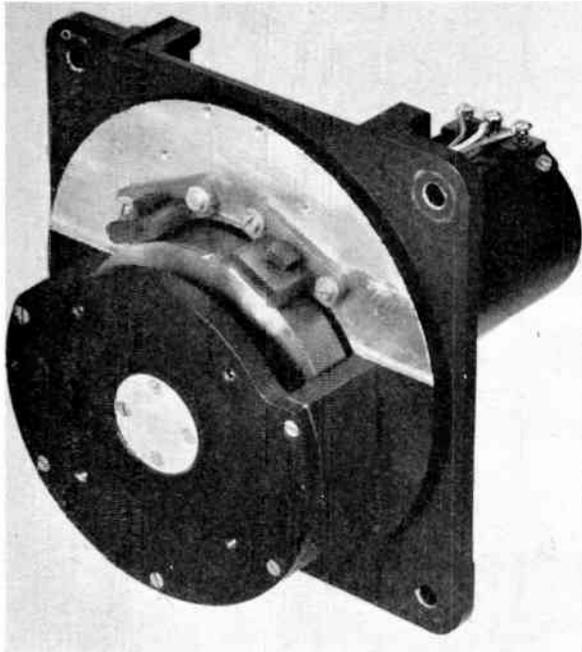


Fig. 8. Magnetic recording drum assembly.

each time-base stroke to be controlled by the replayed seabed echo displayed on the previous stroke. The amount of delay is determined by the time-base range in use and the range control is ganged to the delay control so that the seabed echo appears at the same place on the screen irrespective of the range in use. The bright-up waveform (taken from the time-base circuit), is gated by the control waveform so that the trace is terminated at the end of the replay period even if this occurs during a time-base stroke.

The signal deflection amplifier is designed to apply signals to the cathode-ray tube without losing the brightness contrast between fish and seabed echoes described in Section 3.2. As the amplifier is heavily overloaded by the seabed echoes, this contrast will deteriorate if the rise time exceeds 10 μ s.

A 7-inch diameter cathode-ray tube is used and it is housed, together with its associated circuits and power supplies, in a separate case. This allows the trawler skipper considerable freedom in choosing a position

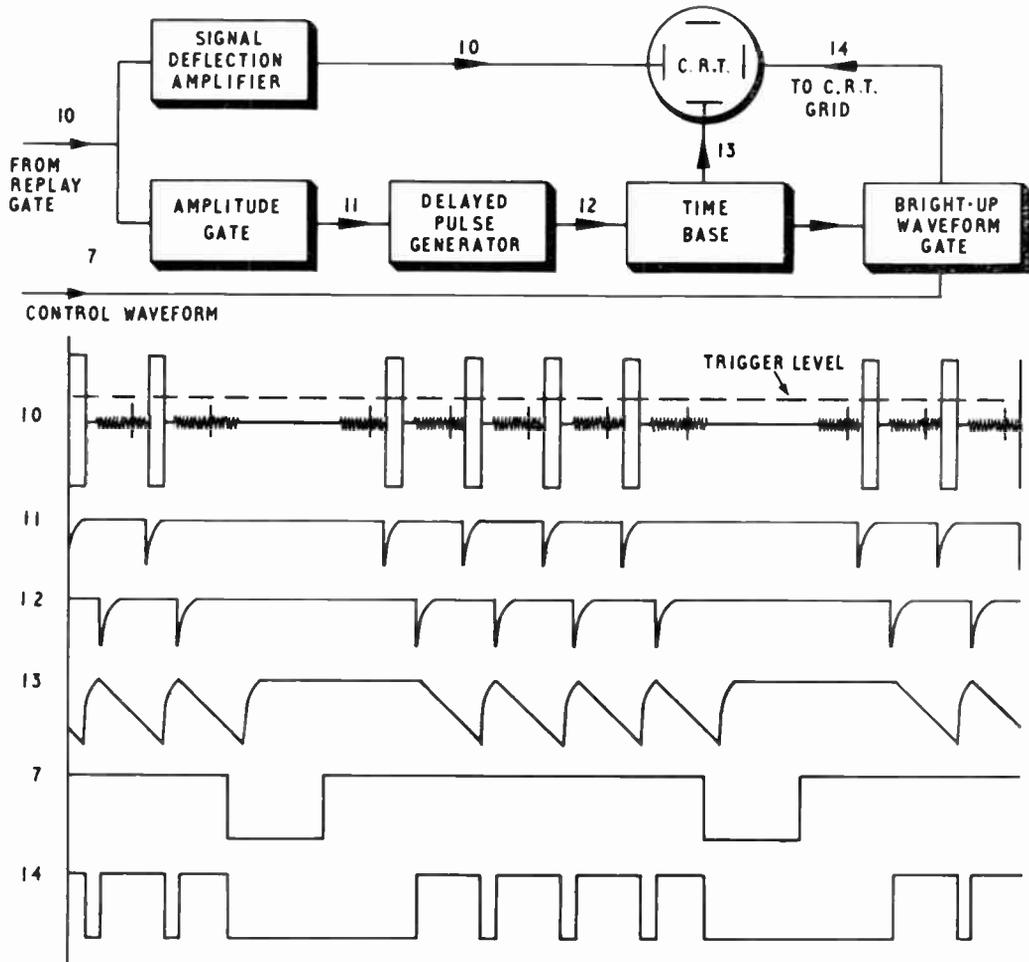


Fig. 9. Schematic and waveforms of c.r.t. display.

for it in the wheelhouse. All the operating controls (including remote sensitivity control of the selection system signal amplifier) are situated on this unit. Two ranges of seabed-locked display (2.5 and 5 fathoms) and two ranges of transmission-locked display (5 and 10 fathoms), operating as described in Section 3.2, are selected by the range switch.

6. Trawler Fish Detection Equipment

The seabed-locked cathode-ray tube display described in Section 5 is part of a comprehensive equipment designed to meet the trawler fish detection requirements. It comprises a high power transmitter of 8 kW output feeding a 6λ by 4λ retractable transducer, a conventional recording echo-sounder with "white-line" facility and a seabed-locked recorder.²⁰

7. Conclusions

The extensive use of cathode-ray tube displays for trawler fish detection equipment indicates that the improved detection of fish close to the seabed which they provide is of value to the trawler skipper. Development of the seabed-locked display has removed the major shortcomings of current instruments and, in conjunction with improvements to other parts of the equipment, should increase the efficiency of fish detection.

A prototype equipment has been used on the Hull vessel M.T. *Westella* for the past two years. The seabed-locked cathode-ray tube display has operated satisfactorily in a wide variety of depths and weather conditions and the evaluation of displayed echoes has been made easier by the improved presentation.

8. Acknowledgments

Engineers who have made major contributions to the developments described in this paper are Mr. N. H. Benning (design of experimental apparatus), Mr. V. C. Buffery and Mr. D. J. Norman (mechanical design) and Mr. P. Evans (circuit development). The extensive experimental work carried out at sea during fishing operations was conducted by Mr. G. H. Ellis. The generous facilities provided by J. Marr & Sons Ltd. on board their trawlers are gratefully acknowledged.

The author is indebted to the Directors, Kelvin Hughes Division, S. Smith & Sons (England) Ltd., for permission to present this paper and would like to thank Mr. W. Halliday, Mr. S. C. Sparling, Dr. R. W. G. Haslett and Mr. G. H. Ellis for valuable discussions during its preparation.

9. References

1. O. Sund, "Echo sounding in fishery research", *Nature*, **135**, No. 3423, p. 953, 1935.
2. W. C. Hodgson and A. Fridriksson, "Report on echo-sounding and asdic for fishing purposes", *Rapp. Cons. Explor. Mer*, **139**, September 1955.
3. I. D. Richardson, D. H. Cushing, F. R. Harden Jones, R. J. H. Beverton and R. W. Blacker, "Echo Sounding Experiments in the Barents Sea", *Fishery Investigations, Series II*, **22**, No. 9, 1959 (H.M. Stationery Office).
4. S. C. Sparling, "Echo Sounder Engineering", Kelvin & Hughes Ltd., 1945.
5. D. G. Tucker, "Underwater echo-ranging", *J. Brit.I.R.E.*, **16**, p. 243, May 1956.
6. J. W. R. Griffiths and I. G. Morgan, "The chemical recorder", *Trans. Soc. Instrum. Tech.*, **8**, p. 62, June 1956.
7. D. G. Tucker, "Detection of pulse signals in noise: trace-to-trace correlation in visual displays", *J. Brit.I.R.E.*, **17**, p. 319, June 1957.
8. S. C. Sparling, R. W. G. Haslett, P. R. Hopkin and J. E. L. Sothcott, British Patent No. 865,111.
9. Atlas-Werke A. G., British Patent No. 758,881.
10. G. Klust, "Über Grossenschatzung von Schleppnetzfangen auf Grund der Echoanzeigen der Elac-Fischlupe", *Arch. Fischereiwiss.*, **3**, No. 3/4, p. 146, 1951.
11. R. W. Woodgate, "Recent Trends and Development in Echo Fishing" p. 488 "Modern Fishing Gear of the World". Ed. H. Kristjonsson, (Fishing News (Books) Ltd., London, 1959).
12. "Echo Fishing with the Kelvin Hughes 'Kingfisher.'" Kelvin Hughes Division, S. Smith & Sons (England) Ltd. Publication No. M.227.
13. D. H. Cushing, F. R. Harden Jones, R. B. Mitson, G. H. Ellis and G. Pearce, "Measurements of the target strength of fish", Sonar Systems Symposium paper, July 1962.
14. R. W. G. Haslett, "Determination of the acoustic back-scattering patterns and cross-sections of fish", *Brit. J. Appl. Phys.*, **13**, p. 349, July 1962.
15. W. Halliday and P. R. Hopkin, British Patent No. 785,001.
16. J. D. Richard, Jr., U.S. Patent No. 2,721,987.
17. D. G. Tucker, "Some new possibilities in civil underwater echo-ranging", *J. Brit.I.R.E.*, **20**, p. 299, April 1960.
18. P. R. Hopkin, British Patent No. 868,146.
19. R. L. Wallace, Jr., "The reproduction of magnetically recorded signals", *Bell Syst. Tech. J.*, **30**, p. 1145, October 1951.
20. R. W. G. Haslett, "A high-speed echo-sounder having seabed lock", Sonar System Symposium paper 15, July 1962, *J. Brit.I.R.E.*, **24**, No. 6, p. 441, December 1962.

*Manuscript received by the Institution on 24th June 1962.
(Paper No. 783/SS14.)*

© The British Institution of Radio Engineers, 1963

A Sub-carrier Regenerator Circuit for Colour Television Receivers

By

M. C. FRENCH (*Graduate*)†

Summary: The three principal circuits used to provide a continuous source of sub-carrier in an N.T.S.C. colour receiver are briefly described. The phase and frequency controlled crystal oscillator circuit, commonly called the automatic phase control circuit, is described in detail.

1. Introduction

In the N.T.S.C. system of colour television there must be provided in the receiver a source of continuous oscillation which is locked in frequency and phase to the sub-carrier burst signal received from the transmitter. This source of continuous oscillation, which may be called the regenerated sub-carrier, is used to operate the synchronous detectors which decode the colour information from the composite video signal.

The received sub-carrier burst signal has three basic characteristics all of which may vary with time. These are frequency, amplitude and phase.

Frequency variations are due to drift in the transmitter generating equipment. Obviously the sub-carrier regenerator must recognize and follow these frequency variations. Drifts in the receiver regenerator oscillator itself may also be included under the general heading of frequency variations and must be corrected.

Amplitude variations due to signal strength variations and other causes must not affect the output amplitude of the regenerator.

Phase variations are caused by noise in the incoming signal which results in changes in its instantaneous phase. It is most important that the phase of the regenerated sub-carrier should be related to the average phase of the incoming burst signal and not to its instantaneous phase, i.e. the sub-carrier regenerator must be insensitive to noise.

Changes in the phase of the regenerated sub-carrier relative to the burst sub-carrier signal will result in colour errors in the picture. It has been found by subjective tests that a phase error of about ± 5 deg still results in an acceptable colour picture. For this reason the total phase error in the regenerated sub-carrier, both static, due to drift in the regenerator and in the transmitted signal frequencies, and dynamic, due to noise, should not exceed ± 5 deg.

The minimum signal/noise ratio with which the regenerator should continue to operate satisfactorily is generally taken as being 1 : 1. At this level of noise other effects, such as loss of line synchronization and brightness variations, make correct functioning of the regenerator unimportant.

† Mullard Research Laboratories, Redhill, Surrey.

The regenerator circuit must pull in, i.e. phase lock the regenerator signal to the burst signal, in 1 second or less, so that when switching to a different channel there is no annoying delay before the colour picture is locked. Generally speaking pull-in time can be reduced at the expense of noise immunity. In designing a regenerator the noise immunity is made as good as is possible compatible with a pull-in time of less than 1 second. A thorough investigation of pull-in time vs. noise immunity for various regenerator circuits has been made by D. Richman.¹

There are three principal circuits which may be used to regenerate the sub-carrier. They are commonly known as the automatic phase control (a.p.c.) circuit, the passive integrator circuit and the d.c. quadricorrelator circuit.

The a.p.c. circuit consists of a crystal oscillator whose frequency and phase are controlled by a phase detector operating a control valve connected to the oscillator. Noise immunity is obtained by passing the output of the phase detector through a low pass filter before applying it to the control valve.

The passive integrator is a parallel resonant tuned circuit using a crystal which rings at the subcarrier frequency when the burst signal is applied. Drifts in the incoming burst signal frequency and in the crystal tuned circuit frequency are not compensated and can result in a large static phase error. Noise immunity is obtained by virtue of the narrow bandwidth of the crystal tuned circuit whose Q factor needs to be of the order of 30 000.¹ The static phase errors can be considerably reduced by the use of a control loop which is very similar to that used in the a.p.c. circuit. The need for a very high Q factor crystal still remains however and, further, the output amplitude of the circuit is not constant since, when the burst is absent, the oscillations in the crystal tuned circuit tend to die away. For these reasons the passive integrator circuit in either its simple or more complex form would appear to be inferior to the a.p.c. circuit.

The d.c. quadricorrelator circuit is an L-C oscillator with a two-mode control system. The L-C oscillator is controlled by the same type of loop as is used in the a.p.c. circuit. When the regenerator is not locked the bandwidth of the low-pass filter is considerably

increased by the bypass action of a second control loop. This results in a fast pull-in time and a phase lock which has poor noise immunity. When the circuit has locked the second control loop bypass action is inoperative and the main control loop operates with a narrow band filter that provides good noise immunity. This type of circuit thus achieves a good pull-in time for greater frequency errors than is possible with the a.p.c. circuit, hence the use of a less stable oscillator. The circuit is, however, rather more complex than the a.p.c. circuit. Since the pull-in time and noise immunity of the a.p.c. circuit can be quite adequate the additional circuit complexity is not justified.

From the foregoing discussion it would seem that the a.p.c. type of regenerator is the best choice. The remainder of this paper is therefore devoted to the design of an a.p.c. type of regenerator circuit.

2. Parameter Relationships in a Linear Circuit

If the circuits of the sub-carrier regenerator, illustrated in Fig. 1, have linear characteristics then a simple relationship between the circuit characteristics, the initial frequency error, and the final phase lock accuracy may be obtained.

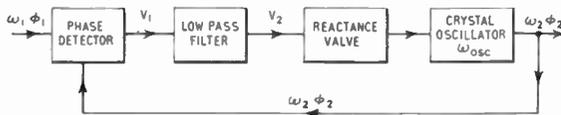


Fig. 1. Basic sub-carrier regenerator circuit.

With reference to Fig. 1,

$$\omega_2 = \omega_{osc} + 2\pi\beta V_2 \quad \dots\dots(1)$$

where β is the reactance valve control sensitivity in c/s per volt, and ω_{osc} is the free running frequency of the oscillator with zero control voltage on the reactance valve.

Also

$$V_1 = \mu t(\phi_2 - \phi_1) \quad \dots\dots(2)$$

where μ is the maximum output voltage of the phase detector, and $t(\phi_2 - \phi_1)$ is its normalized conversion characteristic. If the transfer coefficient of the low-pass filter is T then

$$V_2 = T V_1 = T\mu t(\phi_2 - \phi_1) \quad \dots\dots(3)$$

Now $\omega_1 = d\phi_1/dt$ and $\omega_2 = d\phi_2/dt$ hence, from eqn. (1),

$$d\phi_2/dt = \omega_{osc} + 2\pi\mu\beta T t(\phi_2 - \phi_1) \quad \dots\dots(4)$$

Adding $\omega_1 - d\phi_1/dt = 0$ to the left-hand side of eqn. (4) and re-arranging we have

$$\frac{d}{dt}(\phi_2 - \phi_1) - \omega_{osc} + \omega_1 = 2\pi\mu\beta T t(\phi_2 - \phi_1) \quad \dots\dots(5)$$

Let $\phi_2 - \phi_1 = \phi_D$
and $\omega_{osc} - \omega_1 = \omega_D$

Then substituting in (5)

$$\frac{d\phi_D}{dt} - \omega_D = 2\pi\mu\beta T t(\phi_D) \quad \dots\dots(5a)$$

This is the differential equation of the loop. When the oscillator is in lock $d\phi_D/dt$ is zero, and the transfer coefficient of the low-pass filter becomes unity. Under these conditions eqn. (5a) reduces to

$$-\omega_D = 2\pi\mu\beta t(\phi_D) \quad \dots\dots(6)$$

or $\mu\beta = -f_D/t(\phi_D) \quad \dots\dots(6a)$

2.1. Minimum Values for μ and β

In a practical circuit the oscillator is crystal controlled and the frequency difference, f_D , between the oscillator and the received burst signal is unlikely to exceed 200 c/s. A phase error ϕ_D of about 5 deg results in a picture colour error which is sufficiently small to be acceptable.

If these figures for f_D and ϕ_D are substituted in eqn. (6a) and $t(\phi_D)$, the conversion characteristic of the phase detector is assumed to be linear over the range 0° to $\pm 90^\circ$, then a figure for the product $\mu\beta$ may be obtained:

$$\mu\beta = \frac{200}{5/90} = 3600 \text{ c/s} \quad \dots\dots(7)$$

The reactance valve control sensitivity should be kept fairly low, say 100 c/s per volt, so that valve ageing and other circuit changes will not alter appreciably the free running frequency of the crystal oscillator. If $\beta = 100$, then from eqn. (7):

$$\mu = \frac{3600}{100} = 36 \text{ volts} \quad \dots\dots(7a)$$

These values for μ and β provide a starting point from which circuit design can be commenced, and give the minimum performance figures the circuits must achieve.

3. The Crystal Oscillator

Choice between the various types of crystal oscillator is governed by the need to meet the following four requirements:

- (a) Good frequency and amplitude stability.
- (b) Large output to supply the phase detector.
- (c) Good buffering to prevent contamination of the output by the burst signal.
- (d) Insensitivity to damping, so that a simple triode reactance valve may be used.

3.1. Types of Oscillator

There are four oscillators which fulfil some of the above requirements, though none of them can be said

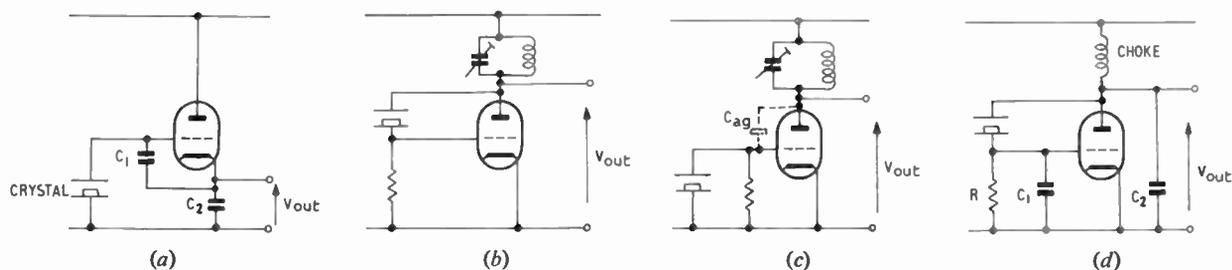


Fig. 2. Types of oscillator.

to meet them all. The basic circuits of these four oscillators are illustrated in Fig. 2.

Figure 2 (a) illustrates the Colpitts oscillator. The stability of this oscillator is very good due to the absence of any tuning inductance save that of the crystal. The output amplitude is sensibly constant with frequency changes, though it is rather small, and, even with a buffer amplifier, difficulty can be experienced in achieving sufficient drive to the phase detector. Since an amplifier is necessary, good buffering against burst contamination is obtained. In order to achieve sufficient frequency control the reactance valve would probably have to be connected across the crystal. This means heavy damping could not be tolerated, and a pentode reactance valve would be necessary.

The oscillators illustrated in Figs. 2(b) and (c) both give large amplitude outputs. This is an attractive feature since the phase detector could be driven directly and an oscillator amplifier would not be needed. A buffer, which is reasonably effective, could be provided between the oscillator output, and the phase detector by the use of a pentode valve. The screen grid circuit could then be used to form the oscillator, with the anode circuit providing a buffered output to drive the phase detector.

Unfortunately the frequency and amplitude stability of these circuits is poor. Small changes in the component values of the anode tuned circuits produce large changes in the frequency and amplitude of the output. Two sketch graphs (Fig. 3) illustrate this point. Figure 3 (a) which is applicable to the circuit of Fig. 2(b) shows that as the capacitance of the anode tuned circuit is reduced, both the frequency and amplitude of the output increase. Oscillation ceases when the capacitance is reduced below a certain critical value. With the circuit shown in Fig. 2(c) a similar effect occurs. Increasing the capacitance in the anode tuned circuit produces an increase in the amplitude, and a decrease in the frequency of the output. This is shown in Fig. 3 (b). Once again at a certain critical value of anode tuned circuit capacitance oscillation suddenly stops.

The dotted curves in the sketch graphs of Fig. 3 show the effect of increasing the damping in the grid circuit. Increasing the capacitance has a similar effect. This is unfortunate because as the reactance valve alters the frequency of operation, so also the amplitude is altered, and to a considerable extent. In order to achieve a large output the anode tuned circuit capacitance must be adjusted unpleasantly near to the critical value at which oscillation ceases.

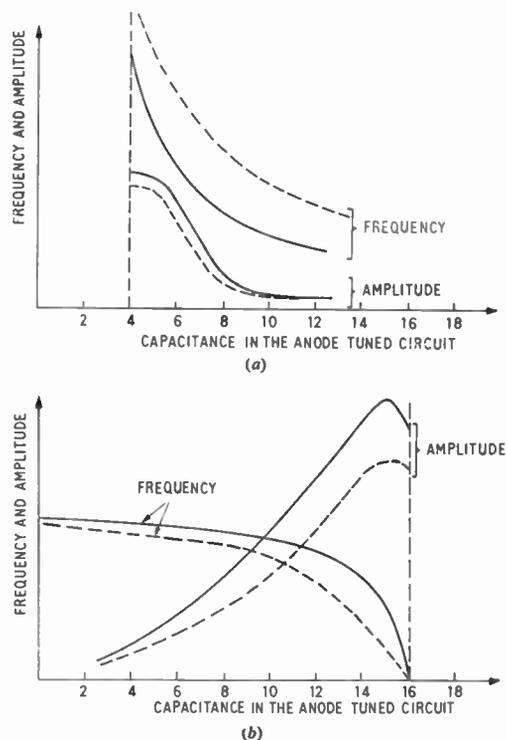


Fig. 3. Amplitude and frequency responses vs. anode tuned circuit capacitance for the oscillator circuits shown in Fig. 2 (b) and Fig. 2 (c).

Although the large output of these two circuits is so attractive it is felt that their poor frequency and amplitude stability make them too unreliable for use in a regenerator.

An improvement in the Colpitts circuit, illustrated in Fig. 2 (a), may be achieved by re-arrangement of the

circuit in the manner shown in Fig. 2 (d). A choke is now needed in the anode circuit. Provided however, its impedance is high compared with that of C_2 the stability of the circuit will be unimpaired. The capacitor C_1 may now be constituted by the reactance valve. The damping of the reactance valve in this position, represented by R in Fig. 2 (d) is not directly across the tuned circuit, and the use of a triode reactance valve is possible. The feedback capacitor C_2 may be reduced and a larger output voltage obtained, again due to reduced damping effects. With the addition of an output amplifier, which provides buffering between the oscillator output and the phase detector, the circuit can be said to fulfil the requirements of the regenerator oscillator.

From among the oscillators which have been discussed this variation of the Colpitts circuit is considered to be the best choice, and is therefore used in the regenerator. Derivation of the oscillation frequency and maintenance condition is useful since it enables the effect of the reactance valve and component variations to be estimated.

3.2. Oscillation Frequency and Maintenance Condition

Considering the general case for a triode valve, illustrated in Fig. 4, and applying four terminal network admittance parameters, the equations are

$$I_1 = (Y_G + Y_F)V_1 - Y_F V_2 \quad \dots\dots(8)$$

$$I_2 = (g_m - Y_F)V_1 + \left(Y_F + Y_L + \frac{1}{r_a}\right)V_2 \quad \dots\dots(9)$$

Y_G , Y_F and Y_L are the admittances between the valve electrodes, g_m and r_a are respectively the mutual conductance and anode resistance of the valve.

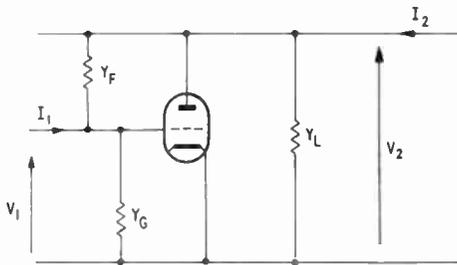


Fig. 4. General four-terminal network for a triode valve.

For oscillation $I_1 = I_2 = 0$ and eqns. (8) and (9) yield

$$\frac{Y_F}{Y_G + Y_F} = \frac{-\left(Y_F + Y_L + \frac{1}{r_a}\right)}{g_m - Y_F} \quad \dots\dots(10)$$

Therefore

$$g_m = -Y_G - \frac{Y_G Y_L}{Y_F} - \frac{Y_G}{Y_F r_a} - Y_L - \frac{1}{r_a} \quad \dots\dots(10a)$$

In the oscillator circuit chosen for the regenerator, and illustrated in Fig. 2 (d)

$$Y_G = j\omega C_1 + 1/R$$

$Y_L = j\omega C_2$ (this neglects the inductive reactance of the choke, and includes its self capacitance together with strays in C_2)

$Y_F = \frac{1}{j\omega L}$ (where L is the effective inductance of the crystal which is estimated in Section 3.4).

Substituting these values in eqn. (10a) gives

$$g_m = -j\omega C_1 - 1/R + (\omega^2 C_1 C_2 - j\omega C_2/R)j\omega L - (j\omega C_1/r_a + 1/Rr_a)j\omega L - j\omega C_2 - 1/r_a \quad \dots\dots(11)$$

$$g_m = -j\omega C_1 - 1/R + j\omega^3 LC_1 C_2 + \omega^2 LC_2/R + \omega^2 LC_1/r_a - j\omega L/Rr_a - j\omega C_2 - 1/r_a \quad \dots\dots(11a)$$

Taking the imaginary terms and equating them to zero:

$$-\omega C_1 + \omega^3 LC_1 C_2 - \omega L/Rr_a - \omega C_2 = 0 \quad \dots\dots(12)$$

$$\omega^2 = \frac{C_1 + C_2 + L/Rr_a}{LC_1 C_2} \quad \dots\dots(12a)$$

Taking the real terms of eqn. (11a) and substituting for ω^2 :

$$g_m = -1/R + LC_2/R \frac{(C_1 + C_2 + L/Rr_a)}{LC_1 C_2} + LC_1/r_a \frac{(C_1 + C_2 + L/Rr_a)}{LC_1 C_2} - 1/r_a \quad \dots\dots(13)$$

which gives

$$g_m = \frac{r_a C_2^2 + LC_2/R + RC_1^2 + LC_1/r_a}{C_1 C_2 Rr_a} \quad \dots\dots(13a)$$

Since $\mu = g_m r_a$ where μ is the amplification factor of the valve

$$\mu = \frac{r_a C_2^2 + LC_2/R + RC_1^2 + LC_1/r_a}{C_1 C_2 R} \quad \dots\dots(14)$$

From this analysis are obtained two equations, namely (12a) and (14), which will be of assistance in assessing the ability of the oscillator to fulfil its function in the regenerator. Before actual values from a circuit can be substituted a crystal must be chosen and a figure for L (its equivalent inductance) must be obtained.

3.3. Choice of Crystal

The crystal used in the reactance valve, oscillator, combination should be so chosen that the following three requirements are complied with:

- (a) With the reactance valve set to the mid-point of its control characteristic the oscillator frequency shall be that of the nominal subcarrier frequency of 2 657 812 c/s.

- (b) With the reactance valve set to either extremity of its control characteristic the frequency change produced in the oscillator shall be ± 300 c/s. (This provides a generous excess over the likely maximum value of f_D mentioned in section 2.1.)
- (c) The reactance valve control sensitivity β shall be about 100 c/s per volt.

Since considerable control in meeting these requirements is available in the design of the reactance valve, and to some extent in the oscillator, the limitations placed on the selection of the crystal are not so stringent as they might at first appear.

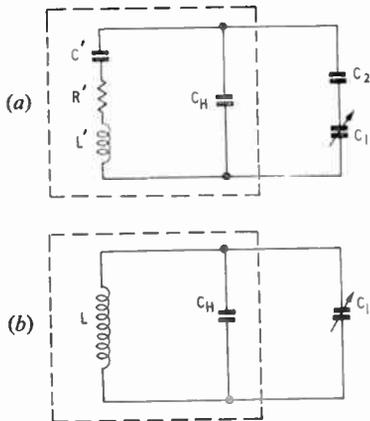


Fig. 5 (a). Simplified equivalent circuit for oscillator tuned circuit components.
(b) Further simplification.

Figure 5 (a) shows a simplified equivalent circuit for the oscillator tuned circuit components. Those within the dotted lines represent the crystal with its holder capacitance C_H . C_1 and C_2 are the tuning capacitances of the oscillator circuit. The capacitance C_1 is constituted by the reactance valve, circuit strays, and the input capacitance of the oscillator valve. The reactance valve element of this capacitance is variable, and so C_1 is shown as a variable capacitance. At frequencies above the crystal series resonant frequency the circuit may be further simplified as shown in Fig. 5 (b). The crystal components C' , and L' may be replaced by an inductance L where

$$L = L' - 1/\omega^2 C' \quad \dots\dots(15)$$

Since the Q factor of the crystal is high the resistance R' is ignored. The ratio of $C_2 : C_1$, which is determined by the feedback necessary to maintain oscillation, is about 10 : 1 in a practical regenerator oscillator. The capacitance C_2 may therefore also be ignored.

For a crystal with a low series resonant frequency the inductance L will be large at the parallel resonant

frequency, which must occur at the subcarrier frequency. Hence the capacitance C_1 must be small. A crystal with a high series resonant frequency will have a small value of inductance L at the parallel resonant frequency. Hence the capacitance C_1 must be large. Further for a given change in the capacitance C_1 , produced by the reactance valve, β will be proportional to the frequency separation between the series and parallel resonant frequencies of the crystal. Provided the crystal chosen results in a value for C_1 which can be readily achieved with a reactance valve and also gives a suitable value for β , then satisfactory circuit design is possible.

The crystal used in the subcarrier regenerator tunes to parallel resonance at the subcarrier frequency when the added shunt capacitance is 20 pF. Series resonance occurs approximately 1 kc/s below the parallel resonant frequency.

3.4. Equivalent Crystal Inductance

The crystal equivalent inductance L (Fig. 5(b)) at parallel resonance may be calculated approximately if the holder capacitance C_H is neglected. At the series resonant frequency of the crystal the equivalent inductance L will be zero. If between these two frequencies L is assumed to vary linearly with frequency, a simple relationship between L and frequency may be obtained.

Value of L at parallel resonance:

$$L = \frac{1}{\omega^2 C_1} = \frac{1}{4\pi^2(2657812)^2 20 \times 10^{-12}} \dots\dots(16)$$

$$L = 177 \mu\text{H} \quad \dots\dots(16a)$$

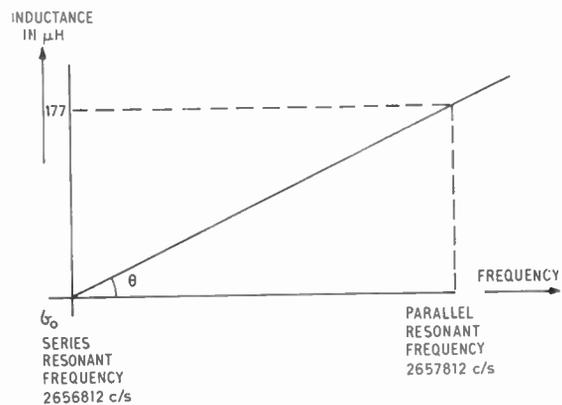


Fig. 6. Equivalent inductance of crystal.

With reference to Fig. 6:

$$\tan \theta = 177/1000 = 0.177$$

$$(f - f_0) \tan \theta = 0.177(f - f_0) = L \text{ in } \mu\text{H} \quad \dots\dots(17)$$

where f is the frequency in c/s, and f_0 is the series resonant frequency of the crystal.

3.5. The Oscillator Circuit

A suitable valve, for both the oscillator and reactance valve stages, is an ECC81 double triode. A reactance valve using half the double triode results in a tolerable degree of damping in the oscillator grid circuit. The valve amplification factor of 60 ensures the reliable maintenance of oscillation.

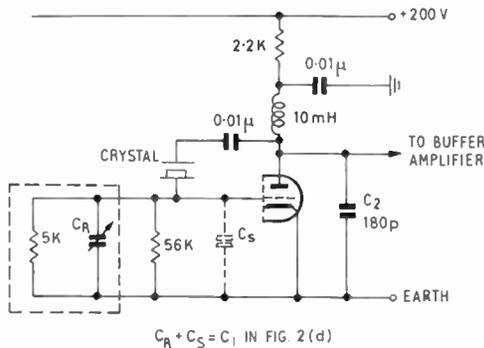


Fig. 7. Oscillator circuit and equivalent reactance valve components.

A circuit diagram of the oscillator is shown in Fig. 7. The two components inside the dotted lines represent the reactance valve equivalent circuit. Justification for this will be given in Section 4.4. The oscillator input capacitance, and strays are shown as C_S . Decoupling is provided in the anode circuit to prevent the oscillator signal reaching the h.t. line.

Using eqn. (14) derived in section 3.2, the maintenance condition for this oscillator will be determined.

$$\mu \simeq \frac{10^4 \cdot 4 \cdot 10^4 \cdot 10^{-24}}{20 \cdot 200 \cdot 10^{-24} \cdot 10^4 \cdot 0.5} \dots\dots(18)$$

$$\mu \simeq 20 \dots\dots(18a)$$

For the purpose of this calculation all the terms in the numerator of eqn. (14) may be neglected except for the term $r_a C_2^2$. The capacitances C_1 and C_2 , and the resistance R are assumed to be 20 pF, 200 pF and 5 kΩ respectively. Since the valve has a μ of 60 there is an adequate safety margin.

The effect of changes in the oscillator circuit components may be investigated using eqn. (12a) in Section 3.2. The term L/Rr_a is approximately 2% of the numerator. Changes in the r_a of the oscillator valve and in the damping produced by the reactance valve will therefore have only a small effect on the frequency of the oscillator. This term can therefore be neglected, and eqn. (12a) becomes

$$\omega^2 = \frac{C_1 + C_2}{LC_1 C_2} \dots\dots(19)$$

Substituting the value obtained for L in eqn. (17) (Sect. 3.4.):

$$\omega^2 = \frac{C_1 + C_2}{[0.177 \cdot 10^{-6} \cdot (f - f_0)] C_1 C_2} \dots\dots(20)$$

Let $0.177 \times 10^{-6} = K$ then:

$$4\pi^2 f^2 = \frac{C_1 + C_2}{K C_1 C_2 (f - f_0)} \dots\dots(21)$$

$$f^3 - f^2 f_0 = \frac{1}{4\pi^2 K} \left[\frac{1}{C_1} + \frac{1}{C_2} \right] \dots\dots(22)$$

Differentiating with respect to C_1 :

$$\frac{df}{dC_1} (3f^2 - 2ff_0) = - \frac{1}{4\pi^2 K C_1^2} \dots\dots(23)$$

Substituting for $1/C_1^2$ from eqn. (22):

$$\frac{df}{dC_1} = - \frac{[4\pi^2 f^2 K (f - f_0) - 1/C_2]^2}{(4\pi^2 K)(3f^2 - 2ff_0)} \dots\dots(24)$$

Similarly

$$\frac{df}{dC_2} = - \frac{[4\pi^2 f^2 K (f - f_0) - 1/C_1]^2}{(4\pi^2 K)(3f^2 - 2ff_0)} \dots\dots(25)$$

Substitution in eqn. (24) for $f = 2657812$ c/s, i.e. the sub-carrier frequency, gives

$$\frac{df}{dC_1} \simeq 40 \text{ c/s per pF} \dots\dots(26)$$

and the same substitution in eqn. (25) gives

$$\frac{df}{dC_2} \simeq 0.01 \text{ c/s per pF} \dots\dots(27)$$

From eqn. (26) the capacitance change in the reactance valve to produce a change in the oscillator frequency of ± 300 c/s can be calculated

$$\Delta C_1 = \pm 300/40 \text{ pF} = \pm 7.5 \text{ pF} \dots\dots(28)$$

Variations in C_2 are shown by eqn. (27) to have a small effect on the oscillator frequency. The self-capacitance of the choke in the anode circuit, the input capacitance to the buffer amplifier, and the input capacitance of circuits driven from the oscillator, will not therefore affect materially the stability. Changes in the anode to grid capacitance of the oscillator valve and circuit will have an effect on the oscillator frequency similar to that of C_1 , i.e. about 40 c/s per pF. Care should therefore be taken to keep this capacitance small and constant.

4. The Reactance Valve

The basic circuit of the reactance valve, with which the oscillator is designed to operate, is illustrated in Fig. 8 (a). To assist in the design of the circuit, and to enable the equivalent damping resistance of the final circuit to be calculated it will be useful to determine the impedance Z_{in} presented to the oscillator by the reactance valve.

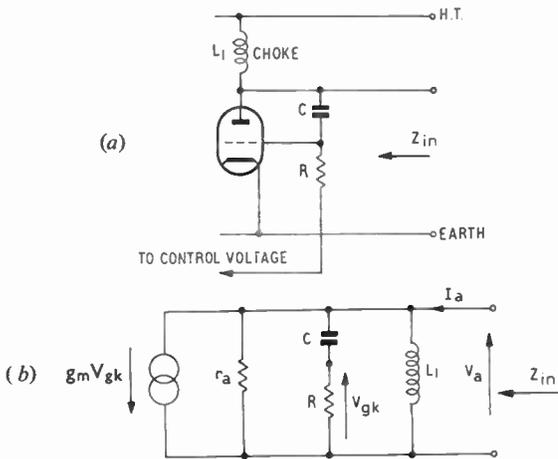


Fig. 8. Triode reactance valve.
(a) Simplified circuit. (b) Equivalent circuit.

4.1. Equivalent Impedance

Figure 8 (b) shows the equivalent circuit of the triode reactance valve. The admittance $1/Z_{in}$ seen by the oscillator is:

$$\frac{I_a}{V_a} = \frac{1}{j\omega L_1} + \frac{1}{r_a} + \frac{1}{R + 1/j\omega C} + \frac{g_m V_g}{V_a} \dots\dots(29)$$

but $V_g = \frac{R V_a}{R + 1/j\omega C}$

therefore

$$\frac{I_a}{V_a} = \frac{1}{j\omega L_1} + \frac{1}{r_a} + \frac{1 + g_m R}{R + 1/j\omega C} \dots\dots(30)$$

$$\frac{I_a}{V_a} = \frac{1}{j\omega L_1} + \frac{1}{r_a} + \frac{(1 + g_m R)(R - 1/j\omega C)}{R^2 + 1/\omega^2 C^2} \dots\dots(31)$$

For satisfactory operation in a practical circuit $R^2 \ll 1/\omega^2 C^2$, i.e. the phase angle between the anode and grid voltages should approach 90 deg.

Hence

$$\frac{I_a}{V_a} \approx \frac{1}{j\omega L_1} + \frac{1}{r_a} + \omega^2 C^2 R + \omega^2 C^2 R^2 g_m + j\omega(C + g_m CR) \dots\dots(32)$$

From eqn. (32) it can be seen that the impedance Z_{in} presented to the oscillator may be represented by the circuit shown in Fig. 9.

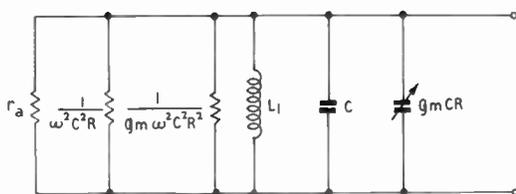


Fig. 9. Representation of the impedance Z_{in} presented to the oscillator by the reactance valve.

4.2. Method used to Design the Circuit

Design of a circuit having the desired control characteristic (oscillator frequency vs. control voltage), by consideration of the equivalent impedance, Z_{in} shown in Fig. 9, is difficult. Curvature in the valve characteristics, stray capacitances, valve grid current flowing through the high source impedance of the control voltage, are some of the factors which make an empirical approach more suitable.

The circuit of the reactance valve used in the regenerator was therefore arrived at by first designing a circuit from consideration of the equivalent impedance shown in Fig. 9. A curve of the control characteristic was then plotted, and the component values of the circuit adjusted to achieve the desired characteristic.

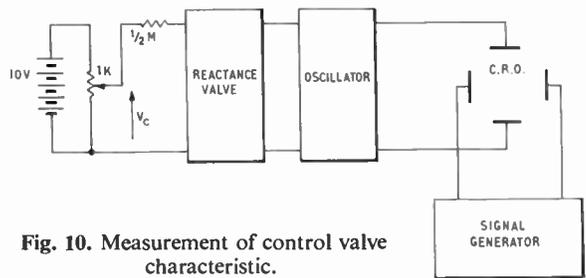


Fig. 10. Measurement of control valve characteristic.

The method employed to measure the control characteristic is illustrated in Fig. 10. The control voltage V_c is derived from a battery, and fed to the reactance valve via a $\frac{1}{2} M\Omega$ resistor, which represents the output resistance of the phase detector from which the control voltage is derived in the complete regenerator. The frequency of the oscillator is measured by comparing it with the output of a signal generator on a cathode ray oscilloscope.

4.3. The Circuit used in the Regenerator

Figure 11 illustrates the circuit used in the regenerator. The product CR , for this circuit, of 4.5×10^{-9} seconds enables sufficient frequency control of the oscillator to be achieved. The individual values for C and R are determined by two factors.

- (i) If the capacitance C is large, and hence R is small, then difficulty in achieving an increase in the oscillator frequency of 300 c/s can be experienced. This is because even when the reactance valve is cut off, and $g_m CR$ is zero, there still remains the capacitance C plus circuit strays in the grid circuit of the oscillator.
- (ii) If the resistance R is large, and hence C is small, then the reactance of the grid to cathode capacitance of the reactance valve cannot be ignored, and results in a reduction of the phase angle between the anode and grid voltages

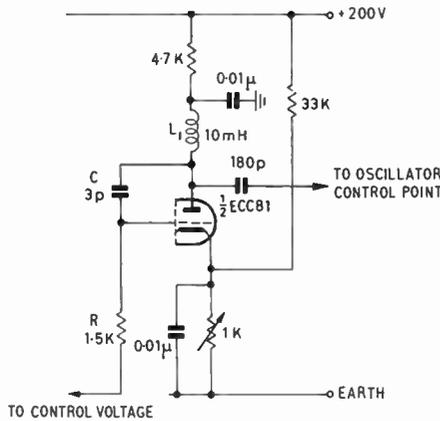


Fig. 11. Reactance valve circuit.

which increases the damping offered by the reactance valve. The values used in this circuit of 3 pF and 1.5 kΩ avoid these effects.

The value of the resistance between the cathode and earth controls the value of β , and is such as to make β about 80 c/s per volt in this circuit. The resistor from the cathode to h.t. enables the valve to be biased so that, with zero control voltage, the oscillator operates at the subcarrier frequency. To allow initial setting of the oscillator to the subcarrier frequency the cathode resistor is variable. The h.t. dropping resistance in the anode is as small as valve dissipation will permit, so that the maximum change in g_m is obtained. The cathode resistor is decoupled at the oscillator frequency for the same reason. As can be seen from Fig. 9 the inductance L_1 must have a high impedance at the oscillator frequency, and its self capacitance should be small.

The control characteristic is illustrated in Fig. 12. The characteristic is slightly unsymmetrical so that as

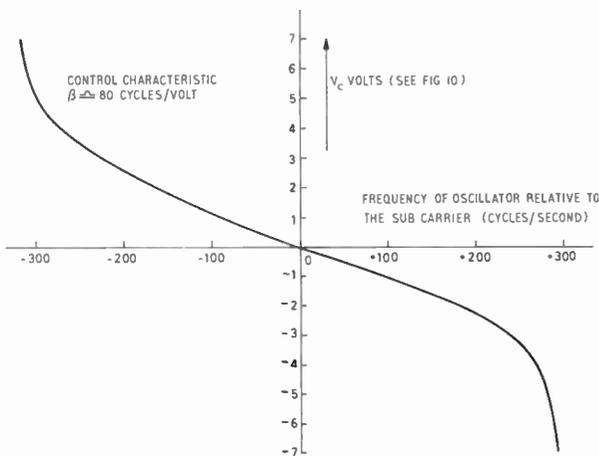


Fig. 12. Reactance valve control characteristic.

the g_m of the valve decreases with age the lower frequency control limit of -300 c/s is maintained.

It was assumed when dealing with the oscillator that the resistive damping offered by the reactance valve was 5 kΩ. This assumption will now be investigated.

4.4. The Equivalent Damping Resistance

Figure 9 shows that there are two damping terms to be evaluated. Substituting for the component values of the circuit in these two terms gives:

$$\frac{1}{\omega^2 C^2 R} \approx 120 \text{ k}\Omega \quad \dots\dots(33)$$

and,

$$\frac{1}{g_m \omega^2 C^2 R^2} \approx 11 \text{ k}\Omega \quad \dots\dots(34)$$

In these two calculations the value of C must include both the circuit capacitance of 3 pF and the valve capacitance of 1.6 pF. The g_m of the valve was assumed to be 7 mA/volt, and the frequency to be the subcarrier frequency of 2.66 Mc/s. The lowest value for the r_a of the valve is 10 kΩ. The minimum equivalent resistance of the reactance valve is therefore about 5 kΩ, and justifies the value assumed when calculating the maintenance condition for the oscillator.

5. The Phase Detector

The purpose of the phase detector is to provide a d.c. correction voltage to the reactance valve, dependent upon the phase error between the oscillator and sub-carrier burst signals, when the regenerator system is in lock. It must also be able to provide a correction voltage when there is a frequency error between the two signals, to enable pulling in to take place.

5.1. Principle of Operation in the Regenerator Circuit

The phase detector, illustrated in Fig. 13, adds vectorially one signal input with each of two antiphase components of the other signal. The two vector sums produced are each peak detected giving d.c. voltages proportional to their peak amplitude.

If one signal is in phase quadrature with each of the two antiphase components of the other, both the peak detectors will give equal outputs. If the two outputs are arranged to be of opposite polarity, then the final output obtained by summing the two outputs will be zero.

Any departure from the phase quadrature condition will increase the output from one of the peak detectors, and decrease the output from the other. The final output will not then be zero, but will have a magnitude dependent upon the departure from the phase quadrature condition, and a polarity indicative of whether a phase advance or lag has taken place.

When the two inputs are at different frequencies the output from the phase detector will vary alternately

positive and negative, since a continual change in the phase error is taking place. The frequency of the output will be the frequency difference between the two inputs.

At first sight it appears that the phase detector is unable to provide a d.c. correction voltage when a frequency error exists. This is indeed the case when the phase detector is considered by itself. When, however, it is considered as part of the subcarrier regenerator circuit, where its output voltage is controlling the oscillator frequency via the reactance valve and the low-pass filter, its output voltage is modified in such a manner as to produce a correction voltage which tends to reduce the frequency error.

The exact manner in which this happens is complicated, even if simple characteristics are assumed. In the case of the regenerator circuit described in this paper, where the output from the phase detector is sufficient to drive the reactance valve well beyond the limit of its control range, an attempt to give a full solution is impractical.

A simple explanation may be given by saying that if a frequency error exists, then there will be an alternating output voltage from the phase detector. This voltage will be fed to the reactance valve via the low pass filter, and will alter the frequency of the oscillator.

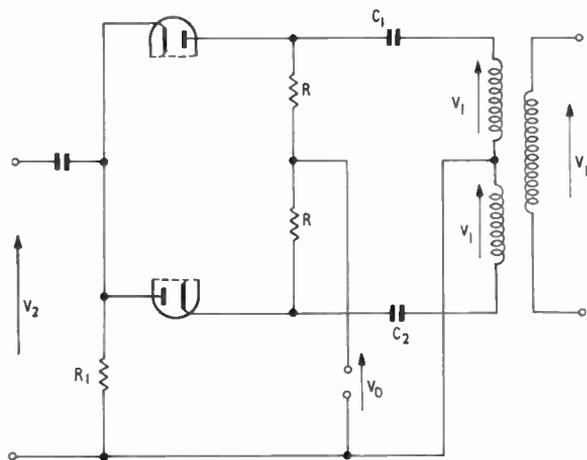


Fig. 13. Basic circuit of phase detector.

When the phase detector output polarity is such that the frequency error is decreased, the time taken for the phase detector output voltage to fall to zero will be greater than when the correction voltage is increasing the frequency error. The output voltage of the phase detector will therefore be modified by the action of the regenerator circuit, and will become asymmetrical, and hence will have a d.c. component. The mean potential across the capacitors of the low-pass filter will tend to increase, and reduce the frequency error.

The action is cumulative so that the frequency error will eventually be reduced to zero. The time taken to accomplish this will be determined by the bandwidth of the low-pass filter, the output amplitude of the phase detector, and the control range and sensitivity of the reactance valve.

5.2. Phase Detector Characteristics

The relationship between the phase difference of the input signals and the output voltage of the phase detector, which may be called its conversion characteristic, is affected by the relative amplitude of the two input signals V_1 and V_2 .

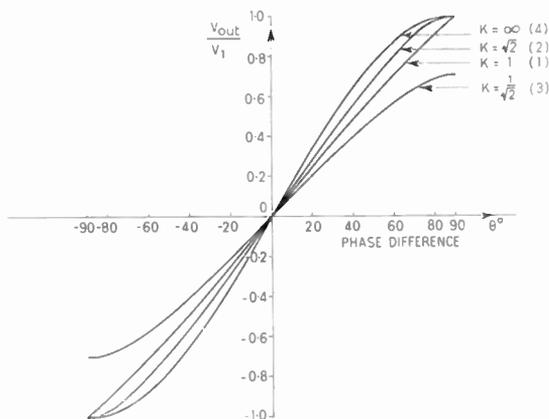


Fig. 14. Phase detector conversion characteristics.

Figure 14 illustrates the conversion characteristic of the phase detector for various values of K , which is the ratio of V_2/V_1 . When $V_2 \gg V_1$, i.e. $K \rightarrow \infty$, the output of the phase detector is a sinusoidal function of the phase difference, with a maximum value equal to V_1 . If $V_2 = V_1$, i.e. $K = 1$, the characteristic is nearly linear, and the maximum output is equal to V_1 and V_2 . As K becomes less than 1 the characteristic tends to become sinusoidal again, but the maximum output is now dependent upon V_2 . The maximum output is always dependent upon the smaller of the two inputs. The conversion characteristic varies from sinusoidal when one input is much larger than the other to an approximately linear function when the two inputs are of equal amplitude.

The derivation of the phase detector conversion characteristics is given in Appendix 1.

5.3. Choice of Input Arrangements

In some regenerator phase detectors V_1 is supplied by the burst signal, and V_2 , which is supplied by the oscillator is assumed to be large compared with V_1 . Under these conditions the conversion characteristic of the phase detector is sinusoidal, and the maximum output is dependent upon the amplitude of the burst signal V_2 .

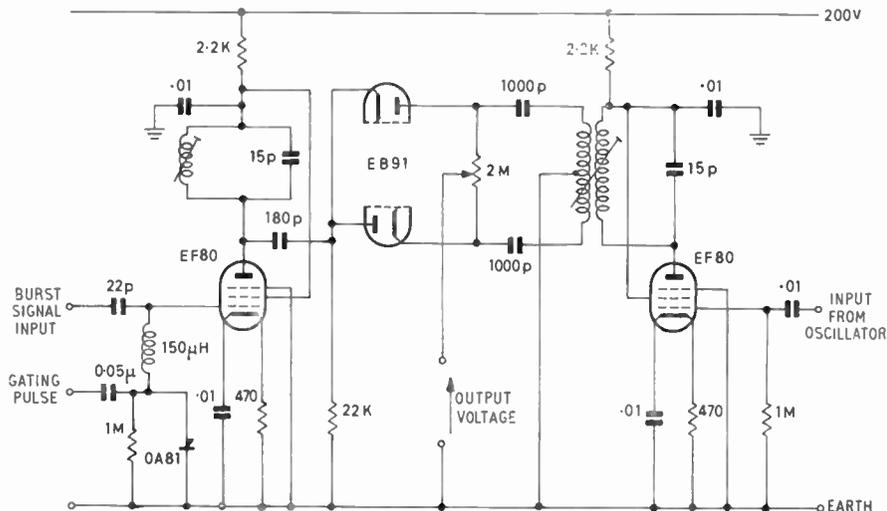


Fig. 15. Circuit of the phase detector and its associated input amplifiers.

In the phase detector of the sub-carrier regenerator described in this paper, V_1 is supplied by the oscillator, and V_2 by the burst signal. There are two advantages in this arrangement. Firstly the burst signal supplies only one input, V_2 , instead of the two antiphase components of V_1 , and it is possible to achieve a larger burst signal in the phase detector. Secondly, when V_1 is supplied by the burst signal the two peak detector capacitors, C_1 and C_2 , are charged via R_1 and the forward resistance of the diodes; whereas when V_2 is supplied by the burst signal the capacitors are charged via the forward resistance of the diodes and the resistance of the transformer windings only. The burst signal is present for only 9 cycles per line period, so the reduction in the charging path resistance is important, as it enables higher peak detector voltages to be achieved. The oscillator signal is present continuously so the resistance R_1 in the charging path for the oscillator signal can be tolerated.

5.4. The Circuit of the Phase Detector and its Associated Input Amplifiers

Figure 15 shows the circuit of the phase detector and its associated input amplifiers.

The amplifiers are of the simple tuned anode type. The grid of the burst amplifier is gated so that the amplifier is only operative when the burst signal is present; this prevents noise and spurious signals reaching the phase detector. The gating pulses are derived in the regenerator circuit from a mono-stable flip-flop which is triggered from the line sync pulse. In a complete receiver the gating pulse could be derived from the line flyback pulse, and the flip-flop would not be needed. The Q factor of the anode tuned circuits should not be greater than about 15, otherwise phase

errors will be introduced, by drift in tuned circuit components, which will not be corrected by the normal action of the regenerator circuit.

The 2 M Ω potentiometer, from which the final output of the phase detector is obtained, enables the outputs of the two peak detectors to be balanced. The 1000pF capacitors of the peak detectors are a suitable compromise between the small time constant required for the charging circuit, and the large time constant required for the discharge circuit.

6. The Low-pass Filter

The purpose of the low-pass filter is to integrate the output from the phase detector, so that information derived from the burst signal is stored during the line period. It further serves to eliminate noise by limiting the bandwidth of the phase detector output voltage which reaches the reactance valve. Detailed design data have been produced by D. Richman.¹

6.1. Description of the Filter

The filter used in this regenerator has a bandwidth which is narrow enough to ensure reliable operation with poor signal-to-noise ratios without impairing the ability of the circuit to pull in from an initial frequency error in an acceptably short time.

The graph of Fig. 16 illustrates the transfer characteristic of the filter for sinusoidal input voltages. R_1 in Fig. 17 represents the output resistance of the phase detector.

At zero frequency, i.e. d.c., the insertion loss of the filter is zero, and the full output voltage of the phase detector reaches the reactance valve, enabling an accurate phase lock to be obtained.

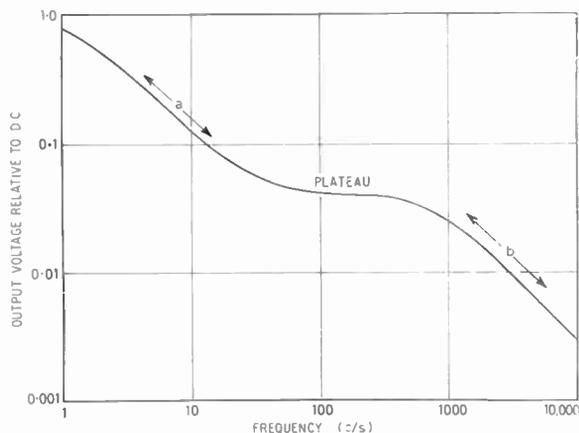


Fig. 16. Low-pass filter characteristic for sinusoidal input.

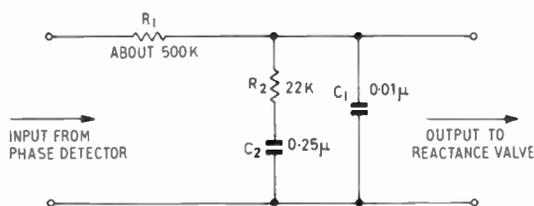


Fig. 17. Low-pass filter circuit.

As the frequency is increased the insertion loss of the filter increases rapidly to about 28 dB at 100 c/s, and then remains roughly constant up to about 300 c/s. The rate and extent of this increase in attenuation is determined by R_2 and C_2 in conjunction with R_1 , and is chosen so that the pull-in time of the regenerator is kept below 1 second.

At frequencies above 300 c/s the regenerator will not be able to pull in because the maximum frequency

deviation of the oscillator is ± 300 c/s, so the capacitor C_1 is used to further increase the insertion loss of the filter at these frequencies.

7. Results

The circuit diagram of the complete regenerator is shown in Fig. 18.

The circuit will lock the phase of the crystal oscillator to the burst signal with an accuracy of better than ± 5 deg, with a stabilization time of less than 1 second, for initial frequency differences between the burst and oscillator signals of ± 250 c/s. For full performance the burst signal input should be not less than 4 volts peak-to-peak amplitude. The circuit is insensitive to noise, and will remain in phase lock when the noise is such as to make the picture quality quite unacceptable.

The phase detector produces an output of approximately ± 45 volts, and its transfer characteristic follows approximately the $K = 1$ characteristic.

8. Conclusions

The chief criteria for the satisfactory design of a sub-carrier regenerator circuit for a colour television receiver of the N.T.S.C. type have been considered. A complete circuit, whose performance may be considered satisfactory, has been suggested.

Further efforts to improve the overall performance of this circuit would be best directed to increasing the phase detector output. Three valve envelopes are used in the suggested regenerator (excluding the gate pulse generator), but this number could be reduced to two by the use of two triode-pentodes instead of the double triode and two pentodes. Experiments in this direction led to difficulties due to interaction between valves in the same envelope, which resulted in contamination of the oscillator output with the burst signal, and frequency modulation of the oscillator by signals of

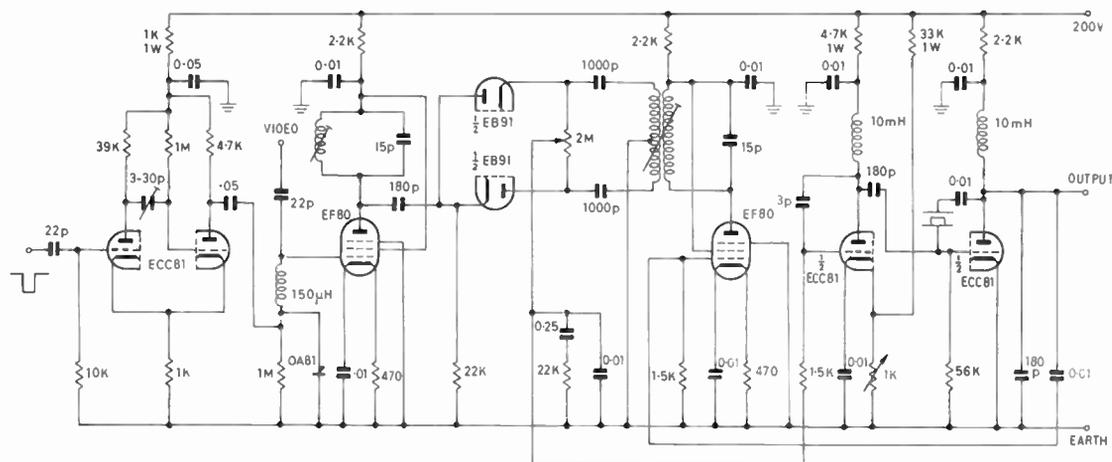


Fig. 18. Sub-carrier regenerator.

sub-carrier frequency reaching the control grid of the reactance valve.

The performance of the oscillator is satisfactory and little further improvement can be made. The reactance valve is of the simplest and most economic type and its performance is adequate. The use of a reversed biased diode instead of a reactance valve was considered, but at the present time no suitable diode is readily available.

9. Acknowledgments

The author would like to thank the Director of Mullard Research Laboratories and the Directors of Mullard Limited for permission to publish this paper.

10. References and Bibliography

1. D. Richman, "Colour carrier reference phase synchronization accuracy in N.T.S.C. colour television", *Proc. Inst. Radio Engrs*, 42, pp. 106-133, January 1954.
2. Hazeltine Staff, "Principles of Color Television". (John Wiley, New York, Chapman & Hall, London, 1956.)
3. D. G. Fink, "Television Engineering Handbook". (McGraw-Hill, New York, 1957.)
4. H. A. Thomas, "Theory & Design of Valve Oscillators", Monographs on Electrical Engineering. (Chapman & Hall, London, 1951.)
5. S. Krishnan, "Diode phase detectors", *Electronic Radio Engr*, 36, pp. 45-50, February 1959.
6. J. E. P. Vigoureux, "The valve maintained quartz oscillator", *J. Instn Elect. Engrs*, 68, pp. 265-95, February 1930.

11. Appendix 1:

Derivation of Phase Detector Characteristics

Consider the vector diagram Fig. 19 illustrating the signal inputs arrangements in a phase detector when the two input signals are at the same frequency but at differing phase angles. The two antiphase components of one input signal are represented by V_1 , and V_2 represents the other input signal. The output

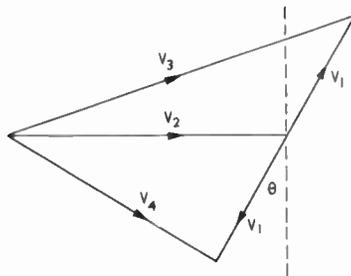


Fig. 19. Signal input arrangements in the phase detector.

voltages of the two peak detectors are represented by V_3 and V_4 . The angle θ shows the departure from the balanced condition when $V_3 = V_4$.

The output voltages of the peak detectors are

$$V_3 = [V_1^2 + V_2^2 - 2V_1V_2 \cos(90 + \theta)]^{\frac{1}{2}} \dots\dots(35)$$

$$V_4 = [V_1^2 + V_2^2 + 2V_1V_2 \sin \theta]^{\frac{1}{2}} \dots\dots(35a)$$

and:

$$V_4 = [V_1^2 + V_2^2 - 2V_1V_2 \cos(90 - \theta)]^{\frac{1}{2}} \dots\dots(36)$$

$$V_4 = [V_1^2 + V_2^2 - 2V_1V_2 \sin \theta]^{\frac{1}{2}} \dots\dots(36a)$$

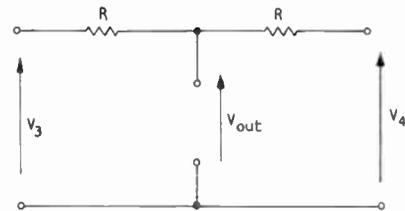


Fig. 20. Resistive summing network of the phase detector.

The outputs of the peak detectors are summed with a resistive network as illustrated in Fig. 20, and the diodes of the peak detectors are arranged to give outputs of opposite polarity. If V_4 is assumed to be negative and V_3 positive then the output voltage will be:

$$V_{out} = \frac{V_3 - V_4}{2} \dots\dots(37)$$

if both resistors are equal.

Subtracting eqn. (36a) from eqn. (35a) and substituting for V_3 and V_4 gives:

$$2V_{out} = [V_1^2 + V_2^2 + 2V_1V_2 \sin \theta]^{\frac{1}{2}} - [V_1^2 + V_2^2 - 2V_1V_2 \sin \theta]^{\frac{1}{2}} \dots\dots(38)$$

$$4V_{out}^2 = 2V_1^2 + 2V_2^2 - 2[(V_1^2 + V_2^2 + 2V_1V_2 \sin \theta) \times (V_1^2 + V_2^2 - 2V_1V_2 \sin \theta)]^{\frac{1}{2}} \dots\dots(38a)$$

Let $V_2 = KV_1$ where K is a constant $\dots\dots(39)$

$$4V_{out}^2 = 2V_1^2(1 + K^2) - 2[V_1^4(1 + 2K^2 + K^4 - 4K^2 \sin^2 \theta)]^{\frac{1}{2}} \dots\dots(40)$$

$$4V_{out}^2 = 2V_1^2[1 + K^2 - (1 + 2K^2 + K^4 - 4K^2 \sin^2 \theta)^{\frac{1}{2}}] \dots\dots(40a)$$

By substituting in eqn. (40a) for various values of K , the transfer characteristic of the phase detector for various amplitude relationships between V_1 and V_2 can be investigated.

Consider the case when $K = 1$, i.e. when $V_1 = V_2$. Equation (40a) then reduces to:

$$V_{out} = V_1\sqrt{2} \sin \theta/2 \dots\dots(41)$$

When $K = \sqrt{2}$, i.e. when $V_2 = \sqrt{2}V_1$, eqn. (40a) reduces to

$$V_{out} = (1.5 V_1^2 - 0.5 V_1^2\sqrt{5 + 4 \cos 2\theta})^{\frac{1}{2}} \dots\dots(42)$$

When $K = 1/\sqrt{2}$, i.e. when $V_1 = \sqrt{2}V_2$ eqn. (40a) reduces to

$$V_{out} = [0.75 V_1^2 - 0.5 V_1^2(1.25 + \cos 2\theta)]^{\frac{1}{2}} \dots\dots(43)$$

Lastly consider the case when $K \rightarrow \infty$, i.e. $V_2 \gg V_1$. Equation (40a) then yields

$$\frac{2V_{out}^2}{V_1^2} = 1 + K^2 - [(1 + K^2)^2 - 4K^2 \sin^2 \theta]^{\frac{1}{2}} \dots\dots(44)$$

$$(1 + K^2)^2 - 4K^2 \sin^2 \theta = (1 + K^2)^2 + \frac{4V_{out}^4}{V_1^4} - \frac{4V_{out}^2}{V_1^2}(1 + K^2) \dots\dots(44a)$$

$$\frac{V_{out}^2}{V_1^2} - \frac{1}{1 + K^2} \cdot \frac{V_{out}^4}{V_1^4} = \frac{K^2}{1 + K^2} \sin^2 \theta \dots\dots(44b)$$

which as $K \rightarrow \infty$ gives

$$V_{out} = V_1 \sin \theta \dots\dots(44c)$$

12. Appendix 2:

Setting-up Procedure for the Sub-carrier Regenerator

Equipment required:

Valve voltmeter with an 0-50 volts a.c. range.

Oscilloscope with X and Y inputs of 1 volt/cm sensitivity.

Sub-carrier test unit: this provides a sine-wave output having an amplitude of 4 V peak to peak which can be gated at line frequency (10.125 kc/s) with a 3 μ s pulse. The frequency of the sine-wave should be variable over a range of ± 300 c/s about the colour sub-carrier frequency of 2 657 812 c/s.

Oscillator Frequency Adjustment

Earth the output from the phase detector. This is achieved by earthing the wiper of the phase detector balance potentiometer.

Remove the gate pulse input from the grid of the burst amplifier.

Connect the c.w. output of the test unit to the oscilloscope X amplifier, and the output of the regenerator to the Y amplifier. Adjust the oscilloscope controls to obtain a Lissajous figure. Set the test unit to the sub-carrier frequency.

Adjust the 1 k Ω potentiometer in the cathode of the reactance valve for a stable Lissajous figure.

Burst and Oscillator Tuned Amplifiers Adjustment

Disconnect the filter from the wiper of the phase balance potentiometer and earth the filter.

Connect the valve voltmeter to the wiper of the phase detector balance potentiometer. Set the valve voltmeter to 50 V a.c. range. Set the frequency of the test unit 100 c/s above or below the sub-carrier frequency.

Connect the gated output of the test unit to the video input.

Adjust the inductance of the burst amplifier tuned circuit for maximum valve voltmeter reading, similarly adjust the tuned circuit of the oscillator amplifier.

Adjustment of Phase Balance Control

Reconnect the filter to the wiper of the phase balance control. Remove the video input and earth this input. Set the test unit to the sub-carrier frequency. Adjust the phase balance control for a stable Lissajous figure.

Adjustment of Burst Amplifier Gating Pulse Width

Reconnect the gate pulse and video inputs to the grid of the burst amplifier. Connect the oscilloscope to the grid of the burst amplifier and adjust the pulse width of the monostable multivibrator until the gating waveform at the grid of the burst amplifier is 10% wider than the burst signal.

Manuscript first received by the Institution on 21st September 1961 and in final form on 21st February 1962. (Paper No. 784/T18.)

© The British Institution of Radio Engineers, 1963

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. Information on translating services will be found in the Institution publication "Library Services and Technical Information".

AERIAL INSIDE A TOWER

An investigation has recently been undertaken by the National Research Council of Canada to determine the radiation characteristics of a dipole antenna located inside a lattice-type supporting tower. Results are presented to show the effect of both a triangular and square tower on the radiation pattern and impedance bandwidth of the antenna. Some elevation pattern measurements were carried out, and in the range of tower parameters considered, the tower had little or no effect on the pattern. However, as the tower size was increased; i.e. radius $> 0.5\lambda$, a considerable portion of the energy was radiated above and below the horizon. This implies a reduction in the gain of the antenna. The results presented demonstrate the feasibility of using a vertical dipole antenna inside a tower for providing a more or less omnidirectional pattern. This configuration should find useful application, particularly in the v.h.f./u.h.f. communications band.

"Vertical dipole inside lattice tower provides omnidirectional pattern", J. Y. Wong. *Canadian Electronics Engineering*, 6, No. 5, pp. 36-9, May 1962.

SWITCHING DIODES

The use of high speed silicon computer diodes as r.f. switches at frequencies up to 1.5 kMc/s has been discussed in an Australian paper. Measurements of diode capacitance and Q are reported, and a theoretical analysis of diode switching performance is given and correlated with measured diode characteristics. It is concluded that with impedance transformations of the order of 2 : 1, this type of diode is capable of being used in 50 Ω circuits as an r.f. switch at a frequency of 408 Mc/s with an insertion of less than 0.3 dB and an isolation of greater than 28 dB.

"The use of high-speed silicon computer diodes for r.f. switching", R. E. Aitchison. *Electrical & Mechanical Engineering Transactions, Institution of Engineers, Australia, E.M.4*, No. 1, pp. 7-10, May 1962.

LOCALIZATION OF STEREOPHONIC SOUND

The phenomenon of directional localization of the sound in stereophony may be explained by assuming that the apparent direction of a source is that of the normal to the wave-front. That direction may be calculated, for the general case of a double-channel stereophonic system, in terms of the conditions in the studio and in the listening room. Some anomalies of the stereophonic wave-front compared with direct listening to a real source and their influence on directional localization were discussed by a Japanese broadcasting engineer in a paper read at the 12th Technical Meeting of the European Broadcasting Union in 1960. This discussion deals essentially with the variation

of the direction perpendicular to the wave-front as a function of frequency in the case where two spaced microphones are used. Finally, the conclusions of the theory are compared with the results of experimental determinations of the direction of the sound image in stereophony.

"On the directional localization of sound in the stereophonic sound field", Y. Makita. *European Broadcasting Union Review*, No. 73A Technical, pp. 102-8, June 1962.

RADIO TELEPRINTER RECEIVER

Automatic frequency control at the receiving end of radio teleprinter links is a special problem in as far as the signal can be considered to be a control signal varying at liberty between two or four values. A proposal is made by a German engineer for employing a double-spot control which is extended by an inoperative interval affecting only one direction of transmission. Theoretical considerations show that such a non-continuous control is a good solution of the control problem. Control oscillations and effects from interfering transmitters as well as from noise are also investigated. Theoretically derived relations are confirmed by experiments. The design of a control circuit is also mentioned.

"Double-spot a.f.c. for radio teleprinter reception", K. Grabe. *Nachrichtentechnische Zeitschrift*, 15, pp. 292-8, June 1962.

TRANSISTOR REGULATED POWER SUPPLIES

Problems in the design of low-voltage regulated power supplies using transistors are discussed by Professor R. E. Aitchison of the University of Sydney in a paper published by the Institution of Engineers, Australia. The requirements for a general purpose standard design of high performance are considered. This is illustrated by a basic series regulator design which can be simply modified to give either a negative or positive output, or a dual positive and negative unit of various voltage and current ratings. The main stabilization factor of the basic design exceeds 10^4 : 1, and the internal impedance is determined primarily by wiring resistances. By the use of a balanced differential input stage for the comparison amplifier and a low temperature coefficient reference system, the temperature coefficient of the basic design is reduced to less than 0.5 mV/deg C. Methods are considered for externally sensing output variations to correct for interconnecting lead resistance, and for protecting the series transistor from overload.

"General purpose regulated power supplies using transistors", R. E. Aitchison. *Electrical & Mechanical Engineering Transactions, Institution of Engineers, Australia, E.M.4*, No. 1, pp. 15-20, May 1962.