

THE RADIO AND ELECTRONIC ENGINEER

The Journal of the Institution of Electronic and Radio Engineers

FOUNDED 1925 INCORPORATED BY ROYAL CHARTER 1961

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

VOLUME 29

APRIL 1965

NUMBER 4

COLOUR BLINDNESS

IN most fields of human endeavour the lessons of history are learned too slowly. A complementary fact is the tardiness shown in the development and universal application of new engineering techniques.

A current example is the inability to secure agreement on a common system of colour television. Four meetings of the International Radio Consultative Committee (C.C.I.R.) have been held (the last in Vienna) to try and reach a decision acceptable to all members of the European Broadcasting Union. At the time of this publication it seems that a final decision may be postponed until next year's Oslo meeting of the International Radio Consultative Committee. Such postponement could, however, lead to some countries blindly following their own inclinations and thus historically repeating the mistakes of fifteen years ago, as a result of which the 405, 625, and 819 line systems were employed throughout Europe.

Although much original work was, in fact, done in Great Britain, the only systems now being considered are the American N.T.S.C., the French SECAM, and the German PAL. It is claimed that by modifying the N.T.S.C. system, both SECAM and PAL have secured technical improvements. Not made clear, however, is the basic patent rights which America holds in respect of both the German and French systems.

A significant report from Vienna is that some measure of agreement has been reached in combining the N.T.S.C. and PAL systems. A straightforward choice may not, however, bring any earlier decision, for there remains an economic battle to secure a vast commercial market. Amidst the turmoil of commercial and political propaganda a blind eye may be turned to the strong opinion of many engineers that more extensive research and development has still to be undertaken in order to provide a satisfactory and reasonably economic colour television service. Commercially, it is well argued that, as with the birth of monochrome television, the cost of further development should at least be partly financed by the first production of colour receivers. It is, of course, outside the bounds of technical argument as to whether the demand for colour receivers will be sufficiently stimulated by programmes more imaginative than those at present transmitted in monochrome, and by making greater utilization of Eurovision links.

Unfortunately, the British Prime Minister's statement that there is to be closer collaboration in electronics between Great Britain and France is perhaps too late to permit of further discussions between the two countries on the adoption of colour standards. In the welter of political considerations we should not neglect a first class opportunity to build a new and significant link for understanding among the nations of Europe. Above all, we must not be blind to the needs of the final link—the viewer. In this connection it is particularly apt to recall the comment made in the Institution's Post-War Report (December 1944) that television standards ". . . should not be frozen at a level which is below the technical and economic limits of the present time." Whilst, therefore, procrastination is not recommended, there is still time to avoid repeating historical mistakes.

G. D. C.

INSTITUTION NOTICES

Institution Dinner

Members are reminded that the Institution Dinner will be held on Thursday, 24th June, at the Savoy Hotel, London. Applications for tickets, which cost £3 15s., *inclusive of wines and liqueurs*, may now be sent to The Secretary, I.E.R.E., at 8-9 Bedford Square, London, W.C.1, marking the envelope "Institution Dinner".

Joint I.E.R.E.-I.E.E. Symposium

Outline programmes for the Joint Symposium on "Microwave Applications of Semiconductors", which is to be held at University College, London, from 30th June to 2nd July, were sent out with the March issue of *The Radio and Electronic Engineer*.

Registration forms are now available on application to the Joint Symposium Secretary, I.E.R.E., 8-9 Bedford Square, London, W.C.1. The registration fee is £10 for members of the I.E.R.E. and of the I.E.E.; the fee for non-members is £13.

The Symposium will include forty papers which will be preprinted for those registering to attend the Convention. The full programme and a synopsis of the papers will be published in the May issue of *The Radio and Electronic Engineer*.

Karachi Section

A meeting of the local section of the Institution of Electronic and Radio Engineers was held in Karachi recently. The following were elected to the Section Committee for the year 1965:

Chairman: Mr. S. A. Aziz (Member); Honorary Secretary: Mr. I. A. Ansari (Associate Member); and Messrs. Sadiq Shah (Associate Member), Lt. Cdr. M. Idris Khan (Associate Member), Sajid Hasan (Associate Member) and S. H. Ansari (Graduate).

London Meeting on Military Electronics

An additional meeting has been arranged in the current session programme of Institution activities. On 26th May an address on "The Impact of Electronics on the Army's Repair Organization" will be given by Major General L. H. Atkinson, O.B.E., B.Sc. (Member). The meeting will be held at the London School of Hygiene and Tropical Medicine, Gower Street, W.C.1, and will start at 6 p.m.

Major General Atkinson has served on the Institution's Council since 1963, and on the Education Committee for the past six years. He is Director of Electrical and Mechanical Engineering in the Army Department of the Ministry of Defence.

The Council has long appreciated that many members of the Institution are concerned with work in the military electronics field and believes that this invitation to General Atkinson to give this address will be of especial interest to those members.

Members will not require tickets to attend the meeting, but those wishing to introduce non-members are asked to apply for cards of invitation for this meeting.

North Eastern Section Meeting

Under the joint sponsorship of the North Eastern Section of the Institution and the Northern Regional Office of the Ministry of Technology, a symposium on "Electronic Control Systems for Industry" will be held at the University of Durham from 8th to 10th September 1965. The object of the meeting is to bring to the attention of industry modern electronic control techniques and ways of increasing productivity and reducing costs, while at the same time improving quality and reliability.

Three fields will be covered: Computer Analysis of Information; Advanced Telecommunication Techniques; and the Use of Automative Systems in Production. The papers and discussions will concentrate on the practical application of these techniques.

Further information is available from the Institution at 8-9 Bedford Square, London, W.C.1.

Post-Graduate Courses

The attention of members is drawn to the following advanced courses in radio and electronics:

The Department of Electrical and Control Engineering of the College of Aeronautics, Cranfield, is holding a short course on "Microwave Laboratory Practice" from 17th to 21st May 1965. Intended for engineers who are familiar with the theory of microwaves, the emphasis will be on experimental techniques. Further information may be obtained from the Registrar, The College of Aeronautics, Cranfield, Bedford.

A one-year M.Sc. and Diploma course in Quantum Electronics is being started by the Department of Electronics of the University of Southampton in October of this year. The aim of the course is to provide an understanding of the principles and operation of existing devices, notably masers and lasers, in such a way that students are more likely to be able to conceive and design new devices. The course is open to University Graduates with suitable science or engineering degrees. Further information is obtainable from the Academic Registrar, University of Southampton.

A High-speed Tunnel Diode Counter

By

Professor

ANDREW D. BOOTH,

D.Sc., Ph.D. (Member)†

AND

T. R. VISWANATHAN,

M.Sc., Ph.D., D.I.I.Sc.‡

Summary: The tunnel diode binary counter circuit reported by Kaenel and subsequently discussed by many authors is described. The mode of operation of the circuit is explained in terms of a piece-wise linear model for the tunnel diode. The switching trajectories obtained using a low frequency analogue simulation of the circuit are used to discuss the bias and trigger conditions to be employed. A method of intercoupling several stages using transistors and tunnel diodes is suggested.

1. Introduction

A circuit configuration (Fig. 1) in which the transitions of a tunnel diode, made bistable by suitable biasing can be brought about by pulses of a single polarity, was reported by Kaenel.¹ The simplicity of this configuration has aroused considerable interest in the counter circuit²⁻⁵ which it makes possible and an attempt is made in this paper to bring out the salient features of the circuit. The circuit action is explained using piece-wise linear model (Fig. 2) and the simplified equivalent circuit of Fig. 3 for the tunnel diode.

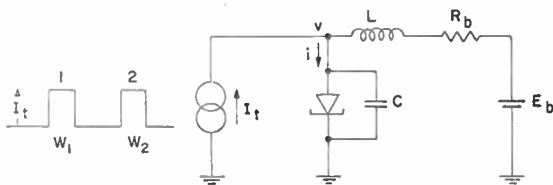


Fig. 1. Circuit configuration of the counter.

There are two distinct modes of operating the circuit as a counter, depending upon the nature of the trigger source from which the trigger pulses are derived. Further, the nature of the trigger source determines the type of coupling to be employed between the trigger source and the tunnel diode. If the trigger source impedance is high, compared to the minimum incremental negative resistance R_N exhibited by the tunnel diode, the trigger is d.c. coupled. This is possible because the presence of the trigger source does not appreciably affect the d.c. bias conditions of the tunnel diode. If, on the other hand, the trigger pulses are obtained from a voltage source, capacitance coupling must be employed to produce the necessary d.c. isolation. Diode coupling cannot be

† College of Engineering, University of Saskatchewan.

‡ Department of Electrical Engineering, University of Waterloo, Waterloo, Ontario.

employed since the trigger point assumes two values of d.c. potential depending upon the state of the tunnel diode. These two situations will now be examined in detail.

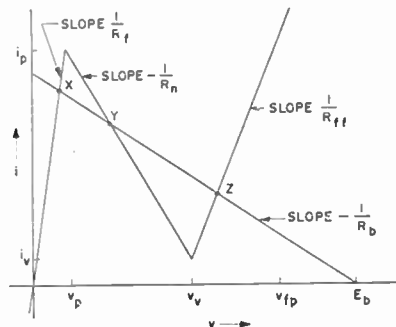
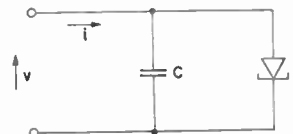


Fig. 2. Simple piece-wise linear model.

Fig. 3. Simplified equivalent circuit of tunnel diode (lead inductance neglected).



2. D.C.-Coupled Current-triggered Operation

The counter operates via a change of state (X to Z) of the bistable tunnel diode circuit shown in Fig. 1 caused by a pulse (termed pulse 1). The tunnel diode is returned to its ground state (X) by the succeeding pulse (termed pulse 2).

The tunnel diode shown in Fig. 1 is biased by the d.c. voltage source E_b and an inductance L and the load line corresponding to the bias arrangement intersects the characteristic close to the peak and valley points as shown in Fig. 4. A train of ideal current pulses of magnitude I_t and widths W_1 is applied to the tunnel diode.

It is easy to understand how a current pulse (I_t) injected into the tunnel diode can produce an X to Z

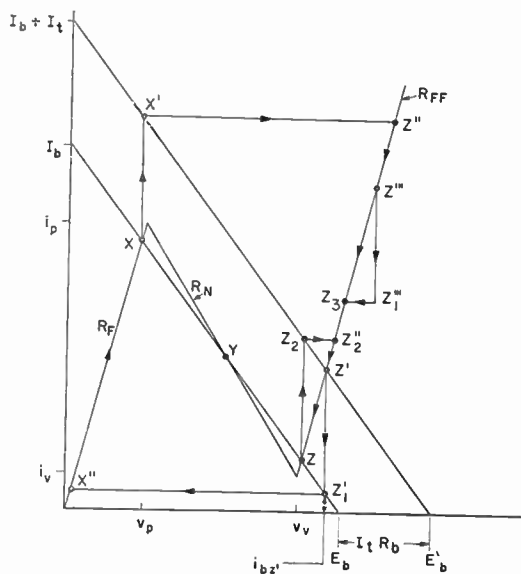


Fig. 4. D.c.-coupled, current-triggered, operation of Kaenel's counter.

transition. However, it is more difficult to see why a second pulse of current injected into the tunnel diode can produce a Z to X transition.

Assume that the system composed of the tunnel diode and its bias and trigger arrangements is occupying the state X. Let the inductive time-constants of the system ($L/R_b + R_F$, $L/R_b - R_N$, $L/R_b + R_{FF}$) be very large compared to the forward and backward switching times of the tunnel diode and let the widths W_1 and W_2 of pulses 1 and 2 be controllable at will.

During the forward edge of trigger pulse 1 the point of operation will move to a point X' vertically above X by an amount equal to I_t . During the pulse period, the system switches over along a trajectory $X'Z''$ which will be virtually parallel to the v -axis. Thereupon, the system slowly moves from Z'' to Z' in an exponential manner. Z' is located by drawing a new load line $X'Z'E'_b$ which is shifted from XYZ vertically by an amount I_t .

At the point Z' , the current i'_z flowing through the tunnel diode can be considered to be composed of two components: (1) I_t due to the trigger source, (2) $i_{bz'}$ due to the bias source. The magnitude of $i_{bz'}$ will depend upon the position of Z' and hence will be a function of I_t for a given bias arrangement.

At the termination of the trigger pulse, I_t is instantaneously reduced to zero and the tunnel diode has only $i_{bz'}$ flowing through it. Thus, the point of operation will move to Z'_1 vertically below Z' by an amount I_t . If $i_{bz'}$ is less than i_v , the system will switch back along $Z'_1X''X$, since the inductance will not let

the current supplied by the bias source $i_{bz'}$ change in a time corresponding to the switching-back time of the tunnel diode. Thus, under the above conditions, corresponding to every pulse the system will execute an X to Z' and a Z' to X transition.

For counter action, the first pulse should take the tunnel diode from X to Z and this can be achieved by so choosing I_t that $i_{bz'}$ is greater than i_v , or by terminating the pulse when the system is at a point Z''' where $i_{bz'''}$ is above i_v . The first possibility is not conducive to effecting a change of state from Z to X by the second pulse. Hence the circuit is made to change from state X to Z on the first pulse by controlling the pulse width.

As the point of operation moves from Z'' to Z' (due to pulse 1) the current supplied by the bias source is decreasing from i_x to $i_{bz'}$ in a manner given by the equation

$$i_b = i_{bz'} + (i_x - i_{bz'}) \exp \left[\frac{-(R_b + R_{FF})t}{L} \right] \dots\dots(1)$$

If the trigger is terminated before i_b reduces below i_v , the system will find a current greater than i_v flowing through it. Hence, the system will move to Z and has no chance of switching back. The corresponding pulse width W_1 is given by

$$i_{bz'} + (i_x - i_{bz'}) \exp \left[\frac{-(R_b + R_{FF})W_1}{L} \right] > i_v \dots\dots(2)$$

i.e.
$$W_1 < \frac{L}{R_b + R_{FF}} \ln \left[\frac{i_x - i_{bz'}}{i_v - i_{bz'}} \right] \dots\dots(3)$$

If I_t is again injected into the tunnel diode (corresponding to pulse 2), the system initially in state Z once again tends to move to Z' along $ZZ_2Z_2'Z'$ provided W_2 is very large compared to the inductive time constant $L/(R_b + R_{FF})$. In the process of moving from Z_2' to Z' , the current i_b supplied by the bias source decreases with time in a fashion given by

$$i_b = i_{bz'} + (i_z - i_{bz'}) \exp \left[\frac{-(R_b + R_{FF})t}{L} \right] \dots\dots(4)$$

In order for a Z to X transition to be possible, W_2 should be large enough for i_b to fall below i_v . Hence for switching back

$$W_2 > \frac{L}{R_b + R_{FF}} \ln \left[\frac{i_z - i_{bz'}}{i_v - i_{bz'}} \right] \dots\dots(5)$$

Let I_t be so chosen that, corresponding to the location of Z' , $i_{bz'}$ is equal to zero and let $i_x = 0.8 i_p$, $i_z = 0.2 i_p$ and $i_v = 0.1 i_p$. Then, from eqns. (3) and (5),

$$W_1 < \frac{L}{R_b + R_{FF}} \ln \left[\frac{i_x}{i_v} \right] \text{ or } \tau \ln 8 \text{ or } 2.1 \tau \dots\dots(6)$$

$$W_2 > \frac{L}{R_b + R_{FF}} \ln \left[\frac{i_z}{i_v} \right] \text{ or } \tau \ln 2 \text{ or } 0.7 \tau \dots\dots(7)$$

where

$$\tau = \frac{L}{R_b + R_{FF}} \dots\dots(8)$$

Thus we see that W_1 has an upper limit and W_2 has a lower limit. In other words, if the pulse width W is so chosen that its magnitude lies between the two limits 2.1τ and 0.7τ for the same pulse, the system counts by switching from X to Z and Z to X alternately. The upper limit for W_1 increases with i_x which can assume a value as high as i_p . Thus, for a tunnel diode with $i_p/i_v = 10$, W_1 has an upper limit of $\tau \ln 10$ (i.e. 2.3τ). The lower limit for W_2 decreases with i_x which can assume a value as low as i_v and, hence, the lower limit for W_2 tends to zero. In other words, the overlap region for the width requirements imposed on pulses 1 and 2 increases as the stable states X and Z move to the peak and valley points respectively. Thus, a 'peak to valley' type of load line is most favourable for the counting action of the circuit.

The above analysis will be true only if τ is very large compared to the switching time associated with the tunnel diode. However, it is necessary to know approximately how large τ should be compared to the

switching times of the tunnel diode. When the system switches from X to Z, it is under the influence of the trigger, whereas when switching back from Z to X, it is left to itself at a point Z'_1 and is made to switch back. Whether the system switches back or not will depend upon the relative magnitudes of switching back time T_{zx} and the inductive time-constant with which the bias current tends to increase in the tunnel diode. An exact determination of the relationship requires a knowledge of the separatrix⁷ of the system and the procedure is quite involved. Using a tunnel diode analogue, an approximate value for τ has been obtained for $i_{bz} = 0$. It turns out to be five times the switching-back time of the tunnel diode. The nature of the trajectories for ideal conditions have also been obtained and are shown in Fig. 5. The scale factors used in the analogue simulation are shown in Table 1.

Table 1

Scale factors used for analogue simulation

	Tunnel Diode	Tunnel Diode Analogue	Scale Factor
Current	1 mA	10 mA	10
Voltage	500 mV	5 V	10
Resistance	ohms	ohms	1
Capacitance	1 pF	0.01 μ F	10 ⁴
Inductance	1 nH	0.01 mH	10 ⁴
Time	1 ns	10 μ s	10 ⁴

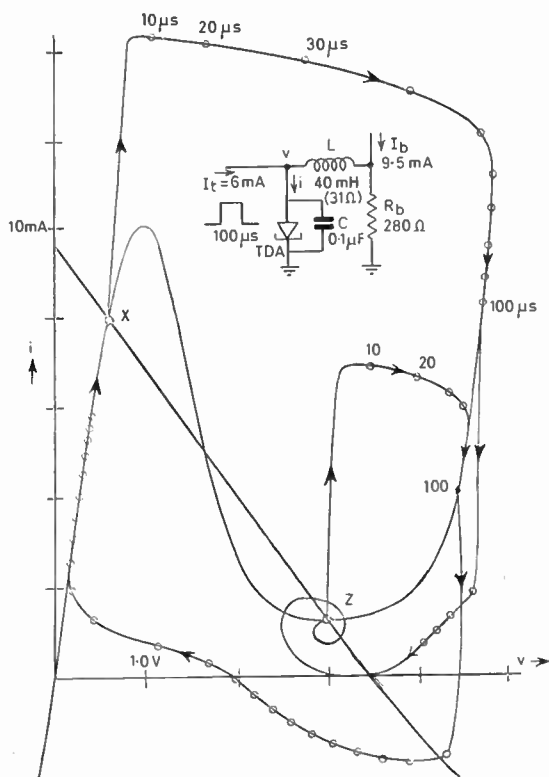


Fig. 5. The trajectories of Kaenel's counter (obtained using the tunnel diode analogue) in the d.c.-coupled current-triggered mode of operation.

In order to count at high repetition rates, τ should be reduced to a minimum since W is dependent on τ and the time between the pulses to allow for the slow movements of the operative point (from Z'_1 to Z and X'' to X) to be completed. This is done using the tunnel diode analogue. As the tunnel diode characteristic is highly non-linear compared with the piece-wise linear model, the above considerations give only an indication by which the orders of magnitude of the component values of circuit parameters and time factors can be determined. The detailed tolerance limits and centre values adopted in the circuit are obtained with the aid of the analogue in which, in order to estimate the high-frequency performance, second-order effects such as lead inductance for the tunnel diode, realistic source and load impedances are included.

3. A.C.-Coupled Voltage-triggered Operation

As in the earlier circuit, let the inductive time-constants of the circuit shown in Fig. 6 be large compared to the switching times of the tunnel diode. Let an ideal voltage pulse E_t be a.c.-coupled to the tunnel diode using a capacitance C_c . The system occupies the state X when E_t is applied, and the point

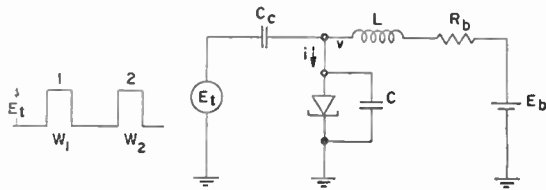


Fig. 6. Voltage triggered mode of operation of counter.

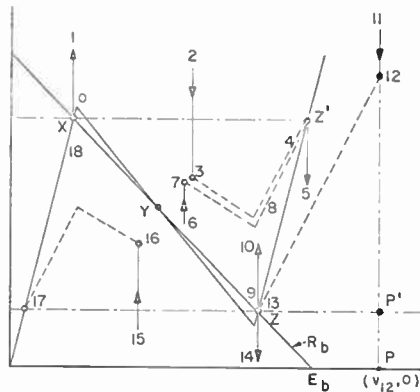


Fig. 7. A.c.-coupled, triggered-operation, of Kaenel's counter.

of operation moves along the trajectory 0123 (Fig. 7) where

$$(v_3 - v_x) = E_t \cdot \frac{C_c}{C_c + C} \quad \dots\dots(9)$$

In the simplified equivalent circuit (Fig. 3), the lead inductance of the tunnel diode is neglected. Hence, during the forward edge of the pulse, unlimited current flows, charging the tunnel diode capacitance C instantaneously to a voltage V_3 . Thus, the trajectory starting from X leads to $+\infty$ and comes back to the point 3.

If $(V_3 - v_x) > v_y \quad \dots\dots(10)$

the system switches over to Z' along $34Z'$. Since the bias current in the tunnel diode has not changed ($i_x = i'_2$), Z' is obtained by drawing a straight line XZ' parallel to the v -axis. The position of the trajectory $34Z'$ is obtained by dividing the distance between the line XZ' and the tunnel diode characteristic in the ratio of the capacitances C_c and C . As the tunnel diode switches, the rates of change of voltage across C_c and C are the same and hence the currents flowing through C_c and C are in the ratio C_c to C .

If the pulse width W_1 is equal to the time taken by the system to switch to Z' along $34Z'$ ($W_1 = T_{3Z'}$), at the occurrence of the back edge of the pulse, the point of operation moves from Z' to 7 along $Z' 567$

such that $(v_2 - v_7) = (v_3 - v_x)$. Let the interval between the pulses be large so that the system moves from 7 to Z' along $78Z'$ and from Z' to Z. The trajectories $78Z'$ and $34Z'$ will overlap though for clarity they are shown distinctly as two paths in Fig. 7. The current supplied by the bias source decreases slowly at a rate determined by the inductance L , R_b and R_{FF} bringing the system from Z' to Z. Now when the forward edge of the second pulse occurs, the system moves from 9 (Z) to 12 along $9 10 11 12$ where

$$(v_{12} - v_9) = (v_3 - v_x) = E_t \cdot \frac{C_c}{C + C_c} \quad \dots\dots(11)$$

The position of the point 12 is obtained as follows. A straight line perpendicular to the v -axis is drawn through a point $P(v_{12}, 0)$. The voltage v_{12} is known from eqn. (11). ZP' is the a.c. load-line since the bias current i_z has not changed. The trajectory 12 13 has to lie between this load-line and the tunnel diode characteristic. Furthermore, the ratio of the vertical distances between the characteristic and the trajectory and the load line must be equal to C_c/C . Thus the point 12 is obtained by dividing the vertical distance between the characteristic and the load line ZP' in the ratio of the capacitances. If the pulse width W_2 is equal to the time taken by the system to move from 12 to 13 or Z ($W_2 = T_{12,Z}$) at the occurrence of the back edge the point of operation moves from Z to 16 via Z 14 15 16 where

$$(v_z - v_{16}) = E_t \cdot \frac{C_c}{C + C_c}$$

If $v_{16} < v_y$, the system switches back to 17 and from there onward to X slowly with time.

Summarizing the conditions, we have:

for switching at the forward edge of pulse 1

$$E > (v_y - v_x) \quad \dots\dots(12)$$

where

$$E = E_t \cdot \frac{C_c}{C + C_c} \quad \dots\dots(13)$$

for not switching back at the back edge of pulse 1

$$(v_z - E) > v_y \quad \dots\dots(14)$$

and

$$W_1 = T_{3Z'} \quad \dots\dots(15)$$

for switching back at the back edge of pulse 2

$$(v_z - E) < v_y \quad \dots\dots(16)$$

$$W_2 > T_{12Z} \quad \dots\dots(17)$$

combining eqns. (14) and (16) we get

$$(v_z - v_y) < E < (v_z - v_y) \quad \dots\dots(18)$$

time between the pulses

$$T_{Z'9} \gg \frac{L}{R_b + R_{FF}} \quad \text{and} \quad T_{17X} \gg \frac{L}{R_F + R_b} \quad \dots\dots(19)$$

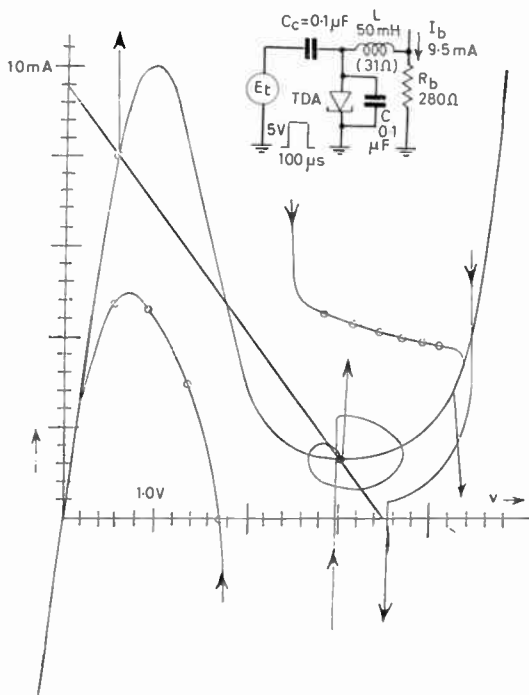


Fig. 8. The trajectories (obtained using the tunnel diode analogue) of the Kaenel's counter in the a.c.-coupled voltage-triggered mode of operation.

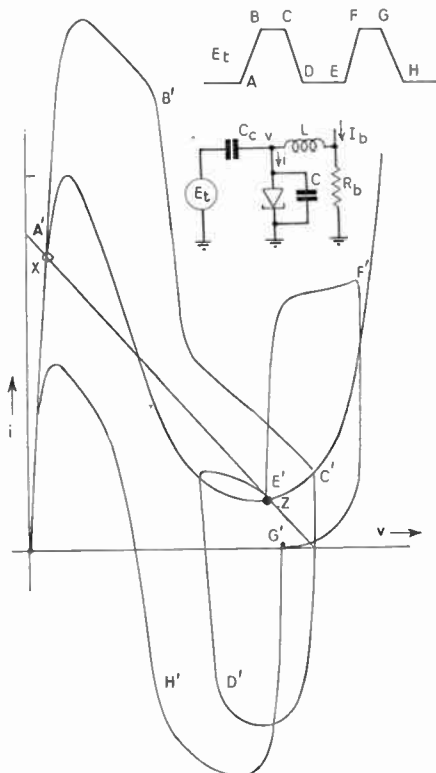


Fig. 9. The nature of the trajectories (of Kaenel's counter) when the a.c.-coupled trigger pulses (voltage) have finite rise and fall times.

Under these conditions, the system exhibits counting properties. Equation (18) shows that it is not difficult to realize the above conditions provided there is enough voltage separation between Z' and Z and, hence, the system can have such properties only for a peak-to-valley load-line (linear or non-linear). W_1 and W_2 can be made equal by appropriate choice of pulse amplitude and locations of X , Y and Z . The time between the pulses tends to be equal as $R_F \approx R_{FF}$ (eqn. (19)).

In the above discussion an ideal trigger source and step-type of pulse shape were assumed. The trajectories obtained in practice using the tunnel diode analogue for capacitance coupled voltage triggered operation are shown in Fig. 8.

When pulses with finite rise and fall times are used, the trajectories obtained using the analogue are shown in Fig. 9. The shape of the pulses employed is also shown. When the pulse starts to rise, corresponding to the point A (Fig. 9), the system occupies state X. During the rising edge AB of the pulse, a finite current (determined by C_c and dE_t/dt) will flow into the tunnel diode (unlike the case when an ideal step is used, $dE_t/dt \rightarrow \infty$, and the trajectory will go to infinity and come back). The system moves to B' during the rise-time of the pulse. After the point B where the pulse attains its maximum amplitude the current flowing through C_c starts to fall ($dE_t/dt = 0$) and during the pulse period BC the system switches from B' to C'. Corresponding to the back edge CD, a current flows out of the tunnel diode and the point of operation moves from C' to D'. At the point D where the pulse amplitude becomes zero, and during the period DE between the pulses, the system moves from D' to E' (or Z). Similarly, corresponding to pulse 2, the system moves back from Z to X along E'F'G'H'A'. When the back edge of pulse 2 is very slow, the system will not switch back from Z to X unless the inductance L , or τ , is increased. This happens since the pulse 2, after reducing the current supplied by the d.c. bias source (i_{bg}) in the tunnel diode below i_v , in the processes of recovering slowly, gives time for the bias current to build up in the tunnel diode to a value greater than i_v . Thus, apart from the switching-back time of the tunnel diode, the fall-time for the back edge of the pulse should also be taken into account in determining the τ necessary for circuit operation. When the pulses have slow back-edges, the increased τ needed would bring down the maximum rate with which pulses can arrive.

In order to increase trigger sensitivity a peak-to-valley non-linear load-line (as shown in Fig. 10) is advantageously employed. For voltages greater than v_z , R_b being very small, L can be small to give the necessary τ needed for circuit operation. This results in increased repetition frequency. Thus, a non-

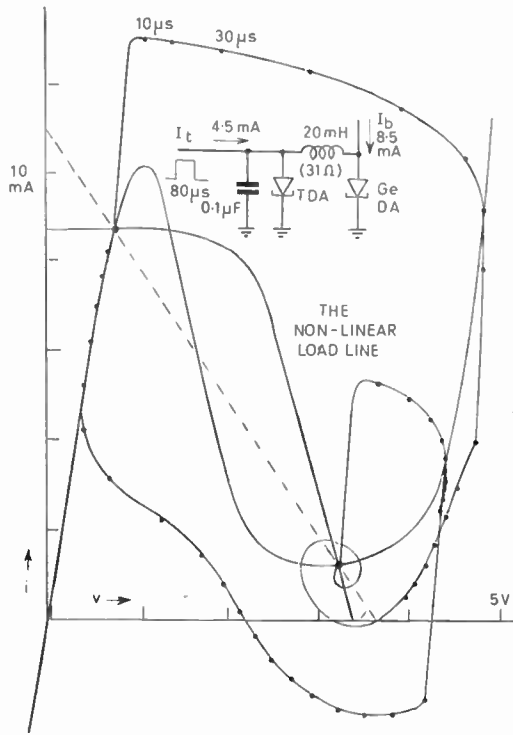


Fig. 10. The trajectories of the modified Kaenel's counter obtained (using the tunnel diode analogue) when a non-linear bias arrangement (load line) and d.c.-coupled current trigger are employed.

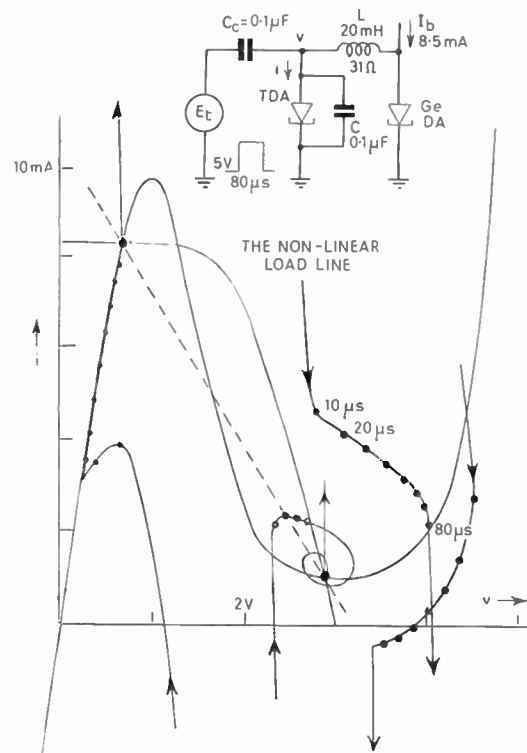


Fig. 11. The trajectories (of the modified Kaenel's counter) obtained when a non-linear load line and a.c.-coupled trigger pulses (voltage) are employed.

linear load-line can improve the circuit action. Examination shows that the bias tolerances improve and the trigger requirements are reduced in a practical circuit. Further, the current change in the tunnel diode as it changes state from X to Z (which is obtained as useful output) tends to be a well defined quantity equal to $(I_b - i_v)$. The trajectories for non-linear biasing in the two modes of triggering are shown in Figs. 10 and 11.

4. Output and Cascading

It now remains to be seen how a useful output can be obtained from the basic counter circuit to trigger at least one more similar stage. In the d.c.-coupled current-triggered operation of the circuit, a current pulse is needed to trigger a given stage. Hence, a current output has to be obtained from the basic circuit. This is easily achieved by using a grounded base transistor⁸ as shown in Fig. 12. Looking into the emitter of the transistor the input characteristic is such that for a d.c. bias, the load-line assumes the desired 'peak-to-valley form'. The composite characteristic of the tunnel diode in parallel with the effective trigger source impedance of approximately 500Ω is

shown by the dotted curve in Fig. 13. The new valley current of the system is denoted by i'_v . The current change occurring in the system as it changes state will be equal to $(I_b - i'_v)$ and flows into the emitter of the transistor. Thus a current output of $(I_b - i'_v)$ is obtained as output at the collector of the transistor (assuming $\alpha = 1$ for the transistor).

Next the collector current waveform at the occurrence of pulse 2 will be examined. If the system is in state Z when I_t is again injected, the load line shifts vertically, by an amount I_t , to the position 2 (shown in Fig. 13) and the corresponding point of intersection Z' is obtained. Since the tunnel diode has a rather broad valley region, i'_z is very nearly equal to $i_z (i'_v)$ and the current in the collector of the transistor must rise to a value $(I_b + I_t - i'_v)$. When the tunnel diode switches back to the state X, the collector current falls to zero.

Thus, during the period of the pulse 2, an extra current of magnitude I_t flows in the collector over and above the current $(I_b - i'_v)$ which is set in the collector of the transistor by pulse 1 (shown in Fig. 12).

The collector current, as such, cannot be used to trigger another stage, since the pulse width which

depends upon the time between the pulses 1 and 2 is not known if pulses occur at random. Even for a fixed frequency of operation, the output of the first stage will have a duty cycle of approximately 1/2 and the situation cannot be taken care of by increasing the inductance values in successive stages. As the inductance is increased, the time between the pulses will also have to be increased for the inductive relaxation of the circuit to be completed. Thus, a pulse of current has to be generated corresponding to the second pulse received by the stage.

This is easily achieved by using another tunnel diode. The peak and bias currents of the tunnel

For the trigger current I_t to be the same irrespective of the state of the tunnel diode TD2 in the second stage circuit for the second stage is so arranged that the d.c. potential of the trigger point P does not change as the tunnel diode in the second stage TD2 changes state. To do this, a resistance R_Q is connected in series with the parallel combination of the tunnel diode TD2 and the transistor VT2 as shown in Fig. 14. R_Q is given by

$$R_Q = \frac{v_b - v_x}{I_b - i_v}$$

for large values of β for the transistor. Thus, as the state of the tunnel diode changes from X to Z, the

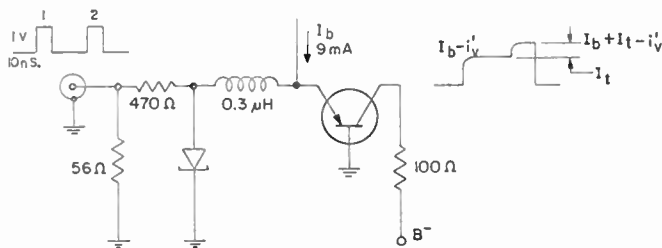


Fig. 12. Circuit arrangement for current output from Kaenel's circuit using a grounded base transistor.

diode TD3 (Fig. 14) are such that as the collector current ($I_b - i_v$) flows (due to pulse 1), the tunnel diode TD3 moves to an operating point arbitrarily close to its peak point. Corresponding to the extra collector current (I_t) which flows at the arrival of the second pulse, the tunnel diode (TD3) switches on. Once again, when the collector current reduces to zero, the tunnel diode switches back. A gallium arsenide tunnel diode may be used for TD3 which yields an output voltage swing of roughly 0.8 V. The trigger current for the second stage is obtained by using a series resistor R_g which is connected to the trigger point P of the succeeding stage.

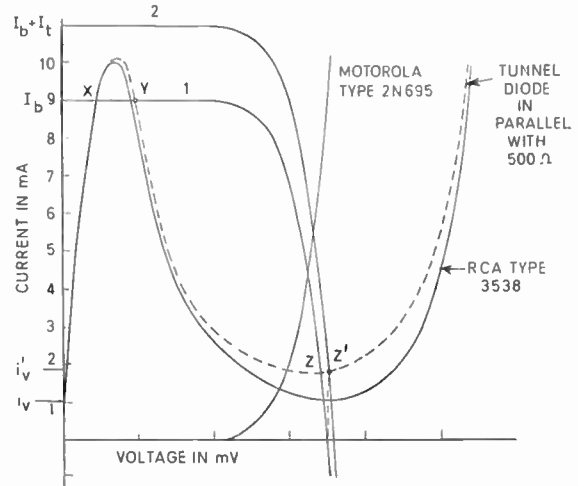


Fig. 13. Analysis of circuit of Fig. 12.

point Q moves down instead of the trigger point P moving up, this leaves the trigger current I_t unaltered.

The complete experimental circuit diagram with bias magnitudes is shown in Fig. 14.

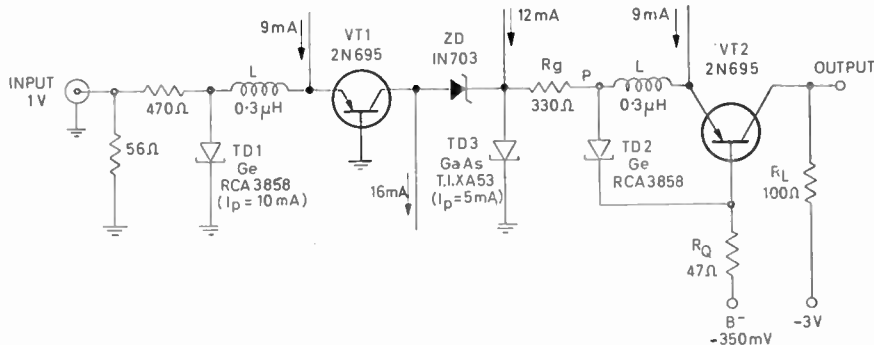


Fig. 14. Method of cascading several stages of Kaenel's counter.

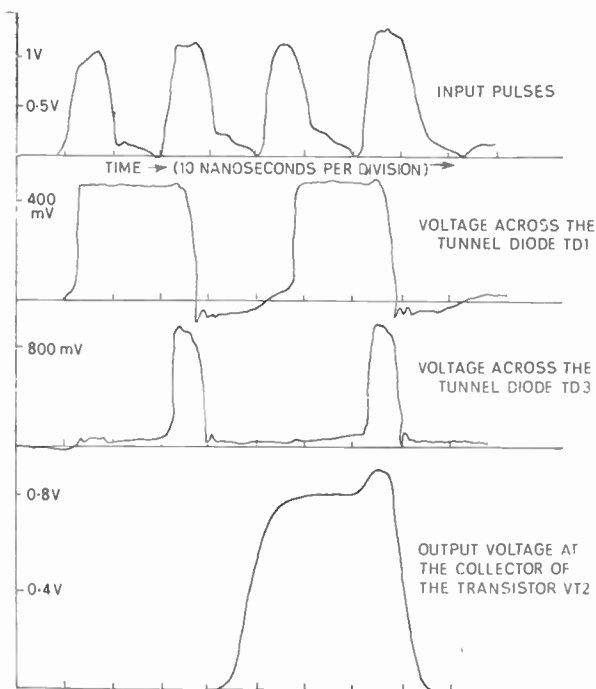


Fig. 15. Waveforms from circuit of Fig. 14.

5. Experimental Results

With 10-mA tunnel diodes (RCA type 3858) in conjunction with 2N695 transistors and gallium arsenide tunnel diodes (T.I. type XA53) for intercoupling, a maximum repetition frequency of 100 Mc/s was observed to be the limit. Since the tunnel diode switching times are in the order of one nanosecond, the τ needed for reliable circuit operation is fixed at 5 nanoseconds. The inductance needed to obtain this τ is estimated as $0.3 \mu\text{H}$. The pulse width under these conditions can vary between 5 and 10 nanoseconds. The circuit was constructed and tested using a four-

pulse generator.⁸ The recording of the waveforms obtained at the various points obtained using the Hewlett Packard sampling oscilloscope (H.P. model 185 B) and X-Y recorder is shown in Fig. 15.

6. Acknowledgments

The authors wish to express their thanks to Mr. H. N. Mahabala for his valuable suggestions and discussions. This work was supported by the National Research Council of Canada (Grant No. A878) and the Defence Research Board of Canada (Grant No. 2804-05). Financial assistance by the N.R.C. in the form of a studentship is gratefully acknowledged.

This paper is based on portions of Chapter 5 of a thesis submitted to the University of Saskatchewan, Saskatoon as partial fulfilment of the requirements for the Degree of Doctor of Philosophy.

7. References

1. R. A. Kaenel, "One tunnel diode flip-flop", *Proc. Inst. Radio Engrs*, 49, p. 622, March 1961 (Letter).
2. J. F. Banzhaf and H. S. Katzenstein, "One tunnel diode flip-flop", *Proc. I.R.E.*, 50, p. 212, February 1962 (Letter).
3. H. Guckel, "One tunnel diode flip-flop h.f. behaviour", *Proc. I.R.E.*, 49, pp. 1685-86, November 1961.
4. U. R. Hanoch, "Tunnel diode binary counter circuit", *Proc. I.R.E.*, 49, p. 1092, June 1961.
5. A. L. Whetstone, S. Kounosu and R. A. Kaenel, "One tunnel diode binary", *Proc. I.R.E.*, 49, p. 1445, September 1961.
6. R. S. C. Cobbold and H. N. Mahabala, "A tunnel diode analogue and its application", *Proc. Instn Elect. Engrs*, 110, pp. 51-63, 1963 (I.E.E. Paper No. 4061E.)
7. H. N. Mahabala, "Oscillation and Switching in Tunnel Diodes", Ph.D. thesis, University of Saskatchewan, 1964.
8. N. Moody *et al.* "A four pulse generator for testing tunnel-diode circuits", *Electronic Engineering*, 35, pp. 72-7, 1963.

Manuscript received by the Institution on 7th December 1964. (Paper No. 970/C78.)

© The Institution of Electronic and Radio Engineers, 1965

The Effect of a Linear Phase Taper on the Near Field of an Ultrasonic Multi-element Array

By

L. KAY, B.Sc.(Eng.), Ph.D. †

AND

M. J. BISHOP, B.Sc.(Eng.) ‡

Reprinted from the Proceedings of the Symposium on "Signal Processing in Radar and Sonar Directional Systems", held in Birmingham from 6th-9th July, 1964.

Summary: Electronic beam deflection of the near field of a 10-element ultrasonic array in water has been studied as a preliminary step towards using the principle in solids. The change in the field pattern arising from the change in the path length from each element is discussed for steady state conditions as the beam is caused to be deflected by a linear phase taper. It was shown that the 'beam' can be deflected a significant amount within the Fresnel region and single discontinuities in the medium lying at an angle to the normal of the array can therefore be examined. Multiple discontinuities, each of different impedance, may give rise to ambiguous results of a more serious nature than those experienced in the far field.

1. Introduction

The growing application during recent years of ultrasonics for the inspection of solids and for medical diagnosis has indicated the potential of the technology, but an examination of the methods used reveals little advance when compared with that made in underwater sound systems where complex signal processing techniques are being employed to improve information quality and rate. It seems reasonable therefore to attempt to exploit these advances in the more difficult technological field of ultrasonic inspection.

One promising advance is that of electronic beam scanning,^{1,2} which in effect is an aperture sampling system,³ and the possibility of applying this to ultrasonic inspection is attractive. The field distribution in the Fraunhofer region, which is commonly referred to as a 'beam', is quite different from that in the Fresnel region and reference to 'beam scanning' is a gross over-simplification of the situation when working in the near field. It is not immediately obvious, in fact, that the direct application of electronic beam deflection techniques can be used. It is first necessary to determine the field pattern in the near-field region of a sectionalized array when a phase taper is used to deflect the direction of the radiated energy.

The near field of a rectangular piston source has already been investigated by Freedman,⁴ and whilst the axial variations in acoustic pressure are less than for the circular piston source, considerable variations do take place with corresponding variations in

pressure at points off the axis as the distance from the source increases. A phase taper applied to the array will clearly modify the field pattern as well as cause the direction of maximum radiation to be deflected from the axis, and the computation of this change is not a straightforward process. Both attenuation and spreading effects have to be considered for each section of the array and although a linear phase taper may be the term applied to the phase difference between the pressure amplitude at the face of each section, it is in fact a uniformly-stepped phase taper which is used in practice. Experimental investigation related to the work by Freedman therefore seemed to be the most appropriate way in which to determine the practical application of the method, since the condition of zero phase taper approximates to a continuous piston source.

2. Equipment

Electronic equipment for producing a linear phase taper was designed according to the schematic diagram of Fig. 1. An electronic delay line is used to feed the sections of the transducer via a frequency changer M2 in each channel. The purpose of the delay line is to introduce a phase difference of equal amount between each section, which can be varied from $\theta = +\pi$ to $\theta = -\pi$ radians. This is achieved by applying a frequency of f_D which can be varied from f_1 to f_2 corresponding to a delay per section of $-\pi$ to $+\pi$ radians respectively. The frequency to be transmitted must be constant, and this is arranged by means of the modulators M1 and M2. The lower side-band of the single side-band modulator M1 has a frequency $f_D - f_0$ which on being modulated with f_D from the delay line, gives $2f_D - f_0$ and f_0 , etc. All frequencies except f_0 are filtered out before amplification and subsequent transmission. Varying f_D thus causes

† Formerly in Department of Electrical Engineering, University of Birmingham; now Head of the Department of Electrical Engineering, Lanchester College of Technology, Coventry.

‡ Formerly in Department of Electrical Engineering, University of Birmingham; now with the Royal Air Force.

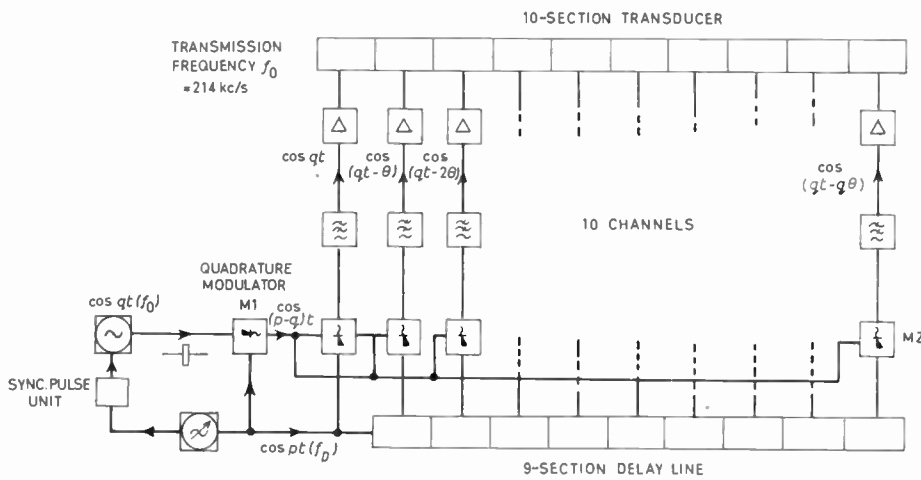


Fig. 1. Beam deflection system.

a linear phase taper to be applied to the transducer array since phase relations are maintained through the frequency changers.

The equipment performance was checked by adding the 10 outputs, via an adding network, and varying the delay line frequency. The result should be the well-known $(\sin nx)/n \sin x$ function shown plotted in

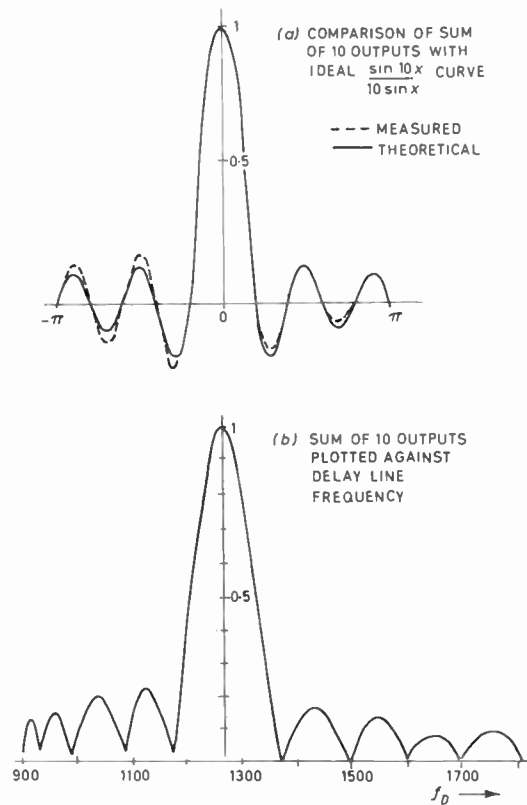


Fig. 2. Electronic performance of beam deflection system.

Fig. 2(a). Superimposed on this is the actual result obtained. In Fig. 2(b) the amplitude is plotted against the delay line frequency. It will be seen that there are minor errors present due almost entirely to the non-linear frequency-phase relationship in the delay line. These errors were accepted as one of the limitations of the system which could be improved with additional effort in design and manufacture. Whilst similar errors have little effect on the formation of a beam in the far field, the authors are not entirely satisfied at this stage that this can also be said about the near field.

An anechoic water tank was used for the field measurements. Although the ultimate aim is to operate either in solids or biological tissue, a liquid medium offered obvious advantages for field plotting. A ten-element transducer (Fig. 3) was designed to produce a Fresnel region which extended almost the length of the tank. The field probe was in the form of a thin rod bent through a right angle at the end so as to point towards the array, and the vertical portion of the rod was thinly coated with sound reflecting material. A barium-titanate transducer was coupled to the rod

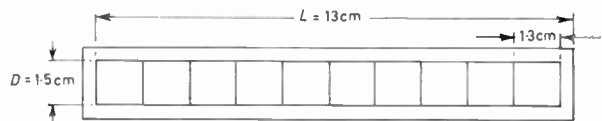


Fig. 3. Ten-section transducer.

above the water surface and the response field of the arrangement is shown in Fig. 4. The mountings of the sectionalized array and the probe were such that the probe could be accurately positioned anywhere in the Fresnel region within an arc of ± 15 deg up to a

distance of 70 cm from the array, and the latter could be rotated through ± 15 deg thus allowing a total coverage of ± 30 deg relative to the axis of the transducer.

3. Experimental Results

3.1. Comparison Between the Theoretical Near-Field Pattern and the Measured Pattern with No Phase Taper

The calculated field patterns for a rectangular piston source obtained by Freedman⁴ were used as the basis for comparison with the measured field patterns of the sectionalized array. Each section of the array had previously been checked independently for both the amplitude and the phase of the radiated signal to ensure that as far as practically possible the whole array approximated very closely to a piston source. Measurements were taken at 22.7 cm, 30 cm and 60.6 cm from the centre of the array corresponding to $X = 0.375, 0.5$ and 1 in Freedman's paper, and these are plotted in Fig. 5 together with the theoretical curves. Whilst agreement between the two sets of curves is far from perfect, it was considered to be sufficiently good to justify the use of the system for measurements when a phase taper was applied. Axial pressure variations are also plotted as the probe was traversed from near the transducer face towards the far end of the tank.

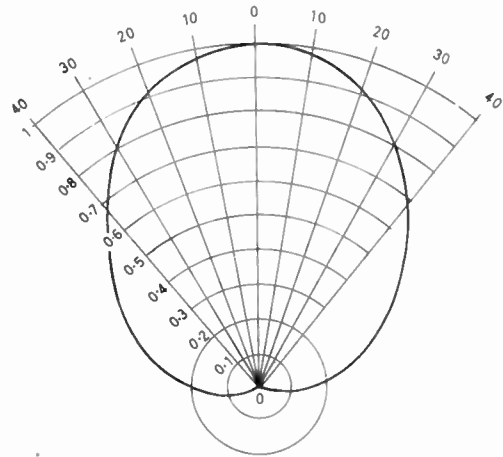


Fig. 4. Probe directivity.

3.2. Measured Field Pattern for a Linear Phase Taper

The curves of Fig. 6 show the results obtained for zero phase taper, a phase taper producing 7 deg deflection of the beam, and one producing 12 deg deflection of the beam when referred to far-field conditions.

Considerable distortion of the field pattern is seen to take place when a phase taper is applied, even at

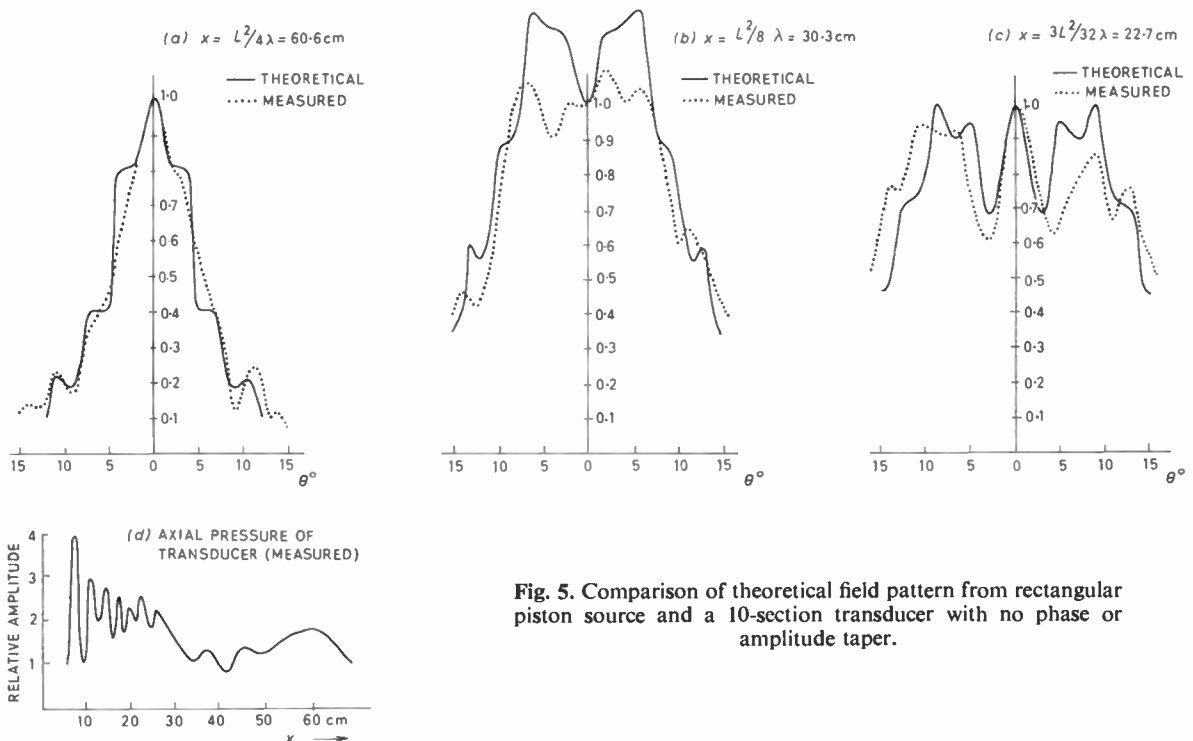


Fig. 5. Comparison of theoretical field pattern from rectangular piston source and a 10-section transducer with no phase or amplitude taper.

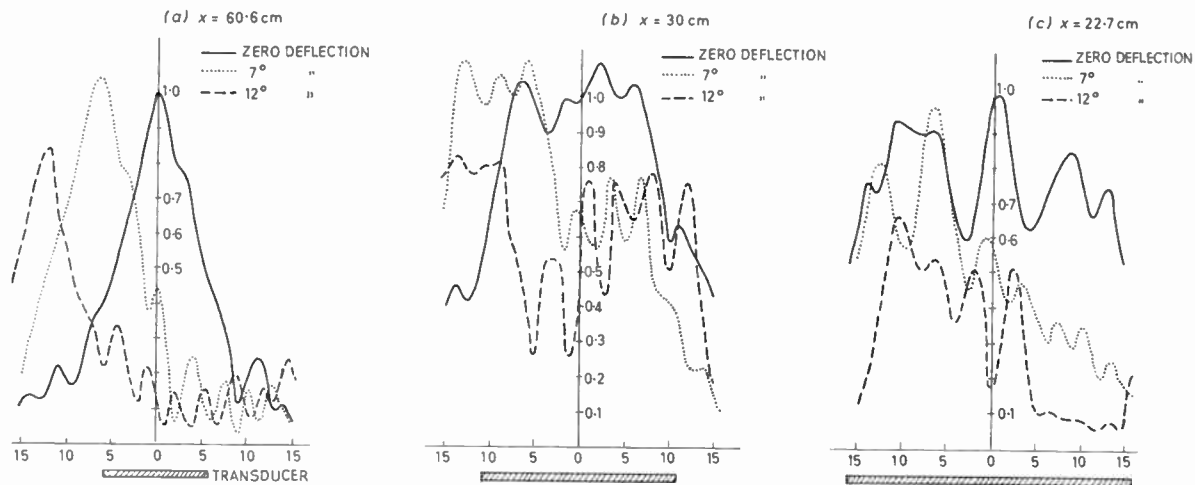


Fig. 6. Near-field pattern for a linear phase taper.

60 cm. It was quite clear that a serious limitation existed for ranges less than half the distance to the last axial pressure maximum, i.e. about 30 cm in this case.

3.3. *Measured Field Pattern for a Linear Phase Taper and a Symmetrical Linear Amplitude Taper*

It is well known however that an amplitude taper across the face of a transducer reduces the effects of diffraction in the far field region and Krautkramer⁵ has shown that pressure amplitude variations on the axis of a circular piston source in the Fresnel region can be considerably reduced by a Gaussian taper. An amplitude taper which reduced the radiated pressure in equal steps from the centre to the ends of the array was therefore used as an approximation to a linear taper where the amplitude of the vibrations at the ends would be zero.

The results obtained are shown in Fig. 7 together with pressure variations on the axis.

3.4. *Measured Field Pattern for a Gaussian Amplitude Taper and Zero Phase Taper*

A Gaussian amplitude taper was also used which had a value for *K* of 0.06735 in the equation

$$\text{pressure amplitude } P = \sqrt{\frac{K}{\pi}} \cdot \exp(-Ka^2)$$

(*a* = distance along transducer from centre)

The results obtained are shown in Fig. 8 together with the variations in axial pressure. Comparing these with the linear amplitude taper for zero phase taper it will be seen that the only significant change is in the pressure variations on the axis, indicating that operation at less than 22.7 cm may be possible.

4. **General Discussion**

The object of the investigation was to determine the change in the near-field pattern when a linear phase taper is applied to a sectionalized array, and relate this to the requirements of a scanning system used for plotting discontinuities in the propagating medium. It is seen from Fig. 6 that for distances greater than 60.0 cm in this case, the field pattern closely approximates to the concept of a beam of energy in the sense that there is a single maximum at 60.6 cm; as the distance increases the beam will more closely approach the well-known far-field pattern for a rectangular array. The conclusion is that electronic scanning of the beam is a feasible proposition when considered at distances greater than the last axial maximum, i.e. $L^2/4\lambda$ corresponding to 60.6 cm for the transducer used.

From Fig. 6, it is also clear however that for distances less than 60.6 cm the field pattern changes shape as the phase taper is increased, and this becomes more marked as the distance from the array is reduced. A serious limitation thus exists in the application of the system to the plotting of discontinuities which exist within this region. Take for example the curves for a distance of 30 cm. When the phase taper corresponds to a beam deflection of 12 deg, two distinct humps are formed which would introduce ambiguous results.

In Fig. 7 this does not occur, showing a distinct improvement as a result of an amplitude taper. A secondary hump is seen in Fig. 7(a) for a deflection angle of 12 deg. This is suspected as being excessive. A computer program is being set up to check this and other points. Nevertheless, operation at distances of 22 cm (approximately $3L^2/32\lambda$) and greater seems feasible.

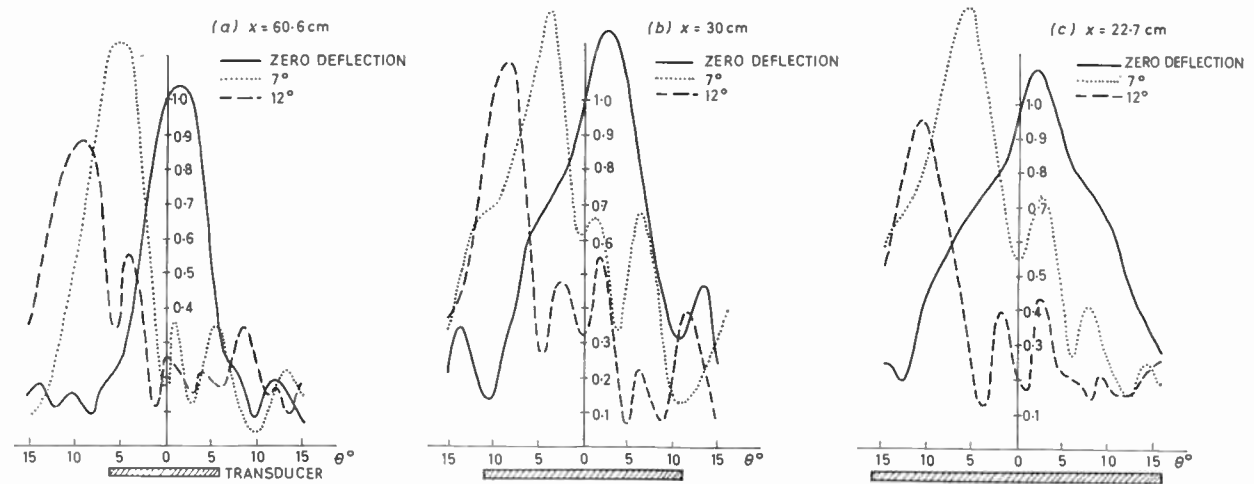


Fig. 7. Near-field pattern for a linear phase taper and a symmetrical linear amplitude taper.

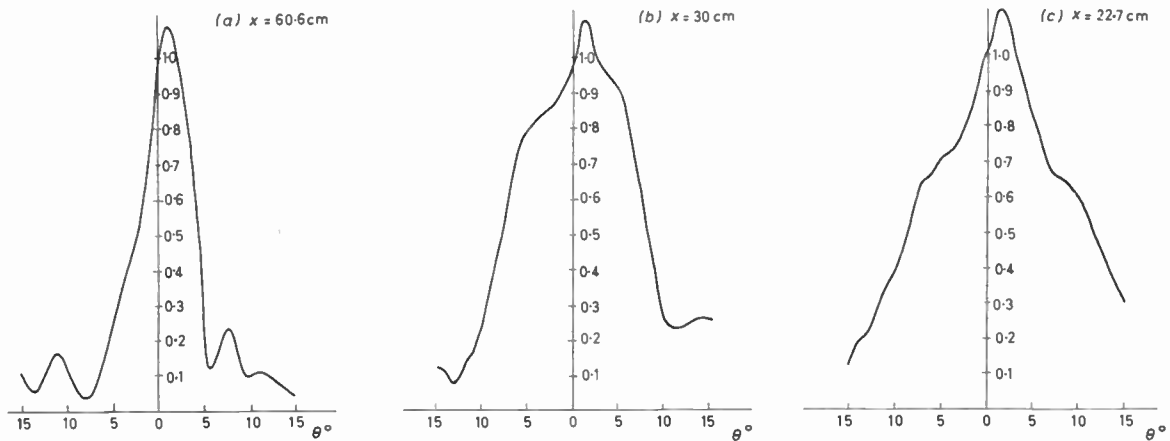
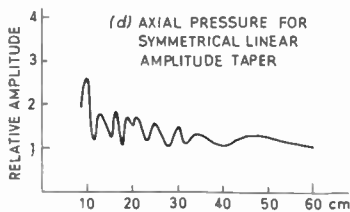
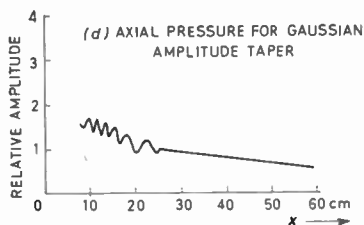


Fig. 8. Near-field pattern for a Gaussian amplitude taper.



An interesting feature about the curves is the width of the beam in terms of the transducer dimensions shown on the curves in relation to the point of measurement. When the field pattern is measured in the far field after applying an amplitude taper, it is well known

that the beam is widened and the secondary lobes reduced. In the near field the opposite takes place so far as the width of the beam is concerned, but the secondary lobes—if one can still use this term—are still reduced. It should therefore be possible to resolve

discontinuities separated by less than the transducer dimensions which has hitherto not been considered possible except under conditions of focusing.

The arrangement thus shows considerable promise for exploring the near field of an array in terms of discontinuities which may exist in the propagating medium. Application to non-destructive testing of solids would be useful in at least two ways. Defects are usually of irregular shape and may lie at an angle to the plane of the surface from which a test is made. The direction of maximum scatter may therefore be other than to the normal to the surface and as a result a large defect may be assessed as unimportant because of the small echo return. Deflection of the beam over an angle of ± 15 deg or so would reduce errors of this kind.

The extent of a defect can also be assessed without the need for movement of the transducer and a more accurate measurement made, since the arc over which returns are obtained can be related to the distance by a suitable display.

5. Acknowledgments

The authors wish to acknowledge the financial assistance of the Department of Scientific and Industrial Research; one of them (M. J. B.) held a D.S.I.R. Research Studentship. They also wish to express their appreciation for the generous facilities made available by Professor D. G. Tucker and the helpful discussions with colleagues in the Department of Electrical Engineering at the University of Birmingham.

6. References

1. D. G. Tucker, V. G. Welsby and R. Kendall, "Electronic sector scanning", *J. Brit. I.R.E.*, **18**, p. 465, 1958.
2. D. G. Tucker et al., "Underwater echo-ranging with electronic sector scanning: sea trials on R.R.S. *Discovery II*", *J. Brit. I.R.E.*, **19**, p. 681, 1959.
3. V. G. Welsby, "The angular resolution of a receiving aperture in the absence of noise", *J. Brit. I.R.E.*, **26**, p. 115, 1963.
4. A. Freedman, "Sound field of a rectangular piston", *J. Acoust. Soc. Amer.*, **32**, p. 197, 1960.
5. J. Krautkramer, "Determination of the size of a defect by the ultrasonic impulse echo method", *Brit. J. Appl. Phys.*, **10**, p. 240, 1959.

Manuscript received by the Institution on 4th April 1964. (Paper No. 971/EA19.)

© The Institution of Electronic and Radio Engineers, 1965

DISCUSSION

Under the chairmanship of Mr. W. K. Grimley, O.B.E.

Dr. M. I. Skolnik: Have you considered the focusing?

Dr. L. Kay (in reply): Focusing of a multi-element transducer is being considered as a further step in the research programme, and there appears to be no reason why this should not be possible. To be of any practical use, however, it will be necessary to vary the focal length and the direction of the focal region, and this will involve considerable complexity in the electronics. The aperture will almost certainly be small in terms of the number of wavelengths and the degree of resolution will, therefore, be severely limited.

Dr. R. Benjamin: Have you considered the use of an external spacer of length comparable to the Rayleigh distance, between the transducer and the object being examined?

Dr. Kay (in reply): A spacer between the transducer and the material under test is already used in many com-

mercial ultrasonic testing equipments for various reasons, and one is to reduce the effect of the near field. A water path is the most common form since this has many advantages and comes under the general heading of immersion testing. In this particular case we want to avoid the use of a spacer if at all possible so that the system will be more versatile.

Mr. R. Blommendaal: Have the authors considered measuring the amplitude distribution parallel to a phase front, instead of measuring it parallel to the transducer? It is my experience that a smoother amplitude distribution will be found when using the first method.

Dr. Kay (in reply): We had not considered measurement parallel to the phase front. This would undoubtedly show a smoother amplitude distribution but would not be related to a practical result. The measurement was made at a constant radius since the displayed information would normally be presented in this way.

Least-Squares Array Processing for Signals of Unknown Form

By

MORRIS J. LEVIN, Ph.D. †

Reprinted from the Proceedings of the Symposium on "Signal Processing in Radar and Sonar Directional Systems" held in Birmingham from 6th-9th July 1964.

Summary: Statistical methods are applied to the estimation of the velocity, arrival-angle and waveform of a signal appearing in an array of sensors in the presence of random noise. The signal is assumed to be a plane wave propagating through a linear, homogeneous, non-dispersive medium so that it is the same in each sensor except for a time delay due to its finite velocity. A novel formulation is introduced which requires no assumptions concerning the signal waveform but permits its estimation along with the vector of time delays per unit distance. The method is appropriate for applications such as seismology and passive sonar in which the signal waveform is unknown, yet cannot be realistically represented as a stationary random process (as is required, for example, by Wiener filtering theory).

A least-squares procedure is described which does not depend on any assumptions regarding the noise. This simple criterion is found to imply time-shift and sum processing which is related to other techniques previously employed. For known noise statistics the mean-square response of the processor to the noise is calculated and the covariance matrix of the estimates is approximated for the high signal-to-noise ratio case. The resulting array pattern is evaluated in terms of the signal spectrum and the array geometry. The results are compared with a more elaborate maximum-likelihood approach based on stationary Gaussian noise with a known spectral density matrix.

1. Introduction

Techniques are considered here for processing the outputs of an array of sensors perturbed by noise to obtain efficient estimates of the waveform of the received signal and its velocity arrival-angle. The accuracy attainable by these techniques is also investigated. Ideally, we would like to invoke statistical theory to determine estimators whose errors are, in some sense, as small as possible on the average. Unfortunately, decision theory has not yet provided a general method for solving this problem directly. Instead, we select a criterion which measures how well any particular choice of the signal waveform and its parameters fit the observed sensor outputs and deduce the processing which provides the values giving the best fit. It is emphasized that the suitability of the criterion depends on the knowledge available concerning the character of the signals and noise. The more detailed and accurate this knowledge is, the more efficient the processing which can be designed.

Various array processing techniques now in use are appropriate for different signal and noise situations.

† Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Massachusetts, U.S.A. (Dr. Levin was killed in a motor accident on 6th February 1965.)

In conventional radar the received signal is a known function except for several unknown parameters. The noise is additive, white, Gaussian and independent at each receiving element, the optimum processing is linear combining followed by likelihood detection or estimation procedures, and the situation is, on the whole, well understood. Various authors have suggested non-linear combining schemes to narrow the angular beamwidth of the array for a given number of elements.¹⁻⁴ However, it is clear that the departure from optimum processing degrades the performance of these schemes in detection and estimation as soon as the noise becomes appreciable. In addition, when more than one signal is present, the overall response becomes confused by cross-modulation terms.^{4,5}

A different situation arises when the signal is a random process having an unknown waveform but a known statistical description. Several early studies of array processing for this situation were conducted by Faran and Hills.⁶ Optimum processing techniques have been deduced and their performances analysed.⁷⁻¹¹ For detection and mapping of sources observed in radio astronomy multiplicative combining schemes have been very valuable.¹² An application to oceanography was described by Munk *et al.*¹³

Another technique, which requires the knowledge of the cross-spectral densities of the noise between all sensor pairs, but makes no assumptions regarding the signal, has been applied by Claerbout.¹⁴ He employed the noise statistics to design a linear processor for the sensor outputs which provides a minimum mean-squared-error prediction of the noise at some one sensor a short interval, say 0.3 second, ahead. This prediction is subtracted from the actual sensor output thus greatly reducing the noise level but distorting the signal waveform after the prediction interval.

The present study was motivated by an investigation of array processing applicable to seismology, a field in which a suitable representation of the signal is not readily apparent. Since the exact shape of the signal at a seismometer cannot be forecast before its arrival, it cannot be handled as a known waveform or even as a known function of some unknown parameters. On the other hand, the finite duration of the seismogram of an event and the variations in its nature as different phases arrive suggest that the representation as a stationary random process does not accurately mirror its essential character either. A more realistic formulation might be as a non-stationary random process, but the difficulty of measuring or postulating a suitable covariance function and the mathematical clumsiness of the analysis render this approach unattractive.

In this paper a novel formulation has been adopted which avoids the above difficulties. It is assumed that a single signal is present and that it is a plane wave propagating in a homogeneous, linear and non-dispersive medium. The signal waveform is taken to be the same in each of the sensors except for a time delay due to the finite propagation velocity. (The analysis applies to general three-dimensional arrays, but in seismology reflections from the earth's surface are present and the plane-wave signal model is accurate only for two-dimensional surface arrays.) The departure from previous treatments lies in the assignment of the signal waveform as a completely unknown time function which is to be estimated along with the components of the velocity vector. Thus no *a priori* assumptions regarding the shape of the signal are made. This model is thus appropriate for other applications such as passive sonar.

Although the least-squares estimates, described below, require no specific assumptions regarding the signal, an evaluation of their performance requires a detailed specification of the noise. In seismology, recent measurements¹¹ have shown that it is realistic to consider the noise as a random function of space and time which is zero-mean and time-stationary during observation intervals of duration many times that of typical signals. The cross-spectral densities between seismometer pairs can be obtained either

from experimental measurements or from calculations based on an assumed physical model. To carry out the statistical error evaluation, it is also necessary to assume that the noise is a multi-dimensional Gaussian process. The maximum-likelihood estimates in Section 5 are also based on this assumption. The observational evidence for the Gaussian assumption is very limited. Since the noise presumably results from the superposition of many individual disturbances after propagation through a linear filter (the Earth) the assumption is quite plausible from a theoretical point of view.

2. The Least-Squares Estimates

The mathematical model on which the least-squares estimates are based is now described. Let the sensors be enumerated by $k = 1, 2, \dots, K$ and let the location of the k th sensor be specified by its coordinates $r_k = [r_{k1}, r_{k2}, r_{k3}]$ in some convenient three-dimensional Cartesian system. Let $s(t)$ be the signal that would be observed by a sensor at the origin in the absence of noise. Then the output of the k th sensor is

$$x_k(t) = s(t - \alpha \cdot r_k) + w_k(t) \quad \dots\dots(1)$$

where $w_k(t)$ is the noise, and $\alpha = [\alpha_1, \alpha_2, \alpha_3]$ is the vector of delays per unit distance suffered by the signal as measured along each coordinate axis. It is assumed that the possible values of α are bounded, that the outputs of the sensors are observed for $0 \leq t \leq T$ and that the signal is non-zero over a finite interval which is entirely contained within $(0, T)$ for any possible α .

It is most convenient to work out the estimation theory in terms of α which can be related to other quantities more commonly employed. For a sinusoidal component of the signal with frequency ω rad/s, the vector wave number is

$$k = \omega\alpha \quad \dots\dots(2)$$

It is also possible to define a velocity vector v which is oriented in the direction of travel of the wave and has a magnitude $|v|$ equal to the scalar phase velocity (speed) of the wave. It is quickly seen that

$$v = \frac{\alpha}{|\alpha|^2} \quad \dots\dots(3)$$

However, the components of the velocity vector along each coordinate axis are not physically meaningful.

When the character of the noise is unknown, a reasonable and mathematically tractable criterion for estimates of α and $s(t)$ is to select them so as to best fit the observed data in the sense of minimizing the sum over all the sensors of the integrated squared differences

$$D(\alpha, s(t)) = \sum_{k=1}^K \int_0^T [x_k(t) - s(t - \alpha \cdot r_k)]^2 dt \quad \dots\dots(4)$$

To avoid end effects put

$$x_k(t) = 0 \text{ for } t \text{ outside } (0, T).$$

It is also assumed that the components of α lie within known finite ranges and that the duration of the signal is limited so that for any possible α , the $s(t - \alpha \cdot r_k)$ are zero for t outside $(0, T)$. Then the limits of integration in (4) can be replaced by $(-\infty, \infty)$ and by a change of variable

$$D(\alpha, s(t)) = \sum_{k=1}^K \int_{-\infty}^{\infty} [x_k(t + \alpha \cdot r_k) - s(t)]^2 dt \dots\dots(5)$$

To obtain the least-squares estimates, $D(\alpha, s(t))$ is first rewritten† in the form

$$\begin{aligned} D(\alpha, s(t)) &= \sum_k \int x_k^2(t + \alpha \cdot r_k) dt + K \left\{ \int s^2(t) dt - \right. \\ &\quad \left. - \frac{2}{K} \int s(t) \sum_k x_k(t + \alpha \cdot r_k) dt \right\} \\ &= \sum_k \int x_k^2(t) dt + \\ &\quad + K \int \left\{ s(t) - \frac{1}{K} \sum_k x_k(t + \alpha \cdot r_k) \right\}^2 dt - \\ &\quad - \frac{1}{K} \int \left\{ \sum_k x_k(t + \alpha \cdot r_k) \right\}^2 dt \dots\dots(6) \end{aligned}$$

For any arbitrary value of α the second term in (6) assumes its minimum value of zero by choosing as the conditional estimate for $s(t)$,

$$\hat{s}(t; \alpha) = \frac{1}{K} \sum_k x_k(t + \alpha \cdot r_k) \dots\dots(7)$$

D now depends only on $\hat{\alpha}$ and is minimized by the value $\hat{\alpha}$, which maximizes the quantity

$$C(\alpha) = \int \left\{ \sum_k x_k(t + \alpha \cdot r_k) \right\}^2 dt \dots\dots(8)$$

Substituting $\hat{\alpha}$ in (7) provides the final least-squares estimate of $s(t)$,

$$\hat{s}(t) = \frac{1}{K} \sum_k x_k(t + \hat{\alpha} \cdot r_k) \dots\dots(9)$$

The residual value of D is then

$$D(\hat{\alpha}, \hat{s}(t)) = \sum_k \int x_k^2(t) dt - \frac{1}{K} C(\alpha) \dots\dots(10)$$

Thus $C(\alpha)$ is obtained by squaring the sum of the time-shifted sensor outputs and integrating while $s(t)$ is the average of the sensor outputs each time-shifted by $\alpha \cdot r_k$.

Relationships of the least-squares estimates to other methods of processing can be seen by writing

† All integrals are definite integrals taken over the limits $(-\infty, \infty)$ unless otherwise noted. All summations are taken over the limits $k = 1, \dots, K$ or $l = 1, \dots, K$ unless otherwise noted.

$$\begin{aligned} C(\alpha) &= \sum_{k,l} \int x_k(t + \alpha \cdot r_k) x_l(t + \alpha \cdot r_l) dt \\ &= \sum_k \int x_k^2(t) dt + \\ &\quad + \sum_{k \neq l} \int x_k(t + \alpha \cdot r_k) x_l(t + \alpha \cdot r_l) dt \dots\dots(11) \end{aligned}$$

Hence, maximizing $C(\alpha)$ over α is equivalent to maximizing the sum of the cross-correlation functions measured between all disjoint pairs of sensors. The maximization of various sums of cross-correlation functions has frequently been suggested as a method for array processing. (See, for example, references 15-18.) However, these suggestions have generally been on an *ad hoc* basis and the fact that the maximization of (8) or (11) provides a least-squares fit to a plane wave signal has not been appreciated. By taking the Fourier transforms of these cross-correlation functions, it can also be shown that the method used by Munk *et al.*¹³ to fit the wave number of a single wave train to the spectral density matrix measured among the elements of an oceanographic array is equivalent to applying the least-squares processing to the sensor outputs after they have been narrow-band filtered.

An example of least-squares processing of seismic data carried out on a digital computer may be of interest. In Fig. 1 are the outputs of five seismometers of an eleven-element linear array recording an earthquake at an epicentral distance of 1850 km. The overall record length displayed is 75 seconds. The one component α is measured along the array. $C(\alpha)$ was computed for four successive sections of the array outputs as indicated in Fig. 1. The first section produced the plot of Fig. 2. The delay increment ('moveout') per point is 25 ms/km. The peak value, corresponding to α , stands out and could be located even more precisely by a finer plot. The next two sections produced similar plots. However, the plot for the fourth section is different (Fig. 3). There are apparently two arrivals with different projected velocities along the array. It is unlikely this conclusion could be reached by visual examination of the original records alone.

It is well known that the least-squares method is equivalent to the maximum-likelihood method when the noise components are Gaussian, independent between sensors, and white. If, instead of being white, they have a common spectral density $P(\omega)$, then the maximum-likelihood estimate of α is obtained by passing the $x_k(t)$ through whitening filters with transfer functions $1/|\sqrt{P(\omega)}|$ before carrying out the least-squares processing. Thus the least-squares procedure, while reasonable for many kinds of noise, is also optimum in the sense of being maximum-likelihood

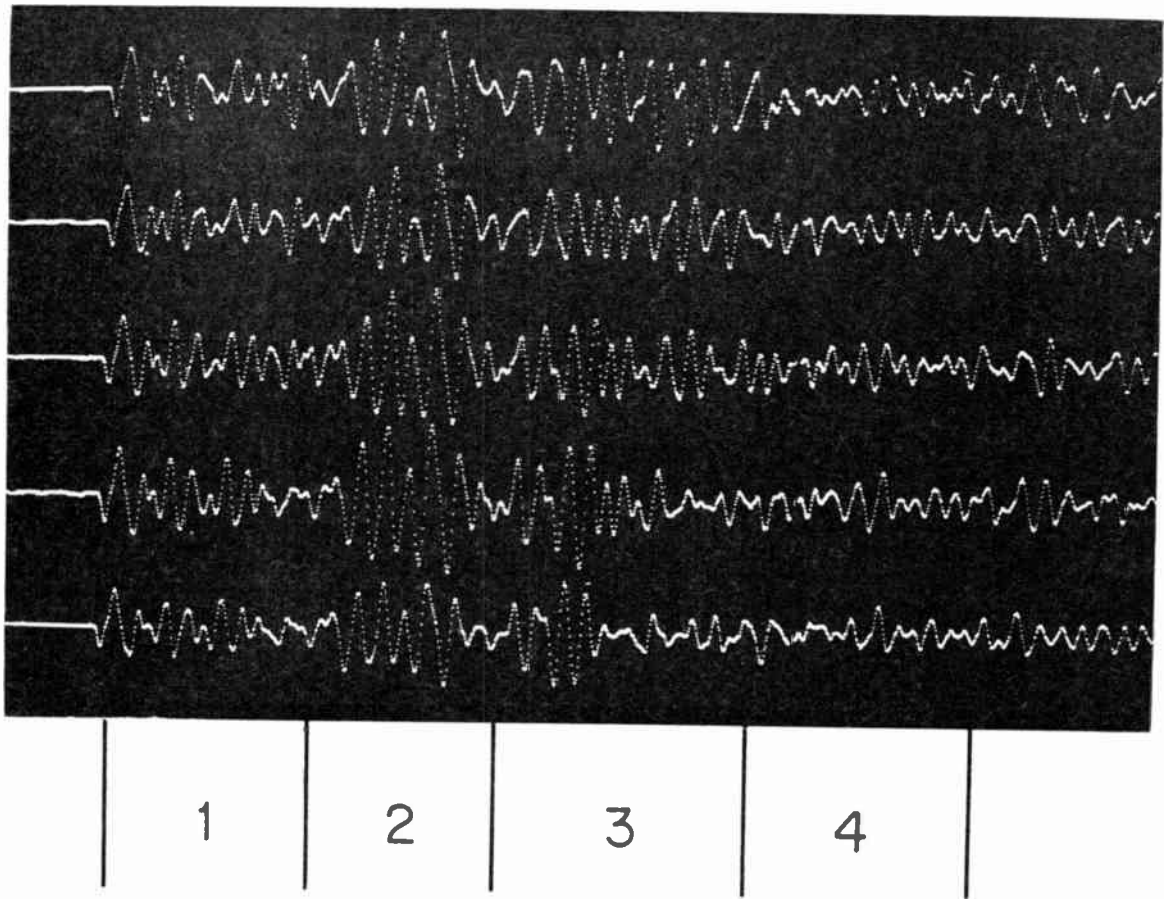


Fig. 1. Earthquake seismograms from five elements of a linear eleven-element array.

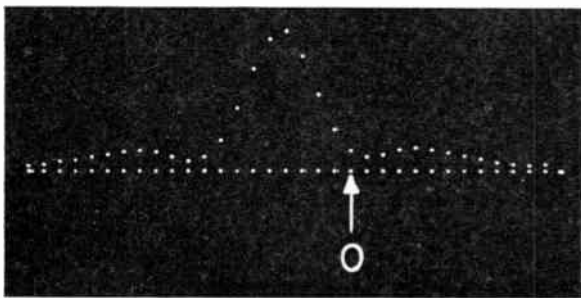


Fig. 2. Array response, $C(\alpha)$, for the first three sections of the seismogram. The α interval is 25 ms/km per point.

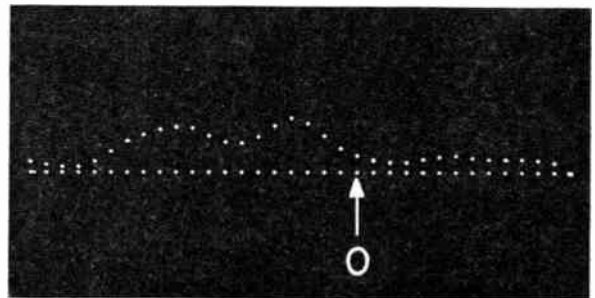


Fig. 3. Array response, $C(\alpha)$, for the fourth section of the seismogram.

for Gaussian noise independent between sensors. In the following sections we slightly generalize the processing to permit the inclusion of a whitening filter by assuming that the summed sensor outputs are passed through a linear filter with transfer function $H(\omega)$ before $C(\alpha)$ is formed.

The question arises as to whether this least-squares approach could be generalized to the estimation of two or more separate plane-wave signals. Indeed, we have been able to work out the analogous two-signal case but the results at present seem too involved to be of practical value, so they are not presented here.

3. Accuracy of the Least-Squares Estimates

The covariance matrix of the three components of $\hat{\alpha}$ (i.e. the moment matrix of the estimation errors) can be approximated for the case of stationary noise and a high signal/noise ratio. The basic idea is to represent $C(\alpha)$ by a quadratic surface in the vicinity of the true parameter values.¹⁹ The displacement of the peak of the surface by the noise is determined, terms of first order in the noise being retained. The details are too complicated to reproduce here so only the final results are presented. Let $S(\omega)$ be the Fourier transform of $s(t)$ and let the cross-spectral density of the noise between sensors k and l be denoted by

$$P_{kl}(\omega) = \int \varphi_{kl}(\tau) e^{-j\omega\tau} d\tau \quad \dots\dots(12)$$

where

$$\varphi_{kl}(\tau) = E w_k(t) w_l(t + \tau) \quad \dots\dots(13)$$

and E is the expectation or ensemble average. The covariance of $\hat{\alpha}_i$ and $\hat{\alpha}_j$ is denoted by

$$\text{Cov } \hat{\alpha}_i, \hat{\alpha}_j = E(\hat{\alpha}_i - E\hat{\alpha}_i)(\hat{\alpha}_j - E\hat{\alpha}_j) \quad i, j = 1, 2, 3 \quad \dots\dots(14)$$

Then after extensive calculations we find for the matrix of covariances

$$[\text{Cov } \hat{\alpha}_i, \hat{\alpha}_j] \simeq [A]^{-1} [B] [A]^{-1} \quad \dots\dots(15)$$

where the matrix (A) has elements

$$A_{ij} = \left\{ \int \omega^2 |S(\omega)|^2 |H(\omega)|^2 \frac{d\omega}{2\pi} \right\} \left\{ \sum_k (r_{ki} - \bar{r}_i)(r_{kj} - \bar{r}_j) \right\} \quad \dots\dots(16)$$

and the matrix (B) has elements

$$B_{ij} = \int \left\{ \omega^2 |S(\omega)|^2 |H(\omega)|^4 \sum_{k,l} (r_{ki} - \bar{r}_i) \times (r_{lj} - \bar{r}_j) \psi_{kl}(\omega) \right\} \frac{d\omega}{2\pi} \quad \dots\dots(17)$$

In these expressions the centre of gravity of the seismometer locations is represented by

$$\bar{r}_i = \frac{1}{K} \sum_k r_{ki} \quad \dots\dots(18)$$

and we also define

$$\psi_{kl}(\omega) = \{ \exp[-j\omega\alpha \cdot (r_k - r_l)] \} \cdot P_{kl}(\omega) \quad \dots\dots(19)$$

These results simplify considerably when the noise is independent between sensors with a common spectral density $P(\omega)$. Let us further assume the presence of a whitening filter $H(\omega) = 1/|\sqrt{P(\omega)}|$. The previous results reduce to

$$[\text{Cov } \hat{\alpha}_i, \hat{\alpha}_j] = \frac{1}{\Omega} [\rho]^{-1} \quad \dots\dots(20)$$

where the matrix (ρ) has elements

$$\rho_{ij} = \sum_k (r_{ki} - \bar{r}_i)(r_{kj} - \bar{r}_j) \quad \dots\dots(21)$$

and

$$\Omega = \int \frac{\omega^2 |S(\omega)|^2 d\omega}{P(\omega) 2\pi} \quad \dots\dots(22)$$

Surprisingly, it has been found²⁰ that in this independent noise case the results are the same as those obtained for the corresponding case in which the signal waveform is known exactly except for amplitude and arrival time. However, this is not true in general when the noise components are not independent between sensors.

A translation and rotation of the coordinate system can now be performed so that for the new coordinates r'_{ki} and all i, j ,

$$\left. \begin{aligned} \sum_k r'_{ki} &= 0 \\ \sum_k r'_{ki} r'_{kj} &= 0 \quad i \neq j \end{aligned} \right\} \quad \dots\dots(23)$$

Then for the estimate of the new delay vector α' , (20) becomes

$$\left. \begin{aligned} \text{Var } \hat{\alpha}'_i &= \frac{1}{\Omega \sum_k (r'_{ki})^2} \\ \text{Cov } (\hat{\alpha}'_i, \hat{\alpha}'_j) &= 0, \quad i \neq j \end{aligned} \right\} \quad \dots\dots(24)$$

The numbers

$$L_i = \sqrt{\frac{1}{K} \sum_k (r'_{ki})^2} \quad \dots\dots(25)$$

are measures of the size of the array in each of the three dimensions. They can be expressed as

$$L_i = l_i D_i \quad \dots\dots(26)$$

where D_i is the overall extent of the array as measured along the i th coordinate axis and l_i is a form factor depending only on the array geometry. For example, if the K sensors form a uniformly spaced linear array along the i axis, with an overall length D_i , then summation of a simple series shows that

$$l_i = \sqrt{\frac{K+1}{12(K-1)}} \quad \dots\dots(27)$$

The quantity Ω can be expressed in terms of the signal/noise ratio at a single sensor

$$\gamma = \int_{-\infty}^{\infty} \frac{|S(\omega)|^2 d\omega}{P(\omega) 2\pi} = 2 \int_0^{\infty} \frac{|S(\omega)|^2 d\omega}{P(\omega) 2\pi} \quad \dots\dots(28)$$

the mean frequency

$$\bar{\omega} = \frac{2}{\gamma} \int_0^{\infty} \frac{\omega |S(\omega)|^2 d\omega}{P(\omega) 2\pi} \quad \dots\dots(29)$$

and the second moment about the mean frequency

$$\mu_2 = \frac{2}{\gamma} \int_0^{\infty} \frac{(\omega - \bar{\omega})^2 |S(\omega)|^2 d\omega}{P(\omega) 2\pi} \quad \dots\dots(30)$$

It is quickly seen that

$$\Omega = \gamma[(\bar{\omega})^2 + \mu_2] \quad \dots\dots(31)$$

In the narrow-band case (appropriate to radar) the width of the spectrum is small compared with the centre frequency and only the $\gamma(\bar{\omega})^2$ term is significant. In the wide-band case (appropriate to seismology), μ_2 will not be negligible although it will usually be smaller than $(\bar{\omega})^2$. For example, if $|S(\omega)|^2/P(\omega)$ is rectangular so that it has a constant level for $|\omega| \leq \omega_a$ and is zero elsewhere then

$$(\bar{\omega})^2 = \omega_a^2/4 \quad \dots\dots(32)$$

and

$$\mu_2 = \omega_a^2/12 \quad \dots\dots(33)$$

so

$$\Omega = \gamma\omega_a^2/3$$

Hence, for a fixed value of γ the wideband character of the signal produces only a 25% reduction of the variance from what it would be if the signal energy were concentrated at the mean frequency.

A further interpretation of Ω is in terms of the output, $u(t)$, of the whitening filter $1/|\sqrt{P(\omega)}|$ when the input is $s(t)$. Then

$$\Omega = \int \left[\frac{d}{dt} u(t) \right]^2 dt \quad \dots\dots(34)$$

This expression shows quantitatively how a signal with a sharply fluctuating waveform (after whitening) produces a smaller variance than a smoother waveform since its passage across the array is more recognizable.

Turning to the errors in the estimated waveform $\hat{s}(t)$ as given by (9), we observe that they arise from two causes. First, the estimate $\hat{\alpha}$ used in (9) is in error and second, even if α were known exactly, there still remains a noise term due to the sum of the additive noise components $w_k(t)$. The covariance matrix for $\hat{\alpha}$ has been discussed above; no simple relationship has been found between this covariance matrix and the error in $\hat{s}(t)$. However, the error in $\hat{s}(t)$ due to the second cause mentioned above can be readily evaluated for any stationary noise. The spectral density of the additive noise component of $\hat{s}(t)$ is readily found to be

$$\zeta(\omega) = \frac{1}{K^2} \sum_{k,i} \psi_{ki}(\omega) \quad \dots\dots(35)$$

so the mean-square value of this noise component is

$$\int \zeta(\omega) \frac{d\omega}{2\pi} \quad \dots\dots(36)$$

Experience with variance calculations of this type shows that these expressions accurately represent the relative effects of the various quantities involved but that the constant factor gives an overall result which is on the optimistic side. However, by careful design one can attain performance which is well within an

order of magnitude of this ideal and sometimes it can be approached quite closely.

4. The Pattern of an Array

This section examines the effect of the signal spectrum on the array pattern in a manner somewhat different from the usual analysis in frequency-wave number space. With least-squares processing an array is pointed in different directions by adjustment of the time shifts $\alpha \cdot r_k$. In considering the response of the array to a signal from a specific direction α_0 as a function of α it is convenient to take as a measure of this response the value of $C(\alpha)$ with noise absent. By (1) and (8)

$$C(\alpha) = \int \left[\sum_k s(t + (\alpha - \alpha_0) \cdot r_k) \right]^2 dt \quad \dots\dots(37)$$

or by Parseval's Theorem

$$C(\alpha) = \int |S(\omega)|^2 |A(\omega; \alpha - \alpha_0)|^2 \frac{d\omega}{2\pi} \quad \dots\dots(38)$$

where

$$A(\omega; \alpha - \alpha_0) = \sum_k \exp[j\omega(\alpha - \alpha_0) \cdot r_k] \quad \dots\dots(39)$$

This expression can be generalized in an obvious way if the sensor outputs are weighted in amplitude.

The array pattern is seen to depend only on the signal waveform, the array geometry and the difference $(\alpha - \alpha_0)$. Corresponding to (7) the spectrum of the summed output is

$$\tilde{S}(\omega; \alpha) = S(\omega) A(\omega; \alpha - \alpha_0) \quad \dots\dots(40)$$

so $A(\omega; \alpha - \alpha_0)$ can be thought of as the frequency response of the array. $|A(\omega; \alpha - \alpha_0)|^2$ has a main lobe of amplitude K^2 at $\omega = 0$, and a side-lobe structure away from this peak. The factor $(\alpha - \alpha_0)$ changes only the scale of $A(\omega; \alpha - \alpha_0)$ relative to the ω -axis, a smaller value expanding $A(\omega; \alpha - \alpha_0)$ about the origin and a larger value shrinking it. If $\alpha - \alpha_0 = 0$, then $|A(\omega; 0)|^2 = K^2$ for all ω .

To study the array pattern, α_0 can be set equal to zero; this represents a signal arriving at all sensors simultaneously. Then for a given orientation of α , $A(\omega; \alpha)$ is the same as for an equivalent linear array in which all the sensors are projected on a base-line having this orientation. The reason for this is that all sensors on a given line perpendicular to this base-line have the same value of time shift $\alpha \cdot r_k$.

Some features of the array pattern are brought out by the example of a uniform linear array with spacing d . It is found that

$$|A(\omega; \alpha - \alpha_0)|^2 = \left[\frac{\sin \omega(\alpha - \alpha_0) Kd/2}{\sin \omega(\alpha - \alpha_0) d/2} \right]^2 \quad \dots\dots(41)$$

This is sketched in Fig. 4 for $K = 6$. The distance between each principal maximum and the adjacent

zero is

$$\omega = \frac{2\pi}{Kd(\alpha - \alpha_0)} \quad \dots\dots(42)$$

and the response is periodic with period

$$\omega = \frac{2\pi}{d(\alpha - \alpha_0)} \quad \dots\dots(43)$$

Pointing the array at a particular signal to maximize $C(\alpha)$ consists of setting $\alpha = \alpha_0$, so that $|A(\omega; \alpha - \alpha_0)|^2$ in (38) has its maximum value, K^2 , at all ω . The higher the frequencies to which $S(\omega)$ extends, the more rapidly $C(\alpha)$ drops off as α departs from α_0 and the better the resolution. These high frequency components also produce ambiguities in $C(\alpha)$ for $\alpha \neq \alpha_0$ as other maxima of $|A(\omega; \alpha - \alpha_0)|^2$ sweep across them. There is always a lobe of $|A(\omega; \alpha - \alpha_0)|^2$ around $\omega = 0$ so components of $S(\omega)$ in this vicinity produce a residual level in $C(\alpha)$ which is present for all values of α . Suitable filtering of the signal before forming $C(\alpha)$ to remove low- and high-frequency components can improve the pattern of the array. However, a filter based on these considerations is not necessarily the same as one which will minimize the variances of the $\hat{\alpha}_i$ as given by (15) in which the noise statistics also play a part.

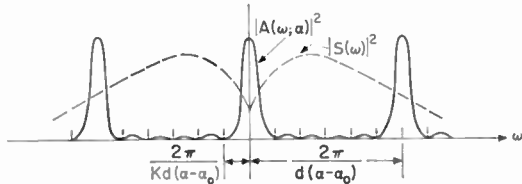


Fig. 4. The array frequency response, $|A(\omega; \alpha - \alpha_0)|^2$, for a uniformly spaced linear six-element array. The total array response, $C(\alpha)$, is the integral of $|S(\omega)|^2 |A(\omega; \alpha - \alpha_0)|^2$.

When there are two signals $S_1(\omega)$ and $S_2(\omega)$ with true time delay vectors, α_{01} and α_{02} , the overall response is

$$C(\alpha) = C_1(\alpha) + C_2(\alpha) + 2\text{Re} \left\{ \int S_1(\omega) A_1(\omega) S_2(-\omega) A_2(-\omega) \frac{d\omega}{2\pi} \right\} \quad \dots\dots(44)$$

where $C_1(\alpha)$ and $C_2(\alpha)$ are the responses to the signals individually and $A_1(\omega)$ and $A_2(\omega)$ are the corresponding array frequency responses. The presence of the cross-product term may make it difficult to interpret the resulting array output. However, this term will be small for values of α for which $A_1(\omega)$ and $A_2(-\omega)$ fail to overlap or if $S_1(\omega)$ and $S_2(-\omega)$ are unrelated so their average product is close to zero.

5. The Maximum-Likelihood Estimates

When more detailed knowledge of the noise process is available, a criterion more specialized than that of least-squares can be applied. Suppose that the noise outputs at the sensors form a time-stationary vector Gaussian random process having zero mean and a known spectral density matrix with elements $P_{kl}(\omega)$ as defined in (12). This information makes it possible to apply the method of maximum-likelihood, which is known to have certain optimal asymptotic properties. In the case of a finite number of parameters maximum-likelihood estimators are equivalent to Bayes' estimators based on uniform *a priori* probability densities. In the present case, however, in addition to the parameter vector α , the signal waveform $s(t)$ is to be estimated and it is not clear how a uniform *a priori* probability density could be assigned to each signal waveform to make possible the calculation of Bayes' estimates.

Maximum-likelihood estimators of the signal waveform $s(t)$ and time delay vector α have been derived in reference 20 with the further simplifying assumption that the observation interval T is long compared with the effective duration of the covariance functions of the noise components. When the noise is independent between sensors with a common spectral density then the maximum-likelihood estimators are the same as the least-squares estimators with a whitening filter. For a more general noise structure the maximum-likelihood estimators incorporate more complicated filtering operations with a consequent improvement in performance. The derivation of these estimates, which is a generalization of that given in Section 2 for the least-squares estimates, will be briefly sketched here.

Let $X_k(\omega)$ be the Fourier transform of $x_k(t)$, let $[Q_{kl}(\omega)]$ be the matrix which is the inverse of $[P_{kl}(\omega)]$ and assume it exists for all ω , let

$$\epsilon_{kl}(\omega) = Q_{kl}(\omega) \exp[-j\omega\alpha \cdot (r_k - r_l)] \quad \dots\dots(45)$$

and let

$$\Delta(\omega) = \sum_{k,l} \epsilon_{kl}(\omega) \quad \dots\dots(46)$$

The logarithm of the likelihood function, which is to be maximized over α and $S(\omega)$, is found to be

$$\Lambda(\alpha; S(\omega)) = -\frac{1}{2} \sum_{k,l} \int \{ \epsilon_{kl}(\omega) [S(\omega) - \exp(j\omega\alpha \cdot r_k) X_k(\omega)] [S^*(\omega) - \exp(-j\omega\alpha \cdot r_l) X_l^*(\omega)] - Q_{kl}(\omega) X_k(\omega) X_l^*(\omega) \} \frac{d\omega}{2\pi} + \text{constant} \quad \dots\dots(47)$$

This is a quadratic function of the $x_k(t)$ and is a generalization of (5). The function $\hat{S}(\omega; \alpha)$ which maximizes $\Lambda(\alpha; S(\omega))$ for any arbitrary α is found

explicitly as

$$\tilde{S}(\omega; \alpha) = \frac{\sum_{k,l} \varepsilon_{kl}(\omega) \exp[j\omega \alpha \cdot r_k] X_k(\omega)}{\Delta(\omega)} \dots\dots(48)$$

This is substituted back into (47) giving

$$\Lambda(\alpha; \tilde{S}(\omega)) = \frac{1}{2} \int \frac{\left| \sum_{k,l} \varepsilon_{kl}(\omega) \exp[j\omega \alpha \cdot r_k] X_k(\omega) \right|^2 d\omega}{\Delta(\omega)} \frac{d\omega}{2\pi} \dots\dots(49)$$

which must be maximized over α by successive approximations to determine $\hat{\alpha}$. Then $\hat{\alpha}$ is substituted back into (48) to give the final maximum-likelihood estimate of $S(\omega)$,

$$\hat{S}(\omega) = \tilde{S}(\omega; \hat{\alpha}) \dots\dots(50)$$

The measurement of $\Lambda(\alpha; \tilde{S}(\omega))$ for each trial value of α can be carried out by passing the sensor outputs through filters (different for each sensor and each value of α , in general) summing, squaring, and integrating over the observation interval. $\tilde{S}(\omega)$ is then found by time-shifting the sensor outputs according to $\hat{\alpha}$, passing them through a different set of filters and summing.

The covariance matrix for $\hat{\alpha}$ can be approximated for the high signal/noise ratio case. There is obtained

$$[\text{Cov } \hat{\alpha}_i, \hat{\alpha}_j] \simeq [M]^{-1} \dots\dots(51)$$

where the elements of the matrix (M) are

$$M_{ij} = \int \left\{ \omega^2 |S(\omega)|^2 \sum_{k,l} \varepsilon_{kl}(\omega) [r_{ki} - \bar{r}_i(\omega)] \times [r_{lj} - \bar{r}_j(\omega)]^* \right\} \frac{d\omega}{2\pi} \dots\dots(52)$$

and

$$\bar{r}_i(\omega) = \frac{1}{\Delta(\omega)} \sum_{k,l} \varepsilon_{kl}(\omega) r_{ki} \dots\dots(53)$$

is a frequency-dependent weighted centre of gravity. Finally, the mean-square response of the array to the noise alone is

$$\int \frac{1}{\Delta(\omega)} \frac{d\omega}{2\pi} \dots\dots(54)$$

This last expression is also valid for non-Gaussian noise.

6. Conclusions

Some tentative conclusions can be drawn from work currently in progress in which these methods are being applied to the processing of seismic signals. The assumption of a plane-wave model is valid for the initial arrivals of the different seismic phases. The least-squares processing has been found useful for analysis of direction of arrival and velocity even when several different signals are present. The design and realization of the maximum-likelihood processing is rather complicated and whether it provides an im-

provement over least-squares depends on the structure of the noise field. A practical method for obtaining the maximum-likelihood estimates by digital computation has been developed and has achieved a worthwhile gain in performance with actual seismic data.

7. Acknowledgments

Dr. E. J. Kelly worked out important parts of the analysis, especially the maximum-likelihood estimates. Dr. P. E. Green, Jr. provided many helpful suggestions. Philip and Laurice Fleck prepared the digitally computed example from data furnished by United Electro Dynamics, Inc., Alexandria, Virginia. The work was supported by the U.S. Advanced Research Projects Agency.

8. References

1. A. Berman and C. S. Clay, "Theory of time-average-product arrays", *J. Acoust. Soc. Amer.*, 29, No. 7, pp. 805-12, July 1957.
2. V. G. Welsby and D. G. Tucker, "Multiplicative receiving arrays", *J. Brit. I.R.E.*, 19, No. 6, pp. 369-82, June 1959.
3. D. G. Tucker, "Sonar arrays, systems, and displays", Lecture 2 in "Underwater Acoustics", ed. by V. M. Albers (Plenum Press, New York, 1963).
4. M. E. Pedinoff and A. A. Ksienski, "Multiple target response of data processing antennas", *I.R.E. Trans. on Antennas and Propagation*, AP-10, No. 1, pp. 112-126, January 1962.
5. R. H. MacPhie, "Comments on 'Multiple target response of data-processing systems'", *I.R.E. Trans. on Antennas and Propagation*, AP-10, No. 5, pp. 642-3, September 1962. (Letter to the Editor.)
6. J. J. Faran, Jr. and R. Hills, Jr., "The Application of Correlation Techniques to Acoustic Receiving Systems", Tech. Memo. No. 28, Acoustics Research Laboratory, Harvard University (November 1, 1952).
7. C. J. Drane and G. B. Parrent, "On the mapping of extended sources with non-linear correlation antennas", *I.R.E. Trans. on Antennas and Propagation*, AP-10, No. 1, pp. 126-30, January 1962.
8. F. Bryn, "Optimum signal processing of three-dimensional arrays operating on Gaussian signals and noise", *J. Acoust. Soc. Amer.*, 34, No. 3, pp. 289-97, March 1962.
9. I. J. Good, "Weighted covariance for detecting the direction of a Gaussian source", "Time Series Analysis", ed. by M. Rosenblatt (John Wiley, New York, 1963).
10. J. B. Burg, M. M. Backus and L. Strickland, "Seismometer Array and Data Processing System", Texas Instruments, Inc., AFTAC Project VT/077, Final Report, Phase I (December 1, 1961).
11. M. Backus *et al.*, "Spatial Characteristics of Ambient Short Period Seismic Noise and Wide-Band Extraction of P-Waves", 33rd Annual Int'l Mtg., Soc. of Exploration Geophysicists, New Orleans, La. (October 20-24, 1963).
12. B. Y. Mills *et al.*, "A high resolution radio telescope for use at 3.5 m", *Proc. Inst. Radio Engrs*, 46, pp. 67-84, January 1958.
13. W. H. Munk *et al.*, "Directional recording of swell from distant storms", *Phil. Trans. Roy. Soc. London*, 255, Series A, No. 1062, pp. 505-84, April 18, 1963.

14. J. Claerbout, "Detection of P-Waves from Weak Sources at Great Distances", 33rd Annual Int'l Mtg., Soc. of Exploration Geophysicists, New Orleans, La. (October 20-24, 1963).
15. D. C. Fakley, "Comparison between the performance of a time averaged product array and an intraclass correlator", *J. Acoust. Soc. Amer.*, 31, No. 3, pp. 1307-14, October 1959.
16. A. Ryall, "Improvement of array seismic recordings by digital processing", *Bull. Seism. Soc. Amer.*, 54, No. 1, pp. 277-94, February 1964.
17. K. McCamy and R. P. Meyer, "A correlation method of apparent velocity measurement", *J. Geophys. Res.*, 69, No. 4, pp. 691-700, February 1964.
18. V. Baranov and C. H. Picou, "Energy and vector record-sections", *Geophysics*, 29, No. 1, 17-37, February 1964.
19. E. J. Kelly, I. S. Reed and W. L. Root, "The detection of radar echoes in noise", Part I, *J. Soc. Indust. Appl. Math.*, 8, No. 2, pp. 309-41, June 1960, and Part II, *J. Soc. Indust. Appl. Math.*, 8, No. 3, pp. 481-507, September 1960.
20. E. J. Kelly and M. J. Levin, "Signal Parameter Estimation for Seismometer Arrays", Tech. Report 339, Lincoln Laboratory, M.I.T. (January 8, 1964, DDC 435489).
21. L. E. Brennan, "Angular accuracy of a phased array radar", *I.R.E. Trans. on Antennas and Propagation*, AP-9, No. 3, pp. 268-75, May 1961.

9. Appendix: Examples of Accuracy Computations

Some further calculations based on eqn. (20) are presented here. In the narrow-band case

$$\bar{w} \simeq 2\pi |v|/\lambda$$

where $|v|$ is the speed of propagation of the signal and λ is the carrier wavelength. Then (22) becomes

$$\text{Var } \hat{\alpha}_i = \frac{\lambda^2}{4\pi^2 |v|^2 \gamma K L_i^2}$$

Recall that

$$\alpha_i = \frac{\cos \theta_i}{|v|}$$

where the $\cos \theta_i$ are the direction cosines of the incident wave. When $|v|$ is known, as for a specified seismic phase and emergence angle, there follows

$$\text{Var}(\cos \hat{\theta}_i) = \frac{\lambda^2}{4\pi^2 \gamma K L_i^2}$$

If the errors are small

$$\text{Var}(\cos \hat{\theta}_i) \simeq \sin^2 \theta_{i0} \text{Var}(\hat{\theta}_i)$$

where the θ_{i0} are the true direction angles. Finally,

$$\text{Std. Dev.}(\hat{\theta}_i) = \frac{\lambda}{2\pi\sqrt{K\gamma} L_i \sin \theta_{i0}}$$

which is essentially the implication of classical diffraction theory since $(L_i \sin \theta_{i0})$ is the projected dimension of the array normal to the true arrival direction θ_{i0} and $K\gamma$ is the total array output power signal/noise ratio. This also agrees with Brennan.²¹

It is interesting to see how well the same array can determine the phase velocity from an estimate of α_i

when the angle of arrival is known. We can put

$$|\hat{\theta}| = \frac{\cos \theta_{i0}}{\hat{\alpha}_i}$$

Hence

$$\text{Var } |\hat{\theta}| = \frac{(\cos \theta_{i0})^2}{\alpha_{i0}^4} \text{Var } \hat{\alpha}_i$$

or

$$\frac{\text{Std. Dev.}(|\hat{\theta}|)}{|v|} = \frac{\lambda}{2\pi\sqrt{\gamma K} L_i \cos \theta_{i0}}$$

Whereas angular accuracy depends upon the extent of the array (measured in wavelengths) normal to the direction of propagation, the accuracy of measurement of phase velocity depends upon the depth of the array (in wavelengths) along the direction of propagation. This means that for a linear array angular measurements are most accurate when the array is broadside to the incoming wave while velocity measurements are most accurate when the array lies along the direction of propagation.

We next consider a wide-band example where (22) can be expressed in terms of an observable output signal/noise ratio. Suppose that the sensors have a band-pass frequency response with unity gain and a noise bandwidth B c/s and that the noise is independent between sensors and has a flat spectral density in this region with a level N_0 watts/c/s. The observed mean square noise output of each sensor is

$$\sigma^2 = 2N_0 B$$

so

$$N_0 = \frac{\sigma^2}{2B}$$

Let

$$s(t) = A \sin 2\pi f_1 t$$

where f_1 lies within the sensor pass-band. It is assumed that f_1 is an integral multiple of $1/T$, or alternatively, that $f_1 \gg 1/T$. Then from (34)

$$\begin{aligned} \Omega &\simeq \frac{1}{N_0} \int_0^T [s'(t)]^2 dt \\ &\simeq BT \frac{A^2}{\sigma^2} (2\pi f_1)^2 \end{aligned}$$

Hence, from (22)

$$\text{Std. Dev.} \hat{\alpha}_i \simeq \frac{1}{(A/\sigma)\sqrt{KBT} (2\pi f_1) L_i}$$

$$\text{Std. Dev.} \hat{\theta}_i \simeq \frac{|v|}{(A/\sigma)\sqrt{KBT} (2\pi f_1) L_i \sin \theta_{i0}}$$

and

$$\frac{\text{Std. Dev.}(|\hat{\theta}|)}{|v|} \simeq \frac{|v|}{(A/\sigma)\sqrt{KBT} (2\pi f_1) L_i \cos \theta_{i0}}$$

The factor (A/σ) is just the observed peak-signal/r.m.s.-noise ratio observed at the output of each sensor. The 'enhancement factor' $\sqrt{(KBT)}$ represents the processing gain which results from the use of K sensors and integration of BT effectively independent samples in each sensor. For the sinusoidal signal $(\bar{\omega})^2 + \mu_2$ is equal to the square of the angular signal frequency $(2\pi f_1)^2$ without directly invoking the narrow-band assumption.

As a numerical example consider a uniform linear array of ten seismometers placed along the earth's surface with an overall length of D km. From (27) the form factor is $l = 0.32$. Assume that the array is broadside to an incoming wave which can be repre-

sented as a sinusoid with $f_1 = 1$ c/s. For a P -wave at third-zone distances a typical horizontal velocity is 10 km/s. Hence,

$$\text{Std. Dev. } \hat{\theta} \simeq \frac{1.57}{(A/\sigma)\sqrt{BTD}} \text{ rad.}$$

For a factor $(A/\sigma)\sqrt{(BT)} = 10$ (which might reasonably be attained with $(A/\sigma) = 3$ and $BT = 10$) an array about 1.57 km long could provide an angular standard deviation of 0.1 rad.

Manuscript first received by the Institution on 2nd April 1964 and in revised form on 1st February 1965. (Paper No. 972.)

© The Institution of Electronic and Radio Engineers, 1965

DISCUSSION

Under the chairmanship of Dr. R. Benjamin

Dr. G. O. Young: Do your estimators lead to Bayes' estimator for α ?

Dr. M. J. Levin (in reply): Generally, maximum-likelihood estimators are equivalent to Bayes' estimators with a uniform *a priori* probability density for the parameter to be estimated. However, in this paper, in addition to the parameter vector α , the signal waveform $s(t)$ is to be estimated and an attempt to assign a uniform *a priori* probability to each possible signal waveform would create difficulties in formulating the Bayes' estimators.

Mr. P. R. Wallis: The point that worries me most is the limitation of the process of least squares to one signal $s(t)$, α , at a time. Is this not rather limiting in the seismic field with so many simultaneous arrivals to consider?

I note also that the accuracy of α depends on the derivative of the signals; would it not be better to carry out some prior filtering operation, e.g. differentiation prior to applying the least squares process?

Dr. Levin (in reply): It appears that at least the initial part of the seismic signal is principally a single plane wave. Later multiple arrivals can be seen as multiple peaks in

$C(\alpha)$ plot. A simultaneous least-squares solution for multiple signals would be very useful but it seems to require a rather complicated trial and error search over all combinations of the time delay vectors.

For the second question, please see the reply below.

Dr. R. Benjamin: For the purpose of determining timing, the optimum voltage weighting of each frequency component of a signal is proportional to signal amplitude times frequency, divided by the noise power. Hence, in this context, there seems to be no obvious merit in trying to preserve the undistorted signal wave-form.

Dr. Levin (in reply): The maximum-likelihood estimates of α are found by maximizing the quantity (49) which can be obtained by filtering the sensor outputs, summing, squaring, and integrating. The filters, which take into account the noise cross-spectral densities, differ, in general, for each sensor and are optimum in the maximum-likelihood sense. The estimation of $s(t)$, eqns. (50) and (48), is optimized separately and is accomplished by summing the sensor outputs after they have been passed through a different set of filters. The two filtering operations are not quite the same, but differ by a factor of $1/\sqrt{\Delta(\omega)}$.

Reproducibility of Signal Transmissions in the Ocean

By

ANTARES PARVULESCU,

L.Sc. †

AND

C. S. CLAY, Ph.D. †

Reprinted from the Proceedings of the Symposium on "Signal Processing in Radar and Sonar Directional Systems" held in Birmingham from 6th-9th July 1964.

Summary: Repeated transmissions of acoustical signals were made over a distance of 20 nm. One transmission was compared with succeeding transmissions with the matched signal technique. The experiments showed 10 average arrivals added coherently. The signal is matched for particular source and receiver positions. Displacement of the source or receiver from these positions reduces the coherent addition of arrivals. The sensitivity to small displacements also depends upon the signal bandwidth. These experiments were done at 400 c/s with a bandwidth of 100 c/s. A displacement of the source of about 0.1 nm decreased the peak of the coherently added arrivals to about 60%.

1. Introduction

If one makes an acoustical impulse at a source in the ocean and receives the signal at a hydrophone, the signal is extremely complex. The complexity of the signal, i.e., 'number of arrivals', is dependent upon the ocean, ocean bottom, ocean surface, source-receiver positions, and many other parameters, some of which are independent of time and some that are dependent upon time.

The crux of the matter that concerns us is the dependence of acoustical propagation upon time and position. If the propagation of the bulk of the acoustical energy is independent of time, then many types of signal processing techniques can be used. On the other hand, considerable difficulties arise if the propagation is dependent upon time and changes in a short period. What can be done with signal processing for the intermediate case simply depends upon how extreme the time dependence is. For the present, the time dependence of acoustical propagation is determined by experiment.

2. Experimental Techniques

The time dependence of acoustical propagation can be studied by observing the propagation between fixed source and receiver in the ocean. Ideally one uses a source that can transmit impulse waves; however, for our purpose, we can use any reasonable $f(t)$ that the source can transmit reproducibly. The received signal for an impulse of the source is recorded. The source is impulsed again, and the first transmission

can be compared with the second transmission. The dependence of the propagation upon changes in the source and receiver positions is measured in a similar way except that the source or receiver is moved between transmissions.

This is the question: how similar are the received signals due to the repeated transmissions of the same $f(t)$ from the source? The cross-correlation of the signals or other signal processing techniques can be used to measure the similarity of the signals. If one

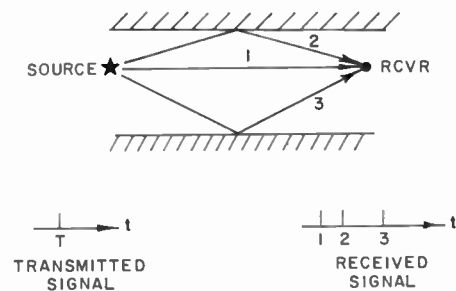


Fig. 1. Received signal for an impulse source and three travel paths.

has good communication between the receiving hydrophone and the source, the choice of the technique is merely a matter of convenience and the equipment that is available. The matched signal technique was used by Parvulescu¹ to study propagation in a room. Similar experiments have been described by Kuttruff.² (Matched filters are related to

† Hudson Laboratories of Columbia University, Dobbs Ferry, New York.

matched signals, and the former are discussed by Turin.³) The method may be illustrated by placing a source and receiver in a medium with several travel paths, as shown in Fig. 1. The source is driven with an impulse, and the received signal is recorded. The recorded signal is played back reversed in time

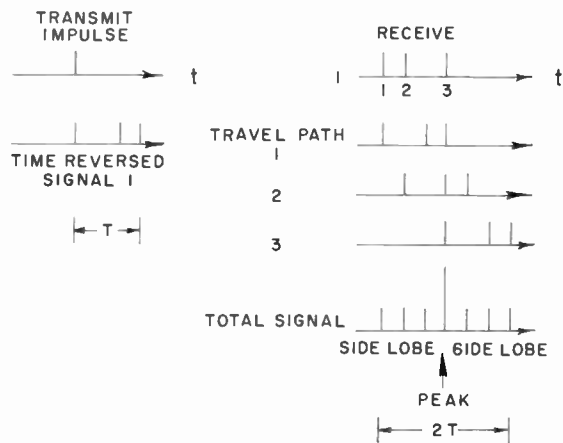


Fig. 2. Matched signal experiment with impulse source and three travel paths. There is no noise, and the signals are assumed to be reproducible for this figure.

through the source. The signal is matched to the medium if one thinks of the medium and its travel paths as a filter. The sequence and signal travelling in each travel path are shown on Fig. 2. The source transmits a delta function, and the receiver observes three arrivals that travel by paths 1, 2, and 3. The

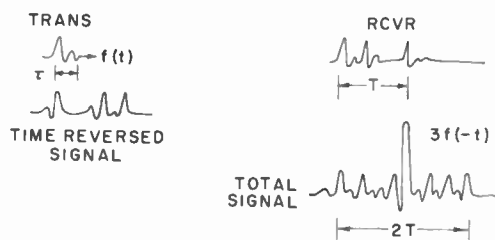


Fig. 3. Matched signal experiment with $f(t)$ source and three travel paths. This figure shows what would be observed if the signals are perfectly reproducible with a source drive $f(t)$.

received signal is recorded and played back time reversed through the source. The signals that travel in each path are shown as well as the sum signal. For this example of three arrivals, the peak is three times as high as the side-lobes. With ten arrivals, the peak would be ten times the side-lobes, provided the arrivals

do not add together in the side-lobes. A second example of the same process for a short transient $f(t)$ is shown in Fig. 3. If the transient $f(t)$ is long and the arrivals overlap, it is apparent that the side-lobes would be much more complicated. The process considered as linear filter theory is the same as that described by Eckart⁴ and Smith⁵ except that the filter now involves the acoustical travel paths and has very complicated distortion.

The peak is due to the proper addition of all the signals. If the travel time of the paths is changed slightly, the peak would be reduced. Using the examples in Fig. 2, assume that the travel time for path 2 is slightly decreased and the travel time for path 3 is slightly increased. Nine impulses and no peak would be observed.

3. Reproducibility of Signals in the Tongue of the Ocean

Essentially the same experiment was done in the Tongue of the Ocean; this is a deep narrow strait between the Islands of Andros and Eleuthera in the Bahamas (lat. 24°N., long. 77°W). The ships were anchored on opposite sides of the Tongue of the Ocean. A hydrophone was on the bottom, and the signal from the hydrophone was transmitted back to the source ship by radio. As shown in Fig. 4(a), the source ship was anchored in deep water and the source was suspended at 500-ft depth. (Originally it was intended to site the source on the bottom, but signals were not received from the source at a position where the source could be placed on bottom.) The source was driven by a short 400-c/s ping, and the received signal was recorded on a loop tape recorder, as shown in Fig. 4(b). The signal was played back in reversed time to yield the matched signal. The matched signal was amplified and used to drive the source, Fig. 4(c).

Figure 5 is a set of oscilloscope photographs made during the experiment of a transmission of a matched signal. The matched signal and the received signal are shown in (a) and (b). For these data the source was pinged once every 6 seconds. The general appearance of the data is similar to a correlation peak with side-lobes. The peak is about four or five times the side-lobe level. The amplitude of the peak is about ten times the amplitude of an arrival for the single pinged source. For presentation, the gain of the amplifier was changed. Data shown in Fig. 5(c) were taken after the ship had drifted to a new position. The side-lobe level is the same, but the peak cannot be seen. There are many more data, but these examples are typical. The change of the source position is estimated to be $\frac{1}{2}$ nm at a distance of 20 nm from the hydrophone.

A laboratory playback of the data is shown in Fig. 5(d). These data were recorded 10 minutes later than the data in Fig. 5(b). The ship position could be about 500 ft different from the position for Fig. 5(b). The replay was on a geophysical camera with an expanded time scale. The peak of the signal on expanded scale is probably a good estimate of the result of transmitting the original 400-c/s ping through the source twice. The width of the envelope peak is about 20 ms null to null and 15 ms for the half-power points. The various arrivals from a shot taken during the experiment lasted about 2 seconds. The voltage signal/noise ratio, σ , as estimated from Fig. 5(a) and similar data was about 2 : 1.

The peak of the signal is a measure of the similarity of the travel paths at the time of the first transmission with the travel paths at some later time. The paths at the later time may differ because of changes in the medium or its boundaries, or may differ because of changes in the position of the source (the receiver was fixed on the ocean bottom). It was not possible to separate these two effects because the ship carrying the source slowly swung about its anchor; nevertheless, the changes in the peak of the correlation function

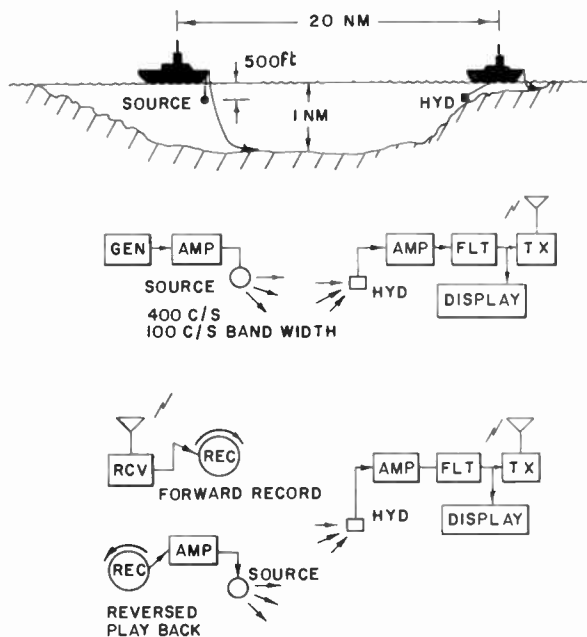


Fig. 4. Signal experiment in the Tongue of the Ocean. The placement of the source and receiver and a simplified diagram of the equipment are shown. The experiments were done with part of the equipment developed for geophysical studies. The source had about 100-c/s bandwidth at 400 c/s and was driven by a 1.5-kW power amplifier. The source could be pinged with a 10 to 15-ms repeatable ping, as well as be driven continuously. The balance of the equipment was standard amplifiers, tape recorders, filters, etc.

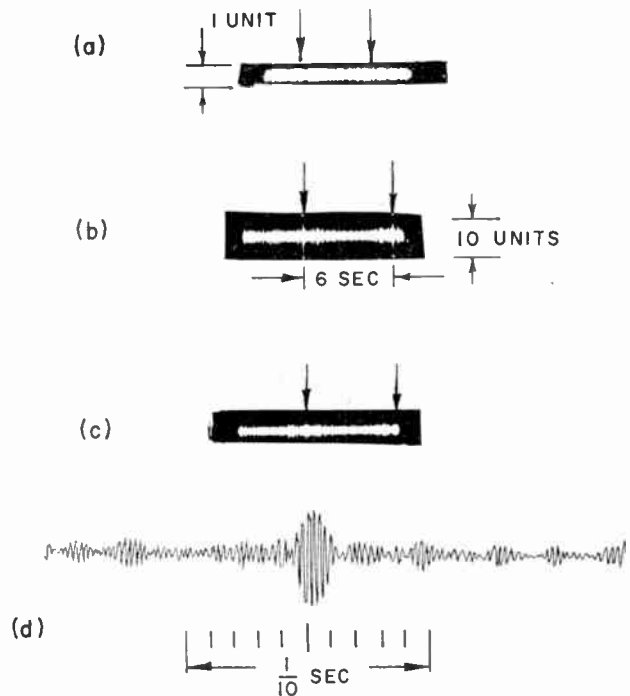


Fig. 5. Data taken to show the reproducibility of signals. The data were taken during the matched signal experiment in the Tongue of the Ocean. The original ping was 15-ms, 400-c/s ping.

- (a) Matched signal plus noise, from single 400 c/s-15 ms ping. Except for the time reversal, this is the signal received at the hydrophone for single pings repeated every 6 seconds. The voltage signal/noise ratio for the ping transmission is about 2 : 1.
- (b) Hydrophone signal for matched signal transmission from the source. This transmission was made about 2 minutes after the first. The amplitudes of the peaks are 10 times the signal shown on Fig. 5(a).
- (c) Hydrophone signal for an unmatched signal, i.e., the ship had drifted to a new position about $\frac{1}{4}$ nm from matched position. This transmission was made after 40 minutes elapsed time.
- (d) Expanded time scale trace of the hydrophone signal for a matched signal. These data were taken after 10 minutes and an estimated source displacement of 500 ft.

were smooth so that we tend to assume that the source motion was by far the greater cause of change. The initial peak was about ten times the amplitude of the direct pings (and also about ten times the amplitude of the side-lobes). The peak decreased from ten to six units after 10 minutes; during this time the ship moved an estimated 500 ft. After 30 minutes the peak was about $1\frac{1}{2}$ times the side-lobe level. The side-lobes were mainly due to noise recorded in the first transmission, and side-lobe levels were nearly constant during the test. The peak can still be observed after 40 minutes (Fig. 5(c)) with an estimated 1500-ft displacement of the source from the original position.

Several experiments to test the effect of a change in source depth were made. The peak remained about the same for source depth change from 135 to 140 ft, while a change of depth from 140 ft to 130 ft caused the peak to be lost in the side-lobes.

4. Signals and Noise in Matched Signal-Type Experiments

The comparison of experimental data with theory requires a theory that includes background noise and estimates of the side-lobe level. The problem is analysed in terms of the ratio of the peak and side-lobes for completely coherent addition of the multipath arrivals. The signal $p_0(t)$ at the receiver, a, due to a source $f(t)$ at b may be expressed by alternate forms of the convolution integral:⁶

$$p_0(t) = \int_{-\infty}^{\infty} s(\tau)p_{ab}(t-\tau) d\tau \quad \dots\dots(1)$$

or

$$p_0(t) = \int_{-\infty}^{\infty} f_s(t-\tau)p_{ab}(\tau) d\tau \quad \dots\dots(2)$$

The acoustical system is considered a linear filter. The impulse response, $p_{ab}(t)$, is the signal at the receiver a due to an impulse source function at b.

The second step consists of using $p_0(-t)$ as the source drive. Substitution of $p_0(-t)$ from (1) into (2) yields, with manipulation, the following:

$$p_{abf}(t) = \int_{-\infty}^{\infty} f_s(-\tau)\psi_{ab}(t-\tau) d\tau \quad \dots\dots(3)$$

where

$$\psi_{ab}(t-\tau) = \int_{-\infty}^{\infty} p_{ab}(\tau')p_{ab}(\tau'-t+\tau) d\tau' \quad \dots\dots(4)$$

In simple cases such as the layered waveguide, ψ_{ab} may be calculated theoretically. However, it may be more practical to determine ψ_{ab} experimentally for transmission between the fixed points a and b.

A detailed picture of the propagation is not available, so in the following a statistical model is used to estimate the various contributions to the side-lobes. The process is basically the same as that shown in Fig. 2 extended to the N -arrival case. The input impulse and the received signal are shown in Fig. 6. The arrivals are assumed to arrive at a random sequence of times. The time duration of the arrivals is T . The amplitudes and phases are dependent upon the propagation; however, for this analysis the amplitudes are assumed to be the average amplitude. The second step is to play the received signal backwards through the source and receive signal shown in Fig. 6(b). The peak is N units high and the side-lobes

are 1 unit high. The duration of the side-lobes is $2T$. There are $N(N-1)$ impulses in the side lobes.

Each impulse in the side-lobe represents an arrival of the input function $f(-t)$ as on Fig. 3. A time-expanded version is shown in Fig. 6(c). Let us assume that instead of an impulse input, we have a function $f(t)$ with duration τ . It is apparent that the number of overlapping functions in the side-lobes is dependent upon the number of side-lobe arrivals in the time τ .

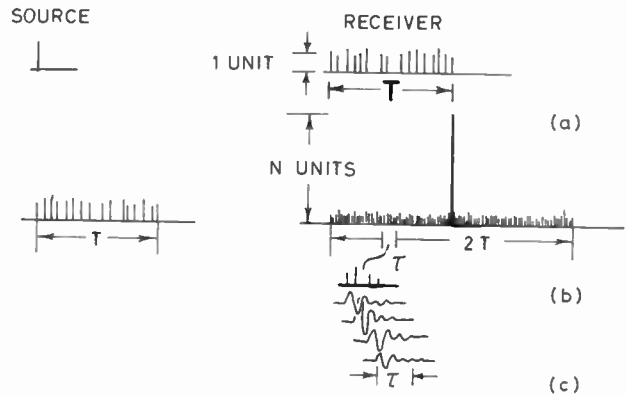


Fig. 6. Matched signal transmission for N arrivals and arbitrary $f(t)$.

- (a) The impulsive source transmission gives N arrivals each lasting time τ . The average amplitude is 1 unit.
- (b) The matched signal transmission: $N(N-1)$ arrivals in the side-lobes and N arrivals in the peak. The peak is N units high.
- (c) The source transmission of $f(t)$ has a duration time τ . $f(-t)$ is received for each of the arrivals shown on (b). (c) is expanded over the time τ to indicate how the arrivals would overlap each other.

There are $N(N-1)$ arrivals in time $2T$. Thus an estimate of the average number of arrivals in τ is:

$$\langle \text{number of arrivals in } \tau \rangle_{\tau} \approx \frac{\tau}{2T} N(N-1) \quad \dots\dots(5)$$

for

$$\tau \geq \frac{2\tau}{N(N-1)}$$

For the purpose of estimating the average level of the side-lobes, the functions $f(t)$ are assumed to be added at random phase, thus the average side-lobe level due to multiple travel paths is

$$\langle \text{side-lobe power due to multiple path} \rangle_{\tau} \approx \frac{\tau}{2T} N(N-1) \langle f^2(t) \rangle \quad \dots\dots(6)$$

A similar argument can be used for the background noise $n(t)$ present at the first transmission. The second transmission sends the noise $n(-t)$ into all N travel paths. The noise is not matched so that it adds incoherently giving an average noise contribution to

the side-lobes of

$$\langle \text{noise power} \rangle_\tau \approx \langle n^2 t \rangle \dots\dots$$

The total estimate of the side-lobe average level is, from (6) and (7)

$$\langle \text{side lobe power} \rangle_\tau \approx \frac{\tau}{2T} N(N-1) \langle f^2(t) \rangle_\tau + N \langle n^2(t) \rangle \dots\dots(8)$$

The noise at the receiver for the second transmission $\langle n^2(t) \rangle$ has been ignored because N is assumed to be large.

The signals adding coherently, i.e. the peak, are $N f(-t)$ and the level of the peak is $N^2 \langle f^2(-t) \rangle$. This is used with (8) to give the ratio of the peak power to average side-lobe power R^2 .

$$R^2 \approx \frac{N^2}{\frac{\tau}{2T} N(N-1) + N\sigma^{-2}} \dots\dots(9)$$

for $\tau > 2T/N(N-1)$, $\sigma^2 \equiv \frac{\langle f^2(t) \rangle}{\langle n^2(t) \rangle}$

For an impulsive source function, R_I^2 is

$$R_I^2 \approx \frac{N^2}{1 + N\sigma^{-2}} \dots\dots(10)$$

$f(t)$ may be considered as an impulsive function, when

$$\tau \leq 2T/N(N-1)$$

The approximate theory may be used to calculate the side-lobe ratio R for comparison with experiment. Recalling the discussion of Fig. 5, the following parameters are assumed:

$$\begin{aligned} \tau &= 0.015 \text{ seconds} & T &= 2 \text{ seconds} \\ \sigma &= 2 & N &= 10 \end{aligned}$$

The value $N = 10$ was chosen because the peak for the matched signal transmission is ten times the amplitude of an average arrival for the single ping transmission. N is the equivalent of ten average arrivals. Substitution of these values in (9) gives

$$R_{\text{calc}} \approx 5.9$$

From Fig. 5(b), the observed value of R is about 6.

The ratio of the mean square of the peak to the mean square level or the side-lobes is dependent upon the number of arrivals N , the signal/noise σ and the time duration of the original source function $f(t)$. It is obvious that both noise and large τ increase the average side-lobe level.

The analysis and examples are given for the case of passing the signal through the filter, i.e. multipath medium, into a noisy receiver and then time reversing the received signal and passing the signal through the filter again. The theory also applies if a matched filter is used at the receiver instead of the second

transmission through the medium. Since the output of the matched filter is the autocorrelation of the signal, it is evident from the examples that matched filter detection is equivalent to cross-correlation of the signal with a stored reference signal, and a similar analysis applies to this case.

5. Conclusions

These experiments showed that the signal transmissions in the ocean are reproducible within limits. The experiments were limited by the ability to fix the source and receiver throughout the tests. The experimental data have been compared with a rather simple theory and are consistent with the theory. The theoretical analysis assumed that all arrivals have the same amplitude and yet this is obviously not true in the experiments. The terms average arrival and average number of arrivals have been used. In this way of thinking, the amplitudes of the arrivals are replaced by the average value. The main effect of the different actual amplitudes is to cause more fluctuations of the terms in the side-lobes. With a 400-c/s signal, about 10 average arrivals were added coherently to yield the correlation maximum.

It is believed that this theory also shows the advantages and the limitations of the particular procedure that was used in this experiment. The side-lobes are small if $f(t)$ is short; however, for a long $f(t)$ the side-lobes can approach the size of the peak. Noise at the receiver also contributes to the side-lobes. The matched signal transmission improved the signal/noise ratio. The voltage ratio for a single ping was 2 : 1, whereas the peak-to-side-lobe ratio was 6 : 1 for a matched signal. For the 6-second ping interval and the noise conditions during the experiment, most of the contribution to the side-lobes is due to noise.

6. Acknowledgments

This experiment required two ships to be anchored under difficult and unusual conditions. The officers and men of both ships are congratulated.

We particularly thank Lt. Harmon, Commanding Officer of the USS *Allegheny*, and Capt. W. Olivey, Master of USNS *Gibbs*.

Mr. P. Weber was responsible for the design and operation of the source at 500-ft depth. The radio transmission link and reversible tape recorder were vital to the experiment, and Mr. H. Gruen, Mr. P. E. Schad, Mr. F. Cole, and Mr. D. D. Mitchell assisted with these. Many Hudson Laboratories' personnel made less obvious, but important, contributions to the experiment.

The work was supported by the Office of Naval Research under Contract Nonr-266(84). It is Hudson Laboratories of Columbia University Contribution No. 194.

7. References

1. Antares Parvulescu, "MESS processing", (A), *J. Acoust. Soc. Amer.*, **33**, p. 1674, 1961.
2. H. Kuttruff, "Raumakustische Korrelationsmessungen mit einfachen Mitteln", *Acustica*, **13**, p. 120, 1963.
3. G. L. Turin, "An introduction to matched filters", *I.R.E. Trans. on Information Theory*, IT-6 pp. 311-29, 1960.
4. Carl Eckart, "The Theory of Noise Suppression by Linear Filters", Scripps Institute of Oceanography Ref. 51-44, 1951.
5. M. K. Smith, "A review of methods of filtering seismic data," *Geophysics*, **23**, pp. 44-57, 1958.
6. Y. W. Lee, "Statistical Theory of Communication", p. 328 (Wiley, New York, 1960).

Manuscript first received by the Institution on 8th June 1964 and in final form on 12th August 1964. (Paper No. 973/RNA40).

© The Institution of Electronic and Radio Engineers, 1965

DISCUSSION

Under the chairmanship of Dr. R. Benjamin

Mr. E. D. Shearman: It may be of interest to mention an investigation which is under way in the Department of Electronic and Electrical Engineering at Birmingham University and which is closely related both to the present paper and to the paper on Moon echoes presented by Mr. Ponsonby† at the beginning of this conference.

In this investigation an X-band microwave Moon-reflection communication link is being set up. By measuring continuously the impulse response of the link, a matched filter is synthesized which corrects the multi-path distortion in the fashion described by Dr. Clay. A significant difference is the rate of change of this distortion

† Informal paper omitted from Proceedings of Symposium.

due to the motion of the observer relative to the Moon. In contrast to Dr. Clay's 10-minute period before the impulse-response changed, our matched filter has to be altered at intervals of $\frac{1}{20}$ th second or so.

Dr. C. S. Clay (in reply): Mr. Shearman has told us of an application of matched filter technique to reduce multipath distortion. In this work the path is to the Moon and back, whereas we had a transmission path of 20 nm of ocean. With changing geometry, the correlation distance depends upon the signal bandwidth. Microwave experiments usually involve large bandwidths and it is surprising to me that their matched filter holds for as long as $\frac{1}{20}$ th second for a Moon echo path.

Transistor Crystal Oscillators and the Design of a 1-Mc/s Oscillator Circuit Capable of Good Frequency Stability

By

P. J. BAXANDALL,
B.Sc.(Eng.) (Member)†

Summary: The paper describes in detail a two-transistor series-resonance oscillator circuit, in which two point-contact diodes in parallel provide amplitude limitation. No transformer is required. The particular AT-cut crystal used has a Q -value of just over 250 000, and the oscillator frequency increases by approximately 1.5 parts in 10^8 per volt increase in the 10 V d.c. supply. By adding a variable-capacitance diode, this variation may be reduced severalfold. The crystal dissipation is less than $1 \mu\text{W}$, giving a low rate of frequency drift due to crystal ageing.

The paper also discusses the relative properties of various well-known oscillator circuits, and gives tables of useful design formulae. One conclusion is that a simple one-transistor Pierce oscillator is capable, when correctly designed, of a much better performance than that usually associated with valve versions of the circuit. Satisfactory operation at a crystal dissipation of about a microwatt is quite feasible without employing additional circuits for amplitude control.

An unusual crystal equivalent circuit is derived for use particularly in parallel-resonance oscillators, and this leads directly to a simple alternative explanation of the Marconi FMQ system for frequency modulating a crystal oscillator.

1. Introduction

In recent years vacuum mounted quartz crystals with Q -values well in excess of a million have become commercially available.^{1, 2, 3}

These very high Q -values have made it easier than in the past to ensure that the frequency stability of a crystal oscillator is determined mainly by the properties of the crystal and its temperature-control oven, rather than by instabilities occurring in the maintaining circuit.

The use of transistors in place of valves has also been of considerable benefit, largely because the very much smaller power consumption and size of a transistor maintaining circuit permit it to be mounted inside the oven, thus avoiding varying temperature gradients in the connecting leads between the crystal and the maintaining circuit, and hence eliminating the associated capacitance and other changes.

The transistor crystal oscillator described in detail in Section 3 of this paper was designed to enable a better frequency stability than usual to be obtained from a 1 Mc/s AT-cut vacuum-mounted crystal of only moderate cost, having a Q -value in the region of 250,000. The basic circuit is shown in Fig. 11, and will be seen to have the convenient feature that no transformers are required.

The crystal is operated at its series resonance frequency, the amplitude of oscillation being controlled by a symmetrical diode limiter. With a crystal having the parameters given in Fig. 1, the power dissipated in

the crystal is only $0.4 \mu\text{W}$, and the frequency increases by approximately 1.5 parts in 10^8 for a 1 V increase in the 10 V d.c. supply. With the modification described in Section 6, the frequency variation per volt change in the d.c. supply may be made much smaller than just mentioned.

At the time this circuit was designed the author believed that a circuit employing the crystal as a simple series-resonant element was the correct choice when it was desired to obtain the best performance from a given crystal—earlier papers on crystal oscillators using valves having given strong support to this belief.^{4, 8}

More recently, however, the author has become aware that the Pierce circuit†⁴ (see Figs. 3 and 4), which is not usually described as employing the crystal as a series-resonant element—though this is a matter of viewpoint—is capable, when transistors are used, of a far better performance than might at first be supposed, despite its simplicity and widespread use in not-very-exacting applications. As will be explained in greater detail later on, the main reasons for this improved performance are:

(a) Higher values of g_m are easily obtained with transistors, enabling the capacitors associated with the crystal to have much larger values than those normally employed in valve versions of the circuit, thus greatly reducing the effects of instabilities in these capacitors

† The circuit of Fig. 3 appears to be almost universally described as a Pierce oscillator, that of Fig. 6 being referred to as a Miller oscillator. Reference 5, however, uses these names in the inverse sense.

† Royal Radar Establishment, Great Malvern, Worcestershire.

and the associated transistor and stray capacitances.

(b) Because of the much more rapid curvature of the I_c/V_{be} characteristic of a transistor, compared with the I_a/V_g characteristic of a valve, a much lower level of oscillation can be reliably obtained in a very simple circuit, greatly reducing the crystal dissipation and consequently the frequency drift due to ageing.

Because of the above considerations, it cannot be claimed that the series-resonance crystal oscillator described in Section 3 is greatly superior in most respects to a well-designed Pierce oscillator, though it does possess the following good features:

(a) Variation in the crystal losses has considerably less effect on frequency than in a simple Pierce circuit. (See also Section 2.3.3.)

(b) Owing to the use of negative feedback, the effect of long-term drift in some of the transistor parameters is reduced.

(c) The circuit does not favour the excitation of crystal resonances at frequencies well below the wanted frequency—an effect which can occur with some crystals in a Pierce oscillator, especially if the circuit is not adjusted for a sufficiently small amplitude of oscillation. (See also Section 2.3.5.)

Before describing the detailed design of the series-resonance oscillator, some further general points relating to crystal oscillators will be discussed. It may be thought that the theory of crystal oscillators has already been so thoroughly investigated^{4,5} that there can be little justification for repeating any of it here. The argument for doing so arises, however, from the belief that much more than the detailed analysis of particular circuits is required for a full and vivid understanding of the subject, and that any treatment is worth while if it helps to make the various details appear as parts of a more coherent whole, or if it emphasizes design aspects which are often overlooked in the more formal and analytical approach.

2. Some Basic Considerations

In order to obtain the best possible frequency stability with a given crystal, it is necessary, first of all, to hold the temperature of the crystal constant to within a very small fraction of 1 deg C.^{1-4,6}

Additionally, the maintaining circuits should satisfy the following requirements:

(a) The phase shift should be small and vary as little as possible† with supply voltage, and with the ageing of transistors and other components.

† If the phase angle, whilst highly stable, were not also small, the crystal circuit would have to operate sufficiently far off resonance to introduce an equal and opposite phase angle, and the oscillation frequency would then become markedly dependent on the crystal Q -value, which is clearly undesirable.

(b) The power dissipated in the crystal should be very small, preferably not more than about a micro-watt, to reduce ageing effects.

(c) Variations in the magnitude or phase angle of the load connected to the output terminals of the complete oscillator circuit should have a negligible effect on frequency.

The higher the Q -value of the crystal circuit‡, the less will be the effect on the oscillator frequency of a given amount of phase change in the maintaining amplifier, the quantitative relationship being given by eqn. (1):

$$\frac{\Delta f}{f_0} = \frac{\Delta \phi}{2Q} \quad \dots\dots(1)$$

where Δf is the small change in frequency caused by a small change $\Delta \phi$ radians in the maintaining amplifier phase angle, f_0 being the oscillation frequency.

Thus, for example, with a crystal circuit having a Q -value of 100 000, a frequency change of 1 part in 10^9 would correspond to a change in phase angle of 2×10^{-4} radian, i.e. about *one hundredth of a degree*.

For really good long-term stability, the maintaining amplifier must therefore meet quite a stringent specification with regard to stability of phase angle, and this makes it desirable to use transistors having a much higher cut-off frequency than would normally be required for the frequency concerned, and preferably to employ negative feedback to give further stabilization of phase angle.

2.1. Ways of Regarding the Meacham Bridge Type of Circuit

It is important to appreciate, in connection with the use of negative feedback, that a circuit of the Meacham bridge type^{7,8} may be regarded, equally correctly, in either of two apparently different ways:

(a) As a passive Wheatstone bridge network in association with a high-gain amplifier having no overall negative feedback (see Fig. 2(a)). The higher the gain of the amplifier, the closer will be the bridge to balance, and the greater will be the rate of change of phase of the bridge output with frequency, thus minimizing the frequency change resulting from a given small change in phase angle of the amplifier.

(b) As an amplifier with overall negative feedback used as the maintaining amplifier of a simple crystal oscillator. To emphasize this point of view, the circuit may be redrawn as in Fig. 2(b). The higher the internal gain of the amplifier, the more negative feedback will there be in operation, thus (exactly as deduced by the reasoning in (a) above) reducing the effect on

‡ The Q -value here referred to is not necessarily that of the crystal itself, since external resistance may be added to the effective series resistance of the crystal.

frequency of a given small change in the internal phase angle of the amplifier.

The present author prefers to regard such circuits in the negative feedback manner, partly because formula (1) is then directly applicable.

2.2. The Crystal Equivalent Circuit

The equivalent circuit of a quartz crystal for frequencies near the wanted resonance is shown in Fig. 1, the values applying to the particular specimen of AT-cut crystal (Marconi QO1655Y) used in the oscillator described in Section 3. (C_0 was determined on a transformer bridge with the crystal in a shrouded 7-pin valve holder, all unused pins and the shroud being taken to the bridge neutral.)

For frequencies very close to the series-resonant frequency of the left-hand branch, this branch has so low an impedance that the presence of C_0 has very little influence on the impedance, Z_{ab} , seen between terminals 'a' and 'b'. Under these conditions the crystal may be regarded, with very little error, as a simple series tuned circuit, $L_x C_x r_x$, of very high Q -value.

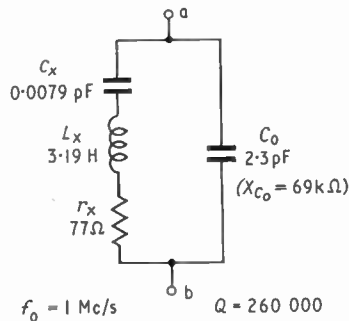


Fig. 1. Equivalent circuit of crystal used.

If the presence of C_0 is taken into account, then, at the exact series-resonant frequency of the left-hand branch, Z_{ab} has a small capacitive phase angle, of magnitude r_x/X_{C_0} . At a very slightly higher frequency than this, Z_{ab} becomes purely resistive, the necessary frequency rise being calculable by substituting r_x/X_{C_0} for $\Delta\phi$ in eqn. (1). For the values shown in Fig. 1, this gives $\Delta f/f_0 = 2.14$ parts in 10^9 . Since C_0 is probably stable over long periods to within $\pm 1\%$, given reasonable layout and construction, variation in C_0 is unlikely to contribute more than about ± 2 parts in 10^{11} instability of frequency in an oscillator employing simple series-resonant operation of the crystal. A 1% variation in r_x will also cause a frequency change of about 2 parts in 10^{11} .

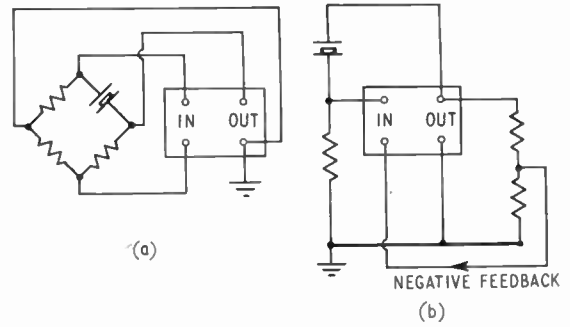


Fig. 2. Alternative ways of drawing Meacham bridge circuit.

When a variable capacitor or inductor is placed in series with the crystal to enable the frequency of a series-resonance oscillator to be adjusted over a small range, the effect of instability of C_0 can become much greater than that mentioned above. This aspect is considered in detail in Sections 2.8 and 3.9.

At a considerably higher frequency than that for series-resonant operation, Z_{ab} once again becomes purely resistive, and has an extremely high value. This occurs when the left-hand part of the Fig. 1 circuit again develops a shunt inductive reactance equal in magnitude to the capacitive reactance of C_0 . For the values shown in Fig. 1, this parallel resonance occurs at a frequency 1720 c/s above the series-resonant frequency. Since this frequency difference is proportional to the reactance of C_0 , it is obvious that any attempt to use this resonant condition (with C_0 unaugmented) in an oscillator would give poor frequency stability, the frequency being far too dependent on the value of C_0 .

In practical crystal oscillators employing parallel resonance, however, C_0 is shunted by additional capacitance of many times its own value, thus bringing the parallel resonant frequency much closer to the series-resonant frequency and giving a great improvement in frequency stability. Such oscillators are considered further in Sections 2.5 and 2.6.

2.3. The Pierce Crystal Oscillator

This very widely used circuit may be drawn in at least three different ways, as shown in Fig. 3, where all details not of immediate relevance have been omitted.

The manner of drawing the circuit can have a considerable influence on the approach adopted in analysing its behaviour. Thus, for example, Fig. 3(b) may lead to the concept of an equivalent parallel tuned circuit, on which the transistor electrodes are well tapped down capacitively, whereas (c) and perhaps (a), lead naturally to the approach given below, which seems to the author to convey the clearest physical understanding of the circuit.

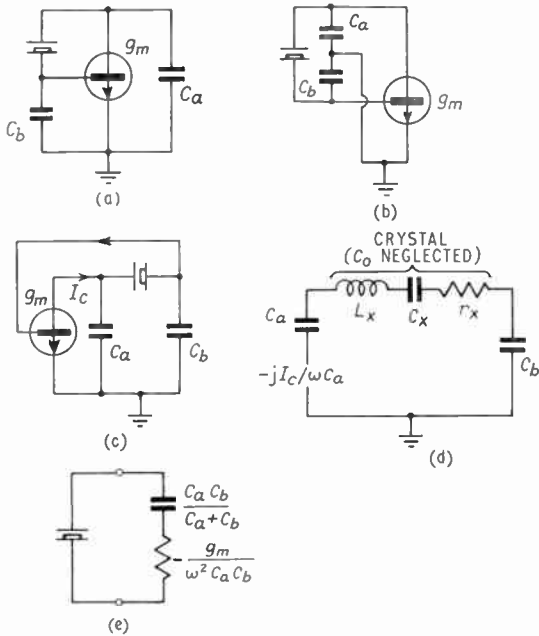


Fig. 3. Circuits relating to Pierce oscillator.

It will be assumed that the transistor can be regarded as a simple device possessing a mutual conductance g_m and infinite input and output impedances. This assumption is not greatly in error in a well-designed circuit of this type. Referring to Figs. 3(c) and 3(d), the combination of C_a and the transistor, with collector current I_c , is equivalent, by Thevenin's theorem, to an e.m.f. $-jI_c/\omega C_a$ acting in series with C_a . If C_a and C_b , and the crystal, are in series resonance, the crystal current I_x will be $-jI_c/\omega C_a r_x$, and it thus lags I_c by 90 deg. The voltage across C_b therefore lags I_c by 180 deg with the result that the net phase shift round the oscillator loop at this frequency is zero. The condition for unity loop gain is

$$(-jI_c/\omega C_a r_x) \times (-j/\omega C_b) \times g_m = -I_c$$

from which

$$r_x = \frac{g_m}{\omega^2 C_a C_b} \quad (C_0 \text{ neglected}) \quad \dots\dots(2)$$

This leads to the conclusion that, as seen by the crystal, the circuit to which it is connected has the impedance shown in Fig. 3(e).

Some important features of the circuit are as follows. (For more quantitative information on some aspects, see Section 2.8.)

(a) The larger C_a and C_b are made, the more closely does the frequency approximate to the series-resonant frequency of the crystal itself, and the less dependent becomes the frequency on the stability of C_a and C_b and the associated transistor capacitances.

(b) Provided g_m can be made large enough, there is no theoretical limit to how large C_a and C_b can be made, in which respect the circuit is much superior to the Miller circuit of Section 2.5.

(c) If the crystal loss resistance r_x varies slightly, g_m will be caused to vary to re-establish equilibrium conditions, but (see Fig. 3(e)) this has no effect on the reactance in series with the crystal and therefore no effect on frequency, in simple theory. In this respect also the circuit is much superior to the Miller circuit of Section 2.5.

(d) The negative resistance acting in series with the crystal (see Fig. 3(e)) becomes rapidly larger in magnitude as the frequency falls, thus favouring the excitation of any crystal modes at frequencies below the wanted mode. This is a weakness of the circuit though it may be overcome by using a parallel tuned circuit in place of the normal resistor to provide the d.c. feed to the collector. Great care should be taken to ensure that such a tuned circuit does not seriously impair the phase angle stability of the maintaining amplifier. In general, the tuned circuit should have no higher a C/L ratio than is necessary for providing the selectivity needed.

(e) By replacing the two capacitors in Fig. 3(c) by small-value resistors, and replacing the transistor by a non-phase-inverting amplifier, the circuit becomes, in essence, the 'series drive' circuit of reference 4. This emphasizes that the relationship between the Pierce circuit and a circuit universally regarded as a true series-resonance crystal oscillator is very close, and

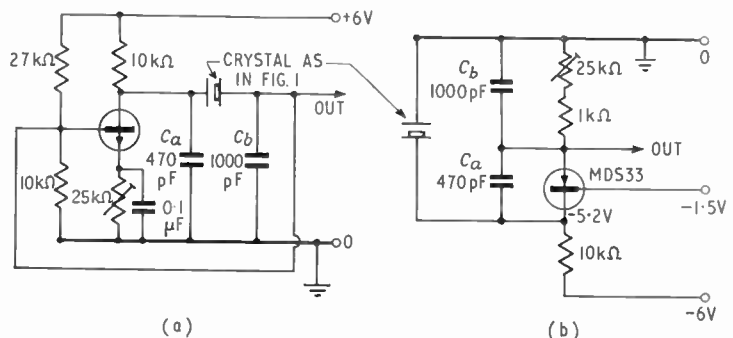


Fig. 4. Practical Pierce oscillator circuits.

that the terms series-resonance oscillator and parallel-resonance oscillator, though widely used, are not altogether satisfactory.

It is interesting to note that the Gouriet⁹ or Clapp¹⁰ L-C oscillator is, in principle, exactly the same as the Pierce oscillator, the crystal in the latter being replaced by an inductor and a capacitor in series.

2.3.1. Stabilization of oscillation level

Consider the practical Pierce oscillator shown in Fig. 4(a). When correctly adjusted, the alternating voltage at the base is much less than the base d.c. bias voltage, so that the mean transistor current is fairly closely equal to the bias voltage divided by R_e . The overall mutual conductance of the transistor, at the fairly small collector currents typically used, is given approximately by:

$$g_m = 40 I_e \quad \dots\dots(3)$$

where g_m is in mA/V and I_e is the emitter current in mA. ($r_{bb'}$ in the transistor equivalent circuit does not have a very significant effect at the small working current involved, which is often less than 1 mA.)

R_e is adjusted so that the loop gain for vanishingly small oscillation amplitudes is considerably greater than unity; oscillation then builds up until the effective g_m has fallen sufficiently to give unity loop gain. Now for a device having the exponential relationship between I_e and base voltage which is implicit in eqn. (3), there is a definite and fixed relationship between the percentage reduction in effective fundamental-frequency g_m and the absolute value of the sinusoidal base voltage causing it, assuming the mean emitter current to be held substantially constant by the d.c. biasing system. When the peak alternating base voltage is several times kT/q , say 100 mV or more, the transistor conducts in pulses, being cut off for the major part of each cycle. It may then be shown that the ratio of peak fundamental component of transistor current to mean value of transistor current approximates to 2, so that the a.c. output remains approximately constant if the input is still further increased, assuming the mean transistor current to remain constant. Under these conditions, the fundamental frequency gain is inversely proportional to input.

When the input is of appreciable magnitude but not so large that the transistor current can be assumed to flow in narrow pulses, the analysis is much more complex, involving Bessel functions. Fortunately the necessary calculations have been carried out by K. Holford¹¹, and the graph of Fig. 5 is based on his work. The asymptote and the three points shown by circles have been calculated by the present author independently as a check.

It will be seen from Fig. 5 that a peak base voltage of 25 mV corresponds to the initial loop gain being

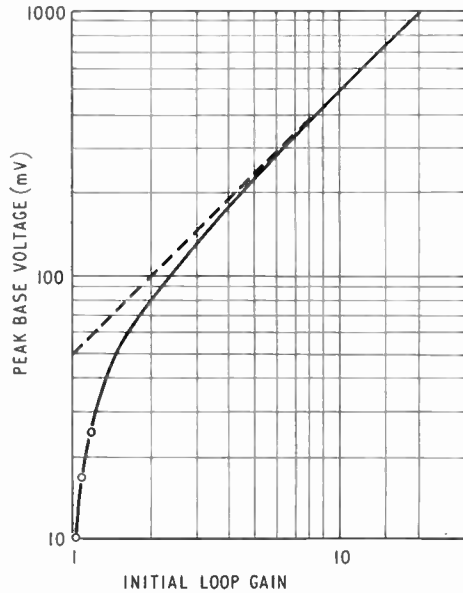


Fig. 5. Variation of peak alternating base voltage with very-small-signal loop gain, for simple transistor oscillator with constant mean transistor current.

about 18% larger than that necessary for oscillation just to occur. Since a good crystal oscillator will be operated from a stabilized d.c. supply, and will be held at a constant temperature, this small level of oscillation can be expected to be maintained quite reliably over very long periods of time.

With $C_b = 1000$ pF (reactance = 160 Ω at 1 Mc/s), a sine wave voltage of 25 mV peak across C_b requires a crystal current of 0.155 mA peak. This current flowing in the crystal series resistance of 77 Ω gives a mean power dissipation in the crystal of 0.92 μ W.

An interesting result of operating the circuit with its initial loop gain so little in excess of unity is that a time interval of about 6 seconds elapses, after switching on the supply, before substantially the full output level is reached. (The circuit of Fig. 11, however, has an initial loop gain greatly in excess of unity and gives much more rapid build-up.)

2.3.2. Choice of capacitor values

The optimum choice of values for C_a and C_b in Fig. 3 or Fig. 4 is by no means a simple matter, the main considerations being as follows:

(a) The larger C_b is made, the larger must the crystal current be to develop the desired alternating base voltage, as discussed above. Doubling C_b thus quadruples the power dissipated in the crystal for the same base voltage. A small value of C_b therefore reduces frequency drift caused by crystal ageing.

(b) If the values of C_b and C_a are both halved, the frequency is made twice as dependent on the percentage stability of the capacitors and four times more dependent on the absolute variations of the transistor input and output capacitances. However, only one quarter of the previous mutual conductance will be necessary for maintaining oscillation, so that the transistor can operate at about a quarter of its previous mean current. This reduction in current will reduce $C_{b'e}$, in the transistor hybrid π equivalent circuit, by a factor which is likely to be considerably less than 4 with the v.h.f. type of transistor preferably used, and in view of the small value of the mean current. Thus the transistor input capacitance has a more significant adverse effect on the frequency stability when C_a and C_b are reduced, though not to the extent which might at first be expected. (For a more detailed treatment of these effects, the formulae given in Section 2.8 should be referred to.)

(c) With $C_b = 1000$ pF and the crystal of Fig. 1, a 1 pF change in transistor input capacitance (or in C_b) gives a frequency change of approximately 4 parts in 10^9 .

(d) It is reasonable to make C_b somewhat larger than C_a , since only C_b is directly associated with the transistor input capacitance, which is likely to be considerably larger than the other unwanted capacitances. The transistor collector/base capacitance comes directly across the crystal, so that its effect is equivalent to a variation in C_0 (Fig. 1) and formula (3) of Table 1 is directly applicable. (It is possible to choose C_a and C_b to make the value of the bracketed expression in this formula zero, but inconveniently large capacitance values are required in the 1 Mc/s design here considered.)

(e) It is possible to improve the performance of this circuit still further by inserting an emitter follower between C_b and the base of the present transistor, thus considerably reducing the transistor capacitance thrown across C_b . C_a and C_b should then preferably be given equal values, which may be larger than before.

2.3.3. Practical Pierce circuits

So far it has been assumed that the emitter is the earthy electrode, but in practice the base or collector may, of course, be made earthy if more convenient. In many practical circuits, moreover, a separate emitter by-pass capacitor, such as that in Fig. 4(a), is not used, C_b being used to perform this function.

Figure 4(b) shows a very simple earthed-base version of fundamentally the same circuit as that of Fig. 4(a). The measured frequency change caused by a 10% increase in both supply voltages was +4.5 parts in 10^9 . Adding 30 ohms of resistance in series with the crystal gave approximately 14 parts in 10^8 rise in

frequency—a much greater effect than that produced by inserting the same resistance in the series-resonance circuit described later (see Section 4.1). This result could be improved by using higher supply voltages and higher values of resistance; the presence of the resistors, neglected in Fig. 3, has the effect of making the frequency depart slightly from the series-resonant frequency of the Fig. 3(d) circuit by an amount dependent on the Q -value of the crystal circuit.

2.3.4. Output waveform purity

In both the Fig. 4 circuits, the oscillator output voltage is the voltage across C_b and is a good sine-wave, since it is produced by the crystal current flow in C_b . The voltage across C_a is much more distorted. The crystal shunt capacitance C_0 largely determines the crystal current at harmonic frequencies whereas r_x determines the fundamental-frequency current.† The reactance of C_0 will normally be much larger than r_x even at harmonic frequencies, however, so that the harmonic current is greatly reduced.

If the collector is made the earthy electrode then no point is available in the circuit from which a low-distortion output may be taken with respect to earth.

2.3.5. Excitation of unwanted low-frequency crystal modes

If R_e is reduced in value to obtain a much larger output than normal, oscillation at a lower frequency crystal mode may build up in addition to the wanted oscillation, and this does indeed happen with the Fig. 1 crystal specimen. The effect is easily mistaken for 'squegging', but is quite different in mechanism.

It was mentioned in paragraph (d) Section 2.3 that eqn. (2) shows that the negative resistance acting in series with the crystal increases rapidly as the frequency is reduced. However, whereas under class 'A' operating conditions, eqn. (2) tells the whole story, when the conditions in a simple Pierce oscillator are class 'B' or class 'C', a further effect comes into play, making the build-up of a low-frequency mode much more likely. This may best be understood by considering the fairly exaggerated case of a Pierce oscillator working under extreme class 'C' conditions. On switching on, g_m is enormously in excess of the minimum value necessary for the wanted oscillation, which therefore builds up rapidly until the transistor is well biased back and conducts only in narrow pulses. The effective g_m at the wanted fundamental frequency, assuming for the moment that this is the only oscillation present, settles down to the value required just to maintain the oscillation.

† If appreciable excitation of a harmonic resonant mode occurs, the harmonic current may be considerably greater than the value determined by C_0 .

Consider now that a trace of oscillation at a much lower frequency is present, due, say, to random disturbances. If this oscillation, at one moment, is acting to bias the transistor on, the wanted oscillation, at a much higher frequency, will be caused to turn on larger current pulses than before, and vice-versa. It is evident that with the transistor conducting only on the extreme peaks of the wanted oscillation, a very small change in bias will have a large effect on the current turned on. Thus the effective g_m available for building up the low-frequency oscillation is much in excess of that effective in maintaining the wanted oscillation.

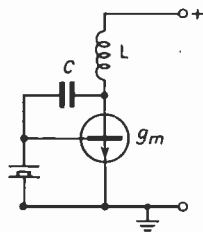


Fig. 6. Basic circuit of Miller oscillator.

If a fairly large amplitude of oscillation is required for a particular application, the danger of unwanted low-frequency oscillation simultaneously occurring may be avoided by inserting some undecoupled resistance, e.g. 100Ω, directly in the emitter lead. For a given angle of flow, the a.c. base voltage is thereby much increased—a large angle of flow increases the margin of safety against excitation of the low-frequency oscillation. The emitter resistor also tends to make the angle of flow independent of the supply voltage, since the effective g_m of the transistor is determined largely by the emitter resistor rather than by the working current; consequently the supply voltage may be varied over a wide range with little danger of causing unwanted low-frequency oscillation. When good long-term frequency stability is the main consideration, however, the emitter resistor should be omitted, thus enabling reliable oscillation to be achieved at the lowest possible level.

2.4. Effect of Distortion on Frequency Stability

In the oscillators described in this paper, amplitude stabilization is achieved by permitting carefully controlled non-linearity distortion to occur in the maintaining amplifier.

Harmonics are therefore present at the output of the amplifier, and are fed back via the highly selective crystal circuit. At the output of the latter they are of very small magnitude, but the vital point is that they are shifted in phase relative to the fundamental

because the crystal is approximately resistive at the fundamental frequency but is reactive at the harmonic frequencies.

In the amplifier the harmonics and the fundamental intermodulate and one of the output intermodulation products is at the fundamental frequency but shifted in phase relative to the normal fundamental output component. The overall effect is therefore equivalent to a slight phase shift in the maintaining amplifier, and is accompanied by a small frequency shift as given by eqn. (1).

Additionally, if the amplifier input circuit loads the crystal circuit appreciably, the latter being reactive at harmonic frequencies, then undesirably phased harmonics may be generated owing to the non-linear input impedance of the amplifier, even though the harmonic e.m.f. at the crystal circuit output might be negligible. This provides a further argument in favour of the emitter follower mentioned in paragraph (e) of Section 2.3.2.

It is thought that in a well-designed circuit employing non-linearity for amplitude stabilization, intermodulation probably contributes negligibly to the overall instability of frequency.

2.5. The Miller Crystal Oscillator

This is another very widely used circuit, the essential features being shown in Fig. 6. As with the Pierce oscillator, the circuit may be drawn in various ways and any of the three transistor electrodes may be earthy. Assuming the transistor may be regarded as a simple device possessing a mutual conductance g_m and infinite input and output impedances, it may be shown that the admittance of the maintaining circuit as seen by the crystal is given by

$$Y = \frac{g_m - \frac{j}{\omega L}}{1 - \frac{1}{\omega^2 LC}} \dots\dots(4)$$

Provided $\omega^2 LC$ is made less than 1 at the required oscillation frequency, the two components G and B of Y consist of a negative conductance proportional to g_m in shunt with a capacitive susceptance whose value is independent of g_m . This parallel combination is equivalent, at the oscillation frequency, to negative resistance and reactance in series, both these series elements being functions of g_m .

The condition for steady oscillation is that the series negative resistance provided by the maintaining circuit must be equal in magnitude to the effective series loss resistance of the crystal. If this series loss resistance varies slightly with time, then g_m will have to vary to maintain equilibrium conditions, and this

will necessarily vary the effective reactance in series with the crystal, and hence the frequency. This dependence of the frequency on variations in the crystal losses is an inherent weakness of the Miller circuit, not possessed by the Pierce circuit.

Another weakness of the Miller circuit is that there is a limit to how large the shunt capacitance represented by eqn. (4) can be made, for, as g_m is varied, the maximum possible value of negative resistance appearing in series with the crystal occurs when the shunt negative conductance is numerically equal to the shunt susceptance B , and this maximum negative resistance is of half the magnitude of the shunt reactance. For oscillation, the following conditions must, therefore, be satisfied:

$$\frac{1}{\omega_0 C_1} > 2r_x \quad \dots\dots(5)$$

where C_1 is the total shunt capacitance appearing across the crystal owing to the action of the maintaining circuit. (In practice C_1 includes C_0 , the internal shunt capacitance of the crystal.)

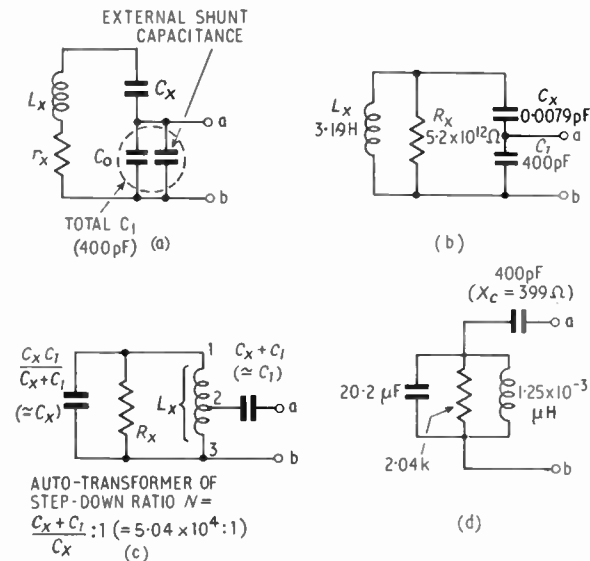


Fig. 7. Alternative equivalent circuit for capacitance-shunted crystal.

2.6. A Different Viewpoint

The Miller circuit is, of course, only one example of a type of maintaining circuit, which presents the crystal with a shunt combination of capacitance and negative conductance.

The afore-mentioned method of regarding the shunt capacitance and conductance of the maintaining circuit as equivalent to a series combination of ele-

ments is not the only way of analysing this type of circuit.

The alternative approach, which is rather instructive, is to consider the shunt capacitance as combined with the crystal equivalent circuit, this combination then being shunted by pure negative conductance.

The crystal equivalent circuit, including the additional shunt capacitance from the maintaining circuit, may be drawn as in Fig. 7(a), the values given applying to the crystal of Fig. 1. L_x and r_x in series are almost perfectly equivalent, because of the very high Q -value, to L_x and an enormously high resistance $R_x (= Q^2 r_x)$ in parallel, as in Fig. 7(b).† Further, this last circuit may be shown to be exactly equivalent to that of Fig. 7(c).

The impedance between 'a' and 'b' in Fig. 7(c) will reach a maximum value at a frequency just below the resonant frequency of the parallel tuned circuit, and will be purely resistive when the parallel tuned circuit exhibits an inductive reactance, referred to the low-impedance terminals 2 and 3 of the auto-transformer, numerically equal to the reactance of $(C_x + C_1)$.

To make the frequency adequately independent of unwanted variations in the capacitance C_1 , C_1 should be made large. With $C_1 = 400$ pF, the circuit values corresponding to those given in Fig. 1 are shown in Fig. 7(d), the parallel tuned circuit impedances of Fig. 7(c) having been divided by N^2 to eliminate the need for the auto-transformer.

Figure 7(d) is, therefore, one way to express the equivalent circuit of the crystal when shunted by 400 pF, and it will be seen that the circuit has a convenient order of impedances for connection to a transistor maintaining amplifier.

Provided eqn. (5) is satisfied, which it easily is for the values shown, there are two frequencies at which Z_{ab} in Fig. 7(d) is purely resistive. One of these gives a low value of Z_{ab} and is the one of interest in so-called series-resonance oscillators, in which the maintaining circuit functions as a negative resistance of the open-circuit-stable, short-circuit-unstable variety. The equivalent circuit of Fig. 7(d) is not the most

† It may be shown that the parallel inductance L_p is actually equal to $L_x (1 + 1/Q^2)$. Thus, with $Q = 260\ 000$, as in Fig. 1, L_p differs from L_x by approximately 1.5 parts in 10^{11} .

The fractional change in L_p resulting from a change δQ in Q is given by

$$\frac{\delta L_p}{L_p} = -\frac{2}{Q^2} \times \frac{\delta Q}{Q}$$

Hence, with $Q = 260\ 000$, as in Fig. 1, a 1% increase in Q , which might occur spontaneously, gives a change in L_p of -3 parts in 10^{13} corresponding to a frequency change of +1.5 parts in 10^{13} . This is so small compared with other effects that it may be neglected. Thus the assumption that L_p is constant and equal to L_x is justified and leads to no significant error.

suitable to use for such oscillators, however. At the other frequency where Z_{ab} is purely resistive, its value is much higher, and this is the condition of interest in so-called parallel-resonance oscillators, in which the maintaining circuit functions as a short-circuit-stable, open-circuit-unstable negative resistance.

For the high-impedance condition it is evident that the parallel tuned circuit must operate below resonance by a considerable fraction of its bandwidth in order to develop a reactance equal and opposite to that of the series capacitance and thus make Z_{ab} purely resistive.

If the crystal Q -value increases slightly, thus raising the dynamic resistance of the parallel tuned circuit, it will have to operate below resonance by a smaller fraction of its bandwidth to develop the required reactance; further the bandwidth itself will be reduced, with the result that an increase in Q -value of $x\%$ will reduce the frequency by which the parallel tuned circuit is below resonance by $2x\%$.

For the values shown, the parallel tuned circuit operates below resonance by about 3.7 parts in 10^7 . A 1% increase in Q -value will therefore give a frequency increase of approximately 0.02×3.7 parts in 10^7 , i.e. 0.74 parts in 10^8 ; this should be compared with approximately 2 parts in 10^{11} for a 1% change in Q in a series-resonance oscillator using the same crystal without C_1 . (See Section 2.2.)

If C_1 in Fig. 7(c) is halved, then the auto-transformer ratio will be approximately halved, thus multiplying the tuned circuit dynamic resistance referred to terminals 2 and 3 by four. The reactance of the capacitance in series with the lead going to terminal 'a' will be multiplied by only two, however, so that the parallel tuned circuit will now have to operate only approximately half as far off resonance and the effect on frequency of a change in Q -value will be reduced by a factor of two. The effect on frequency of unwanted variations in C_1 is made greater as C_1 is reduced, however, so that a suitable compromise must be struck. This inability to make the frequency nearly independent of Q variations and capacitance variations, at one and the same time, is a feature of the type of oscillator here discussed, unless the modification described in the next Section is employed.

2.7. The FMQ System^{12,13}

Referring again to Fig. 7(d), if a reactive element could be connected in series with the lead going to terminal 'a' having a reactance equal in magnitude but opposite in sign to that of the series capacitance, then direct access to the parallel tuned circuit could, in effect, be obtained. The frequency could then be linearly modulated over a considerable range by con-

necting a linearly-variable susceptance between 'a' and 'b', the dynamic resistance being unaffected by the frequency variation, thus making it easy, in principle, to obtain f.m. without a.m.

The ideal element to connect in series with the lead to terminal 'a' is thus a negative capacitance, but an inductor having the same reactance is a very satisfactory, and simpler, substitute, in view of the fact that the frequency deviation obtainable in such systems does not normally exceed something in the region of 0.1%. This is the basis of the FMQ system used so successfully in v.h.f. f.m. broadcasting transmitters,^{12,13} though the explanation usually given is rather different. It is easily shown from Fig. 7(c) that to make the mean frequency equal to the natural frequency of the crystal, i.e. as determined by $\omega^2 L_x C_x = 1$, the mean modulating capacitance which must be connected between 'a' and 'b' is of value $(C_1 + C_x)$, which is nearly equal to C_1 .

When an inductor is used as described above, it is necessary to take precautions to prevent parasitic oscillation occurring at a frequency determined by the inductance and the associated capacitances, rather than by the crystal. This and other details are very clearly described in references 12 and 13.

2.8. Tables of Formulae

In this Section is presented a number of formulae found useful for design purposes. $\Delta\omega$ in these formulae represents the departure of the angular frequency from the true mechanical resonant value ω_0 , where $\omega_0^2 L_x C_x = 1$.

The formulae of Table 1 relate to the circuit of Fig. 8(a) in its series, or low-impedance, resonant condition, i.e. Z_{ab} purely resistive and of low value.

C_f may be a frequency-adjusting capacitor in a simple series-resonance oscillator, or it may consist of $C_a C_b / (C_a + C_b)$ in a Pierce oscillator (Fig. 3). In a

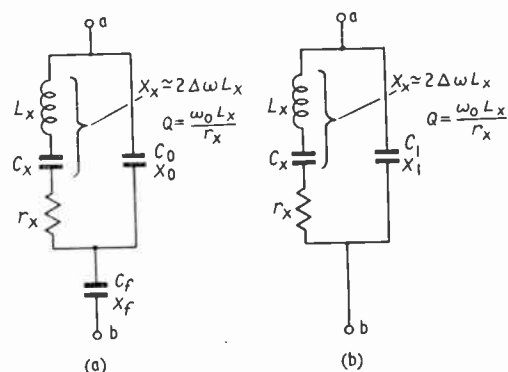


Fig. 8(a). Circuit to which the formulae of Table 1 relate.
 (b). Circuit to which the formulae of Table 2 relate.

Table 1

Figure 8(a) circuit, in series-resonant condition

1.	$\Delta\omega \simeq \frac{\omega_0 C_x}{2(C_f + C_0)} \left[\frac{C_f C_0}{Q^2 C_x^2} + 1 \right]$
	(The $C_f C_0/Q^2 C_x^2$ term is significant only at very large C_f values, where r_x/x_f begins to become appreciable in relation to x_0/r_x)
2.	$\frac{\delta(\Delta\omega)}{\omega_0} \simeq \frac{\delta r_x}{r_x} \cdot \frac{C_0}{Q^2 C_x} \left(\frac{C_0}{C_f} + 1 \right)$
3.	$\frac{\delta(\Delta\omega)}{\omega_0} \simeq \frac{\delta C_0}{C_0} \cdot \frac{C_x C_0}{2(C_f + C_0)^2} \left[\frac{C_f^2}{Q^2 C_x^2} - 1 \right]$ ($C_f^2/Q^2 C_x^2 = 1$ corresponds to $ X_x \simeq r_x$)
4.	$\frac{\delta(\Delta\omega)}{\omega_0} \simeq \frac{\delta C_f}{C_f} \cdot \frac{C_x C_f}{2(C_f + C_0)^2} \left[\frac{C_0^2}{Q^2 C_x^2} - 1 \right]$ (The $C_0^2/Q^2 C_x^2$ term will normally be quite negligible)
5.	$Z_{ab} \simeq r_x \left(1 + \frac{C_0}{C_f} \right)^2$ provided, as will normally be the case, $ C_f \gg \left \frac{C_0^2}{QC_x} \right $

simple series-resonance oscillator having no frequency adjusting capacitance, $C_f = \infty$. On putting $C_f = \infty$ in the formulae of Table 1, simplified formulae for the latter case are obtained.

If C_f is replaced by an inductor, the Table 1 formulae may still be used if the inductance is represented by the equivalent negative value of C_f .

The formulae of Table 2 relate to the Fig. 8(b) circuit in its parallel, or high impedance, resonant state, i.e. Z_{ab} purely resistive and of high value.

C_1 will usually be of much larger value than C_0 (Fig. 8(a)), but provided $|X_1| > 2r_x$, there will be two frequencies at which Z_{ab} is purely resistive. When $|X_1| = 2r_x$, the series and parallel resonance frequencies coincide and when $|X_1| < 2r_x$, Z_{ab} is not resistive at any frequency. (It is interesting to note that this same relationship $|X_1| > 2r_x$ was reached by a different approach in Section 2.5—see eqn. (5).)

In deriving the formulae in Table 2, the assumption $X_1^2 \gg r_x^2$ has been made, the formulae becoming inconveniently cumbersome without this assumption.

Table 2

Figure 8(b) circuit, in parallel-resonant condition

1.	$\Delta\omega \simeq \frac{\omega_0 C_x}{2C_1}$
2.	$\frac{\delta(\Delta\omega)}{\omega_0} \simeq - \frac{\delta r_x}{r_x} \cdot \frac{C_1}{Q^2 C_x}$
3.	$\frac{\delta(\Delta\omega)}{\omega_0} \simeq - \frac{\delta C_1}{C_1} \cdot \frac{C_x}{2C_1}$
4.	$Z_{ab} \simeq \left X_1 \cdot \frac{QC_x}{C_1} \right $
	$X_1^2 \text{ assumed } \gg r_x^2$

The fractional errors produced are of the same order as r_x^2/X_1^2 .

The Table 1 formulae assume $X_0^2 \gg r_x^2$, but this condition is normally satisfied by such an enormous factor that the error introduced is quite minute. Other slight approximations made in deriving the formulae also result in very little error. (When C_f is small, r_x may be neglected in deriving formulae 1, 3 and 4 of Table 1. When C_f is large, a simple binomial expansion may be used in solving the quadratic equation giving the frequencies for Z_{ab} to be real. The complete formulae given combine the results of these two methods and are of good accuracy for both high and low values of C_f .)

The formulae are given in the form thought most useful for design purposes; for example, Table 1 formula 3 gives the fractional change in frequency for a given fractional change in C_0 , but may, of course, be simplified slightly, if preferred, to:

$$\frac{d\omega}{dC_0} = \frac{\omega_0 C_x}{2(C_f + C_0)^2} \left[\frac{C_f^2}{Q^2 C_x^2} - 1 \right] \dots\dots(6)$$

3. Design of a Series-Resonance Crystal Oscillator

3.1. Methods of Amplitude Stabilization

By using a thermistor, or some other form of a.g.c. system, to control the oscillation amplitude, an almost complete absence of harmonic distortion can be secured.

The use of non-linearity to control the amplitude tends to lead to simpler circuit designs, however, and has been adopted in the present instance. The instability attributable to the presence of harmonics (see Section 2.4) is believed to be negligible.

3.2. Diode Limiter Characteristics

Figure 9 shows the result of a measurement on the limiting properties of a pair of point-contact diodes. A better limiting characteristic than this is exhibited by a pair of junction diodes, but the point-contact diodes have the over-riding advantage of much lower shunt capacitance and were chosen for this reason.

When used in an oscillator circuit, some extra capacitance will inevitably be thrown across the limiter diodes and part of this capacitance, contributed by the transistor(s), will be voltage-dependent. In order to minimize the resultant phase shift and to keep the change of phase shift with variation in supply voltage small, the diodes must be operated at a sufficiently high current, so that their effective fundamental-frequency resistance is low compared with the reactance of the parallel capacitance. A current of approximately 0.4 mA r.m.s., giving an effective limiter resistance of 500Ω, was adopted for the oscillator described below.

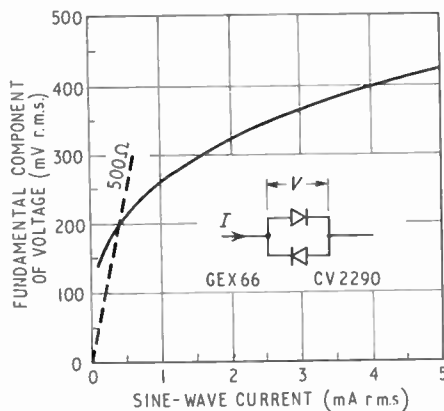


Fig. 9. Limiting characteristic of a pair of point-contact diodes.

3.3. Possible Circuit Configurations

In (a) and (b) of Fig. 10 are shown two circuits for series-resonance oscillators, in both of which the maintaining amplifier has a considerable amount of negative feedback† and in which the limiting action is provided by a simple diode limiter.

A further feature common to these two circuits, however, is that the alternating current fed to the limiter diodes is not greatly different in magnitude from that flowing in the crystal itself.

As explained in Section 3.2, the limiter current must be adequately large to minimize the effects of

† Figure 10(a) can alternatively be regarded as a nearly-balanced bridge circuit, the concept of negative-feedback not then needing to be invoked. To appreciate this viewpoint, the circuit should be redrawn with the emitter earthed.

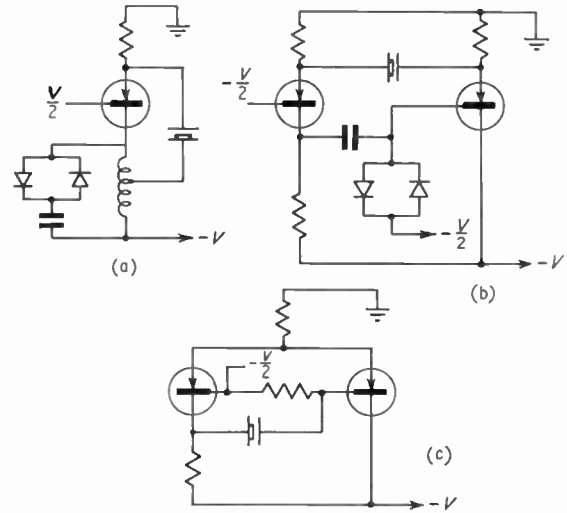


Fig. 10. Simple series-resonance oscillator circuits.

capacitances across the limiter, so that the crystal current must also be moderately large when these simple circuits are used. A very low crystal dissipation is desirable, however, in the interests of good long-term frequency stability.

The very simple circuit of Fig. 10(c), employing a long-tailed-pair as a limiter, also requires a moderately large crystal current for satisfactory limiting action.

Thus what is wanted is a circuit in which the limiter current greatly exceeds the crystal current and the circuit described below fulfils this requirement excellently.

3.4. The Basic Circuit Used

Referring to Fig. 11, it will be seen that the two transistors are d.c. coupled with overall negative feed-

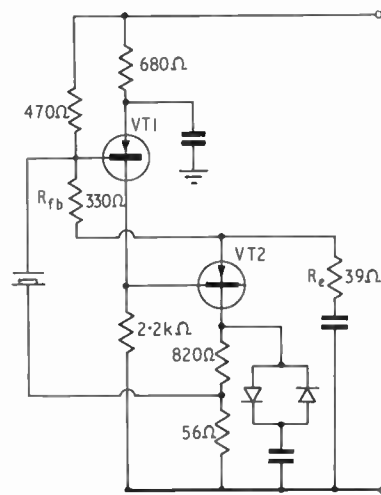


Fig. 11. Basic circuit of the author's series-resonance oscillator.

back via R_{fb} . This feedback is operative at zero frequency and gives tight control of the d.c. sit points.¹⁴ It is also operative at the oscillation frequency and results in a low impedance (about 4 ohms) looking into the VT1 base circuit, i.e. this point is a 'virtual earth'.¹⁵

The alternating voltage at VT2 emitter is approximately equal to the crystal current I_x multiplied by R_{fb} . The alternating current in VT2 is given approximately by:

$$I_{VT2} = \frac{I_x R_{fb}}{R_e R_{fb} / (R_e + R_{fb})}$$

i.e.

$$I_{VT2} = I_x (1 + R_{fb} / R_e) \dots\dots(7)$$

In the final design, R_{fb} / R_e is 8.5, so that the limiter current is approximately 9.5 times the crystal current.

3.5. Effect of Supply Voltage on Phase Angle

Owing to the use of v.h.f. transistors, and because of the considerable amount of negative feedback, the gain and phase stability of the amplifier are of quite a high order. (A 1-V change in the 10 V d.c. supply gives about 0.17 deg change in phase angle.)

The main cause of the slight remaining variation in phase angle with supply voltage is the voltage dependence of the collector/base capacitances of the transistors.

The collector/base capacitance of VT1 (Fig. 11) comes effectively across R_{fb} , since the emitter of VT2 has almost the same alternating voltage on it as has the collector of VT1.

The effect of the collector/base capacitance of VT2 is rather more difficult to appreciate. If there were infinite loop gain, the effect of the overall negative feedback on VT1 and VT2 would be to determine precisely, for a given input current to VT1 base, the amplitude and phase of VT2 emitter current. Regarding VT2 as being an ideal transistor with the collector/base capacitance connected externally, the collector current would also be closely determined. The collector alternating voltage being much larger than the base voltage, the division of the collector current of the ideal transistor between the collector load circuit and the collector/base capacitance would be approximately as if this capacitance were in parallel with the collector load. Thus the effect of the collector/base capacitance of VT2 on the phase angle of the main-

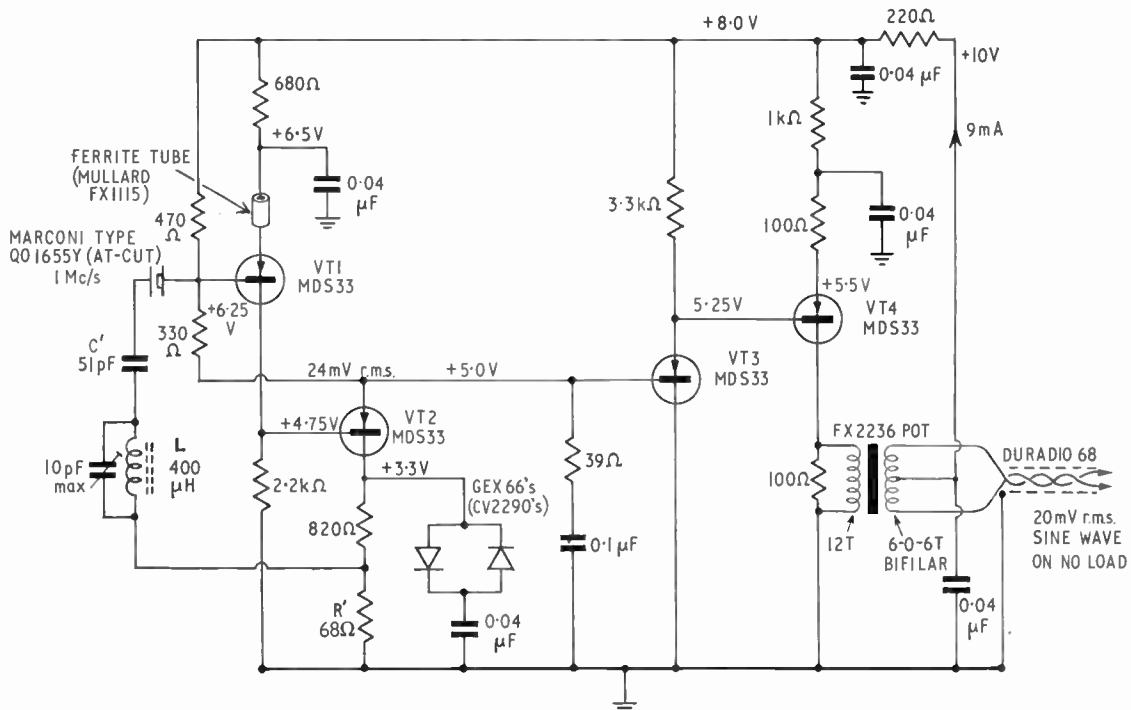


Fig. 12. The author's series-resonance crystal oscillator. (The Germanium transistors used, made by Semiconductors Ltd., are of microalloy-diffused type, having a minimum f_i rating of 300 Mc/s. Many other suitable transistors are now available: if silicon transistors are used, the emitter resistor of VT1 should be reduced in value to maintain VT2 collector at about +3.3 volts.)

taining amplifier would be approximately as if the capacitance appeared directly across the collector load circuit.

However, if the loop gain is not assumed infinite, then the phase lag introduced by VT2 collector/base capacitance is increased in the following way. There is a voltage gain in the region of 10 between base and collector of VT2, so that the collector/base capacitance appears considerably amplified (Miller effect) across VT1 collector load and gives a relatively large phase lag. This lag is inside the forward path of the feedback amplifier so its effect on the overall phase angle is greatly reduced by the feedback. As the supply voltage is reduced, however, the feedback loop gain falls off, so that the lag becomes more effective. In the practical design, this mechanism contributes rather more overall lag than the first mechanism mentioned above.

Adding a 22 pF capacitor between collector and base of VT2 gave a 2.7 times greater frequency change than adding the same capacitor across VT2 collector load. This is about what the above theory would predict.

3.6. Adjustment to Suit Crystal Specimen

Whilst the circuit as finally designed (see Fig. 12) will, without adjustment, perform fairly satisfactorily with crystal specimens having a wide range of effective series resistance, nevertheless it is preferable to set the circuit up to obtain the intended a.c. levels with the particular crystal employed. This is best done by suitably altering the tapping point on VT2 collector load. Thus a crystal with a low effective series resistance will be fed from a low source resistance, and vice versa, a state of affairs tending to give a constant percentage degradation of Q -value.

3.7. Suppression of Parasitic Oscillation

It will be seen from Fig. 12 that a ferrite tube is placed on VT1 emitter lead. This constitutes a one-turn inductor whose purpose is to give attenuation of loop gain at high frequencies in such a manner as to obtain a good margin of feedback stability. Without the ferrite tube, a parasitic oscillation at many megacycles occurred with some transistor samples, and this was accompanied by a several times increase in the rate of change of the wanted frequency with supply voltage.

The by-pass capacitor in VT1 emitter circuit has been chosen, in association with the inductance value, so that the two elements are series resonant at 1 Mc/s, thus providing a very low impedance, of small phase angle, in VT1 emitter circuit, without requiring an exorbitant capacitor value.

It would probably be satisfactory to omit the ferrite

tube in versions of the circuit using transistors of lower cut-off frequency.

3.8. Means for Frequency Adjustment

It was decided to design the oscillator so that it could be set precisely to 1 Mc/s with any crystal sample within the maker's frequency tolerance of ± 50 c/s.

Referring to the equivalent circuit of Fig. 1, the impedance of the left-hand part is so low at frequencies within ± 50 c/s of the crystal series-resonant frequency that the presence of C_0 across the crystal does not exert a major influence. Consequently it is a good approximation to say that if a small amount of reactance is added in series with the crystal terminals, it simply adds to the reactance of L_x and C_x and hence modifies the resonant frequency slightly. The reactances of C_x and L_x , for the type of crystal used, are numerically approximately 20 M Ω each. Consequently, the addition in series with the crystal of 2 k Ω of capacitive reactance will change the total capacitive reactance by 1 part in 10^4 and will therefore increase the frequency by approximately 50 c/s. Similarly, 2 k Ω of series inductive reactance will lower the frequency by approximately 50 c/s.

By connecting a fixed inductor of 400 μ H ($X_L = 2510 \Omega$ at 1 Mc/s) in series with the crystal, together with a series capacitor, the total series reactance added may be varied from +2 k Ω to -2 k Ω to give a variation of ± 50 c/s, by adjusting the reactance of the series capacitor over the range 510 ohms to 4510 ohms, which requires a capacitance range of 312 pF to 35.3 pF. In the oscillator as constructed, the approximate capacitance required was obtained by soldering in an appropriate fixed capacitor, a fine adjustment, covering approximately ± 6 c/s, being provided by a small trimmer capacitor across the inductor, as shown in Fig. 12. The range of this fine adjustment is substantially unaffected by the value of the fixed series capacitor used.

The arrangement adopted was used because it was thought desirable to avoid a variable inductor, which might have poor stability unless of special construction, and a large-value variable capacitor for the coarse adjustment was avoided both for stability reasons and because of its inconveniently large physical size.

Care was taken in the construction of the inductor to keep the stray capacitance as small as possible, also in the interests of long-term stability. A single-layer winding of 44 s.w.g. enamelled wire was used on a 0.4 in diameter former, with the dust core permanently waxed in after setting to 400 μ H. The shunt stray capacitance of this coil is approximately 1 pF only. The Q at 1 Mc/s is about 70.

3.9. Reasons for Variation in Amplitude with Frequency Setting

The amplitude of the sine-wave output increases by about 10% as the crystal series capacitor is varied from its minimum value (giving 50 c/s above the true crystal frequency) to its maximum value (giving 50 c/s below the true crystal frequency). There are two reasons for this effect, both of which involve stray capacitances.

The first significant stray capacitance is that across the crystal. The equivalent circuit for the crystal plus frequency-adjusting series reactance is shown in Fig. 13(a). The impedance of the complete network must be purely resistive at the oscillation frequency, assuming negligible phase shift in the maintaining amplifier. When X_x is nearly zero (frequency near to crystal resonant frequency), the presence of C_0 has very little effect indeed, since its reactance is numerically very much greater than r_x . When the magnitude of X_f is made large, causing the crystal frequency to change and X_x to become correspondingly large, C_0 then appears across a much higher impedance than before and its effect is quite significant. The series resistive component r' of the impedance of the above network is given by formula 5 of Table 1; since C_f in the present situation is always large compared with C_0 , a simple binomial expansion may be used, yielding the result, in reactance form:

$$r' \approx r_x(1 + 2X_f/X_0) \dots\dots(8)$$

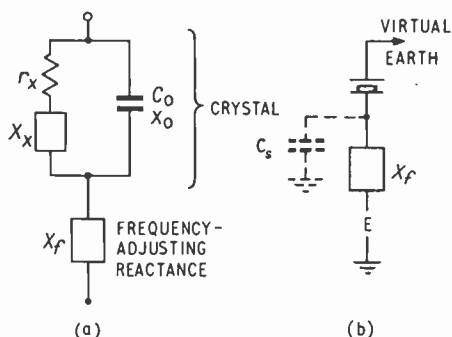


Fig. 13. Diagrams illustrating behaviour of frequency-adjusting circuit.

At 50 c/s above the crystal resonant frequency, X_f is approximately $-2\text{ k}\Omega$, X_0 being about $-69\text{ k}\Omega$ (see Fig. 1). Equation (8) then gives $r' = 1.06r_x$. At 50 c/s below the crystal frequency, X_f is $+2\text{ k}\Omega$, giving $r' = 0.94r_x$. Thus over the complete tuning range required there is a theoretical variation from this cause of about 12% in the effective series resistance of the crystal and its frequency-adjusting reactance, the lowest resistance occurring at the low-frequency end of the tuning range. The percentage variation in

the total series resistance of the crystal circuit in the oscillator is, of course, less than the above figure, since there is some additional series resistance of constant value contributed by the collector circuit of VT2 and by the imperfect virtual earth at VT1 base.

The second mechanism giving a variation in amplitude with frequency setting involves the inevitable presence of stray capacitances to earth from the junction of the crystal and the frequency-adjusting reactance X_f —see Fig. 13(b). Here E is the effective e.m.f. obtained from the limiter circuit. If X_f is made negative (capacitive) then a capacitive potential divider is formed by X_f and C_s , with the result that the effective e.m.f. driving the crystal is reduced by the presence of C_s . If, however, X_f is positive (inductive), the presence of C_s increases the effective e.m.f. driving the crystal. Another effect is involved, however, which works in the opposite direction. In series with E (but not shown in Fig. 13(b)) is some resistance, i.e. that seen looking back into the resistive tapping on the limiter circuit. When X_f is negative, this resistance appears with reduced magnitude in series with the crystal, and vice versa; the theory is that just given in connection with the effect of crystal shunt capacitance. This tends to offset the previous effect, but in any case, with good layout and construction, the effect of C_s is small compared with that due to the crystal shunt capacitance C_0 .

Thus, to minimize the amplitude variations, the layout and construction should be such as to augment the above-mentioned capacitances as little as possible.

If the stray capacitance across the crystal is increased too far, there is a danger of a wrong mode of oscillation occurring, involving a series $L-C$ tuned circuit, the crystal merely behaving as a small capacitance in this circuit. With the crystal capacitance C_0 not appreciably augmented, however, the effective series loss resistance of this circuit at resonance is too great for the unwanted oscillation to occur, with the type of crystal employed.

If a crystal with a considerably higher C_0 were used, it might then be necessary to connect a resistor, having a value of the order of $100\text{ k}\Omega$, across the crystal in order to suppress the unwanted oscillation.

Some variation in amplitude (about $\pm 8\%$) also occurs when the fine tuning trimmer is operated. The coil has series resistance (approximately 36 ohms), so that the equivalent circuit of the coil shunted by the trimmer is of the same form as the top part of the Fig. 13(a) network. The effective series resistive component of the total impedance thus increases as the trimmer capacitance increases, giving a reduction in oscillation amplitude.

A fine frequency control almost free from amplitude variation could be provided, if required, by shunting the trimmer across another capacitor in series with the crystal, thus still leaving the coarse tuning capacitor free to be changed without affecting the sensitivity of the fine tuning control.

If, in addition to capacitances across the crystal, and to earth from the junction of the crystal and the tuning reactance, there are shunt resistive losses associated with these capacitances, then a further mechanism for amplitude variation with series tuning capacitance value is introduced.

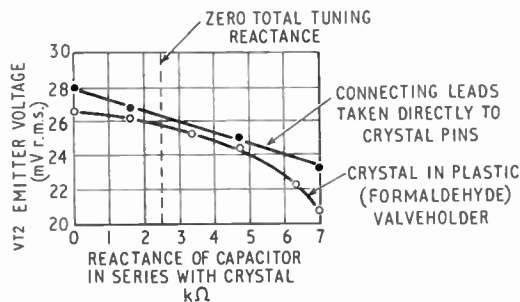


Fig. 14. Curves showing effect of insulation losses on behaviour of frequency-adjusting circuit.

Thus, for example, suppose a loss resistance R exists across the crystal, due, perhaps, to a crystal socket of low-grade dielectric. The value of this shunt loss resistance will normally, of course, be very high compared with the total crystal reactance, even when the crystal has been detuned 50 c/s from its natural frequency. The effective series resistance, equivalent to this constant shunt resistance, is then easily shown to be approximately proportional to the *square* of the total crystal reactance. This effect thus gives a reduction in oscillation amplitude when the crystal is detuned *either* side of its natural frequency. The effects discussed earlier in this Section, however, give a variation of effective series resistance which is an approximately *linear* function of the tuning reactance.

Figure 14 shows convincing experimental evidence of the functioning of these different mechanisms.

In systems designed to give wide-band linear frequency modulation of a crystal oscillator, it is necessary to employ special circuit arrangements^{12, 13} to avoid the amplitude variation caused by the presence of unwanted shunt capacitances. The present results emphasize the further necessity, in such systems, to employ only low-loss insulating materials.

Even in oscillators such as that described here, required merely to supply a constant output frequency with good stability, it is desirable to minimize shunt

insulation losses, since variation in such losses, e.g. with humidity, will have some effect on frequency.

3.10. Output Isolating Circuit

The sine-wave output from the crystal oscillator could be taken directly from VT2 emitter (Fig. 12), but the frequency would then be affected to some extent by variations in the reactive loading imposed on this point.†

By the addition of the transistors VT3 and VT4, the frequency is made virtually independent of any likely changes in the magnitude or phase of the output load. The switching on and off of a 100 Ω reactive load gives a frequency change not exceeding about 1 part in 10¹⁰.

It will be seen that arrangements are provided for supplying the d.c. power along the same balanced r.f. cable as is used to convey the 1 Mc/s output. The 10-V supply is connected to the centre-tap of a transformer winding at the other end of the line.

4. Some Performance Measurements

4.1. Measurements on the Complete Oscillator

The full-line curve in Fig. 15 shows the variation in frequency with d.c. supply voltage for the complete circuit of Fig. 12, the crystal constants being approximately as in Fig. 1. At the normal working voltage of 10 V, the variation is about -1.5 parts in 10⁸ per volt. (The broken-line curve applies when the modification described in Section 6 is incorporated.)

Adding 30 Ω of series resistance in the crystal circuit was found to reduce the frequency by about 1 part in 10⁸ (see also Section 2.3.3).

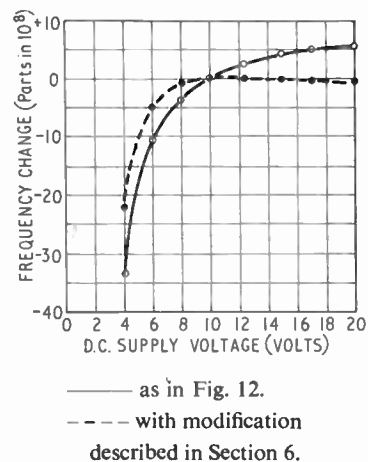


Fig. 15. Variation of frequency with supply voltage for circuit of Fig. 12.

† The output should be taken via a series resistor, otherwise a few tens of pF's of capacitive loading may cause the negative feedback loop to become unstable, resulting in v.h.f. oscillation.

By voltage measurements it was deduced that the current in the 330Ω feedback resistor, was 0.074 mA r.m.s. This is also, approximately, the crystal current, giving a dissipation in the 77 Ω effective series resistance of the crystal of 0.42 μW.†

Figure 16 shows the variation in output voltage as a function of supply voltage, both when the output is taken from VT4 collector and when it is taken from VT2 emitter. The difference between these outputs is due to the finite g_m of VT3 and VT4, the presence of $r_{bb'}$ in each, and the existence of shunt core loss in the output transformer.

The effect on frequency of warming the transistors was investigated and was found to be well under 1 part in 10^9 per deg C.

4.2. Measurement of Crystal Parameters

The effective series resistance of the crystal may be determined as follows:

(a) The limiter diodes are replaced by a potentiometer, whose value is adjusted so as just to cause oscillation.

(b) Resistance r_s is added in series with the crystal, stopping oscillation.

(c) Resistance R_p is added across the crystal, the value being adjusted so as just to produce oscillation again.

Then it is easily shown that

$$r_x = \frac{r_s}{2} (1 + \sqrt{1 + 4R_p/r_s}) \quad \dots\dots(9)$$

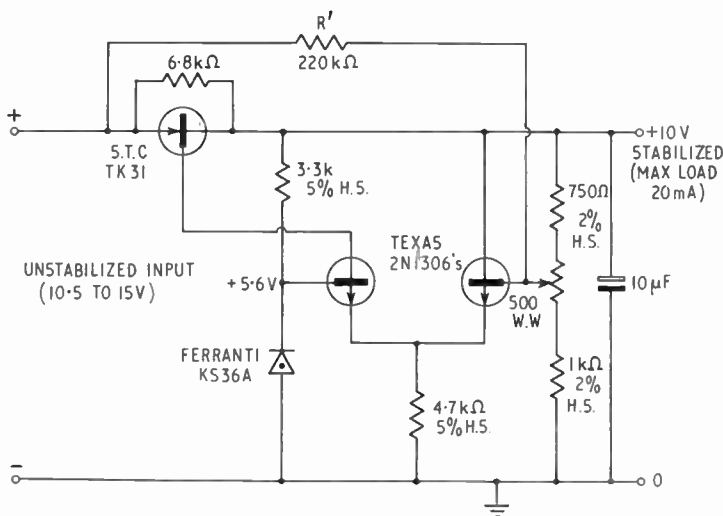


Fig. 17. Voltage stabilizer for crystal oscillator.

† The signal level nevertheless exceeds the noise level in the circuit by an enormous factor, giving good spectral purity of output.

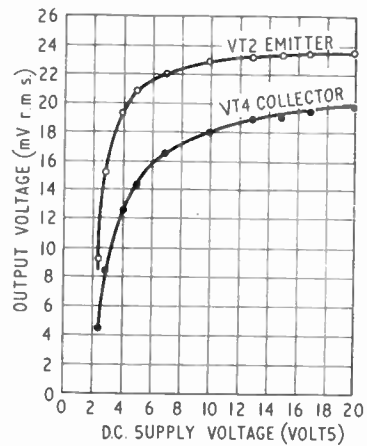


Fig. 16. Variation of alternating voltages in Fig. 12 circuit with d.c. supply voltage.

This measurement is preferably done with the series tuning elements of Fig. 12 removed and replaced by a suitable d.c. blocking capacitor.

The values of L_x and C_x for the crystal (see Fig. 1) may be deduced by determining the frequency change caused by inserting a known capacitive reactance in series with the crystal. Several known values of reactance should preferably be used, and a graph should be plotted. This will be a straight line unless excessive values of series reactance are used. The fractional frequency change will be half the fractional change in total capacitive reactance, so that the reactance of C_x , and hence C_x , may be deduced. L_x is then easily calculated.

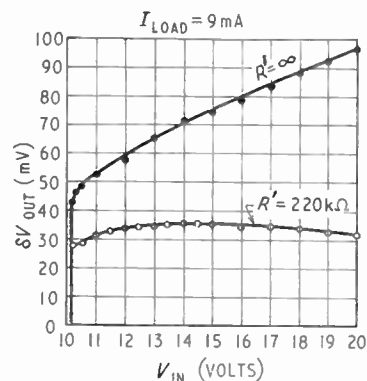


Fig. 18. Curves showing performance of Fig. 17 circuit.

5. Voltage-Stabilizing Circuit

The circuit of Fig. 17 provides a stabilized 10 V supply for the oscillator from a battery voltage of nominally 12 V. It reduces the effect on frequency of normal battery-voltage variations to less than 1 part in 10^{10} , even when the oscillator modification described in Section 6 is not employed.

An unusual feature is that the series transistor has its *emitter* connected to the unstabilized supply, instead of the collector as in more normal practice. This innovation, which has considerable advantages, was first seen by the author in a circuit designed at Durham University,¹⁶ though the author's colleague R. C. Bowes also devised it, quite independently.

The performance of the Fig. 17 circuit is shown in Fig. 18, both with and without the 'feed forward' resistor in operation. It should be noted that good stabilizing action is obtained until the input voltage falls to within about 0.2 V of the output voltage.

The 6.8 k Ω resistor across the series transistor is to ensure that the circuit starts up correctly when the supply is switched on. Without it, it would appear to be possible for the circuit to remain indefinitely with none of the transistors conducting. In the version built, however, the transistor leakage currents were sufficient to make the circuit start up even without the 6.8 k Ω resistor.

The stabilized output voltage was found to vary with the temperature of the whole stabilizer circuit by well under 1 mV/deg C.

The low-frequency output impedance is about 4 Ω .

6. Oscillator Modification to Reduce Frequency Variation with Supply Voltage

As stated in Section 3.5, the slight variation in maintaining amplifier phase angle which occurs when the supply voltage changes is due almost entirely to the presence of collector/base capacitance in the transistors.

When the supply voltage falls, the collector/base capacitances increase, and that of VT2 is also rendered more effective by the fall in loop gain which occurs.

Compensation of the above effects requires the addition to the circuit of some element which will give an increasing phase lead as the supply voltage falls.

A simple and effective method is to connect a variable-capacitance diode† across the 820 Ω resistor in VT2 collector (see Fig. 12). The broken-line curve

† This is a junction diode used in the back-biased condition, giving a capacitance inversely proportional to the square root of the applied voltage. Commercial diodes intended for use in this manner are available.

in Fig. 15 shows the effect of connecting a Ferranti ZC10A in this position. With the supply at its normal value of 10 V, this diode has approximately 3 V across it and its mean capacitance is then approximately 20 pF.

There are obviously several possible ways of making such compensation adjustable. A satisfactory arrangement is to choose a diode (or two in parallel if necessary) which would give over-compensation if connected as discussed above; this diode is then connected between VT2 collector and the slider of a potentiometer of about 1 k Ω connected across the lower of the two resistors in VT2 collector circuit.

7. Conclusions

The series-resonance oscillator circuit (Figs. 11 and 12), when used in conjunction with the voltage stabilizer circuit of Fig. 17, reduces frequency instability due to likely variations in the unstabilized supply voltage, and due to drift in the effective series resistance of the crystal, to less than 1 part in 10^{10} , even with a crystal of only moderate Q -factor. The power dissipated in the crystal is less than one microwatt, giving a low rate of frequency drift due to crystal ageing.

The cause of the change in frequency with maintaining circuit supply voltage is found to be mainly the change in transistor collector capacitances with voltage, and this effect may be compensated by means of a variable-capacitance diode, as described in Section 6.

The well-known Pierce circuit, using only one transistor, is shown to be capable of a far better performance than might at first be expected, the variation in frequency with supply voltage to the oscillator circuit, using the version shown in Fig. 4(b), being about 5 parts in 10^9 for a 10% supply voltage change.

The Miller circuit of Fig. 6 is shown to be much less desirable than the Pierce circuit in that its frequency is much more dependent on variations in the crystal losses, but the Pierce circuit, in its turn, is inferior, in this respect, to the series-resonance oscillator of Figs. 11 and 12.

Operation of the simple Pierce circuit of Fig. 4(b) at a crystal dissipation of only about one microwatt is found to be perfectly practicable without using special a.g.c. arrangements.

The close relationship between the Pierce crystal oscillator and the Gouriet/Clapp L - C oscillator is pointed out, and the influence exerted by the manner of drawing the circuit diagram on the way one thinks about the functioning of the circuit is emphasized, it being argued that the circuit is most simply regarded as a form of series-resonance oscillator.

An unusual equivalent circuit for a crystal is derived in Section 2.6, and is particularly appropriate for use in parallel-resonance oscillators such as the Miller. The operation of the FMQ frequency-modulation system^{12, 13} is shown to be very simply explained with the help of this equivalent circuit.

The theory of circuit arrangements permitting adjustment of the oscillator frequency is discussed in detail in Sections 3.8 and 3.9, including the reasons for variation in amplitude with frequency setting.

A practical method for determining the crystal equivalent circuit parameters, using the series-resonance oscillator circuit itself, is given in Section 4.2.

Tables of convenient design formulae derived by the author are given, and these enable the effect on frequency of variations in crystal shunt capacitance, series tuning capacitance, and effective series resistance to be determined quickly for both the series-resonant and the shunt-resonant condition.

8. References

1. J. F. Mercurio, "Stable, low-cost 1 Mc/s oscillator", *Electronics*, **32**, No. 6, p. 50, 6th February 1959.
2. K. H. Sann, "Miniaturized High-Precision Crystal Oscillator", Diamond Ordnance Fuze Laboratories Report No. TR/878, 1960.
3. P. L. Fleck, "A Transistorized Frequency Standard", M.I.T. Lincoln Laboratory Group Report No. 34/85, January 1960.
4. J. L. Creighton, H. B. Law and R. J. Turner, "Crystal oscillators and their application to radio transmitter control", *J. Instn Elect. Engrs*, **94**, Pt. III A, No. 12, p. 331, 1947.
5. H. Stanesby and P. W. Fryer, "Variable-frequency crystal oscillators", *J. Instn Elect. Engrs*, **94**, Pt. III A, No. 12, p. 368, 1947.
6. L. B. Turner, "Constant temperature: a study of principles in electric thermostat design", *J. Instn Elect. Engrs*, **81**, p. 399, 1937.
7. L. A. Meacham, "The bridge stabilized oscillator", *Proc. Inst. Radio Engrs*, **26**, p. 1278, October 1938.
8. W. A. Marrison, "The evolution of the quartz crystal clock", *Bell Syst. Tech. J.*, **27**, p. 510, 1938.
9. G. G. Gouriet, "High-stability oscillator", *Wireless Engineer*, **27**, p. 105, 1950.
10. J. K. Clapp, "An inductance-capacitance oscillator of unusual frequency stability", *Proc. Inst. Radio Engrs*, **36**, p. 356, 1948.
11. K. Holford, "Transistor L-C oscillator circuits", *Mullard Tech. Commun.*, **5**, No. 41, p. 17, December 1959.
12. W. S. Mortley, "FMQ", *Wireless World*, **57**, No. 10, p. 399, October 1951.
13. W. S. Mortley, "Frequency modulated quartz oscillators for broadcasting equipment", *Proc. Instn Elect. Engrs*, **104**, Pt. B, No. 15, p. 239, May 1957.
14. J. S. Murray, "Transistor bias stabilization", *Electronic Radio Engr*, **34**, p. 161, 1957.
15. 'Cathode Ray', "Virtual earth", *Wireless World*, **67**, No. 11, p. 595, November 1961.
16. F. J. U. Ritson and R. C. Foss, "Transistor power supplies with limited overload current", *Electronic Engng*, **34**, No. 414, August 1962.

*Manuscript first received by the Institution on 11th December 1963 and in final form on 1st December 1964.
(Paper No. 974.)*

© The Institution of Electronic and Radio Engineers, 1965

The Generation of a Selected Harmonic or Sub-harmonic by means of a Single Externally-driven Switch

By

D. P. HOWSON, M.Sc.

(Associate Member)†

Summary: It is shown that it is possible to generate a selected harmonic with high efficiency using a circuit containing a single externally-operated switch driven by the signal source. Higher efficiencies are obtainable at the cost of less convenient circuit impedances if the switch is operated in quadrature rather than in phase with the controlling input. The circuit cannot function with ideal switches, and the efficiencies quoted are for practical switches. For simplification, however, ideal tuning is assumed. A circuit in which the switch is controlled from both input and output is also considered, and shown to give good results although it is not self-starting. Finally, applications of the circuits to sub-harmonic generators are mentioned.

1. Introduction

In a recent paper¹ Tucker has shown that highly efficient generation of selected harmonics and sub-harmonics is possible using a lattice of externally-driven switches. (The efficiency considered was the efficiency of conversion of the *available* source power to a harmonic frequency.) The new principle involved was essentially that by controlling the switches directly from the signal source and not by the voltages appearing across them (as in normal rectifier circuits) the limitations associated with rectifier harmonic generators could be avoided. The efficiency was significantly higher than that reported for varactor or Boff diode multipliers, and gave rise to hopes that it might be possible to develop an alternative form of high-frequency device. However, three basic problems remained to be solved before this could become a possibility: the provision of a high-frequency externally-driven switch, the development of a circuit which did not require a high ratio of source to load resistance for the generation of high-order harmonics, and the development of an unbalanced circuit requiring a smaller number of switches—preferably one only. The solution of the first of these problems is in the field of device design, and some possibilities were enumerated in the previous paper. The solution of the two other problems is the subject of the present paper, and stems from some previous work on single-balanced modulators.² In this it was shown that modulators using only one time-varying resistance could be made highly efficient if the resistance is continually switched between two widely differing values, remaining, however, at one of these for a large proportion of each cycle. Such a circuit is unrealizable, theoretically, with a perfect switch changing from open-

circuit to short-circuit, which explains why the analysis in this paper differs in this important respect from the previous paper on this subject.¹

It will be assumed throughout this work that the amount of power that has to be taken from the signal source in order to actuate the switch is negligible. This assumption is theoretically sound, and it should be possible to approach the condition in practice. Analysis will be concentrated upon circuits in which the switch is in series with two anti-resonant circuits tuned to input and wanted harmonic frequencies. There will be in every case corresponding circuits in which a switch is in parallel with two resonant circuits.² In all cases the filters used to select fundamental and harmonic frequencies in the generator will be considered to be 'ideal'. This means that they will be considered capable of completely short-circuiting all unwanted voltages, whilst presenting a very high impedance to the wanted signals. Since practical filters will fall short of these requirements, in some measure,³ the harmonic generator will have less efficiency than is shown by the analysis to follow.

2. Harmonic Generator Controlled by the Input Signal

Figure 1 gives the circuit of a harmonic generator using a single externally-controlled switch, which changes in conductance from g_f to g_b at times dependent on the phase and magnitude of the input signal. Hence the time-varying conductance of the switch may be written

$$g(t) = g_0 + \sum_{n=1}^{\infty} 2g_n \cos n(\omega_p t + \theta) \quad \dots(1)$$

If the input signal is $I \cos \omega_p t$, then the current through the switch will be at all harmonic frequencies and may be written

$$2i = \sum_{r=1}^{\infty} [i_r \exp(jr\omega_p t) + i_r^* \exp(-jr\omega_p t)] \quad \dots(2)$$

† Electronic and Electrical Engineering Department, University of Birmingham.

The voltage across the switch will be at input and wanted harmonic frequency only, and can therefore be written

$$2v = v_0 \exp(j\omega_p t) + v_0^* \exp(-j\omega_p t) + v_m \exp(jm\omega_p t) + v_m^* \exp(-jm\omega_p t) \dots(3)$$

Evaluating the equations⁴ for current balance at frequencies ω_p and $m\omega_p$ (strictly at $\exp(j\omega_p t)$ and $\exp(jm\omega_p t)$) we have respectively

$$i_0 = v_0 g_0 + v_0^* g_2 \exp(j2\theta) + v_m g_{m-1} \exp(-j(m-1)\theta) + v_m^* g_{m+1} \exp(j(m+1)\theta) \dots(4)$$

and

$$i_m = v_0 g_{m-1} \exp(j(m-1)\theta) + v_0^* g_{m+1} \exp(j(m+1)\theta) + v_m g_0 + v_m^* g_{2m} \exp(j2m\theta) \dots(5)$$

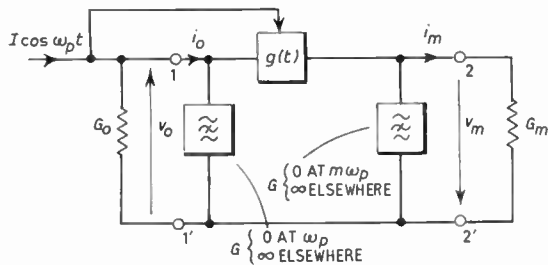


Fig. 1. Harmonic generator controlled from the input.

In general these equations may be evaluated by splitting the voltage terms into real and imaginary parts (see Appendix), finally arriving at four equations with real coefficients. However when $\theta = 0$, and both terminations are pure conductances, so that

$$G_0 v_0 = (I - i_0) \dots(6)$$

and

$$-G_m v_m = i_m \dots(7)$$

the equations reduce to

$$I = v_0(G_0 + g_0 + g_2) + v_m(g_{m-1} + g_{m+1}) \dots(8)$$

$$0 = v_0(g_{m-1} + g_{m+1}) + v_m(G_m + g_0 + g_{2m}) \dots(9)$$

Hence

$$v_m = \frac{(g_{m-1} + g_{m+1})I}{(G_0 + g_0 + g_2)(G_m + g_0 + g_{2m}) - (g_{m-1} + g_{m+1})^2} \dots(10)$$

The efficiency, η , of generation of harmonic power is defined as the ratio of the output power to the available input power. Therefore

$$\eta = 4G_0 G_m \left[\frac{v_m}{I} \right]^2 \dots(11)$$

$$\eta = \frac{4G_0 G_m (g_{m-1} + g_{m+1})^2}{[(G_0 + g_0 + g_2)(G_m + g_0 + g_{2m}) - (g_{m-1} + g_{m+1})^2]^2} \dots(12)$$

For maximum efficiency, the harmonic generator must be conjugately matched to source and load.

Hence⁵

$$G_0 = g_0 + g_2 - \frac{(g_{m-1} + g_{m+1})^2}{G_m + g_0 + g_{2m}} \dots(13)$$

and

$$G_m = g_0 + g_{2m} - \frac{(g_{m-1} + g_{m+1})^2}{G_0 + g_0 + g_2} \dots(14)$$

Solving (13) and (14) for G_0 and G_m ,

$$G_0^2 = (g_0 + g_2)^2 - \frac{(g_0 + g_2)}{(g_0 + g_{2m})} (g_{m-1} + g_{m+1})^2 \dots(15)$$

$$(g_0 + g_{2m})G_0 = (g_0 + g_2)G_m \dots(16)$$

Substitution of these values in the efficiency eqn. (12) gives eventually

$$\eta = \frac{g_0 + g_2 - G_0}{g_0 + g_2 + G_0} \dots(17)$$

so that the efficiency tends to 100% when $G_0 \ll (g_0 + g_2)$. In Fig. 2 efficiency is plotted as a ratio of G_0/g_0 , assuming $g_0 \approx g_2$. From (15) η can be high if $g_0, g_2, g_{m-1}, g_{m+1}$, and g_{2m} are of the same magnitude—which occurs with impulsive switching.

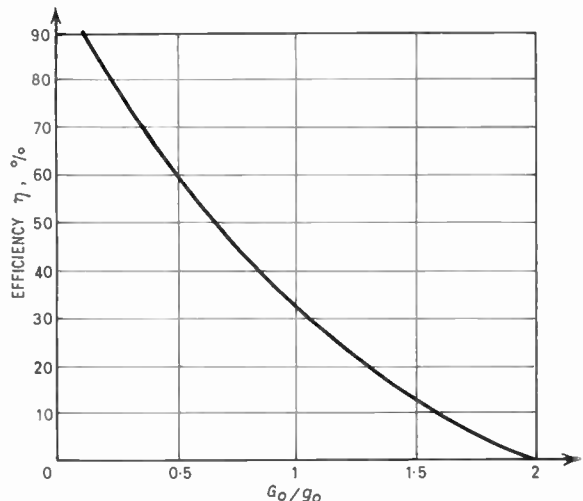


Fig. 2. Efficiency as a function of G_0/g_0 , for $g_0 \approx g_2$.

To show this let s be the fraction of the cycle during which the switch has a high conductance, g_f . Then²

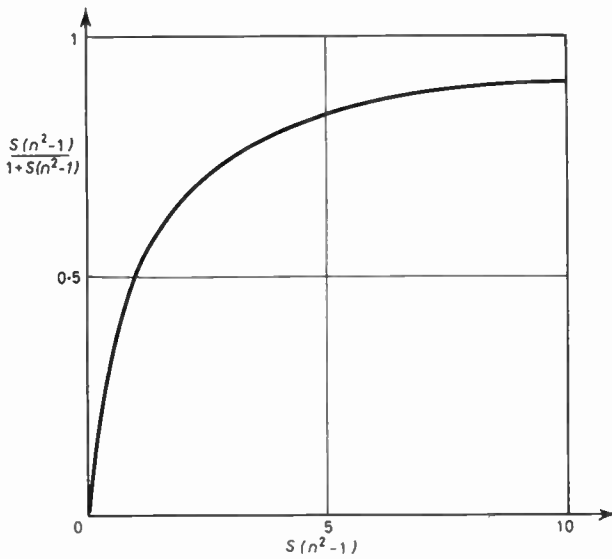


Fig. 3. Graph of $\frac{s(n^2 - 1)}{1 + s(n^2 - 1)}$ against $s(n^2 - 1)$.

$$g(t) = g_b + s(g_f - g_b) + \frac{2}{\pi}(g_f - g_b) \sum_{r=1}^{\infty} \left(\frac{\sin r\pi s}{r} \right) \cos r\omega_p t \dots\dots(18)$$

where g_b is the low conductance taken by the switch for the rest of the cycle. Note that

$$g_r = \frac{1}{\pi}(g_f - g_b) \cdot \frac{\sin r\pi s}{r} \dots\dots(19)$$

and

$$g_0 = g_b + s(g_f - g_b) \dots\dots(20)$$

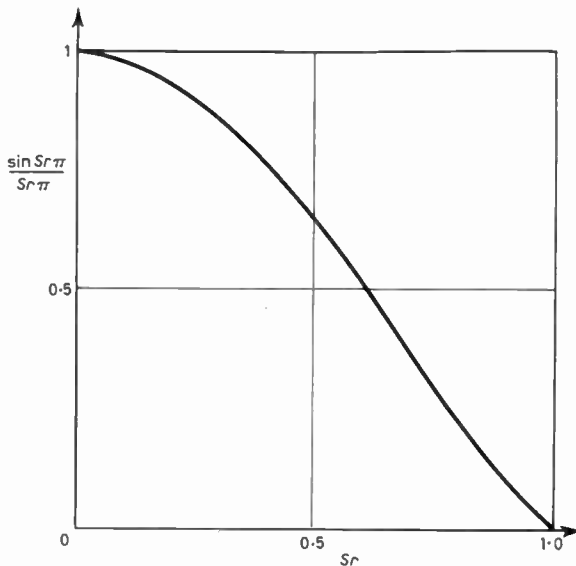


Fig. 4. Graph of $\frac{\sin sr \pi}{sr \pi}$ against sr .

so that

$$\frac{g_r}{g_0} = \left\{ \frac{s(n^2 - 1)}{1 + s(n^2 - 1)} \right\} \cdot \left(\frac{\sin r\pi s}{r\pi s} \right) \dots\dots(21)$$

where

$$n^2 = \frac{g_f}{g_b} \dots\dots(22)$$

The first bracket in (21) is therefore a function of the mark/space ratio of the switching signal, and of the quality of the switch. The value of the quantity in the bracket is plotted as a function of $s(n^2 - 1)$ in Fig. 3. It can be seen that for a high-quality switch the function approaches unity, as $s(n^2 - 1)$ is large for all except extremely small values of s .

The second bracket in (21) is a function of s and of the harmonic of the switching function being evaluated. It is plotted in Fig. 4 as a function of sr .

Using these curves, the efficiency of a harmonic generator for which $s = 1\%$ was calculated, for varying values of n^2 , and is presented in Fig. 5. Some experimental points for $n^2 = 0.5 \times 10^3$ are also shown, and can be seen to give close agreement with the theoretical results. To obtain these results, an

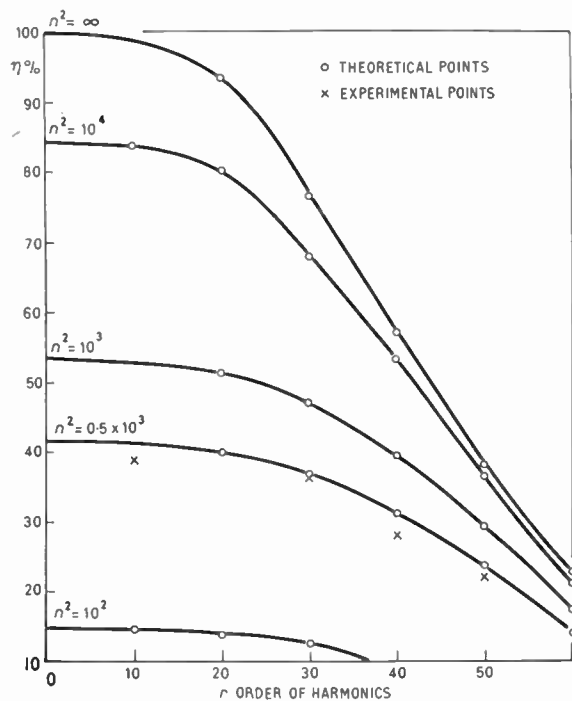


Fig. 5. The efficiency of harmonic generator with $\theta = 0$, for $s = 0.01$.

ASZ 21 transistor was used as the switch, and the circuit was operated with input frequencies around 2 kc/s.

The transistor was switched by a negative-going pulse derived from an external pulse generator, synchronized to the input frequency—see Fig. 6.

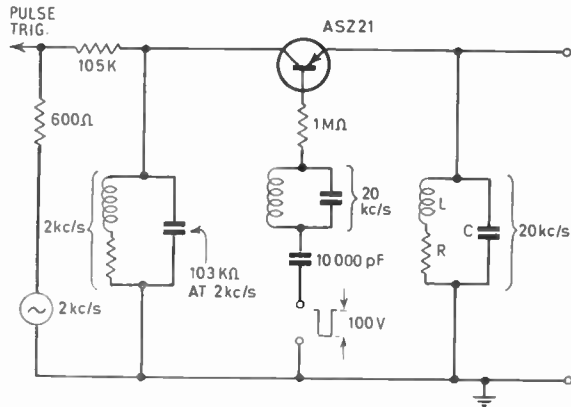


Fig. 6. Experimental circuit.

The rejector circuit in the base of the transistor was added to reduce to negligible proportions the voltage across the output tuned circuit due to the pulse current alone. The 1 MΩ resistor and 10 000 pF capacitor in the base circuit were added to prevent the input voltage being rectified by the collector-base junction, and to minimize the shunting effect of the impedance of the pulse source.

Figure 7 shows the voltage across the output tuned circuit for the generation of the tenth harmonic, with $s = 0.1$. The harmonic amplitude was assessed by taking the average of the decrement visible, after calibration of the oscilloscope. Since in the cases of interest, when the efficiency was high, the output contained only small amounts of unwanted modulation products, it was felt that the method was sufficiently accurate. Figure 8 shows an expanded view of the output waveform; the distortion visible on one peak of the waveform marks the part of the cycle over which the transistor was conducting. 62.5 per cent of the output power is available in the load resistance.

2.1. Switch Operated in Quadrature with Input Signal ($\theta = \pi/2$ rad)

Under this condition, (4) and (5) reduce to

$$I = v_0(G_0 + g_0 - g_2) - jv_m(g_{m+1} - g_{m-1}) \dots(23)$$

$$0 = jv_0(g_{m+1} - g_{m-1}) + v_m(G_m + g_0 - g_2) \dots(24)$$

as shown in the Appendix, as long as m is a multiple of

four, and the harmonic generator is 'cold-tuned'. The input admittance looking into (1, 1') is

$$Y'_0 = g_0 - g_2 - \frac{(g_{m+1} - g_{m-1})^2}{G_m + g_0 - g_2} \dots\dots(25)$$

and the output admittance looking into (2, 2')

$$Y'_m = g_0 - g_2 - \frac{(g_{m+1} - g_{m-1})^2}{G_0 + g_0 - g_2} \dots\dots(26)$$

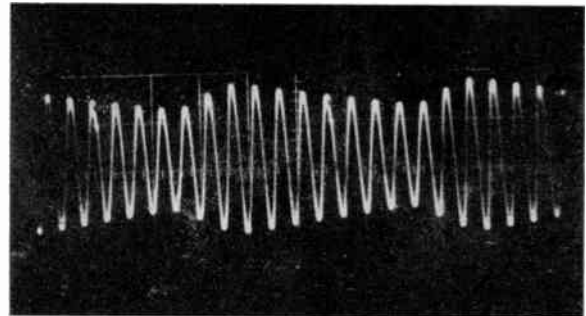


Fig. 7. Output waveform for 10th harmonic.

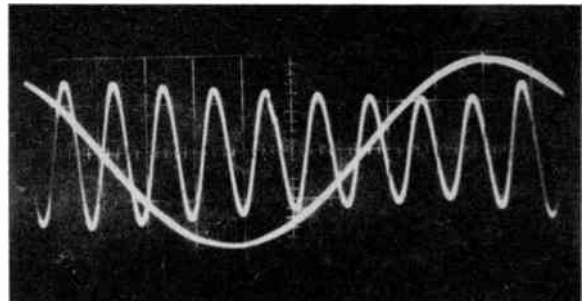


Fig. 8. Output waveform for 10th harmonic, showing transistor conducting.

These are pure conductances in each case, just as when $\theta = 0$. If the generator is conjugately matched to source and load, to ensure maximum efficiency, we find

$$G_0^2 = (g_0 - g_2)^2 - \left\{ \frac{g_0 - g_2}{g_0 - g_2} \right\} (g_{m+1} - g_{m-1})^2 \quad (27)$$

and

$$(g_0 - g_2)G_0 = (g_0 - g_2)G_m \quad \dots\dots(28)$$

Defining efficiency in the same way as before

$$\eta = \frac{4G_0G_m(g_{m+1} - g_{m-1})^2}{[(G_0 + g_0 - g_2)(G_m + g_0 - g_2) - (g_{m+1} - g_{m-1})^2]^2} \dots\dots(29)$$

Table 1

The efficiency of the generation of the *m*th harmonic, for *s* = 0.1

<i>m</i>	$\theta = 0$		$\theta = \pi/2$ rad	
	Theoretical	Experimental	Theoretical	Experimental
4	60%	60%	89%	—
8	2.6%	2.4%	48%	47
12	< 2%	—	6.5%	—

which reduces, when the generator is conjugately matched, to

$$\eta = \frac{g_0 - g_2 - G_0}{g_0 - g_2 + G_0} \dots\dots(30)$$

so that high efficiencies are only obtained when *G*₀ is very much less than *g*₀ - *g*₂, instead of *g*₀ + *g*₂ as for $\theta = 0$.

From (30) and (17) it can be seen that for a particular efficiency

$$(G_0)_{\pi/2} = \left\{ \frac{g_0 - g_2}{g_0 + g_2} \right\} (G_0)_0 \dots\dots(31)$$

so that for *s* = 0.1, for example,

$$(G_0)_{\pi/2} = 0.033(G_0)_0 \dots\dots(32)$$

However, for any particular switch, higher efficiencies may be obtained when $\theta = \pi/2$ rad than when $\theta = 0$. This is illustrated for *s* = 0.1 and *n* large in Table 1. The same form of experimental circuit was used as was discussed earlier in Section 2.

3. Harmonic Generator Controlled by Input and Output Signal

A disadvantage of the generators discussed in Section 2 is the very small mark/space ratio required of the switching signal in order to generate high-order harmonics. It is possible to relax this requirement considerably if the switching signal to be used is a combination of the input and output signals, a possible system being shown in Fig. 9. Some loss of switching power will occur in the auxiliary modulator, but since it has been assumed throughout this work that a negligible proportion of the switched power is required to operate the switch, this loss will in principle have little effect on the overall efficiency.

A harmonic generator of the type considered in this Section may not be self-starting; in practice, this would depend on the behaviour of the auxiliary modulator when only the input at ω_p is available.

From Fig. 9 the time-varying conductance of the switch may be written

$$g(t) = g_0 + \sum_{n=1}^{\infty} 2g_n \cos n((m-1)\omega_p t + \theta) \dots\dots(33)$$

If *m* is chosen so that

$$m + 1 \neq k_1(m - 1), \quad 2m \neq k_2(m - 1) \dots\dots(34)$$

where *k*₁ and *k*₂ are integers (which means that *m* must be greater than 3), the equations of current balance are

$$i_0 = v_0 g_0 + v_m g_1 \exp(-j\theta) \dots\dots(35)$$

$$i_m = v_0 g_1 \exp(j\theta) + v_m g_0 \dots\dots(36)$$

These equations are very similar to those previously reported for a double-tuned modulator,² and may be written

$$I = v_0(G_0 + g_0) + v_m g_1 \exp(-j\theta) \dots\dots(37)$$

$$0 = v_0 g_1 \exp(j\theta) + v_m(G_m + g_0) \dots\dots(38)$$

where the terminations are 'cold-tuned' as before.

For conjugate matching at input and output ports it is easily shown that *G*₀ = *G*_{*m*} and

$$G_0^2 = g_0^2 - g_1^2 \dots\dots(39)$$

The efficiency of harmonic generation is given by

$$\eta = \frac{4G_0 G_m g_1^2}{[(G_0 + g_0)(G_m + g_0) - g_1^2]^2} \dots\dots(40)$$

which reduces, under conditions of conjugate matching to

$$\eta = \frac{g_0 - G_0}{g_0 + G_0} \dots\dots(41)$$

so that, as before, when *G*₀ ≪ *g*₀, $\eta \simeq 1$. From

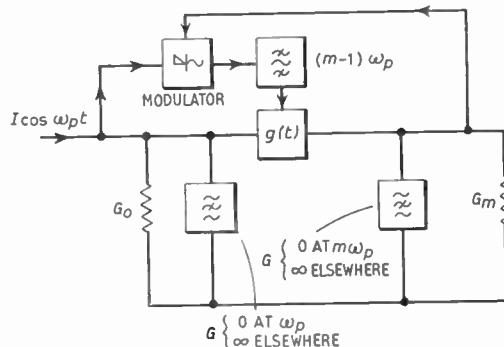


Fig. 9. Harmonic generator with switch controlled from output and input.

previous work, the efficiency can be deduced for varying mark/space ratios of the switching signal. From (39) and (41) it can be seen that the results will be independent of the order of harmonic to be generated. Figure 10 presents a graph of these results. It can be seen that high efficiency is attainable without exaggerated mark/space ratios, albeit at the expense of more complicated circuitry and possible lack of self-starting.

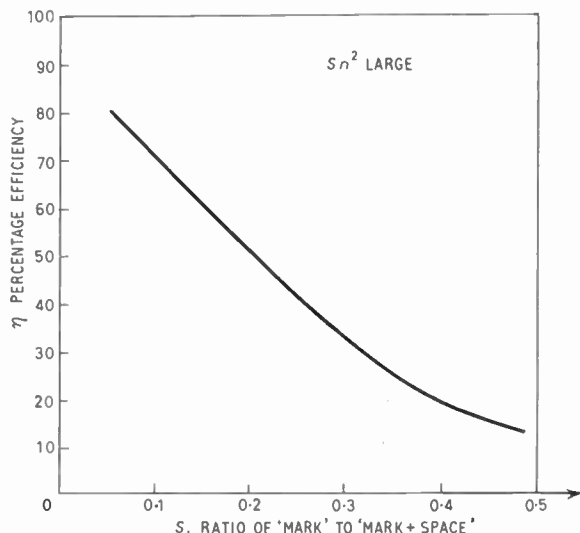


Fig. 10. Efficiency of generator controlled from input and output.

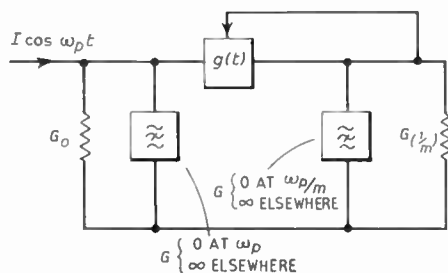


Fig. 11. Sub-harmonic generator.

4. Sub-Harmonic Generators

Because of their linearity,¹ harmonic generators of the types discussed in Sections 2 and 3 may be used as sub-harmonic generators of equal efficiency when input and output are reversed. Since in all these circuits the switching is controlled at least in part from the new output, the sub-harmonic generators are in general not self starting. Figure 11 shows a circuit derived from those discussed in Section 2. In this context it should be noted that the generators of Section 3 in which switching is controlled from input and output appear attractive, since they achieve good results with simply attainable mark/space ratios.

5. Conclusions

It has been shown that high-order harmonics may be efficiently generated by circuits containing one externally-operated switch. The efficiency has been

shown to be a function of the ratio of the 'closed' to the 'open' conductance of the switch, of the mark/space ratio of the switching signal, and of the order of harmonic to be generated, if the switch is controlled from the signal source alone. It has been shown that the terminating conductance has to be much smaller than the sum of the d.c. and second harmonic components of the Fourier series (for the conductance of the switch) in order to attain high efficiency when the switch is operated in phase with the input signal. Even higher efficiencies are attainable when the switch is operated in quadrature with the input signal, but the terminating conductances have been shown to be much smaller than the previous case.

Generators have been also considered in which the switch is controlled from both the input and output signals. Here it has been shown that efficiency is not a function of the order of harmonic generated. High efficiencies are attainable without exaggerated mark/space ratios, but the circuits are not self-starting.

Finally, it has been noted that all the types of generators described can be used as sub-harmonic generators, but none of them are self-starting.

6. Acknowledgments

The author wishes to thank Professor D. G. Tucker for his helpful discussions on this work, and for some initial calculations that he made to test the validity of the ideas. Also Mr. D. Kulesza of the same Department, for the experimental work described in this paper, and for the production of Figs. 5-8.

7. References

1. D. G. Tucker, "Highly-efficient generation of a specified harmonic or sub-harmonic by means of switches", *The Radio and Electronic Engineer*, 28, p. 25, 1964.
2. D. P. Howson, "Single-balanced rectifier modulators", *Proc. Instn Elect. Engrs*, 109C, p. 357, 1962, (I.E.E. Monograph No. 500 E, January 1962).
3. D. P. Howson and D. G. Tucker, "Rectifier modulators with frequency-selective terminations", *Proc. Instn Elect. Engrs*, 107B, p. 261, 1960 (I.E.E. Paper No. 3051E, January 1960).
4. D. G. Tucker, "Circuits with time-varying parameters", *J. Brit. J.R.E.*, 25, p. 263, 1963.
5. R. S. Engelbrecht, "Parametric energy conversion by non-linear admittances", *Proc. Inst. Radio Engrs*, 50, p. 312, 1962.

8. Appendix

Taking eqns. (4) and (5) together with (6) and (7), and setting

$$v_0 = a_0 + jb_0 \quad \dots\dots(42)$$

$$v_m = a_m + jb_m \quad \dots\dots(43)$$

we have, on equating real and imaginary parts,

$$I = a_0(G_0 + g_0 + g_2 \cos 2\theta) + b_0 g_2 \sin 2\theta + a_m(g_{m-1} \cos(m-1)\theta + g_{m+1} \cos(m+1)\theta) + b_m(g_{m-1} \sin(m-1)\theta + g_{m+1} \sin(m+1)\theta) \dots\dots(44)$$

$$0 = a_0 g_2 \sin 2\theta + b_0(G_0 + g_0 - g_2 \cos 2\theta) - a_m(g_{m-1} \sin(m-1)\theta - g_{m+1} \sin(m+1)\theta) + b_m(g_{m-1} \cos(m-1)\theta - g_{m+1} \cos(m+1)\theta) \dots\dots(45)$$

$$0 = a_0(g_{m-1} \cos(m-1)\theta + g_{m+1} \cos(m+1)\theta) - b_0(g_{m-1} \sin(m-1)\theta - g_{m+1} \sin(m+1)\theta) + a_m(G_m + g_0 + g_{2m} \cos 2m\theta) + b_m g_{2m} \sin 2m\theta \dots\dots(46)$$

$$0 = a_0(g_{m-1} \sin(m-1)\theta + g_{m+1} \sin(m+1)\theta) + b_0(g_{m-1} \cos(m-1)\theta - g_{m+1} \cos(m+1)\theta) + a_m g_{2m} \sin 2m\theta + b_m(G_m + g_0 - g_{2m} \cos 2m\theta) \dots\dots(47)$$

When $\theta = 0$, v_0 and v_m are real, since b_0 and b_m can be seen to be zero from inspection of (44)–(47). In the case for which $\theta = \pi/2$ rad, and m is a multiple of four, these equations reduce to

$$I = a_0(G_0 + g_0 - g_2) - b_m(g_{m-1} - g_{m+1}) \dots(48)$$

$$0 = b_0(G_0 + g_0 + g_2) + a_m(g_{m-1} + g_{m+1}) \dots(49)$$

$$0 = b_0(g_{m-1} + g_{m+1}) + a_m(G_m + g_0 + g_{2m}) \dots(50)$$

$$0 = -a_0(g_{m-1} - g_{m+1}) + b_m(G_m + g_0 - g_{2m}) \dots(51)$$

Hence, in general, $b_0 = a_m = 0$. Thus we may write $v_0 = a_0$, and $v_m = j b_m$, when eqns. (48) to (51) reduce to those given in the main text.

Manuscript first received by the Institution on 24th July 1964 and in final form on 23rd November 1964. (Paper No. 975.)

© The Institution of Electronic and Radio Engineers, 1965

STANDARD FREQUENCY TRANSMISSIONS

(Communication from the National Physical Laboratory)

Deviations, in parts in 10^{10} , from nominal frequency for **March 1965**

March 1965	GBR 16 kc/s 24-hour mean centred on 0300 U.T.	MSF 60 kc/s 1430–1530 U.T.	Droitwich 200 kc/s 1000–1100 U.T.	March 1965	GBR 16 kc/s 24-hour mean centred on 0300 U.T.	MSF 60 kc/s 1430–1530 U.T.	Droitwich 200 kc/s 1000–1100 U.T.
1	– 150.3	– 149.4	+ 1	17	– 150.5	– 148.0	– 8
2	– 150.7	– 151.2	+ 3	18	—	– 149.0	– 9
3	– 150.5	– 151.1	+ 4	19	– 150.7	– 150.0	– 8
4	– 149.3	– 151.3	+ 4	20	– 149.7	– 147.9	– 8
5	– 151.0	– 150.6	+ 3	21	– 148.8	– 150.8	– 8
6	– 151.2	– 150.6	+ 3	22	– 150.6	– 150.0	– 7
7	– 150.9	– 151.4	—	23	– 149.6	– 149.2	– 8
8	– 151.0	– 151.4	– 11	24	– 149.3	– 150.8	– 8
9	– 151.1	– 150.5	– 14	25	– 150.8	– 150.9	– 7
10	– 150.1	– 145.1	– 11	26	– 150.4	—	– 6
11	– 148.9	– 150.3	– 10	27	– 150.6	– 150.4	– 7
12	– 151.6	– 150.5	– 11	28	– 150.8	– 150.8	– 7
13	– 150.9	– 151.3	– 10	29	– 150.9	—	– 8
14	– 150.4	– 149.2	– 8	30	– 149.0	– 150.1	– 6
15	– 149.9	– 148.7	– 9	31	– 150.3	– 150.9	– 8
16	– 152.1	– 147.3	– 9				

Nominal frequency corresponds to a value of 9 192 631 770 c/s for the caesium $F, m(4,0) - F, m(3,0)$ transition at zero field.
Note: The phase of the GBR and MSF time signals was retarded by 100 milliseconds at 00 00 U.T. on 1st March 1965.

DISCUSSION

on

“Multi-channel Open-wire Carrier Telephone System”†

Mr. K. G. T. Bishop (*Communicated*):‡ I found Fig. 7 which shows the near-end crosstalk attenuation of a repeater section particularly interesting. The response shown is similar to that obtained from crosstalk fault location tests on coaxial trunk cables.

In this method of fault locating, an artificial fault of similar magnitude to the real fault is introduced at a known point, usually at one end of the repeater section. The near-end crosstalk attenuation is measured over the working frequency range and if the phase change introduced by the two faults is assumed to be equal, the crosstalk currents from the two faults will at certain frequencies be in phase. The troughs in Fig. 7 may be indicating this condition.

The phase difference between the currents from the two faults is given by:

$$2Bx = \frac{4\pi fx}{v}$$

$$\text{Since } B = \frac{2\pi f}{v}$$

where B = phase constant of line rads/second

x = distance between faults in metres

f = frequency, c/s

v = velocity of propagation metres/second

The phase change between adjacent peaks must equal 2π radians.

Therefore

$$2\pi = \frac{4\pi f_1 2x}{v} - \frac{4\pi f_2 2x}{v}$$

Hence

$$x = \frac{v}{2(f_1 - f_2)} \text{ metres}$$

Applying this formula to Fig. 7 it appears that two faults exist on the line 545 metres apart. It would be interesting to know if in fact two faults did exist on the line.

Messrs. Munro, Anderson and Mackenzie (*in reply*): We have not previously had experience of using crosstalk methods for the location of faults on coaxial cables. However, the last formula Mr. Bishop states is that normally used in variable frequency reflection testing to give the distance to a fault condition on open wire lines. The method is not one using crosstalk but consists of applying

directly to the faulty line a variable frequency oscillator, preferably of similar internal impedance to the line and measuring the voltage across the oscillator terminals. In a well-terminated line in good order the voltage reading varies regularly as the impedance of the line. In conditions where a fault exists an oscillatory pattern develops with varying frequency depending on the relative phase and amplitude of the reflected voltage. The formula stated by Mr. Bishop then gives the location of the fault once the frequency separation between successive minima or maxima is known.

The graph shown on Fig. 7 is typical of the near end crosstalk type unbalance where there are in effect a considerable number of like sections. In this case the aim was to gain wide bandwidth without absorptions and a minimum crosstalk attenuation of 65 dB at 552 kc/s so the type of unbalance was not significant when viewed with the line attenuation of the sections concerned and this is shown on Fig. 8.

In fact for location of faults on open wires it is not now customary to use the variable frequency method since the pulse-echo-fault-locator presents the information accurately and rapidly as shown on Fig. 5.

Mr. Bishop: When I wrote of the possibility of faults existing on the line I was using the term relatively. The target of 65 dB crosstalk attenuation is, of course, quite adequate for a repeater section and has been achieved over the whole frequency range as Fig. 7 shows.

Crosstalk faults do occur between pairs in coaxial cables usually because of a break in the outer conductor of one of the pairs. This is often due to an outer conductor ferrule used for connecting two lengths of cable becoming unsoldered and the crosstalk may be serious enough to render some channels unworkable. The indirect cause may be subsidence of the sub-soil placing the cable under strain and it is the practice of good cable maintenance to obtain a location whilst the fault is in its incipient stage and before a complete breakdown occurs.

For this purpose the crosstalk/frequency test has been devised and the art in applying it lies in choosing a value of artificial fault that has the same effect on crosstalk as the real fault. In practice a few inches of twisted jumper wire connected between centre conductors ensures sufficient coupling.

The pulse-echo test set described in reference 3 of the authors' paper may be modified for locating high crosstalk between pairs. If the link between the hybrid centre and the attenuator input is removed and the second coaxial pair is connected to the attenuation input, the fault point will show as a spike on an otherwise linear display.

† A. Munro, G. H. Mackenzie and C. W. M. Anderson, *The Radio and Electronic Engineer*, 28, No. 2, pp. 75-86, August 1964.

‡ Contribution received by the Institution on 18th September 1964.

Hybrid Positive and Negative Parameter Delay-line Synthesis

By

C. I. JONES, Ph.D. †

AND

Professor

E. M. WILLIAMS, Ph.D. ‡

Summary: A network comprising conventional inductors and capacitors in combination with negative inductors and capacitors permits a synthesis method for delay networks which cannot be achieved with positive parameters alone. The negative inductors and capacitors in such hybrid networks are provided by conventional circuit elements in conjunction with active devices. An example is given; in comparison with a simple constant- k network this design has both an increased delay per section for a comparable frequency range and a lower insertion loss as a result of the use of negative resistance for loss compensation.

1. Introduction

Methods for the synthesis of two-terminal circuits which have the properties of negative capacitance or negative inductance have been known for many years. As described by Verman and others¹⁻⁵, for instance, these methods employ active circuit elements in conjunction with suitable passive feedback networks or a negative-resistance element in conjunction⁶ with appropriate positive circuit parameters. Relatively little attention has been given to possible improvements in the properties of the networks which can be synthesized when negative parameter elements are available to the designer in addition to the classical positive inductances and capacitances. One would suspect that application of such elements would provide for the evasion of some of the conventional synthesis restraints which are associated with the availability of positive-parameter elements only. D. J. Storey and W. J. Cullyer, for instance, have shown⁷ that this is, in fact, the case in the synthesis of low-pass audio filters.

This paper describes the results of an investigation of the advantages of the use of a combination of negative and positive parameters, or a hybrid-parameter network, instead of positive parameters only in the design of a delay line, or cascade of delay networks.

2. Review of the Constraints upon Hybrid Positive and Negative Parameter Networks

Negative inductance and capacitance parameters are defined as the properties of two-terminal circuit elements for which voltage and current relations are as indicated in Fig. 1. In the frequency domain, for instance, a negative inductance yields the voltage-current time-phase relationship characteristic of a capacitor (i.e. current leading voltage) but has a reactance which increases as frequency increases, as

† Consulting Engineer, 335 Locust Street, Pittsburgh, Pa., U.S.A.

‡ Department of Electrical Engineering, Carnegie Institute of Technology, Pittsburgh, Pa., U.S.A.

does that of an ordinary inductor. In the time domain, the direction of induced e.m.f., as a result of changing current, is opposite to that of an ordinary inductor. Figure 1 also shows elementary impedance-converter circuits which would simulate a negative inductor and negative capacitor, respectively. One can show from energy considerations that: (1) neither 'negative' element can retain its 'negative' parameter property as the frequency of an applied e.m.f. approaches zero, although for practical purposes, the lower frequency limit might be made very small; (2) the low-frequency limit on realization is inherent in the nature of the negative parameter regardless of the scheme employed for its synthesis; and (3) a two-terminal negative-parameter property cannot be synthesized without the use of at least one active device and energy source within the two-terminal element.

Hybrid networks, comprising combinations of negative and positive parameters, have the following restrictions upon the impedance functions which can be realized: (1) as with conventional networks, the impedance function must be rational and real for real values of s ; and (2) although poles can be synthesized on the right half-plane, this would result in instability. No such limitation as that imposed by Foster's reactance theorem on conventional reactive two-terminal networks, i.e. the necessary alternation of poles and zeros, is encountered with hybrid networks. As will be seen in the synthesis to be described, this latter freedom is of considerable assistance in the development of a simple rationale for the solution of the delay-line problem.

3. Delay-Network Synthesis

In a general sense, the desirable property for most delay lines is a combination, in the frequency domain, of a constant delay and reasonably constant image-impedance over a frequency band ranging from approximately d.c. to a cut-off frequency, with at the same time as small an insertion loss as possible.

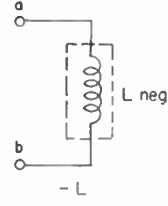
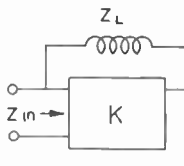
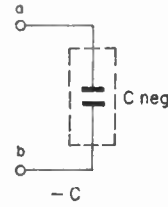
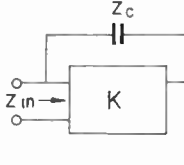
ELEMENT	f DOMAIN	t DOMAIN	TYPICAL CIRCUIT
	$X_{ab} = -2\pi f L \text{ neg} $	$e_{ab} = - L \text{ neg} \frac{di_{ab}}{dt}$	 $Z_{in} = -\frac{Z_L}{K-1}$
	$X_{ab} = \frac{1}{2\pi f C \text{ neg} }$	$i_{ab} = - C \text{ neg} \frac{de_{ab}}{dt}$	 $Z_{in} = -\frac{Z_C}{K-1}$

Fig. 1. Properties of negative inductances and capacitances in the time and frequency domains and simple feedback-amplifier circuit arrangements for synthesis.

Although figures of merit are not ordinarily used, it is clear that the merit of any particular design is (1) in proportion to the product of time delay and the useful frequency band, and (2) also dependent upon an insertion loss (affected both by the real component of the image-transfer constant and the mismatch between image-impedance and a constant source impedance) which is small and preferably relatively constant with frequency.

The simplest of delay networks is a low-pass cascade of sections of T or π constant- k sections. The limitations of such simple cascades are well known; no opportunity is offered in the constant- k section for shaping the phase characteristics in order to provide constant delay over the pass-band. As a result only a small portion of this band is useful. To evade this difficulty, many modifications have been proposed, such as an m -derived section, a type B phase-compensating section, a generalized bridged-T section, or a capacitance-shunted section. Numerous synthesis procedures have been employed, including synthesis directly in the time domain. Generally, dissipative effects have been neglected in the design calculations and the performance-degrading effects of actual dissipation have been determined experimentally; dissipative effects are minimized by selection of components, within size and economic limitations.

The contribution of possible hybrid-parameter networks to the problem of the design of a delay

section can be developed as follows. In the T section of Fig. 2 we will assume that Z_1 is an inductor and that Z_2 is to be so synthesized from reactive elements as to provide a transmission time for the section which is independent of frequency (i.e. a phase-shift proportional to frequency) over a frequency range $\omega = 0$ to $\omega = \omega_0$.

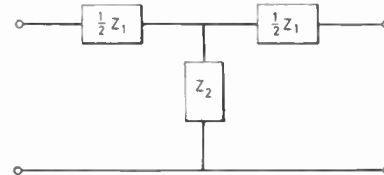


Fig. 2. Conventional T-section with the usual designations $Z_1/2$ and Z_2 for the series and shunt elements.

For this T-network, it is well known that the image transfer constant γ (with a real component α and imaginary component β) is given by

$$\cosh \gamma = 1 + \frac{Z_1}{2Z_2} \quad \dots\dots(1)$$

or

$$\cosh \alpha + j\beta = 1 + \frac{R_1 + jX_1}{2(R_2 + jX_2)} \quad \dots\dots(2)$$

in which $Z_1 = R_1 + jX_1$, and $Z_2 = R_2 + jX_2$. When real and imaginary components are separated:

$$\cosh^2 (\frac{1}{2}\alpha) \sin^2 (\frac{1}{2}\beta) = \frac{1}{2} \left[\sqrt{\left[\frac{R_1 R_2 + X_1 X_2}{4(R_2^2 + X_2^2)} \right]^2 + \left[\frac{X_1 R_2 - X_2 R_1}{4(R_2^2 + X_2^2)} \right]^2} - \frac{R_1 R_2 + X_1 X_2}{4(R_2^2 + X_2^2)} \right] \quad \dots\dots(3)$$

$$\sinh^2 (\frac{1}{2}\alpha) \cos^2 (\frac{1}{2}\beta) = \frac{1}{2} \left[\sqrt{\left[\frac{R_1 R_2 + X_1 X_2}{4(R_2^2 + X_2^2)} \right]^2 + \left[\frac{X_1 R_2 - X_2 R_1}{4(R_2^2 + X_2^2)} \right]^2} + \frac{R_1 R_2 + X_1 X_2}{4(R_2^2 + X_2^2)} \right] \quad \dots\dots(4)$$

When (3) and (4) are solved simultaneously for α and β and the conditions applied that $\alpha = 0$ and $\beta = \pi (\omega/\omega_0)$, for linear phase, one obtains:

$$X_2 = \frac{-X_1}{2\left(1 - \cos \frac{\pi\omega}{\omega_0}\right)} \quad \dots\dots(5)$$

$$R_2 = \frac{-R_1}{2\left(1 - \cos \frac{\pi\omega}{\omega_0}\right)} \quad \dots\dots(6)$$

Neglecting dissipation for the moment, eqn. (1) would be satisfied by a circuit element, the reactance of which is

$$X_2 = \frac{-sL_1}{2\left(1 - \cosh \frac{\pi\omega}{\omega_0}\right)} \quad \dots\dots(7)$$

expressed as a function in the s plane (L_0 , an inductor, provides the reactance X_1). This function will have a pole at every point at which the real part of $\cosh (\pi s/\omega_0) = 1$ and the imaginary part of $\cosh (\pi s/\omega_0) = 0$, simultaneously or, writing $s = \sigma + j\omega$, a pole for

$$\cosh \frac{\pi\sigma}{\omega_0} \cos \frac{\pi\omega}{\omega_0} = 1 \quad \dots\dots(8)$$

and

$$\sinh \frac{\pi\sigma}{\omega_0} \sin \frac{\pi\omega}{\omega_0} = 0 \quad \dots\dots(9)$$

These equations are satisfied simultaneously for $\sigma = 0$, $\omega = 2n\omega_0$ (n is an integer). The function (7) would have zeros when

$$\cosh \frac{\pi s}{\omega_0} = \infty \quad \dots\dots(10)$$

an equation which is satisfied for no possible value of ω if σ is finite. Consequently, the poles and zeros of the function (7) would not alternate and this function could not possibly be realized in a positive-parameter synthesis. This particular fundamental difficulty does not prohibit realization of the constant-delay in a hybrid positive-negative parameter synthesis, which can provide any number of poles in sequence without intermediate zeros. However, although the poles can be selected, the entire reactance function, with suitable values at every frequency between singularities, cannot be precisely realized. The desired reactance behaviour, given in equation (7), is not a rational function, a requirement which must be met by all physically-realizable networks, whether or not negative parameters are used. The desired behaviour can be met approximately, nevertheless, with any degree of accuracy, depending upon the circuit complexity. One general approximation procedure for this purpose is based on the use of the series expansion for $\cosh \pi\omega/\omega_0$ in (7) which yields

$$X_2 = \frac{sL_1}{2\left[\left(\frac{\pi}{\omega_0}\right)^2 \frac{s^2}{2!} + \left(\frac{\pi}{\omega_0}\right)^4 \frac{s^4}{4!} + \dots\right]} \quad \dots\dots(11)$$

If the series in the denominator of (11) is terminated after any arbitrary number of terms, the reactance is realizable. However, a more satisfactory approximation, since it results in fewer components, is based on the collocation method which follows.

The desired reactance X_2 can be regarded as being provided by a frequency-dependent capacitor, of capacitance C_f . If we specify a value C_0 at a cut-off frequency ω_0 , for the T-section of Fig. 1, we have, with a fixed inductance L_1 for the component in Z_1

$$\omega_0 = \frac{2}{\sqrt{L_1 C_0}} \quad \dots\dots(12)$$

$$C_0 = \frac{4}{\omega_0^2 L_1}$$

for the capacitance at the cut-off frequency.

If the phase shift is to be linear as a function of frequency up to the cut-off frequency, the reactance X_2 , as $\omega \rightarrow 0$ approaches that of a capacitance

$$C_f(0) = \frac{\pi^2}{4} C_0 \quad \dots\dots(13)$$

The actual values required for this frequency-dependent capacitance in order to provide linear phase shift over the whole range from $\omega = 0$ to $\omega = \omega_0$ are shown in Fig. 3. The design problem can be reduced to one

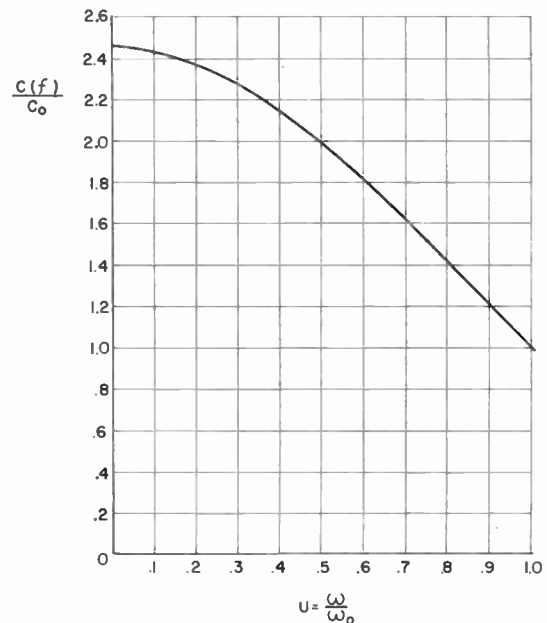


Fig. 3. Characteristic of a frequency-dependent capacitor C_f which would satisfy eqn. (7) and provide constant phase-delay over the pass band.

of synthesizing a reactance which matches that of this required capacitance at as many points as possible. The simplest case, that of the constant- k section, matches this at one point, necessarily that at $\omega = 0$. The next two approximations, in order of complexity, make use of the hybrid positive-negative parameter networks of Fig. 4, which match the required characteristic at two or three points, respectively. The first has poles at zero and infinity and the second has an intermediate pole; neither has zeros. The intermediate pole in the second network, incidentally, lies between the pole given by the series approximation (11) and

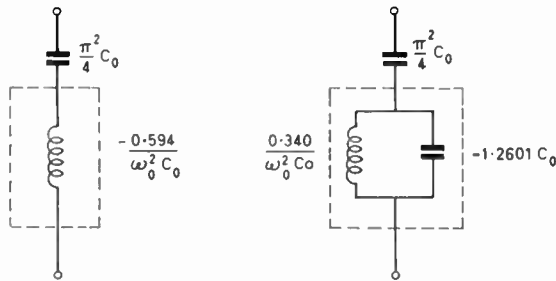


Fig. 4. Hybrid networks which provide second and third order approximations for the desired equivalent capacitance of Fig. 3. The upper network matches the curve of Fig. 3 at $\omega = 0$ and $\omega = \omega_0$. The point of maximum deviation from the desired value is at $\omega = 0.6\omega_0$. The lower network matches the curve of Fig. 3 at $\omega = 0$, $\omega = 0.6\omega_0$ and $\omega = \omega_0$

the second pole of the exact expression (7). The resulting phase characteristics and delay times are shown in Fig. 5. It is immediately apparent that the use of either hybrid network results in a delay section with 57% greater delay than the constant- k network even though the cut-off frequency is the same. This is the case because once one chooses the cut-off frequency in a constant- k section, the delay is immediately determined; the usable value of delay, of course, is dependent upon the initial slope of the phase shift as a function of frequency in the vicinity of $\omega = 0$. In the constant- k section, the slope of the phase characteristic in the vicinity of cut-off is much steeper than that in the hybrid network but this large delay per section is not useful for ordinary delay-line purposes because of the high dispersion in this frequency region. The hybrid networks, indeed, have less delay in this dispersive region but higher delays in the vicinity of $\omega = 0$.

4. Design Example

A delay line based on the simpler of the hybrid networks of Fig. 5 was constructed. The required negative inductance is provided by a transistor negative-impedance converter.^{3, 4} In the initial design, the network was assumed to be entirely

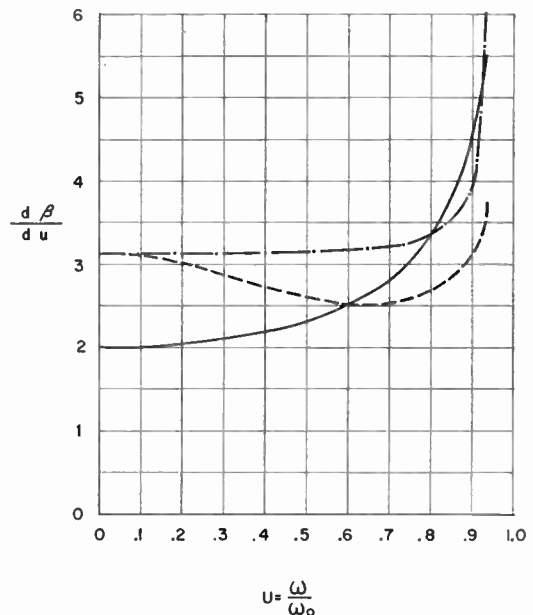


Fig. 5. Transmission time as a function of frequency in a T-section with a constant series inductance and a shunt arm comprising (a) (solid line) constant capacitance, as in a constant- k section (b) (lower broken line) the second-order approximation of Fig. 4, and (c) (upper broken line) the third-order approximation of Fig. 4.

reactive. Of course, a certain amount of dissipation is unavoidable; however, since active elements are incorporated in each section some compensation for losses can actually be realized. The exact expression for the necessary relationship between the resistances of Z_1 and Z_2 of Fig. 1 was given by (6). At low frequencies, R_1 is the d.c. resistance of the inductor, Z_1 , and may be made arbitrarily small. At the higher frequencies R_1 tends to rise, while R_2 tends to approach $-2 R_1$. The required negative resistance,

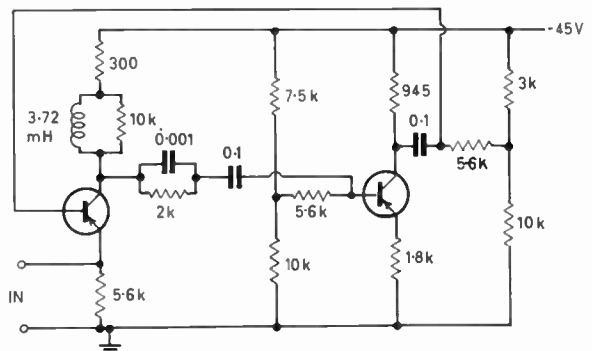


Fig. 6. Negative-parameter, second-order approximation, circuit element for use in a delay line. This uses a compensated transistor negative-impedance converter. The significant parameters in this converter are the 3.72 mH inductor and its shunt and series resistors. The impedance of these is effectively divided by 4, owing to circuit gain, at the input terminals.

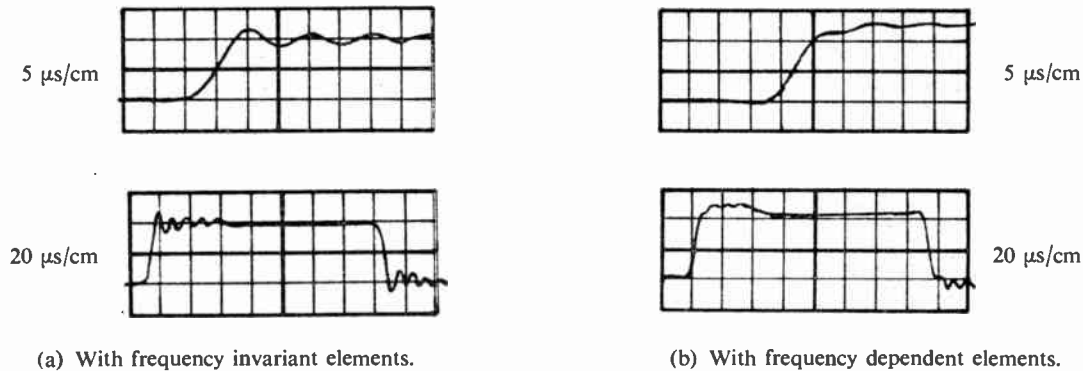


Fig. 7. Comparison of the experimental performance of a constant- k delay line (traces (a)) and a delay line with the shunt frequency-dependent element (traces (b)) shown in Fig. 6. In both cases six sections were employed. A rectangular pulse was applied to the delay-line input at the start of each time base. The delay-line output is displayed with a time scale of 5 ms per division in the upper pair of traces and 20 ms per division in the lower pair of traces.

with an approximate reproduction of the curve defined by (6) for $\omega < \omega_0/2$ may be obtained by using an inductor, in the negative inductance circuit, which has effective resistance characteristics similar to that of the inductor Z_1 . Constants of the entire negative-parameter circuit are shown in Fig. 6 and the experimentally determined pulse characteristics of a delay line comprising six such sections are shown in Fig. 7. Much of the complexity of the negative-impedance converter was due to a supplementary problem, that of achieving stability; owing to the phase-lag at higher frequencies the uncompensated circuit tended to oscillate at a frequency outside the pass band. Use of tunnel diodes in such a synthesis procedure could result in considerable circuit simplification.

In the synthesis example cited, no attention has been given to the characteristic impedance of the resulting delay line. This impedance cannot be specified and designed independently, nor can the inevitable plunge of characteristic impedance at cut-off frequency to zero, for the T-section, or a rise to infinity for a π -section, be avoided. However, the impedance for the case given is inherently 'flatter' than that for the constant- k section. This is the case because the image impedance of a constant- k section is given by

$$Z_0 = \sqrt{\frac{L}{C} \left(1 - \frac{\omega^2}{\omega_0^2} \right)}$$

and for the hybrid unit by

$$Z_0 = \sqrt{\frac{L}{C_f} \left(1 - \frac{\omega^2}{\omega_0^2} \right)}$$

In the latter case C_f is a variable; C_f decreases as ω increases and results in a more constant Z_0 over the useful portion of the frequency range than the Z_0 for the constant- k section. Thus, the use of hybrid parameters in this instance yields three advantages over a cascade of constant- k sections: (1) a greater

delay per section, (2) a greater product of time-delay and useful bandwidth, and (3) a reduced insertion loss, both because of some loss compensation by negative resistance and because of less impedance mismatch with a resistive load. To offset these advantages, of course, there is the much greater complexity of the hybrid system and (of possible importance in some instances) a limited dynamic range.

5. Acknowledgment

This is based on part of a dissertation submitted by C. I. Jones in partial fulfilment of the requirements for the degree of Doctor of Philosophy at Carnegie Institute of Technology. The work was supported in part by the Office of Naval Research under Contract Nonr 760 (09).

6. References

1. L. C. Verman, "Negative circuit constants", *Proc. Inst. Radio Engrs*, **19**, pp. 676-81, April 1931.
2. E. W. Herold, "Negative resistance and devices for obtaining it", *Proc. I.R.E.*, **23**, pp. 1201-23, October 1935.
3. J. L. Merrill, "Theory of the negative-impedance converter", *Bell Syst. Tech. J.*, **30**, No. 1, pp. 88-109, January 1951.
4. J. G. Linvill, "Transistor negative-impedance converters", *Proc. I.R.E.*, **41**, pp. 725-9, June 1953.
5. A. C. Bartlett, "Boucherot's constant-current networks and their relation to electric wave filters", *J. Instn Elect. Engrs*, **65**, pp. 373-6, March 1927.
6. B. van der Pol, "A new transformation in alternating current theory and an application to the theory of audition", *Proc. I.R.E.*, **18**, pp. 221-30, February 1930.
7. D. J. Storey and W. J. Cullyer, "Network synthesis using negative-impedance converters", *Proc. Instn Elect. Engrs*, **111**, pp. 891-906, May 1964. (I.E.E. Paper No. 4454E)

Manuscript first received by the Institution on 7th February 1964 and in final form on 17th July 1964. (Paper No. 976.)

© The Institution of Electronic and Radio Engineers, 1965

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.

DISTORTION OF BINARY SIGNALS IN WIDE-BAND SYSTEMS

In data transmission and more particularly in tele-control systems, binary signals with a d.c. content are frequently transmitted on lines without d.c. paths, the transmission path being blocked by inserted transformers or RC-amplifiers. Often, existing lines (telegraphy and telephony cables) are readily available, and the problem is to develop a transmission system with a minimum outlay for equipment at the transmitting as well as the receiving end, optimum channel utilization being of second order importance. A paper by a German engineer outlines the conditions under which a d.c. transmission is possible and the information rate and protection against transmission errors which can be achieved. Since overshoot has adverse effects, the problem of finding a transmission function with an optimum reduction of such effects, is also investigated.

"Information rate and reliability of transmission of binary signals with a d.c. content when subjected to distortion in wide-band systems", J. Engel. *Nachrichtentechnische Zeitschrift*, 17, No. 6, pp. 301-11, June 1964.

DOUBLE F.M. TRANSMISSION

The double-frequency modulation method is increasingly being applied to the transmission of music channels and television sound signals over point-to-point radio links and satellite links. In this technique the signals to be transmitted are successively passed through two frequency modulators and the low frequency signal is extracted at the receiving end by two-stage frequency demodulation.

The advantages and disadvantages of this modulation method are discussed in a German paper and a practical example is presented. The interference occurring at the receiver output when a sinusoidal interference is superimposed on the r.f.-signal is investigated for the case in which the low-frequency band carries no information. It is found that such an interfering signal produces a similar sinusoidal interference tone in the low-frequency band if the interfering signal lies in the vicinity of a spectrum line of the r.f.-signal with simple frequency modulation when the low frequency band carries no information. The frequency of the interference note is equal to the frequency difference between the interfering signal and the adjacent spectrum line. Its amplitude with respect to the useful low-frequency signal is calculated.

"Sinewave interference in double-FM(FM-FM)", E. Metzger. *Nachrichtentechnische Zeitschrift*, 17, No. 12, pp. 615-20, December 1964.

TUNNEL DIODE MIXERS

The static characteristic of a tunnel diode can be represented by a number of currents with different physical origins. The authors of this German paper have derived analytical representations for the voltage dependence of these currents. By comparison of these approximations with measured characteristics a useful analytical representation for the static characteristic and the differential conductance of a tunnel diode is obtained.

The approximation for the static tunnel diode characteristic is used to calculate the mixer conductances, the power gain and the noise behaviour of a tunnel diode mixer stage. For a simplified mixer equivalent circuit the calculated mixer conductances are graphically represented for different oscillator amplitudes as a function of the diode operating point. The equivalent noise currents of the tunnelling current in a mixer stage are calculated and represented in the same way.

The noise figure of a tunnel diode mixer stage is given. For a mixer stage with a tunnel diode ZJ 56 A the noise figure for different operating points, oscillator voltages and source conductances is numerically evaluated from the analytical representation of the static characteristic of the diode.

"Calculation of the mixer conductance and noise of a tunnel diode mixer using the static diode characteristic", H. Melchior and M. J. O. Strutt, *Archiv der Elektrischen Übertragung*, 18, No. 8, pp. 455-64, August 1964.

LOW NOISE MAGNETIC FILM STORE

The write noise voltage appearing in the read-out channel of a word-organized magnetic film memory limits the shortest cycle time of the memory. A circuit is described in a recent German paper which substantially protects the read-out amplifier from the write noise voltage and which itself has a recovery time of 20 ns. It contains a balanced arrangement of tunnel diodes which limits the write noise voltage in a non-linear manner and amplifies the read-out signal as active elements. This makes possible the use of a common bit sense line.

Measurements show a bit noise rejection ratio of more than 35 dB even with 10% unsymmetry of the write noise voltage. When operated in connection with a memory model up to 20 Mc/s repetition frequency is obtained for write/read and up to 74 Mc/s for non-destructive read-out alone.

"A hybrid circuit for nanosecond pulses to reduce the write noise in a magnetic film store", D. Seitzer, *Archiv der Elektrischen Übertragung*, 18, No. 10, pp. 577-84, October 1964.