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 27 MAY 1964 NUMBER 5

A SOCIETY FOR ELECTRONIC AND RADIO TECHNICIANS

A PUBLICATION of considerable interest to the professional electronic and radio engineer has just been published in Great Britain by the Radio Trades Examination Board. Its main purpose is to recommend the formation of a 'Society of Electronic and Radio Technicians'.

The Institution has for long striven to obtain better educational facilities for the electronic and radio technician. Indeed, in 1930 the Institution itself first held an examination for 'Radio Service Work', and it was this interest and experience which finally gained the support of the radio manufacturers associations and subsequently led to the formation of the Radio Trades Examination Board.

Over the past six years the Board has had discussions with various bodies interested in the desirability of helping to found a Society for Electronic and Radio Technicians. It is envisaged that membership will be graded according to academic achievement *and* responsibility of experience, but only those who possess either the R.T.E.B. Final Certificate, or who have passed examinations of a similar or higher standard, e.g. the Higher National Certificate with radio and/or electronic endorsements will be eligible for membership. Whilst the Board is sponsoring meetings of prospective members to discuss the project, it is recommended that the affairs of such a Society should be run by its members.

The usefulness and often the efficiency of engineering equipment ultimately depends on the ability of the technician. Industrial development in all fields is creating an increasing demand for men trained in *practical* application. These points were well brought out in an editorial published in the Institution's *Journal* in June 1949, and in the Crowther Report* which contained the following definition:

> "The Technician is one who is qualified by specialist technical education and practical training to apply in a responsible manner proven techniques which are commonly understood by those who are expert in a branch of engineering, or new techniques prescribed by a professional technologist. His work involves the supervision of skilled craftsmen, and his education and training must be such that he can understand the reasons for and the purpose of the operations for which he is responsible".

The Institution's policy was again well ventilated by Mr. J. Langham Thompson in his Presidential Address† in which he also advocated that Technicians' Associations in the various disciplines might well collaborate in the same way as the Chartered Engineering Institutions in Great Britain are now proposing to work together under the aegis of the Engineering Institutions Joint Council.

The President's remarks were reflected in the views expressed in the report of the Institution's Education and Training Committee 'The Training of Radio and Electronic Technicians',‡ which emphasized the role of the technician and distinguished his work from that of the fitter or mechanic. The report also criticized the proposal that technicians should be described as "Technician Engineers". In the view of the Council of the Institution, the efforts of the Chartered Engineering Institutions Joint Council to seek fuller recognition of the term 'engineer' will not be helped by inventing a contradiction in terms such as 'Technican Engineer'.

The proposal of the Radio Trades Examination Board to use the title 'The Society of Electronic and Radio Technicians' deserves widespread support for its clarity of description and of purpose.

G.D.C.

* Report of Central Advisory Council for Education, H.M.S.O. 1959.

† *The Radio and Electronic Engineer*, 27, No. 1, p. 7, January 1964

‡ *Proc. I.E.R.E.*, **2**, No. 2, pp. 31–4, March–April 1964.

INSTITUTION NOTICES

Abstracts of Papers published in the Journal of the Brit.I.R.E. 1952 to 1963 inclusive

The seventh edition of a publication, giving abstracts of nearly 900 papers, articles and reports arranged according to the Universal Decimal Classification, has been published by the Institution. The abstracts have been compiled in sufficient detail to determine whether reference to the complete paper will give the searcher the information he requires. Institution reports on technical and educational matters are included under appropriate headings as well as editorial articles discussing matters of general and scientific interest. A subject index with cross-references is included to assist the engineer who is not familiar with the U.D.C. system. An index to authors is also given.

Many of the papers and reports listed are still available and an indication of their availability is given. Copies of "Abstracts" may be obtained from the Institution, price 10s. 6d. (7s. 6d. to members of the Institution) including postage.

Change of Address—an Urgent Request

All members are earnestly requested not to delay in advising the Institution at 9 Bedford Square, London, W.C.1, immediately they change their address.

It is not necessary for members to notify local section secretaries in Great Britain of their change of address. Such notifications are automatically sent from Head Office to the local sections. Because of delays in post, however, overseas members should advise their Local Secretary at the same time as they advise London.

Failure to notify the Secretary at 9 Bedford Square causes considerable delay in members receiving Journals, notices of meetings and other separate communications. In addition there are extra postal charges involved due to letters being returned to the Institution.

Group Provident Scheme

The British United Provident Association (B.U.P.A.) is a service which operates in the United Kingdom for subscribers who wish to insure themselves for private hospital treatment under the National Health Service. The Institution has established a Group Scheme for members with the co-operation of the B.U.P.A. and details of annual subscription rates, medical benefit, etc., may be obtained from the Group Secretary, B.U.P.A., 9 Bedford Square, London, W.C.1. Participation in the Scheme is also open to members of European nationality who are resident in Commonwealth countries and those interested should also write to the Group Secretary.

Cancellation of Lectures by Visiting Russian Engineer

Owing to unforeseen circumstances, Professor Mark Aizerman of the Institute of Automatics and Telematics, Moscow, is unable to visit London in May and June. The lectures which he was to have delivered at the Institutions of Mechanical and Electrical Engineers on the 28th May and 2nd June are therefore cancelled. Particulars of these lectures were given in the April issue of *The Radio and Electronic Engineer*. It is hoped that Professor Aizerman will be able to visit this country later in the year, or early next year.

Proceedings of the Symposium on "Sonar Systems"

The papers presented at the Symposium on "Sonar Systems" in Birmingham from 9th to 11th July 1962 have now been published in collected form with reports of the associated discussions. The volume may be purchased from the Publications Department of the I.E.R.E., 8–9 Bedford Square, London, W.C.1, at a charge of £3 including postage. A list of papers which were presented at this Symposium was published in the Brit.I.R.E. *Journal*, for June 1962 (page 507).

"Guidance for Authors"

Members and others who are contemplating writing papers for *The Radio and Electronic Engineer* are reminded that the Institution has published a leaflet with the above title which may be obtained on application to the Secretary of the Programme and Papers Committee. The leaflet contains information on the requirements of style and presentation of papers, and advice on the preparation of illustrations.

It is helpful if an intending author first submits a synopsis of the paper to the Committee as this enables the contents of the paper to be discussed in advance of carrying out detailed preparation.

Authors are asked to submit at least two copies of their papers (with prints of the illustrations) in order that the process of consideration by the Papers Committee's referees (normally three in number) can be completed with the least possible delay.

I.E.C. Catalogue of Publications

A new catalogue listing all available I.E.C. publications up to 31st Decembei 1963 has been published by the International Electrotechnical Commission. Synopses are given of all publications including those which were under revision at that time. Copies of the catalogue, and all I.E.C. publications, may be obtained from British Standards Institution, Sales Department, 2 Park Streeet, London, W.1.

Joint Symposium on
"SIGNAL PROCESSING IN RADAR AND SONAR DIRECTIONAL SYSTEMS"

with special reference to systems common to Radar, Sonar, Radio Astronomy, Ultrasonics and Seismology

UNIVERSITY OF BIRMINGHAM, 6th–9th JULY 1964

Organized by the Institution and The Department of Electronic and Electrical Engineering, University of Birmingham

Synopses of some of the Papers to be presented at the Symposium

Statistical Optimization of Antenna Processing Systems

G. O. YOUNG. (*Hughes Aircraft Company.*)

This paper treats the antenna as an information processing device. By representing the antenna as a spatial (as well as temporal) frequency filter, information theory concepts can be applied so as to maximize the information content or rate at the system output or to minimize the information content in the system error. Rather than attempt to satisfy classical antenna design criteria such as maximization of gain, minimizing of sidelobe level, etc., the approach used here is to optimize the system by maximizing the useful information rate of the receiver subject to the physical constraints of the system.

A convenient way of evaluating and comparing different signal processing systems is to determine their respective output information rates. Various non-linear as well as linear processing systems can be compared on this basis.

For linear, one-way antennas, the optimum antenna in most cases is one which has uniform shading. For linear, two-way antennas, the optimum shading is non-uniform. The theory is illustrated with a non-reciprocal two-way array whose parameters are adjusted so as to yield the optimum antenna system according to the above criteria.

A Post-amplification Radar Receiving Array with I.F. Multiple-beam-forming Matrix

S. PICHAFROY, J. SALOMON AND J. HURBIN. (*Compagnie Française Thomson-Houston.*)

At the present time array radars are being subjected to a whole series of studies which should culminate in the optimization of data processing and obtaining a performance which is conventional radar processes.

The type of array radar outlined in this paper utilizes individual receiving aerials followed by amplifiers and an intermediary frequency phasing matrix. This matrix simultaneously issues data corresponding to orthogonal aerial lobes covering a wide angular sector.

The general principles involved will be outlined as well as supporting evidence and current developments in this connection.

Multiplicative Processing Antennas for Radar Application

A. KSIENSKI. (*Hughes Aircraft Company.*)

Several non-linear antenna systems were investigated in detail and their responses compared. The response of these antennas was computed for two targets with the following parameter variations: (1) target angular separation varying from zero to one null-to-null beamwidth; (2) target correlation varying between zero and unity; (3) target relative phase varying between zero and 360°. The resulting data are presented in the form of resolution curves and pointing errors, where the pointing error is given by the angular deviation of the peaks of the antenna response from the actual target locations. The results indicate that non-linear processing improves resolution, defined as the ability to separate two closely-spaced targets, for all levels of correlation between the target returns. The amount of improvement beyond that of a linear array varies somewhat with the particular array configuration and, in general, depends on the sacrifice in gain, or signal/noise ratio.

The theoretical studies were accompanied by an experimental program which both confirmed the theoretical results and demonstrated the practical feasibility of non-linear processing antennas for radar applications.

A Note on Multiplicative Receiving Systems and Radar

R. BLOMMENDAAL. (*Nederlandsch Radar Proefstation.*)

A multiplicative system is examined which consists of a 21-element array with a tapered amplitude distribution and a monopole. Radar applications are considered in the following respects:

(1) The influence of the distance between the phase-centres on beamwidth and side-lobe level.

(2) Multiple target response for several relative strengths, phases and angular separations.

The results are evaluated using an analyser for complex Fourier series and compared with similar situations in linear arrays. An estimate is given of the unwanted phenomena to be expected. Further experimental research seems to be justified and the difficulties of such experiments are outlined.

Phased Array Radar Systems

K. F. MOLZ. (*Bendix Corporation.*)

Phased array systems provide a number of features which make them particularly suitable for long range, high target density, radar applications such as satellite tracking and space surveillance. These include long range detection capability, high transmitted power, inertialess electronic beam steering, multiple receiving beam capability, high reliability and extreme flexibility.

Basic principles of phased array system operation are presented and a 90-element linear array feasibility model and a large scale, u.h.f. planar array are described. Several potential applications of phased array systems for space surveillance are discussed.

An Analogue Polarization Follower for Measuring the Faraday Rotation of Satellite Signals

GOTTFRIED F. VOGT. (*U.S. Army Electronics Research and Development Laboratory, Fort Monmouth, New Jersey.*)

The conventional method for measuring polarization changes of a satellite signal is to determine the time elapsed between two zeros of the fading caused by the polarization pattern of a dipole antenna. This method is often inaccurate and falls short when measuring the fine-structure of the angular change at close time intervals. The receiving system described in this paper avoids these disadvantages by employing an electronic inertialess scanning method for the antenna pattern control and by using servo systems to track automatically the incident polarization rotation. Simultaneous recordings can be made of one or two satellite signals in the v.h.f. range at different frequencies. Actual measurements will be presented and discussed.

A Side-Lobe Suppression System for Primary Radar

J. CRONEY AND P. R. WALLIS. (*Admiralty Surface Weapons Establishment.*)

This paper describes work carried out by the Admiralty in 1955 and is thought to represent the earliest application of side-lobe suppression techniques by signal processing in primary radar. The system described was applied to an S-band radar, and two versions were used. Basically the method consists of mounting a receiving aerial which is omni-directional in azimuth above the radar aerial, and feeding the received signal to an exactly similar receiving channel as that employed in the radar. These identical receiving channels are superhet systems comprising crystal mixer, head i.f. amplifier, and logarithmic main i.f. amplifier, the crystal mixers being excited by a common local oscillator. The logarithmic amplifiers preserve amplitude differences between signals up to the highest levels encountered, thus signal processing may be applied at their outputs without vitiation by signal limiting effects.

The gain of the omni-directional aerial was made at least equal to that of the worst side-lobe of the radar aerial, and if possible somewhat greater, to give scope for base clipping of the noise of the omni-channel until the amplitude of the radar aerial's worst side-lobe response and the response of the same side-lobe in the omni was reduced to equality. After this base clipping process, subtraction of the twin channel outputs in a video amplifier cancelled the side-lobe signals but preserved the main lobe radar signal at a level dictated by the excess of the main lobe over the worst side-lobe (about 20 dB in this case). The base clipping process reduces the noise contribution of the omni-channel. In the second version a fraction of the radar channel's received S-band signal, was tapped off and fed through an adjustable attenuator and phase shifter into the omni-channel, to give r.f. cancellation of the omni-signal in the direction of the main beam of the radar aerial. This produced a sharp cusp in the omni pattern which rotated with the main radar beam, and avoided the 20 dB amplitude limiting of main beam signals after subtraction.

Illustrations of the large reductions in side-lobe clutter and side-lobe interference from other radars are given in the paper, and the action of the system in the presence of jamming is considered.

Cross Correlation Radar Systems

R. H. MacPhie. (*University of Waterloo, Ontario.*)

This paper is a mathematical analysis of a new cross-correlation radar which uses two antennas for transmitting and two for receiving, four antennas in all. To distinguish the two transmitted signals the carrier frequency of the signal fed to one of the transmitting antennas is shifted slightly. The two transmitting antennas have directivity patterns $A_1(u)$ and $A_2(u)$. After striking the targets the two signals return and are received by the other two antennas whose patterns are $B_1(u)$ and $B_2(u)$. The terminal voltages of these two antennas are then cross correlated.

It is shown that the greater the flexibility of the cross-correlation pattern (four pattern functions instead of one) results in improved side-lobe levels when both have the same beamwidth measured to the first null. It is also shown that the effect of remote, active, noise sources is virtually eliminated in the time-averaged cross-correlation system's output.

In the second part of the paper, a method of mapping radar targets is proposed. Since the various targets will generally produce echoes which are partially coherent, a cross-correlation radar system is necessary to map the distribution. A Fourier analysis of the system output shows that one can define a *principal solution* to the mapping problem which is a generalization of the well-known principal solution encountered in mapping active radio sources. Coherent and incoherent target distributions give rise to characteristic types of outputs. An empirical test for coherence or incoherence is proposed.

Planar Arrays with Unequally Spaced Elements

Merrill I. Skolnik and John W. Sherman. (*Electronic Communications, Inc., Timonium, Maryland.*)

The application of the optimization technique known as *dynamic programming* to the design of thinned planar arrays with unequally spaced elements is described. Two different approaches were investigated. In one method the planar aperture was considered to consist of a number of concentric ring arrays and dynamic programming was applied to determine the radii of the rings which yielded radiation patterns with the minimum peak side-lobe. In the other method the planar array was divided into an even number of equal sectors with elements identically located in each sector. Both methods were selected to avoid excessive calculational times. With the aid of large digital computers designs were obtained and radiation patterns computed for 40λ diameter circular planar arrays and for various degrees of thinning. The results differed from those obtained by other unequally spaced array design methods in that the side-lobes of the radiation pattern tended to decrease with increasing distance from the main beam.

The Use of an Effective Transmission Pattern to Improve the Angular Resolution of Within-Pulse Sector-Scanning Radar or Sonar Systems

D. C. Cooper. (*Electronic and Electrical Engineering Department, University of Birmingham.*)

The paper describes the application of a double transmission technique to linear arrays in order to obtain a directional response with a small beamwidth. An ideal system is analysed and experimental results obtained with a model of such a system are given.

The theoretical and experimental results show that a reduction of the array beamwidth by a factor of two or four may be achieved by the use of additive or multiplicative received signals processing respectively.

A practical time spaced transmission scheme is proposed and its application to within pulse scanning radar or sonar systems is considered.

Optimum Line and Crossed Arrays for the Detection of a Signal on a Noise Background

Professor H. S. Heaps (*Nova Scotia Technical College, Halifax, Nova Scotia*), and C. Wadden (*Naval Research Establishment, Dartmouth, Nova Scotia*).

The paper relates to the design of hydrophone or antenna arrays together with the associated signal processing networks.

For the detection of weak signals in a noise background the outputs from the individual receivers may be combined by means of linear amplifiers and time delays. For a signal of known time dependence and spatial form there is an optimum set of processing networks which maximizes the signal/noise ratio at the output.

A general theory, developed in a previous paper, is applied to consideration of line and crossed arrays. Several types of noise are considered. The analysis is also extended to treat multiplicative arrays. Application of the theory is to consideration of a crossed array in which some of the outputs are combined linearly and then multiplied at the output stage.

Theoretical and Experimental Properties of Two-element Multiplicative Multi-frequency Receiving Arrays, Including Superdirectivity

B. S. McCartney. (*Formerly Electronic and Electrical Engineering Department, University of Birmingham.*)

The paper presents the theory and properties of the directional responses of two-element multiplicative arrays, sometimes termed correlation arrays, when receiving multi-frequency signals. The results of an experimental investigation employing band-pass signals are given and agree well with the predicted directional responses. Interesting modifications to the straightforward multi-frequency arrays are demonstrated, including a split-beam response, array with one channel clipped, and superdirective arrays.

Superdirectivity is examined in terms of the improved effective aperture, for which a quantitative definition is proposed. Some degree of multi-frequency superdirectivity was obtained but further gain of effective aperture is restricted by the accuracy to which the amplitudes of the received frequencies may be controlled.

Deflection of an Ultrasonic Beam in the Near Field Region

L. Kay and M. J. Bishop. (*Formerly Department of Electrical Engineering, University of Birmingham.*)

Electronic beam deflection, of the near field directional pattern 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 directional pattern, arising from the change in the path length from each element of the array to the source or discontinuity is discussed for both steady state and transient conditions as the beam is caused to be deflected by a linear phase taper across the array. It is shown that the beam can be deflected by a significant amount within the Fresnel region of the field. This enables single discontinuities in the medium to be examined. However, multiple discontinuities, each of different impedance, may give rise to ambiguous results of a more serious nature than those experienced in the far field.

Directional Pattern Synthesis of Circular Arrays

G. Ziehm. (*Atlas-Werke, Bremen.*)

Much research work has been done in recent years on linear arrays especially those of the broadside-type. A detailed paper was published by Dolph in 1946. The so-called Dolph-Chebychev amplitude distribution of the radiating elements optimizes the relation between half-power-beamwidth and side-lobe-level.

No comparable work of circular arrays is known to the author though this type of array has advantages for some practical applications. In this paper the closed and not-closed (open) circular array will be considered.

The first problem is to get the general relations between the desired far-field pressure distribution and the complex amplitudes by which the identical radiators of the array must be excited.

A solution can easily be evaluated if some simplifying assumptions are introduced. The result is outlined in form of a linear-equation system in which the number of equations depends upon the diameter/wavelength ratio of the array. This type of solution enables an easy numerical treatment of the whole problem by means of a digital computer.

In the next step, a definition of the ideal radiation beam-pattern will be given. The quality of the agreement between the ideal and attainable pattern mainly depends also upon the diameter/wavelength ratio and some other parameters.

Finally the paper offers some practical details how the feeding distribution of the array-elements could be influenced. It is shown that a directional pattern-synthesis leads in the end to the problem of linear network-synthesis.

The correctness of the idea has been confirmed under laboratory conditions.

The Effect of Noise on the Determination of Direction in a Multiplicative Receiving System with particular reference to the effect of clipping in the input channels

C. R. Fry and Professor D. G. Tucker. (*Electronic and Electrical Engineering Department, University of Birmingham.*)

The signal/noise performance of a multiplicative receiver is examined theoretically and experimentally with particular regard to the accuracy with which the direction of the signal can be determined, and to the improvement in this accuracy which may be obtainable by the use of signal-clipping in the input channels. Subjective results are presented as well as measurements based on r.m.s. noise levels. It is found that, contrary to popular belief, the accuracy is the same whether the peak or the null of the directional pattern is used, and that when the peak is used a further substantial improvement in accuracy is obtained by the use of clipping. In all this work it is assumed that the input signal/noise ratio exceeds unity.

U.D.C. 621.375.1

A Broadband Balanced Idler Circuit for Parametric Amplifiers

By

J. D. PEARSON, M.Sc.†

AND

K. S. LUNT, B.Sc.†

Summary: The bandwidth of the idler circuit of a parametric amplifier is discussed, and reductions in bandwidth associated with stray reactances of the diode encapsulation and filter networks required to complete the idler circuit are described. A balanced idler circuit using two diodes is shown to approach the theoretically maximum idler bandwidth. A 3000 Mc/s amplifier is described having a 10% bandwidth at 20 dB gain.

1. Introduction

One of the limiting factors governing the use of parametric amplifiers has been the small bandwidth available from most amplifiers. Because of this, considerable work has been done on travelling-wave structures and on the use of external loading of the idler circuit. The theoretical bandwidth of an optimized cavity type amplifier is much larger than generally found in practice, and therefore for a lot of applications it should not be necessary to have to resort to the more complicated travelling-wave system or external loading of the idler circuit.

The gain-bandwidth product of a simple parametric amplifier is dependent on the bandwidths of the unpumped signal and idler circuits. In order to obtain low noise figures the signal circuit is heavily loaded by the source impedance and is therefore inherently wide band. The idler circuit is not loaded externally and therefore has a comparatively narrow bandwidth. It follows from the expression for the bandwidth of an amplifier[1] that the bandwidth is limited by the narrowest circuit in the system which, in most cases, will be the idler circuit.

The maximum bandwidth of the idler circuit can be expressed in terms of the characteristics of the semiconductor junction and the stray elements in the encapsulation. The necessity to isolate the three associated circuits of a parametric amplifier, i.e. the signal, idler and pump, has led to the employment of amplifiers using isolating filters with associated stored energy.

A balanced type of idler circuit will be described which confines the idler frequency to the diodes themselves and eliminates the necessity for the use of filters. Thus the theoretical maximum idler bandwidth can be achieved. A balanced type of amplifier has been previously demonstrated by Hayasi and Kurokawa[2] but the method is different from that described in this paper. Other authors[3, 4] have reported on amplifiers using balanced circuits, but no details on the form of circuit are given.

† Ferranti Ltd., Electronics Department, Wythenshawe, Manchester 22.

2. Idler Circuit Bandwidth

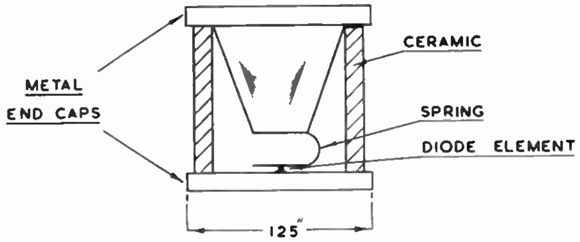
It is possible to calculate the pump frequency which will give the minimum noise figure for a parametric amplifier (see, for example, Greene and Sard[5]). With the variable capacitance diodes now available, the optimum pump frequency is greater than 10 000 Mc/s, thus the optimum idler frequency is high and it will be shown this can be incompatible with obtaining the maximum idler bandwidth. To obtain the maximum bandwidth the idler frequency should be chosen to be the self-resonant frequency of the diode and this is lower than the optimum idler frequency.

The semiconductor *p–n* junction may be represented as a resistance and capacitance in series, in the region where the d.c. conduction is small, that is, under the conditions where parametric amplification occurs. Thus it is theoretically possible to obtain any desired resonant frequency using a semiconductor junction and an inductance. The Q of such a circuit is $1/\omega C_0 r$, a function only of the diode junction at the chosen frequency. In this circuit no excess stored energy is required, the stored electrical energy in the diode capacitance being balanced by the magnetic energy in the inductance.

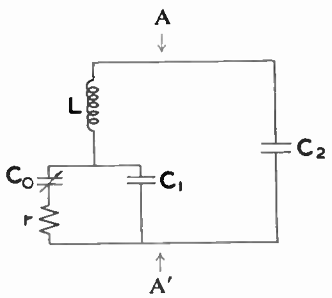
In practice the diode element is contained in an encapsulation which has associated stray reactive elements, and these will increase the minimum Q of the complete diode. A sketch of the internal construction of a pill diode is shown in Fig. 1(*a*). Figure 1(*b*) shows an approximate lumped equivalent circuit of such a diode. An idler circuit may be constructed containing such a diode by the addition of reactive elements placed across the terminals AA′ of Fig. 1(*b*). The maximum value of the idler frequency occurs when an open circuit is placed across the diode terminals AA′, in which case the resonant frequency is given by

$$\omega_s^2 = \frac{C_0+C_1+C_2}{C_2} \cdot \frac{1}{L(C_0+C_1)}$$

The addition of a reactive network containing a capacitance across AA′ to produce a resonant idler circuit will contain more stored energy than the theoretical minimum where ideally the electrical stored energy is restricted to the capacitance in the diode itself. The addition of a pure inductor to complete the circuit adds only the desired magnetic energy to balance the electrical energy in the diode, but reduces the resonant frequency. The maximum bandwidth of a circuit employing the encapsulated diode is obtained by applying a short circuit across the diode in AA′. Such a circuit is resonant at an angular frequency given by



(a) A variable capacitance diode in a pill encapsulation.



(b) Equivalent circuit of a pill diode.

Fig. 1.

$$\omega^2 \simeq 1/L(C_0+C_1) \text{ for } \omega C_0 r \ll 1$$

which is defined as the self-resonant frequency of the diode,[6] and the Q of such a circuit is

$$Q = \frac{C_0+C_1}{C_0} \cdot \frac{1}{\omega C_0 r}$$

In practice C_1 is less than C_0 and C_2, typical values being $C_0 = 0{\cdot}5$ pF, $C_1 = 0{\cdot}1$ pF, $C_2 = 0{\cdot}2$ pF (Ferranti ZC25B). The maximum bandwidth of the idler circuit occurs at the series self-resonance frequency of the diode and its value is only slightly smaller than the theoretical maximum for the semiconductor junction.

The self-resonant frequency of a pill diode when the direct voltage across the semiconductor junction is zero is in the region of 8000 Mc/s; this is lower than the optimum idler frequency for diodes of the quality readily available. However, with this value of idler frequency, good noise figures are theoretically possible, ranging from less than 1 dB at 400 Mc/s to less than 3 dB at 3000 Mc/s.

The difficulty in taking advantage of the maximum bandwidth available from the use of a diode at the self-resonant frequency is the production of the short-circuit to complete the circuit. Johnson[6] pointed out the advantages of the self-resonant idler condition but did not consider how the short-circuit may be produced without affecting the signal and pump circuits.

3. Balanced Idler Circuit

The application of a physical short-circuit across the diode, although producing the optimum idler circuit for the diode, prevents the coupling of a signal or pump circuit, since they will also be shorted out. One of the ways out of the difficulty is to replace the physical short-circuit by a series LC circuit resonant at the self-resonance frequency of the diode. This will produce the desired short-circuit at the idler frequency without preventing coupling of the signal and pump frequencies to the diode. In practice the series circuit would take the form of an open-circuited quarter-wave line or a short-circuited half-wave line placed in parallel with the diode. At the idler frequency the series resonant circuit stores energy and will reduce the bandwidth of the complete circuit. The Q of the idler circuit is then

$$\frac{C_0+C_1+C_3}{C_3} \cdot \frac{C_0+C_1}{C_0} \cdot \frac{1}{\omega C_0 r}$$

where C_3 is the capacitance in the added series circuit. If a transmission line of characteristic impedance Z is used to complete the circuit, the Q of the idler circuit is given by

$$Q = \frac{C_0+C_1}{C_0} \cdot \frac{1}{\omega C_0 r} \cdot \left[1 + \frac{\pi}{4}\omega Z(C_0+C_1)\right]$$

for an open-circuited quarter-wave line, and

$$Q = \frac{C_0+C_1}{C_0} \cdot \frac{1}{\omega C_0 r} \cdot \left[1 + \frac{\pi}{2}\omega Z(C_0+C_1)\right]$$

for a short-circuited half-wave line.

Thus, if the reduction in bandwidth of the idler circuit is to be small, the capacitance of the added circuit should be large or, in the case of the circuit being completed by a transmission line, the characteristic impedance of the added line should be small.

The idler circuit can be completed without an increase in circuit Q by placing two diodes side by side, with opposite polarity with respect to one another as shown in Figs. 2(a), (b) and (c). The signal and pump circuits are so arranged that the signal and pump voltages appear across each encapsulation at any instance in the same direction. The mixing action of the signal and pump voltages, combined with the reversed directions of the semiconductor junctions, is such that the idler voltage is produced in opposite phase in each diode. The idler circuit consists of the two diodes in series, in contrast to the signal and pump circuits where the diodes are effectively in parallel.

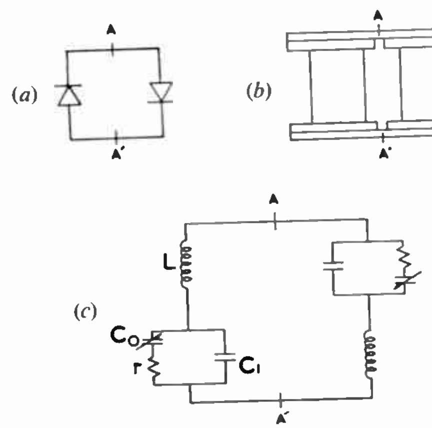


Fig. 2. Balanced pair of diodes.

In a coaxial system the signal and pump circuits are connected to the diodes at their junction (point A in Fig. 2). Figure 3 shows a waveguide system in which the diodes are placed side by side across the waveguide cross-section, which may be of reduced height. The pump and signal power is fed along the waveguide in the H_{01} mode, thus creating equal voltages across each diode.

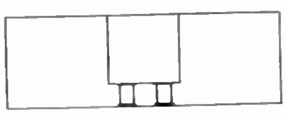


Fig. 3. Balanced pair of diodes mounted in a waveguide.

In both the coaxial and waveguide arrangements, the idler voltage has zeros at the junction points A and A′. Thus idler power cannot flow along the coaxial signal or pump feed, or alternatively with waveguide feeds, the reversed idler voltages cannot excite the H_{01} mode in the waveguide; the waveguide dimensions are chosen to place all higher modes beyond cut-off at the idler frequency. Thus isolation of the idler circuit has been obtained without the use of additional tuned circuits, no excess stored energy is required, and the idler Q of the diode pair is equal to that of a single diode.

It is shown in the Appendix that the double diode arrangement may be analysed by equations similar to the single diode system, the two diodes acting in parallel in the signal circuit and in series in the idler circuit. Thus the well-known expressions for noise figure, bandwidth, etc., for the single diode amplifier apply directly to the double diode case.

4. Practical Considerations

When applied to microwave frequencies, some modification of the basic idea is to be expected. The self-resonant frequency of pill diodes (Ferranti type ZC25B) were measured by placing them in the centre of reduced height (0·1 in) X-band waveguide, and finding the frequency at which the transmission of power passed the diode was a minimum. This was found to vary from 7500 Mc/s to 8500 Mc/s. Diodes were then selected to form matched pairs having as near as possible equal self-resonant frequencies.

When used in amplifiers employing a coaxial signal circuit and a waveguide pump circuit, the idler frequency was found to be approximately 6500 Mc/s, and to have a bandwidth greater than 1000 Mc/s. The reduction in idler frequency from the measured self-resonant frequency may be explained by the additional inductance between the metal ends of the pills. Thus a more exact lumped equivalent circuit of the two diodes used in the idler circuit is as shown in Fig. 4. Some stored energy at the idler frequency occurs now in the stray capacitance C_2, giving a slight reduction in the bandwidth.

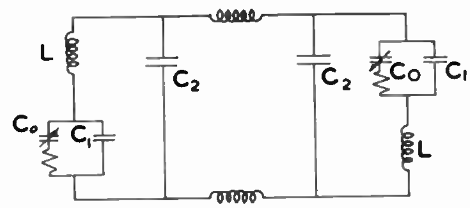


Fig. 4. Equivalent circuit of the pair of diodes at microwave frequencies.

5. Experimental Results

An amplifier was constructed for 2940 Mc/s using a pair of ZC25B diodes mounted across a narrow height X-band waveguide. The signal circuit was coaxial and incorporated a quarter-wave matching section to obtain a coupling of approximately 10 : 1 (over-coupled). When pumped at 9500 Mc/s a bandwidth of 100 Mc/s at 20 dB gain was obtained and the noise figure was 2·9 dB, which included an 0·25 dB circulator loss and a contribution from the second stage (a balanced mixer with a noise figure of 10 dB). By varying the pump frequency from 8100 to 10 500 Mc/s the midband signal frequency could be varied from 2610 Mc/s to 3120 Mc/s with bandwidths from 70 to 120 Mc/s at 20 dB gain.

Because of the relatively small change in signal frequency with changes of pump frequency it was thought that the bandwidth of the idler circuit must be larger than that of the heavily-loaded signal circuit.

Another form of signal circuit was then used with a Chebyschev double quarter-wave transformer in place of the single quarter-wave section to obtain approximately the same coupling (8 : 1). Using this system a bandwidth of 280 Mc/s was obtained at 20 dB gain; a plot of the frequency response is shown in Fig. 5.

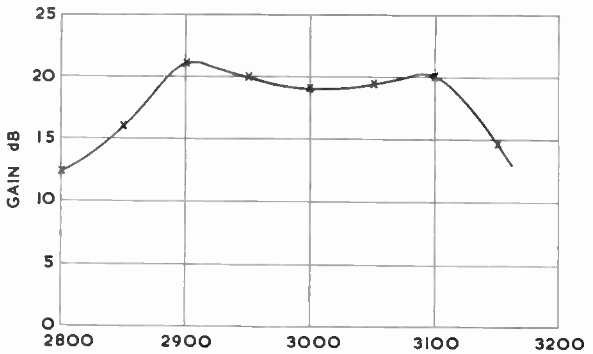


Fig. 5. Gain–frequency response of an S-band parametric amplifier.

6. Conclusions

A method has been found of producing a near-ideal idler circuit at 6500 Mc/s. An extension of the principle to higher idler frequencies is dependent on obtaining diodes with lower series inductance. This will lead to amplifiers having very large bandwidths together with very low noise figures.

7. Acknowledgments

The authors wish to thank the Manager of the Electronics Department of Ferranti Limited for permission to publish this paper.

8. References

1. H. Heffner and G. Wade, "Gain, bandwidth, and noise characteristics of the variable-parameter amplifier", *J. Appl. Phys.*, **29**, pp. 1323–31, September 1958.
2. S. Hayasi and T. Kurokawa, "A balanced-type parametric amplifier", *Trans. Inst. Radio Engrs (Microwave Theory and Techniques)*, **MTT-10**, pp. 185–90, May 1962.
3. J. Kliphuis, "C-band nondegenerate parametric amplifiers with 500-Mc bandwidth", *Proc. Inst. Radio Engrs*, **49**, p. 961, May 1961.
4. M. Grace, "An extremely wideband tuneable S-band parametric amplifier", *Proc. Inst. Radio Engrs*, **49**, p. 1940, December 1961.
5. J. C. Greene and E. W. Sard, "Optimum noise and gain-bandwidth performance for a practical one-port parametric amplifier", *Proc. Inst. Radio Engrs*, **48**, pp. 1583–90, September 1960.
6. Kenneth M. Johnson, "Broad-band cavity-type parametric Amplifier Design", *Trans. Inst. Radio Engrs (Microwave Theory and Techniques)*, **MTT-9**, pp. 187–94, March 1961.
7. R. L. Steven, "Octave tuning parametric amplifiers", *Microwave J.*, **5**, pp. 167–71, October 1962.

9. Appendix

The arrangement of the two diodes in the amplifier is shown in Fig. 6. At the signal frequency the two diodes are in parallel and are resonated by an inductance; at the idler frequency the two diodes in series form the complete idler circuit. The equivalent

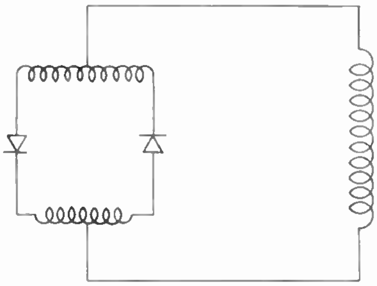


Fig. 6. Equivalent circuit of a pair of diodes in a parametric amplifier.

circuits at the signal and idler frequencies are shown in Fig. 7. In the analysis of a parametric amplifier it is usual to consider the diode as a device with two terminals, at the signal and idler frequency[5]; this form of analysis will be followed in the balanced diode arrangement, and Figs. 7 and 8 define the voltages and currents appearing across the diode terminals at the signal and idler frequencies. In order to follow the normal notation in this appendix the operating junction capacitance of a single diode is defined as C_0, and the amplitude of capacitance variation of the pump frequency by C_1. The currents entering the diode junctions at the signal frequency are given by

$$i_1' = j\omega_1 C_0 V_1 + j\omega_1 C_1 (V_2')^*$$
$$i_1'' = j\omega_1 C_0 V_1 + j\omega_1 C_1 (V_2'')^*$$

where the asterisk shows the conjugate.

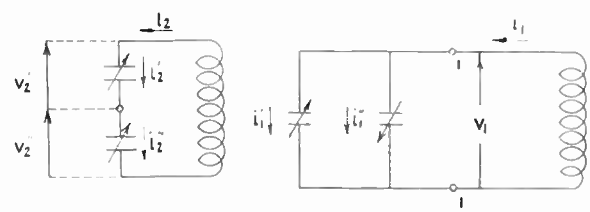


Fig. 7. Equivalent circuit at the signal and idler frequencies.

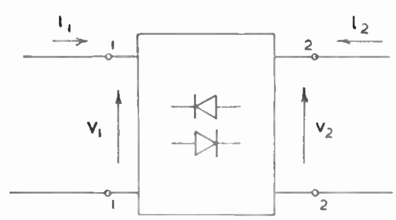


Fig. 8. Diagram showing the current and voltages associated with the balanced amplifier.

At the idler frequency the currents entering the diode junctions are given by

$$i_2' = j\omega_2 C_0 V_2' + j\omega_2 C_1 V_1^*$$
$$i_2'' = j\omega_2 C_0 V_2'' + j\omega_2 C_1 V_1^*$$

Since in the signal circuit the two junctions are in parallel, and in the idler circuit the junctions are in series, it follows that

$$i_2 = i_2' = i_2''$$
$$i_1' + i_1'' = i_1$$
$$V_2' = V_2'' = V_2/2$$

From the equations above it may be shown

$$i_1 = j\omega_1 2C_0 V_1 + j\omega_1 C_1 V_2^*$$
$$i_2 = j\omega_2 \frac{C_0}{2} V_2 + j\omega_2 C_1 V_1^*$$

These equations are similar to the normal expressions for a single-diode parametric amplifier except in the first term on the right-hand side of the equations. These show the two diodes as parallel capacitances in the signal circuit and series capacitances in the idler circuit. Since the linear terms in the analysis for gain and noise figure are cancelled out by the susceptance of the tuning inductors, it follows that the expressions for noise figure and gain is the same in the single diode and balanced idler, parametric amplifiers.

Manuscript received by the Institution on 13th December 1963. (Paper No. 902.)

STANDARD FREQUENCY TRANSMISSIONS

(Communication from the National Physical Laboratory)

Deviations, in parts in 10^{10}, from nominal frequency for **April 1964**

April 1964	GBR 16kc/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.	April 1964	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	− 150·2	− 149·4	+ 8	16	− 150·2	− 150·9	− 4
2	− 151·7	− 151·2	+ 9	17	—	− 149·4	− 4
3	− 151·4	− 150·8	+ 8	18	− 150·2	− 150·0	− 3
4	− 150·9	− 150·3	+ 9	19	− 149·9	− 149·9	− 1
5	− 150·5	− 149·9	+ 8	20	− 150·0	− 150·1	− 1
6	− 150·4	− 150·2	− 2	21	− 150·0	− 150·3	− 3
7	− 149·9	− 151·7	− 4	22	− 150·0	− 150·8	− 3
8	− 151·0	− 150·4	− 6	23	− 150·3	—	− 2
9	− 149·5	—	− 4	24	− 150·4	− 150·6	− 3
10	− 150·7	− 150·8	− 4	25	− 150·7	− 151·2	− 1
11	− 150·1	− 149·9	− 4	26	− 150·9	− 151·9	− 2
12	− 149·9	− 150·4	− 3	27	− 151·7	—	− 2
13	− 149·8	− 149·6	− 5	28	− 151·2	− 149·8	− 3
14	− 149·1	− 150·4	− 4	29	− 148·9	− 149·8	− 2
15	− 149·8	− 150·6	− 3	30	− 149·7	− 150·8	− 1

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. The phase of the GBR/MSF pulses will be retarded by 100 milliseconds at 0000 UT on 1st April, 1964.

New Caribbean Communications Scheme

Equipment for a multi-channel tropospheric scatter and microwave link system, connecting the Windward and Leeward Islands in the Eastern Caribbean and providing improved telephone communications within the eastern Caribbean and ultimately to the Commonwealth and to the rest of the world, is to be installed on the islands of Antigua, St. Lucia and Barbados. This system will link with a co-axial telephone cable to be laid from Antigua to St. Thomas in the Leeward Islands and from there by a new American cable to Florida. In Barbados the link will be integrated with an existing tropospheric scatter system between Mount Misery on the island of Barbados and Blanchisseuse in Trinidad. A later phase of the project is expected to cover a microwave link between Blanchisseuse and La Basse in Trinidad.

This scheme will provide comprehensive communication facilities between Trinidad and Antigua—over a total distance of 530 miles. It will be covered in four major steps, linking Antigua, St. Lucia, Barbados and Trinidad. The complete system will provide well over 36 000 voice-channel miles between the communications centres.

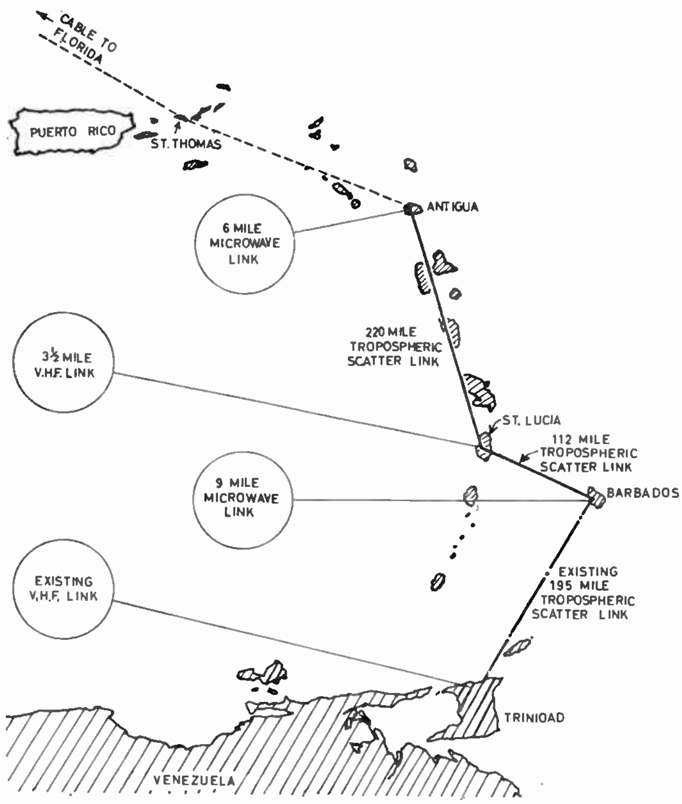
Antigua The first step across the island of Antigua will be covered by a short 7000 Mc/s microwave link joining the cable terminal to Hughes Hill, the tropospheric scatter station. It will be a twin-path system carrying 64 channels.

A 900 Mc/s tropospheric scatter link will join Hughes Hill to Petit Monier on the island of St. Lucia, 220 miles south of Antigua. This link will carry 63 channels, each of 3 kc/s bandwidth and will use quadruple diversity and 1 kW transmitters. Two 30-ft diameter antennae are used at each end of the link with parametric amplifiers to give a low noise factor over this comparatively long section of the route.

St. Lucia From Petit Monier there will be a short $3\frac{1}{2}$-mile v.h.f. spur across the island to a central telegraph office at Castries. This will be a 32-channel twin-path link using equipment that is at present operating between Mount Misery and Carrington on Barbados.

The next section will cover a distance of 112 miles linking Petit Monier to Mount Misery (an existing tropospheric scatter site in Barbados) by a 2000 Mc/s tropospheric scatter link with quadruple diversity. This section will use 5 W transmitters with 20-ft diameter antennae and will carry 80 channels.

Barbados A twin-path 7000 Mc/s microwave link carrying 92 channels, will replace the existing v.h.f. link between the Mount Misery tropospheric scatter site and the receiving station at Carrington, 9 miles away across the island. A passive reflector will be installed on the route three miles from the Carrington terminal.



Tropospheric scatter and line of sight radio systems

The last 195 miles are covered by the existing 900 Mc/s tropospheric scatter link between Mount Misery and Blanchisseuse Road, Trinidad. The link capacity will be increased from 12 to 64 channels.

Trinidad A further development of the scheme will replace the existing v.h.f. link, between the scatter site at Blanchisseuse Road and the telegraph office at Port of Spain, with a microwave link using two passive reflectors.

In order to provide the maximum channel carrying capacity, and to ensure compatibility with the connecting cable systems at Antigua, most of the links will use 3 kc/s spaced channelling equipment.

This project is being carried out for Cable and Wireless by the Marconi Company who are supplying the tropospheric equipment and microwave links. Sub-contractors include Pye Telecommunications (microwave links); Radio Electronic Laboratories of America (parametric amplifiers); Telephone Manufacturing Company (3 kc/s spaced channelling equipment); Associated Electrical Industries (carrier generators and group translating equipment) and Electronic Speciality Company of America (aerial systems).

There are many studies of tropospheric scatter propagation applied to communications systems. See, for example, in the literature, M. Telford, "Tropospheric scatter system evaluation", *J.Brit.I.R.E.*, **18**, No. 9, p. 511, September 1958.

U.D.C. 620.1:621.37/8:621.48

Automatic High-speed Measuring Systems for Complex Products and Shapes

Interdependent Computation and Cybernetic Inspection Machinery

By

JOHN A. SARGROVE

(*Member*)†

Presented at the Convention on "Electronics and Productivity" in Southampton on 19th April 1964.

Summary: A high-speed self-adaptive system of inspection is described and the reasons for evolving it are explained. Sequential measurement on a transfer-line is used to measure complex-shaped parts reliably to effective accuracies between ± 1 to 2 microns. Each measured dimension is stored up in analogue memory devices and later read-out as the object reaches the sorting or pass/reject gates. The decision to pass or reject the object is carried out in a self-adaptive logic system in which the subsidiary dimensions are compared with the dominant dimensions and the upper and lower tolerance limits of the subsidiary dimensions are automatically shifted instantaneously for each object as the dominant dimension varies within tolerance, thus achieving critical inspection for true shape.

Machines are described for measuring conical and spirally grooved complex shaped objects such as twist-drills in more than one place. Also measurements of more abstract factors are referred to, such as inspection at speeds of 180 measurements per minute involving computing for 'density' and pass/reject action on this factor. Practical results have been obtained in machines which have inspected millions of objects. The cost of such apparatus is justified by reduction of scrap and manual inspection labour in production of precision parts.

1. Introduction

Most modern production processes have reached a degree of automation where a great deal of tiresome labour has been reduced to a minimum. With modern production equipment we get a much greater production output from a given team of human workers. Thus, the labour content of any unit produced, so far as cost is concerned, is much lower than it was before the automation equipment was introduced. However, the financial investment in special production equipment has risen appreciably, and thus it is very important to run this production equipment a maximum number of hours per day, so that the 'overall cost' of the product will be less than with the older methods of production techniques.

With the older methods of production the human worker was more closely in charge of the work, and thus to a certain extent could act as his own inspector and control the quality of his work according to his own craftsmanship, temperament and conscientiousness.

With the automated equipment, only a certain degree of automatic *quality-control* is practicable, and although this can be designed to function very efficiently nonetheless the finished product still requires inspection. In many modern factories this leads to the situation that the ratio of human inspectors is very high in relation to the human beings in the Production Department. It is obvious that the automation of inspection work is just as logical as the automation of machining, or in fact of production processes of any kind.

During the second World War it became necessary in the production of high precision components, such as machine gun ammunition, to evolve high-speed equipment for checking all relevant dimensions of, for instance, machine gun bullets and cartridge cases. An early electronic machine of this type designed by the author was an automatic hopper-fed inspection machine, and contained a device for automatically and correctly orientating all the cartridge cases before feeding them into the inspection chamber or magazine. The objects were momentarily halted for inspection by electronic means which took a decision and opened one of three chutes to discharge the object according to sorted characteristics. All this took place at a continuous inspection rate of 180 objects per minute.[1] However, we must not forget that this early machine

† Automation Consultants and Associates Ltd., Weybridge, Surrey.

was designed for a single purpose for inspecting a unique-sized object. For an adjustable machine to deal with many sizes of objects one could not have achieved such speeds as early as 1942 (Fig. 1).

Today a multitude of objects having cylindrical diameters are dealt with by inspection machines of high precision which take simultaneous measurements on as many as 60 places (partly diameter and partly length). They are mainly used in automobile factories, for instance on crank shafts and cam shafts and similar complex parts.[2] The individual measurements and the tolerances are independent of each other.

2. Sequential Multi-Point Inspection

When the shape of the object to be measured is not a simple sphere or cylinder, but is much more complicated, such as for instance a jet nozzle, a milling cutter, a twist drill or such a complex shape such as a turbine blade, it is very difficult to construct a machine in which the object can be presented to the measuring probes (or transducers) for the simultaneous measurement of a number of dimensions. Thus, if one attempted to devise a machine for simultaneous measurements of such complex parts the automatic decision to 'pass' or to 'reject' the object from the production line or process might quite easily be based on errors of positioning the object and not necessarily on the inaccuracies of the various dimensions themselves. This difficulty has led to the evolution of the concept of sequential multi-point inspection in which at least part of the 'use function' of the object (such as a tool) is reproduced in the measuring set-up, so that the ultimate performance of the part can be ascertained and the 'decision' to pass or reject be arrived at accurately.

2.1. *Dynamic Inspection*

The above problem, particularly in the inspection of cutting tools, has led to the evolution of inspection machines in which the complex-shaped object is

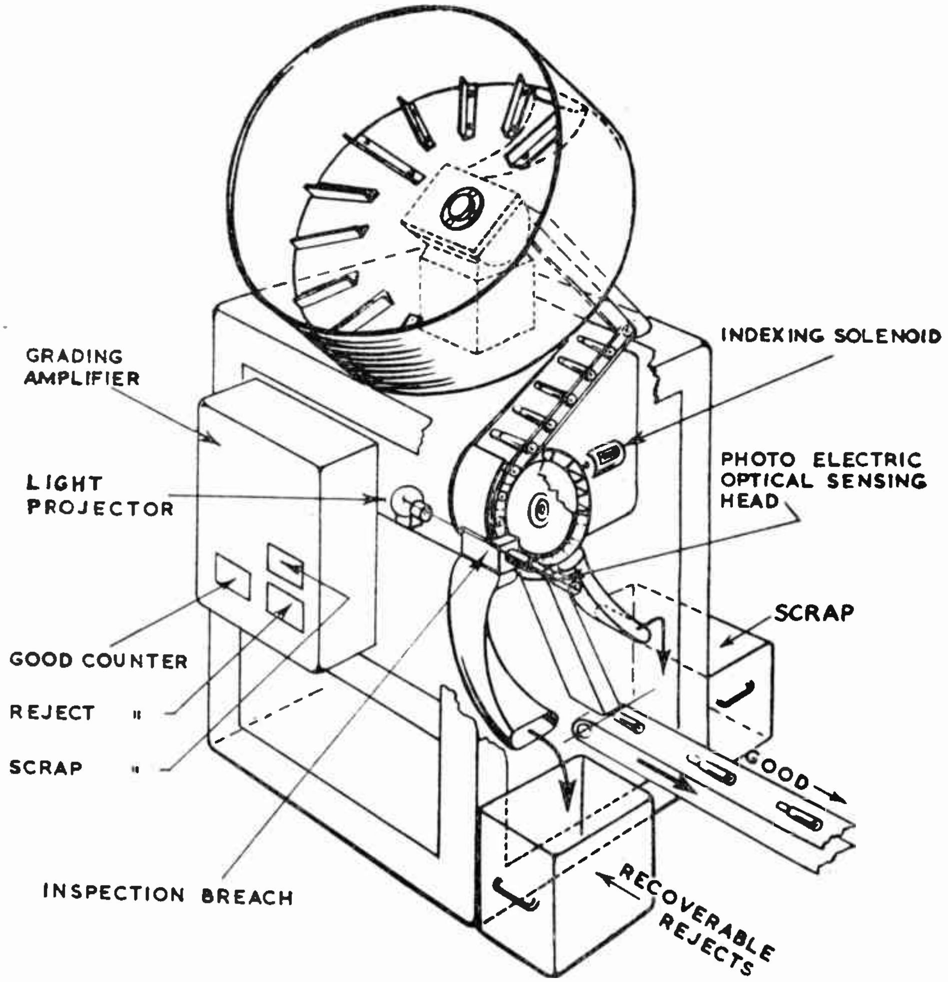


Fig. 1. An early electronic inspection machine (1942). High-speed hopper-fed cartridge inspection machine.

actually moving within the various sequentially-arranged inspection measuring stages, and each peak dimension of the cutting edges of the tool is measured by suitable contact type electro-magnetic transducers. The measurement information from each succeeding stage of measurement on the same object is stored up in a memory device suitably designed so as to be able to present to the 'decision-device' all the measurements taken on the same component at the point of time immediately following the last measurement on the same object. The decision device then controls the 'pass/reject gates' or other 'sorting-gates'.

As several individual objects are passing through the inspection machine following each other, it is obvious that several memory devices have to be associated with each measuring stage, so that the measured information for each object can be stored separately. The electronic circuit has to be so designed as to present the relevant measurement information relating to any one object simultaneously to the 'decision logic system'.

In this 'systems concept' one can carry out 100% production control at high speed, instead of quality control by manual inspection or sampling which have been the accepted practices in the past. This is particularly important if the part being inspected is intended for handling by an automatic assembly machine. In this case it is quite obvious that, unless 100% of produced parts are inspected prior to assembly, the automatic equipment could be subject to frequent stoppages due to jamming of an oversize object, for instance when being pushed into a standard size hole.

2.2. *'Cybernetic' or Interdependent Inspection*

We are now leading up to the concept of one dimension being interdependent with another—for instance, this can be of great use in the automatic assembly of roller-bearings or ball-bearings where selective assembly becomes possible, and one can finish up with a roller assembly having a closer tolerance than that to which the individual rollers can be manufactured. All such schemes have to rely on interdependent measurement logic. Such interdependent inspection logic devices are sometimes also referred to as 'auto-computing inspection devices'.

For instance, if we are inspecting the angle of taper it is not enough to ascertain this by measuring, say, two diameters at a given distance from each other and placing fixed tolerances on each diameter. This argument becomes clear if we suppose one diameter is on the high side but still within tolerance, and the other diameter is on the low side but still within tolerance: the result would be that the angle of taper is quite incorrect. Obviously to obtain the correct angle both have to be high or both low. This concept of ascertaining the contour or shape of a form becomes even more important where it is necessary to consider the relationship of two or more dimensions. Thus, in a nozzle for instance (see Fig. 2) whatever be the dimension of the value of the differences $(b-a)$ and $(c-a)$ should be preserved within the limits of the permitted error. Thus, if the dimension a is at the lower limit of tolerance then dimensions b and c should be correspondingly near the lower limits. Thus by using interdependent inspection logic in the design of the electronic 'decision' circuit we can obtain an intelligent

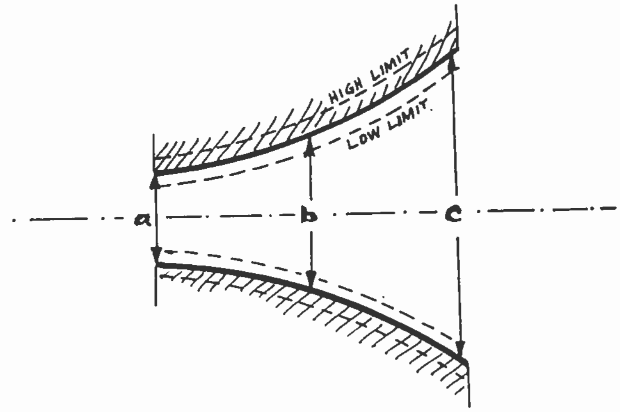


Fig. 2. Interdependent logic used to determine the shape of a curved object. Example is a nozzle.

assessment of *quality of shape*. In many cases this philosophy actually enables one to widen the manufacturing limits at the individual measurement point and thereby make manufacture more easy to carry out and therefore more economical. By using the intelligent sorting ability of the interdependent inspection device one can produce a more consistent functional result in the finished product.

In almost every case the mechanical parts of the inspection machines and the special handling systems have to be evolved to suit the particular object to be inspected. In addition careful attention has to be paid to shapes and movement characteristics of the object to be inspected. The design of such inspection machines is not an easy matter and only practical experience in the combination of mechanical and electronic systems can lead to the design of successful high speed precision inspection equipments. Two examples have been chosen to show how the above principles can be applied in particular cases.

3. High-speed Gauging of Twist-Drills

It would be difficult to find an example of a more complex shaped object, which combines more measurement inspection problems, than that of a range of twist drills.

Both the International Standards Organization and the British Standards Institution specify certain indispensable measurements which have to be made on twist drills during manufacture before they can be marked with numbers or characters designating their specific size. These are as follow:

(*a*) The maximum cutting diameter at the 'tip' (Fig. 3). This, in the smaller-sized drills, must not exceed the stated diameter at all and must never be smaller than by approximately $1\frac{1}{2}$ hundredths of a millimetre (approx. 0·0005 in). As will be seen from the end view of a twist drill (Fig. 4(*a*)) it is no easy matter to measure this dimension exactly across the cutting edges of the drill tip. From the exaggerated long view (Fig. 4(*b*)) it is clear that the micrometer shoes must partly overlap the cutting edges to give the true cutting diameter. To achieve this measurement it is best to roll the twist-drill between truly parallel surfaces and measure the maximum displacement of these surfaces during rolling, i.e. the peak-diameter. The accuracy of discrimination between those between acceptable limits and those outside must be to accuracies at least 10 times better, i.e. to 0·0015 mm (or 0·00005 in).

(*b*) If we only had this one dimension to measure at speed we would in any case be faced with a difficult problem due to the non-cylindrical shape. We also have to check the so-called 'back-taper angle'. In the standard specifications it is laid down that the slope of back-taper must never exceed 1/1000 and should never be less than 1/2000. (The purpose of the back-taper on a twist drill is to avoid the main body of the drill rubbing on the sides of the wall of the hole being drilled, since if the drill had a parallel body it would rub, heat up and jam, and possibly break). It is convenient to ascertain the 'back-taper' by another measurement of peak diameter about halfway along the spiral fluted part of the drill. Such a point *F* along the spiral flutes of the drill is shown in Fig. 4(*b*). It also has to be measured dynamically while the drill is rolling between parallel surfaces in a similar manner to that used at point *T*.

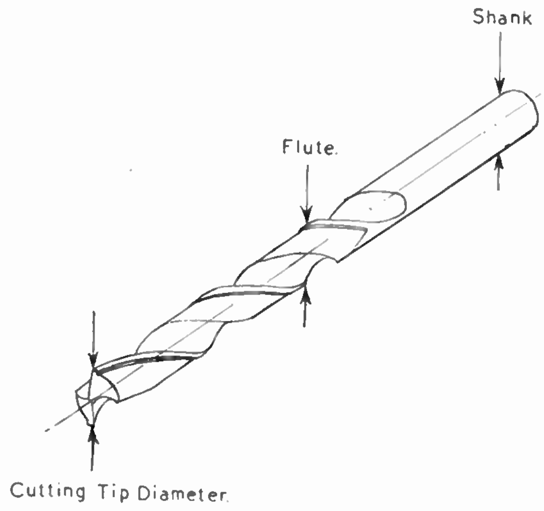
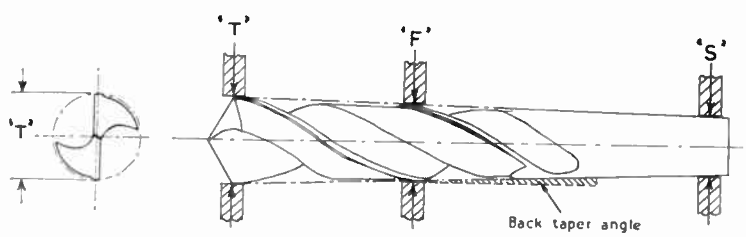


Fig. 3. Problems with a very complex object such as a drill.

(*c*) We also have to measure a true cylindrical diameter at a point along the 'shank' shown in Fig. 4(*b*) as point *S*. This latter measurement is not difficult to ascertain.

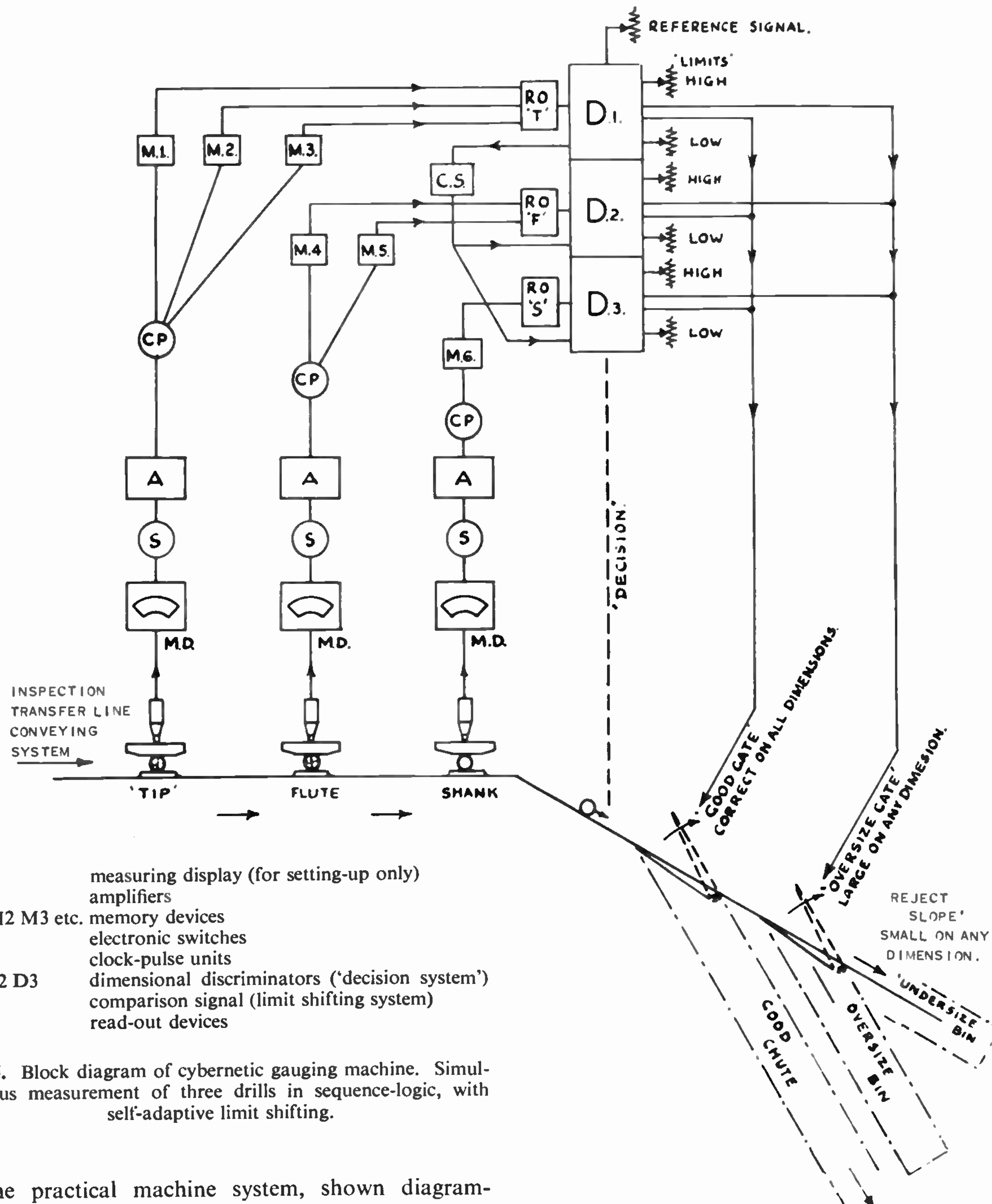
During the many years of research work which preceded the design of a successful twist-drill inspection machine it became manifest that one could not measure all these dimensions simultaneously, either with the drill in a static position or dynamically while rolling. This difficulty is due to the fact that a coincidence of the peak-diameters at all three places in any one angular position on the drill does not occur, and hence by attempting a simultaneous measurement method one could not arrive at a correct 'decision' as to whether a drill was acceptable or not.

The practical machine was thus designed so as to measure the dimensions *T*, *F* and *S* on any one drill in sequence in a specially designed 'measuring-transfer-line', all stages having analogue storage of the peak-diameter measurements.



(*a*) end view of cutting tip; (*b*) typical measuring points.

Fig. 4. Greatly exaggerated 'back-taper slope' on a drill.



MD	measuring display (for setting-up only)
A	amplifiers
M1 M2 M3 etc.	memory devices
S	electronic switches
CP	clock-pulse units
D1 D2 D3	dimensional discriminators ('decision system')
CS	comparison signal (limit shifting system)
RO	read-out devices

Fig. 5. Block diagram of cybernetic gauging machine. Simultaneous measurement of three drills in sequence-logic, with self-adaptive limit shifting.

The practical machine system, shown diagrammatically in Fig. 5, was thus designed to measure the dimensions T, F and S on any drill in a queueing sequence in a specially designed 'inspection-measurement transfer-line'. In the bottom left corner of Fig. 5 this is shown as consisting of three measuring stages following each other from left to right, the parallel motion shoes whose suspension is mechanically designed to enable the shoes to move upwards completely parallel to the anvil base, better than 10 microinches parallel (better than 0·4 microns).

The measurement transducers are of the balanced inductive carrier-frequency type and use 10 kc/s carrier frequency at several hundred volts peak to peak amplitude. The unbalance is measured by phase-displacement ring-demodulators also fed from the same oscillator in the same channel. A display meter is provided in each channel for setting-up purposes only which has sufficient sensitivity to cover twice the specified range of error deviation, and this part of the system has a repeatability better than 1% (where 1% represents 10 microinches, i.e. the whole range of the carrier frequency system is completely linear over 0·001 in).

The carrier frequency amplifier has very high gain suppressed by heavy negative feedback, and it is this which ensures the good stability of the measurement system. These features are diagrammatically shown in Fig. 5 which also shows the display muting switch S.

The measurement units are followed by individual high stability low-gain d.c. amplifiers bringing the measurement range to a preset analogue voltage range of 20 V. Thus 1% of the total range represents 0·2 V. In the practical system this value is held over long periods quite satisfactorily by using heavy negative feedback and other well-known stabilizing techniques.

Because the drills progress through the measurement transfer-line in a queue, the measurement information has to be marshalled into a number of appropriate memory devices so that just before the 'decision' to pass or to reject is taken, all the relevant information concerning a particular drill can be brought together electrically, without any interference by dimensional remembered information concerning other drills passing along the transfer-line. A suitable clock pulse-generator (CP in Fig. 5) feeds the appropriate information into the various memory devices, three for the 'tip' channel, two for the 'flute' channel and one for the 'shank' channel. The information concerning 'tip' dimension has to be remembered for three seconds, (the rhythm of the machine being one drill per second throughput) before the drill reaches the mechanical point in the transfer-line where the acceptable drills can be diverted mechanically from the reject specimens.

3.1. *Interdependent Computation in the 'Decision-device'*

As mentioned above, during normal use the machine measures three different individual drills in a queue which are continuously changing. Every drill passes the measuring stages in sequence, the first one measuring the 'tip' dimension, the second measuring the 'flute' dimension, whilst the last stage measures the 'shank' diameter. Before a particular drill reaches the mechanical point in the machine where a 'decision' to pass or to reject can take place, a particular drill has progressed forward for a total of three seconds through three stages. Hence, we require six 'memory devices'; M1, M2 and M3 concerned with the 'tip' dimensions of three drills, M4 and M5 concerned with the 'flute' dimensions of two drills and M6 a very short term memory for the 'shank' dimension of the drill just before it reaches the 'decision' point in the machine.

The relevant information relating to one particular drill will for instance be contained in M1, M4 and M6 and these have to be marshalled through the appropriate 'read-out' devices to reach the 'decision-system' which is a small simple analogue computer designed to compare the three dimensional analogue voltages with each other all concerned with a particular drill.

The read-out information is repeatably accurate to better than ½%, i.e. better than 5 microinches in mechanical equivalent (0·2 microns).

By suitable switching logic the three relevant analogue dimensional values are presented to the decision device simultaneously. The 'decision-logic' works as follows:

The 'tip' dimension is discriminated on its own according to pre-set discriminator limits; to comply with the previously mentioned B.S. and I.S.O. specifications, it has a top limit which is accurate to within 0.00005 in on the nominal drill diameter, and a bottom limit setting at 0·0005 in undersize ± 0·00005 in. These are in electrical terms equal to one volt discrimination steps at each limit. Thus the discriminator devices should be at least five times more accurate.

To achieve such accuracy it was found necessary to use a form of blocking oscillator circuit which acts as a trigger device at less than 0·1 V upward steps. It will be readily seen that this trigger sensitivity readily meets the requirement, in fact it is better than required, by a factor of ten. There are six such trigger discriminators associated with the dimensional discriminators D1–D3 in Fig. 5. The odd-numbered triggers T1, T3, T5 operate for 'good' dimensions, the even-numbered T2, T4, T6 for 'oversize' dimensions.

To meet the requirements of the B.S.I. and I.S.O. specifications referred to above, one requires the 'flute' and 'shank' dimensions to be related to the actual 'tip' dimension and thus we have resort to a method of interdependent-logic as follows:

In principle the logic is set up by suitable relays as shown in Fig. 5 and controlled by clock-pulse signals to gate them to the final signal hold relays which control the output 'sorting-gates' of the machine. It will be seen, that the 'good-output gate' is governed by three series contacts, via the triggers T1, T3 and T5, thus if all these are 'made' then they energize the 'good-hold-relay'.

Correct size range: If all measured peak-diameters are within tolerance and also the 'back-taper slope' is within tolerance then the resultant action is 'acceptance' and the drill is passed out via a conveying chute to other operations.

Oversize: If any peak-diameter measured is larger than tolerance, or the back-taper is insufficient (the

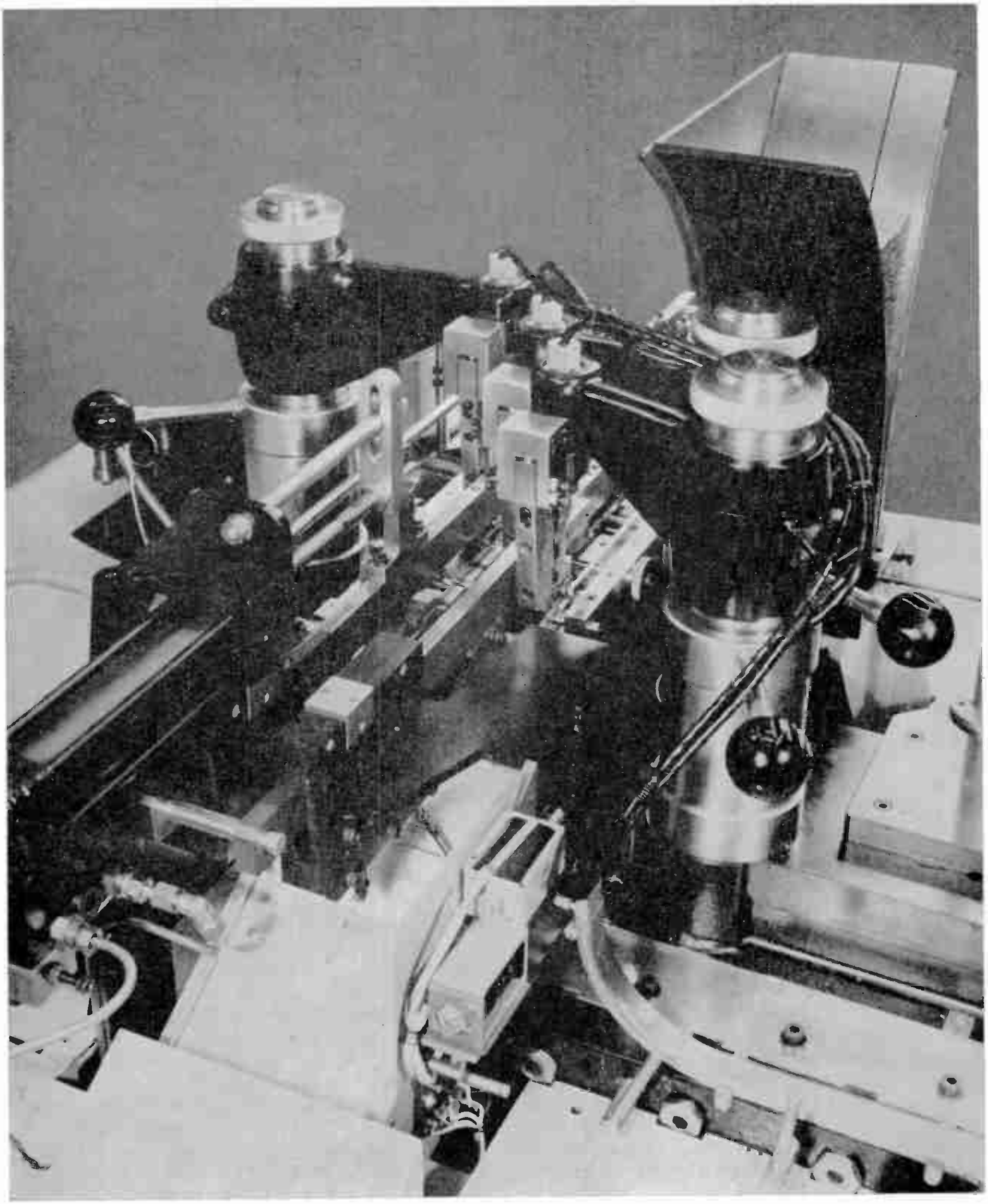


Fig. 6. The high-speed twist drill inspection machine.

'tip' peak-diameter taking precedence), then the resultant action is 'rejection' into the 'oversize' bin of the machine. From this bin these reject drills are taken to be reground to correct back-taper and measured again. Any of the oversize triggers, T2, T4 or T6, make their associated contacts, energizing the 'large-gate-hold' relay.

Undersize: If any diameter measured is smaller, or the back-taper is excessive (again the 'tip' diameter taking precedence) the resultant action is 'rejection' into another reject bin marked 'small'. From this bin the drills are later taken to be reground to a smaller nominal sized drill, and handled with those other drills.

If any of the six relay contacts are not triggered at all, then obviously one dimension is undersized or the taper is wrong. Clearly, from the diagram Fig. 5, neither of the 'gate-hold circuits' will make, and thus both sorting gates will remain inactive and such a drill will roll uninterruptedly into the 'small' or 'undersize' bin, as both gate flaps will remain in their inactive or down positions. If anything in the machine goes wrong the gates remain inactive and all drills are rejected.

It has been found most convenient to adopt analogue-computation techniques in this machine system and the whole circuit system is a novel combination of peak-diameter measurement of a fluted object in three places, analogue memory devices and sorting logic devices based on a 'decision-device' in which '*interdependent logic*' is used (see Sect. 3.2). In practice all these devices work very reliably and the accuracy achieved is higher than that required to meet the B.S. and I.S.O. specifications.

3.2. *Interdependent Logic for Determining 'Back-taper'*

To obtain the taper of an object, provided it is a plain conical surface, it is possible to take two dimensional measurements simultaneously and by electronic means very simply to compute the difference and thus obtain the angle of taper of the cone, but without displaying actually either dimension. Only the difference is displayed and can be calibrated as an angle or as a tangent slope. However this is not the same problem as computing the interdependence of three 'peak-diameters' of a fluted object with one dimension T the peak tip diameter taking precedence and which in turn is compared with an absolute reference. The method used in a practical drill gauging machine is to derive from analogue information from the peak tip diameter not only its own discrimination signal for the cutting diameter, but also a comparison signal which is fed into specially-designed bridge-discriminator circuits relating both the F peak flute diameter and the S diameter and which compare these with the T peak-tip-diameter (ref. 3). Thus the comparison signal CS (Fig. 5) derived from peak tip information provides automatic limit shifting for both tolerances F and S as T varies from drill to drill, each piece of information relative to a particular drill being stored in the appropriate memories and read out at the correct moment when the drill has reached the decision point in the transfer-line before the sorting gates. This is quite instantaneous and the final decision initiates the operation of the output sorting gates.

3.2.1. Analogue computation

To accomplish the task set out above the circuit presents the 'tip' dimension to one side of a bridge circuit whilst the 'flute' dimension is presented to the other side of the bridge, in appropriate polarity to be able to deduct the 'flute' analogue from the 'tip' analogue and to divide the whole by a factor proportional to the length apart of the two above measurements along the drill. The same is done for the 'shank' dimension as shown in Fig. 5.

3.2.2. Erasure of information

These computing bridges are operated by buffer stages keyed to the same clock-pulse generator which operates in step with the mechanical part of the machine; suitable cancelling pulses are introduced into the memory devices as well as the trigger devices which are reset automatically to a clear state (i.e. an untriggered state), in the 'gate-logic' system.

3.3. *Choice of Computing System*

It is perhaps appropriate to comment on the reasons that led to the use of an analogue computing system instead of a digital one, which some designers might have favoured. The first reason is that to use a digital system requires an analogue to digital converter from the transducer and then the reverse process before using the information in an appropriate manner. In the opinion of the author, it seems wrong to take an analogue quantity derived from the transducers and turn it into digital form and back again for sorting purposes. Obviously to do a number of such computations almost simultaneously and bearing the cost of analogue-to-digital conversion equipment in mind, one would obviously try and design the equipment to contain only one such converter. In view of the fact that for each decision many computations are required the electronic switching arrangement itself would add further complications. It is doubtful whether this would be the right approach. There is also degradation of information at each successive change with loss of overall accuracy. It is for these reasons that the far simpler purely analogue system of computation has been chosen. In practice, this concept has proved its value.

3.4. *Measuring Heads (Carrier frequency displacement types)*

The practical circuit is in the form shown in the block diagram (Fig. 5).

The drills are rolled from stage to stage by a transfer mechanism of special design so that the perception-measuring transducers can measure each peak-diameter at least once. With a given length of measuring anvil and parallel motion measuring shoes, which actuate the electro-magnetic balanced two-coil transformers in turn, the smallest sized drills will roll several revolutions while rolling forward, while the largest sized drill in the range will be measured at least once. The parallelism of the measuring shoes has to be better than 1/100th of a thousandth of an inch, and the finish of the measuring anvil and shoe is in the order of a few microinches. All relevant metal parts are made from an alloy having good temperature stability and all rolling surfaces are of ground and lapped sintered tungsten-carbide.

3.5. *Duration of Peak Diameter Measurement*

In rolling a drill the maximum diameter at any point measured is only under the measuring shoe for the time that the largest diameter is in a generally vertical line. At a machine rhythm of one drill per second, the peak diameter measurable depends on the overall diameter; and in practical cases with a 1·5 mm nominal diameter drill, this is about 2 milliseconds. With a 10 mm nominal diameter drill it is measurable for about 15 milliseconds.

To enable a balanced transformer transducer faithfully to reproduce such short duration measuring shoe displacements, a carrier frequency of 10 kc/s was chosen. With this carrier frequency there are not less than twenty actual cycles within the peak-diameter. The subsequent circuits are all made capable of accepting a rise-time of under one millisecond. All the carrier-frequency equipment is push-pull to have short and long-term stability. The oscillation generator is amplitude stabilized, and demodulation is by means of a ring demodulator. The accuracy of all this equipment is better than 0·5% of the whole measuring range which in the practical case amounts to an accuracy of 5 microinches (or about 1/8th of a micron).

3.6. *The Amplifiers and Memory Input Circuits*

The gain of the total pre-memory amplifiers is such that a full-scale range of 0·001 in (about 25 microns) is amplified to 20 V. There is extensive use of negative feedback both at 10 kc/s and d.c. Various forms of amplitude clippers and other stabilizing means are employed. The drift stability obtained is the order of 0·1%.

3.7. *Peak-reading Memory Input and Analogue Memories*

Input from a cathode follower is by silicon diodes with a conductivity ratio of 1 to 10^6. The capacitance storage memories are of polystyrene capacitors, having an insulation resistance of 10^{13} ohms, i.e. much better than required for the few seconds storage while the drills roll through the system. In practical cases even if the machine is stopped during operation, the stores will not lose the analogue voltage to which they are charged to any significant degree for several minutes.

As there are several stores for each stage (three for each), high insulation switching circuits are used running at an exchange frequency of three times less than the machine rhythm, thus every store in the tip channel is charged with one piece of diameter information every three drills. When the first drill reaches the third stage of the machine the stores are switched with another set of high insulation relays to the read-out stage.

3.8. *Read-out from Stores*

This is done by a circulating relay system bringing each capacitor to the read-out circuit in turn. The read-out amplifier is push-pull throughout with negative feedback of about 10 000 times resulting in a unity gain system. The input resistance of this circuit is of the order of 10^{12} ohms. It is followed by further emitter followers to provide adequate buffering before the discriminators.

3.9. *Processing and Comparison of Analogue-remembered Information*

It is at this point where the three systems are brought to bear on each other. The individual channels of remembered information are compared in six trigger discriminator circuits, forming the decision system as a whole.

Two of these bridges compare the actual read-out analogue drill tip diameter value (0–20 V d.c. for a total range of 0·001 in dimension or 25 microns) with a stabilized reference voltage and the trigger discriminators actuate suitable hold circuits with three kinds of information. If both triggers are 'off' it indicates a drill whose tip diameter is below the lower acceptance limit. If one is triggered it lights a green lamp showing the value to be within the acceptance limits for the drill. If both triggers are fired the drill tip is above the acceptance limit and a yellow lamp lights. This information is available for the 'gate logic' as hold signals.

Two further bridge discriminators compare the actual 'tip' information for the drill in question with the read out 'flute' peak diameter and act in a similar manner. It is here that we see 'interdependent logic' in action. The trigger levels for the 'flute' are actually shifted by the actual 'peak tip diameter'.

Two further bridge discriminators compare the actual 'tip' information for the drill in question with the read-out 'shank' diameter and act in a similar manner. Here again we see that 'interdependent logic' is used, actually shifting the limits for the 'shank' trigger levels dependent on the appropriate 'peak tip diameter' information.

To this end the 'tip' information is remembered and used as a positive voltage while the 'flute' and 'shank' information is remembered and used as a negative voltage. They are compared in a d.c. bridge.

3.10. *Sorting Gate Control Logic*

From the previous section we have seen that we can get nine kinds of answers from the six trigger circuits. But for the purposes of 'gate-actuation' these are sorted out into three groups as follows:

Acceptance. If *all systems* show green lights, then a 'good' gate hold circuit is energized;

Large rejects. If any system shows a yellow light, then at that measurement the drill is too large or the taper is too large or too small, and the 'large-reject-gate hold circuit' is energized. (This can occur if the drill is barrel-shaped.)

Small rejects. If *any system* shows no lights, then the duct leading to the small reject bin is left open. This occurs when either the drill is too small at any of the measurements in question or the taper is too small or too large, or the drill is hollow ground.

3.11. *Acceptance of Slightly-bent Drills*

As drills are actually made of ground wire, which at some stage of its manufacture is coiled up, even after much corrective straightening treatment a small bow of perhaps 0·001 mm along its total length could occur. This does not matter because if the drill is correctly sharpened at the point of tip, it will still drill a straight hole. The mechanical features of the gauging block in which the three sequential measuring anvils are located with appropriate rails for rolling are so designed that the drill rolls only along two surfaces at any one time. Thus the measuring arrangements deliberately ignore any small bowed drill. An excess bow and excess diameter reject mechanism is incorporated to safeguard the measuring heads.

3.12. *Performance Achieved in the Drill Gauging Machine*

3.12.1. Stability

In the actual machine the problem of stability has been tackled with great care, and apart from the usual warming-up period of about 15 minutes after switching on, the stability is exceptionally good. This has been achieved by the logical choice of high stability materials and components with good temperature as well as ageing qualities and also by using bridge circuits wherever appropriate and possible.

3.12.2. Speed

With drills in the size range up to about 5 mm diameter it has been found convenient to operate the machine at an inspection rhythm of 60 drills per minute; for three measurements per drill this is 180 measurements per minute.

3.12.3. Sensitivity and range

The total range of a practical drill inspection machine is about 6 mm, but varies with the type of machine, the range being less for the smaller-sized type of machine and more for the larger machines (up to 13 mm).

The sensitivity of a practical machine is such that its full scale per setting is about 0·025 mm (a total of 0·001 in) and as there are 100 divisions on the scale and it can discriminate for decision purposes $\pm$ 1 division, this represents a sensitivity of $\pm \frac{1}{4}$ micron (or 0·00001 in). However, due to minute residual bounce phenomenon when the parallel measuring surfaces diverge at the instant of a drill rolling in between them and its flutes hitting the measuring shoes, there remains a resultant uncertainty of $\pm \frac{1}{2}$ micron (or about $\pm$ 0·00002 in). Thus one could state that the usable sensitivity, if required, could be better than $\pm$ 1 micron. However this resultant sensitivity is well above that required to discriminate drills to meet the I.S.O. specification, and the practical effective sensitivity is customarily stated to be between 1 and 2 microns (in inches this is about $\pm$ 0·00004 in and 0·00008 in).

3.12.4. Durability and reliability

Such drill gauging machines, as described above, have been developed and have been in use in a large twist drill manufacturing factory, having carried out many millions of measurements to date. There seems no reason why with good maintenance these machines should not last indefinitely.

The causes of failure, although very few, have been mainly mechanical and consisted in the main of wear of the mechanical surfaces, due to microscopic pieces of abrasive which are present in a factory using a large number of grinding operations. These particles are in the micron range of size and are very difficult to eliminate. However, one should not exaggerate the incidence of failure from any cause. The reliability is of the same order as that of many digital computers with attached mechanical print-out facilities. In other words we have reached a really practical stage in reliability.

4. Automatic Inspection involving 'Density'

One can of course design for more complex shaped objects, for instance, aero-foil shapes and other shapes obeying a mathematical law, as used in turbine blades and the like, but obviously these require many more transducers than the above machine described. Here again it is considered that interdependent measurement logic would give a better indication of gas flow surface shape than does working to discrete tolerances at all points measured.

A very interesting application of the sequential auto-computing philosophy is in the design of another machine which inspects objects not only for measurable dimensions, but also for a more abstract property, such as for instance 'density'.

Some years ago, the author was presented with a problem of inspecting tablet-shaped pellets moulded at the rate of about 200 pellets per minute. These have to be monitored for diameter (d), height (h) and also density (D). It is obvious that $\pi d^2 h/4 = V$, and that $D = W/V$ where V = volume and W = weight of the object. Thus such a problem can be solved by using a transfer line type automatic inspection system coupled directly to a tablet production machine. The transfer line for the inspection system is designed to measure height and diameter in succession, followed by a high speed tablet weighing stage.

It is obvious that in this machine the measurement of diameter was, comparatively, an easy task as, unlike the problem in the drill gauging machine referred to above, a tablet is a complete cylinder. But the design of the weighing equipment to work accurately in a few milliseconds is not so easy. However, by equipping these various measuring and weighing stages with memory facilities to hold the measured information from each stage for each tablet until the particular tablet reaches the decision point on the transfer-line and also by interdependent computing facilities at the decision point one can obtain automatically the ratio W/V very quickly. Thus the 'decision' device can be made to operate the 'pass/reject gates' when either dimension is too large or too small, or the ratio of W/V is incorrect. Further details of this concept are given in ref. 4.

5. Cost of Automatic Inspection Equipment

It is obvious that automatic inspection equipment is comparatively expensive, the cost depending entirely on the complexity of the inspection task involved. But it would be wrong to generalize and to say that all automatic inspection equipment is expensive.

There are a large number of very simple inspection devices which cost less than £1 per device. For instance, the simple 'thread-break-monitor' as used on automatic knitting and automatic weaving machines each on one thread. These are simple 'yes/no' switches, kept 'open' while all threads are present. In principle they still fall within this general inspection category as even a thousand monitors can be arranged with a stop/hold circuit to work when one or more are 'closed'. It is also possible in rather more complex textile machines to combine these simple devices in an interdependent system, if required. An example would be on a fully fashioned garment knitting machine.

At the other end of the cost scale are the electronic direct reading spectrographs which give a complete typed certificate of the metallurgical analysis of a sample within a matter of minutes. These instruments have the greatest possible importance in metallurgical research and development of very complex alloy steels etc. It is possible, on the other hand, to generalize on cost in this sense: the cost of the equipment is always recovered by very considerable savings elsewhere in the production processes or by a reduction of scrap.

6. General Organizational Considerations relating to Inspection Systems

6.1. *Preliminary Technical Study*

Before embarking on a development project for an automatic inspection system it is necessary to make a very careful study of all the factors concerned with the production of the particular components concerned. It is wrong to think of the automatic inspection device as merely doing something more quickly than human beings can do. The availability of an automatic inspection device in the right place in an automatic transfer line, or other production system, might make a whole series of subsequent operations, such as re-working, unnecessary. The indirect savings can in many cases far outweigh the direct savings on labour.

6.2. *Development Costs*

Obviously in the intelligent assessment an estimate of the probable cost of an inspection machine one must also include the cost of eliminating 'teething troubles' with the new equipment. To keep these to the minimum, it is a good thing to carry out the development work stage by stage, and to use and combine as many known techniques as possible.

6.3. *Cost Accountancy*

One has to endeavour, in co-operation with production personnel and planning staff working under the overall management, to arrive at an economic balance, after the feasibility of the proposals have been accepted. To mention just a few points—it is frequently very difficult to know in a factory what is the 'cost of quality' by the older methods. Also it is difficult to know what is the cost of the real 'reject-percentage'. It is equally difficult to know what are the hidden costs due to troubles which arise when quality is only controlled by sampling techniques. How much is the cost of reworking and stoppages on production lines elsewhere due to unknown quality factors?

Further, the objective facts discovered in the preliminary survey must be assessed against the economic implications so far as the entire firm is concerned. To this end one must have collaboration of the works cost accountants who have sufficient production experience to enable them to be helpful.

Managements will also wish to work out forward costings based on different rates of market growth, i.e. production growth and so on.

6.4. *Long Term Consideration*

Having obtained sufficient facts management is enabled to act sufficiently quickly to adopt a short as well as a long term development plan in regard to automatic inspection equipment. This will fit logically into the general re-equipment plan of the works in general. For what is the use of a most elaborate and clever inspection equipment if other parts of the factory are equipped with obsolete machinery? Having obtained the development sanction from the Board of Directors the technologists responsible for the development of the actual machines must always bear the long-term growth potential of the company in mind, and thus plan the technical equipment to handle more products than are at present going through the works. This involves operational research studies as well as technical studies.

6.5. *Good Maintenance*

It is just as important to train absolutely first class maintenance engineers (competent in electronic and mechanical aspects of the inspection machines) as to develop the equipment in a satisfactory manner. This training programme can usually be satisfactorily accomplished by careful selection of the right type of man already in the works and given the opportunity and adequate facilities.

7. Future Developments

The author's observations lead to the conclusion that there will be considerable development of all types of automatic inspection devices as more and more productive automation is used in industry. This applies particularly to simple 'yes/no' devices such as the simple thread break monitor mentioned in Section 5—literally millions of these simple devices will be used in modern textile machinery. Such devices could well be used in other fields, for instance on coil winders for the electrical industry, on fine filament handling equipment in the electric lamp and similar industries, and in industries using rope and wire drawing equipment.

The very high precision measurement type of inspection machines will also have an increasingly wide use in the precision mechanical industries such as in the production of ball and roller bearings, watchmaking industries, screws and nuts manufacture and the tool industries for making drills, reamers, bushes, cutters and the like, ground steel stock, both round and flat, etc.

The more abstract types of inspection machines such as are described in Sections 3 and 4 will obviously also have an ever-increasing importance as they would liberate the highly trained personnel so frequently employed on routine laboratory tests for sampling product quality. These people could be more aptly employed in the laboratories on direct research work.

Lastly, mention should be made of the automatic performance and behaviour testing devices such as automatic 'climatic' and 'electrical-loading' testing machines (e.g. ACTE, an automatic components test equipment[5]) and the automatic component testing and assembly techniques.[6]

8. Conclusions

Using the methods described above, automatic inspection machines of high accuracy, stability and reliability have been achieved. These machines are obviously the children of two totally different disciplines, electronic and mechanical engineering working in combination. It is to be hoped that a sufficient number of people will be trained in the future in both these arts so that the many problems requiring to be tackled in industry can be more easily solved.

9. Acknowledgments

The author acknowledges with thanks the permission of the Sheffield Twist Drill & Steel Co. Ltd. to publish the results achieved on automatic gauging of twist drills.

The kind permission by the Ministry of Supply to refer to the work done for various departments on other inspection machines and devices is also acknowledged.

10. References

1. J. A. Sargrove and P. Huggins, "Automatic inspection (the anatomy of conscious machines)", *J. Instn Prod. Engrs*, 34, No. 9, pp. 563–74, September 1955.
2. J. A. Sargrove, "Automatic inspection—cybernetic machines", *J. Brit.I.R.E.*, 24, No. 3, pp. 241–9, September 1962.
3. J. A. Sargrove and Sheffield Twist Drill and Steel Co. Ltd. British Patent Application 14898/61.
4. J. A. Sargrove and D. L. Johnston, "Automatic inspection as the key control element in full automation", *J. Brit.I.R.E.*, 17, No. 10, pp. 529–35, October 1957.
5. J. A. Sargrove, "Automation in component testing", *Proceedings of International Symposium on Electronic Components*, R.R.E. Malvern, Paper 22, pp. 127–36, September 1957.
6. D. Hinchcliffe, J. R. W. Smith and C. H. King, "Automatic component assembly in the telephone industry", *J. Brit.I.R.E.*, 24, No. 3, pp. 199–208, September 1962.

Manuscript first received by the Institution on 6th February 1963 and in final form on 11th November 1963. (*Paper No. 903.*)

DISCUSSION ON

"An Engineering Approach to the Design of Transistor Feedback Amplifiers"†

Mr. G. May (*Communicated*): There is an unfortunate omission in this extremely useful paper, namely a method of arriving at the value of r_C. The footnote on page 129 specifically warns one that r_C is not the r_c of the equivalent T circuit.

The author does mention at the foot of page 130 that r_C is unlikely to be less than 20 kΩ but he does not give any hint as to how the value is derived in the general case.

The value of r_C is required in order satisfactorily to assess expressions (6*f*), (7), (8a) and (8b).

Perhaps Dr. Cherry could be asked to supply the missing data.

Dr. E. M. Cherry (*in reply*): I am pleased that Mr. May has provided me with this opportunity to comment on the use of the hybrid π equivalent circuit in my paper.[1] Considerable confusion surrounds the values of the hybrid π parameters, as relatively few manufacturers give them. At the outset, it is worth emphasizing that an accurate knowledge of the values of r_B, r_C, r_E and β_N is not necessary for circuit design using the principles outlined in the paper. In particular, r_B occurs only as a very small correction whilst r_C does not occur at all. Nevertheless, it is worth giving the formulæ necessary for finding the values of all parameters from published data in other systems

The hybrid π circuit was used in the paper as, in the writer's opinion, it is by far the most useful equivalent circuit of a transistor. To be sure, it is more complicated in its exact form shown in Fig. A than many other low-frequency equivalent circuits. However, both r_{ce} and $r_{cb'}$ can be omitted in almost all practical cases, and the resulting simple circuit is accurate to 1 or 2%. Further, the hybrid π circuit forms the basis of many useful high frequency equivalent circuits. It can be derived either from device physics[1] or charge control theory,[2] and is one of the most fundamental equivalent circuits in use. Both methods of derivation show that there are relations between the elements, so that three parameters plus a knowledge of the quiescent operating point completely characterize a transistor. The three parameters chosen in the paper were $\beta_N = \alpha_N/(1-\alpha_N)$, r_B and r_C, together with a parameter r_E which depends only on the quiescent point and is identically the same for all transistors:

$$r_E = \frac{kT}{qI_E} \qquad \ldots\ldots(1)$$

k is Boltzmann's constant ($1{\cdot}38 \times 10^{-23}$ joule/°K)

q is the electronic charge ($1{\cdot}60 \times 10^{-19}$ coulomb)

T is the absolute temperature

I_E is the d.c. emitter current.

The notation r_B, r_C and r_E with single subscripts was used for economy of print, and capital subscripts were used to distinguish the parameters from r_b, r_c and r_e which are commonly used in equivalent T circuit data. Comparison of Figs. A and B shows the inter-relations between the elements.

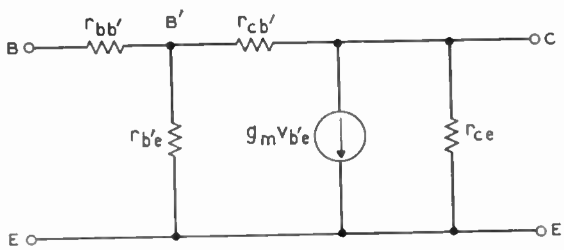


Fig. A. The hybrid π equivalent circuit: common notation.

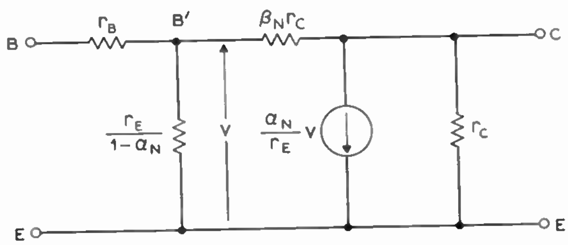


Fig. B. The relations between the elements of the hybrid π equivalent circuit.

An important feature of the hybrid π equivalent circuit is that its parameters are simply related to device physics and geometry. It follows that with the exception of β_N, its parameters are reproducible and predictable and with quite a high order of accuracy. Certainly, they are more reproducible than the mutual conductance g_m of a vacuum tube. It further follows that they have simple, known laws of variation with temperature and quiescent point. If their values are known for a particular transistor under any reference set of conditions, it is easy to find their values for any other temperature and quiescent point. Specifically, r_E is defined by eqn. (1) above, and is known identically

† E. M. Cherry, "An engineering approach to the design of transistor feedback amplifiers", *The Radio and Electronic Engineer*, (*J.Brit.I.R.E.*), **25**, p. 127, February 1963.

for all transistors. The value of r_B depends on a geometrical factor and on the resistivity of the base material; there is no simple expression for its value, but its variation with temperature follows the known variation law for resistivity given in Table A. The value of r_C is given by

$$\frac{1}{r_C} \simeq \left(\frac{\partial I_C}{\partial V_{CE}}\right)_{V_{B'E}} = -\left(\frac{\partial I_C}{\partial b}\right)_{V_{B'E}} \left(\frac{\mathrm{d}b}{\mathrm{d}V_{CB'}}\right) \quad \text{......(2)}$$

For a uniform base transistor, this may be evaluated as

$$r_C = \frac{b}{I_E}\sqrt{\frac{2V_{CE}}{\varepsilon\rho\mu}} \quad \text{......(3)}$$

V_{CE} is the d.c. collector-to-emitter voltage

b is the base width

ε is the permittivity of germanium (or silicon), $1{\cdot}4 \times 10^{-12}$ farad/cm

ρ is the resistivity of the base material

μ is the minority carrier mobility in the base.

The variation of β_N follows most conveniently from

$$\beta_N = \frac{\tau}{\tau_1} \quad \text{......(4)}$$

where τ is the life time of minority carriers in the base, and τ_1 is the mean time taken by a minority carrier in crossing the base. The transit time τ_1 is reproducible, and is given for a uniform base transistor by

$$\tau_1 = \frac{b^2}{2D} \quad \text{......(5)}$$

D is the diffusion constant for minority carriers in the base. Unfortunately, the life time τ depends on a number of factors; it is not reproducible, and varies unpredictably with temperature and minority carrier concentration (i.e. quiescent point). Life time increases with increase of temperature, and empirically β_N varies roughly as the square of absolute temperature. In contrast, the variation of τ with minority carrier concentration is quite small. Typically, a ten-to-one change in emitter current produces less than a two-to-one change in β_N, and to a reasonable approximation, β_N is independent of quiescent point.[3] These variation laws for r_B, r_C, r_E and β_N are summarized in Table A, and typical values for the parameters of various types of transistor are given. It is most unlikely that the parameters of a transistor will differ by more than a factor of two from these typical values.

There remains, then, the problem of finding the parameters r_B, r_C, r_E and β_N at any one quiescent point and temperature from data supplied by the manufacturers. Table B gives conversion formulæ for the three most common data systems, h parameters, T parameters, and g or y parameters. These conversion formulae give the hybrid π parameters at the quiescent point and temperature of the original data. Notice that some manufacturers give some parameters a negative sign; all quantities substituted into Table B should be positive.

There are four possible sources of error in Table B. In decreasing order of importance, these are the following:

(i) Strictly, Table B gives the hybrid π parameters of a particular transistor in terms of its h, T or g parameters. The data supplied by manufacturers are usually the averages of the parameters of a number of transistors, rather than

Table A

Variation of Hybrid π Parameters with Quiescent Point and Temperature

Parameter	Variation with Quiescent Point		Variation with Absolute Temperature	Typical Value at I_E = 1 mA, V_{CE} = 6 V, T = 16° C		
	I_E	V_{CE}		Low Power Audio Transistor $b = 4\times10^{-3}$ cm $\rho = 2\Omega$ cm	Uniform Base H.F. Transistor $b = 10^{-3}$ cm $\rho = 0{\cdot}5\Omega$ cm	Graded Base H.F. Transistor
r_B	—	—	$\propto T^2$	300 Ω	100 Ω	50 Ω
r_C	$\propto I_E^{-1}$	$\propto\sqrt{V_{CE}}$	—	250 kΩ	100 kΩ	> 500 kΩ
r_E	$\propto I_E^{-1}$	—	$\propto T$	25 Ω	25 Ω	25 Ω
β_N	—	—	$\propto T^2$	20–200	35–350	20–500

Table B

Values of Hybrid π Parameters in Terms of other Common Data Systems

Hybrid π parameter	h parameters	T parameters	g or y parameters
r_B	$h_{ie}-h_{fe}\left(\frac{kT}{qI_E}\right)$	$r_b-\frac{1}{2}\frac{\alpha}{1-\alpha}\left(\frac{kT}{qI_E}\right)$	$\frac{1}{g_{11}}-\frac{g_{21}}{g_{11}}\left(\frac{kT}{qI_E}\right)$
r_C	$\frac{2}{h_{oe}}$	$2r_c\left(\frac{1-\alpha}{\alpha}\right)$	approximately $\frac{1}{g_{22}}$
r_E	Identically $\left(\frac{kT}{qI_E}\right)$ for all transistors. $\frac{kT}{q}=\begin{cases}25\text{ mV at }16^\circ\text{ C}\\26\text{ mV at }28^\circ\text{ C}\\27\text{ mV at }40^\circ\text{ C}\end{cases}$		
β_N	h_{fe}	$\frac{\alpha}{1-\alpha}$	$\frac{g_{21}}{g_{11}}$

the parameters of an average transistor. It is possible for errors to occur in the expression for r_B due to inconsistencies introduced in this averaging process; r_B is obtained by subtracting two nearly-equal quantities, and a small error in either will produce a gross error in r_B. The typical values in Table A can be used as a check. Fortunately, manufacturers often give the value of r_B as well as the main data in h, T or g parameters; a variety of notation is used—base spreading resistance, ohmic base resistance, $r_{b'}$ and $r_{bb'}$ are the most common.

(ii) There is a small resistance in series with the emitter lead of a transistor as shown in Fig. C. This emitter spreading resistance $r_{E'}$ is similar in nature to the base spreading resistance r_B, being due to the finite resistivity of the emitter region. However, the emitter of a transistor is much more highly doped than the base, and $r_{E'}$ is less than r_B by an order of magnitude. The effect of $r_{E'}$ is equivalent to a reduction in the d.c. emitter current I_E; it may be shown that

$$I_{E(\text{effective})}=\frac{I_{E(\text{true})}}{1+r_{E'}(q/kT)I_{E(\text{true})}} \qquad \ldots\ldots(6)$$

Typically, $r_{E'}$ is less than 1 ohm, and may be neglected until I_E is 10 mA or more. If $r_{E'}$ is significant, Table B will give an absurdly high value for r_B.

(iii) The collector-base junction of a transistor may have an appreciable leakage resistance, particularly if its surface is poor. As shown in Fig. C, this leakage resistance is in shunt with $r_{cb'}$, and reduces its value below $\beta_N r_C$. If this leakage resistance is significant, Table B will yield an effective value for r_C which is a reasonable compromise between accuracy and simplicity in the equivalent circuit. It will lie below the typical values given in Table A.

(iv) At very high and very low emitter currents, the mutual conductance generator g_m departs from α_N/r_E, but its value lies in the range

$$\frac{\alpha_N}{r_E}\geq g_m\geq\frac{\alpha_N}{2r_E} \qquad \ldots\ldots(7)$$

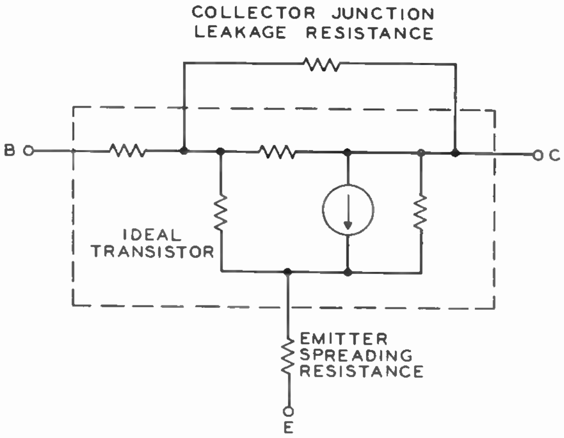


Fig. C. Emitter spreading resistance and collector junction leakage resistance.

At high currents, this effect is due to increased injection of majority carriers back from the base into the emitter[3]; its onset appears to lie well above any normal operating current. At low currents the effect is due to trapping of carriers in the emitter depletion layer[4]; it is more significant for silicon than germanium transistors, but even with a silicon transistor, it is unlikely to appear until the emitter current is reduced well below 100 μA. In all cases, the reduction of g_m is heralded by a marked reduction in β_N.

This mention of errors should not be taken as a limitation of the hybrid π equivalent circuit. Effects (i) and (ii) above are a nuisance when r_B is to be calculated from published data, but as emphasized in the introduction to this note, the exact value of r_B is inconsequential in circuit design. Effects (iii) and (iv) appear more of academic interest than practical importance; the writer has never encountered a design difficulty which can be ascribed to either.

References

1. L. J. Giacoletto, "Study of p–n–p alloy junction transistor from d.c. through medium frequencies", *R.C.A. Review*, **15**, p. 506, December 1954.
2. R. D. Middlebrook, "A modern approach to semiconductor and vacuum device theory", *Proc. Instn Elect. Engrs*, **106B**, Suppl. 17, p. 887, 1959.
3. W. M. Webster, "On the variation of junction-transistor current amplification with emitter current", *Proc. Inst. Radio Engrs*, **42**, p. 914, June 1954.
4. C. T. Sah, R. N. Noyce and W. Shockley, "Carrier generation and recombination in p–n junctions, and p–n junction characteristics", *Proc. Inst. Radio Engrs*, **45**, p. 1228, September 1957.

Manuscript received by the Institution on 18th February 1964 (Contribution No. 75.)

Editorial note: Dr. Cherry's original paper is now out of print. A reprint is being produced and an announcement about its availability will be made shortly.

U.D.C. 621.382.3

Transistor Current-switching Circuits

By

Y. N. BAPAT, Ph.D.†

Summary: The paper gives the analysis and d.c. design criteria with experimental results for high-speed current-switching circuits. Some improved forms of the basic current-switching circuits have been suggested. These modified circuits considerably ease the stringent tolerance requirements of the components and the power supplies of the original circuits in addition to reducing the number of components and power supplies used in the basic building block of the circuit. D.c. circuit analysis is presented for the modified single-transistor-type and also for the modified complementary-transistor-type circuits. Design equations and the pyramiding factors for various types of circuits have been discussed. Transient processes in current-switching circuits are considered in detail. It is shown that the major factors limiting the minimum delay per stage obtained is the emitter-base transition capacitance in addition to the rise-time and hence the inherent frequency response of the transistors. A half-adder circuit and a wave-form generating circuit using current-switching mode is described.

1. Introduction

In recent years, the transistor has replaced the vacuum tube for most switching applications and there are various types of high-speed switching circuits.[1] A very useful class of switching circuits known as current-switching circuits or current-mode logic (c.m.l.) circuits has been devised by Yourke.[2] This type of switching circuit, which avoids the saturation region and hence the hole storage phenomenon associated with it, uses low voltage swings and low impedance levels. It fully utilizes the inherent bandwidth of the triode and attains very high switching speeds approaching to the theoretical minimum rise and fall times for a given transistor. A number of modifications[3, 4, 5, 6] have been suggested to simplify the logical operation and circuit techniques and a number of advantages have been described in the literature for this type of circuitry. Investigations have shown that the current-switching circuit possesses the following disadvantages:

(1) Relatively large number of transistors required.

(2) Large number of power supplies required.

(3) Stringent tolerance requirements of the components and the power supplies used in the circuits.

(4) Large power dissipation in transistors.

The purpose of this paper is to point out the limitations of the circuits proposed previously to show that most of the above-mentioned disadvantages can be overcome by suitable circuit modifications; further, a more detailed analysis and design philosophy and the experimental results of the c.m.l. circuits is presented. The last mentioned disadvantage is inherently present in all types of c.m.l. circuits as the saturation region operation has to be avoided. But its associated effects, namely variation of transistor parameters with temperature has little effect on the c.m.l. circuit operation. The only limitation it produces is on the pyramiding factor due to increase in I_{co} as shown below.

† Assistant Professor, Electrical Engineering Department, Indian Institute of Technology, Bombay.

2. Classification and D.C. Circuit Analysis of C.M.L. Switches

The current switching circuits can be broadly classified into two categories as follows:

(1) Circuits using complementary transistor types, i.e. p-n-p and n-p-n.

(2) Circuits using only single transistor type and voltage-translating blocks (such as Zener diodes).

It must be added that the basic philosophy of switching a well-defined current in circuits with emitter-degeneration remains the same in both types. As it will be seen below the first type is more suitable for computer application and also has much simpler circuitry and consumes lower power than the second type. The second type is perhaps advisable in cases when only one type of transistor is available in that frequency range or in some other special circuits[7] using diodes to perform logical operations.

2.1. *Complementary Transistor Circuits*

Such circuits for a p-n-p and a n-p-n transistor are shown in Fig. 1, and we will consider first the p-n-p circuit. Voltage source-resistance combinations $E_1 - R_e$ and $E_1' - R_2$ form current sources. When the input signal is positive, VT1 is switched 'off', VT2 conducts and current of approximately $0{\cdot}5\ I_e$ flows through R_{l1}, the load resistance of VT2 and output 2 swings positive. Meanwhile $0{\cdot}5\ I_e$ current flows

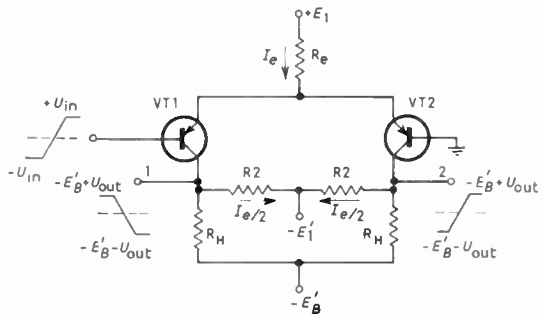
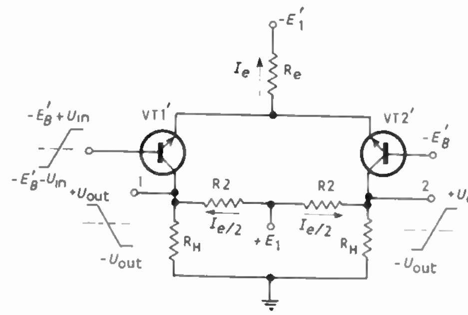


Fig. 1. (*a*) p-n-p current switch.



(*b*) n-p-n current switch.

through R_H of VT1 circuit towards voltage source $-E_1'$ swinging output 1 negative relative to the voltage $-E_B$. When the input signal is negative VT1 is 'on' and VT2 'off' and hence output 1 swings positive and output 2 negative relative to $-E_B'$. The n-p-n switch works in an analogous manner. It may be noted that the output of a p-n-p transistor circuit can be directly connected to the input of an n-p-n switch and vice-versa.

Analysis of the circuits in Fig. 1 gives the following relations:

For a p-n-p switch

$$U^+_{\text{out}} = \frac{R_2 I_e \alpha - (E_1' - E_B')}{R_2 + R_H} R_H \qquad \text{......(1)}$$

$$U^-_{\text{out}} = \frac{E_B - E_1'}{R_2 + R_H} R_H \qquad \text{......(2)}$$

and for a n-p-n switch:

$$U^+_{\text{out}} = \frac{E_1 R_H}{R_2 + R_H} \qquad \text{......(3)}$$

$$U^-_{\text{out}} = \frac{R_2 \alpha . I_e - E_1}{R_2 + R_H} R_H \qquad \text{......(4)}$$

Consider 'd.c. worst case design' conditions for the n-p-n circuit. An external load current equal to $n.I._b$ need be considered only for the negative output, i.e. U^-_{out} from the 'on' transistor. Then we must have the following conditions satisfied:

$$U^+_{\text{out}}(1 \pm \delta U^+_{\text{out}}) \geq U_{\text{in(min)}} \qquad \text{......(5)}$$

and

$$U^-_{\text{out}}(1 \pm \delta U^-_{\text{out}}) \geq U_{\text{in(min)}} \qquad \text{......(6)}$$

where $U_{\text{in(min)}}$ is the minimum input voltage necessary for stable and fast switch-over of current I_e from one triode to the other under all circuit and ambient-conditions variations.

We have from eqns. (3) and (4):

$$\delta U^+_{\text{out}} = \frac{-\delta E_1 R_H + \delta R E_1 R_H}{E_1 R_H} \qquad \text{......(7)}$$

and

$$\delta U^-_{\text{out}} = \frac{-E_1 \delta E_1 R_H + \delta E_1' R_H R_2 \alpha . I_e + 2 R_2 R_H \delta R . I_e . \alpha - R_2 R_H . \delta R n I_b}{R_2 R_H(\alpha . I_e - n I_b) - E_1 R_H} \qquad \text{......(8)}$$

It is assumed that $R_2 \gg R_H$ and that δU_{out} and δR denote the fractional changes in U_{out} and the resistances, and δE_1 denotes the absolute change in E_1. These equations indicate very low tolerance limits allowable for the circuit resistances and the voltage sources. In general, the stability of the circuit can be considered to be poor, for example, let E_1 increase, then U^-_{out} of the n-p-n switch decreases in magnitude and this may affect the pyramiding factor n of the switch and also the switching speed of the n p-n-p loads. At the same time U^+_{out} increases and hence the symmetry of the output wave-forms around the reference signal (earth in case of n-p-n switch) is spoiled.

The various circuit component values can be found as follows:

Choose

$$E_1 = E_B' = E_K \qquad \text{......(9)}$$

and $E_1' = 2E_K$ so as to keep $R_2 \gg R_H$ ......(10)

where E_K is the collector voltage.

Then

$$R_e = \frac{E_1 - U_{eb}}{\alpha . I_e} \qquad \text{......(11)}$$

and

$$R_2 = \frac{R_H}{2}\left(\frac{E_1' - E_B' + U^-_{\text{out}}}{\alpha . I_e . R_H - U^-_{\text{out}}} + \frac{(E_1' - E_B') - U^-_{\text{out}}}{U^-_{\text{out}}}\right) \qquad \text{......(12)}$$

These design equations give a set of typical circuit values as follows: $E_K = 6$ V, $I_e = 6$ mA, $R_H = 240$ ohms, $U_{\text{out}} = 0{\cdot}6$ V, $E_1 = 6$ V, $E_1' = 12$ V, $E_B' = 6$ V, $R_1 = 950$ ohms, $R_2 = 2$ kilohms.

2.2. *Single Transistor Type Circuits*

A p-n-p transistor switch using a reverse-biased Zener diode voltage-translating block to make the

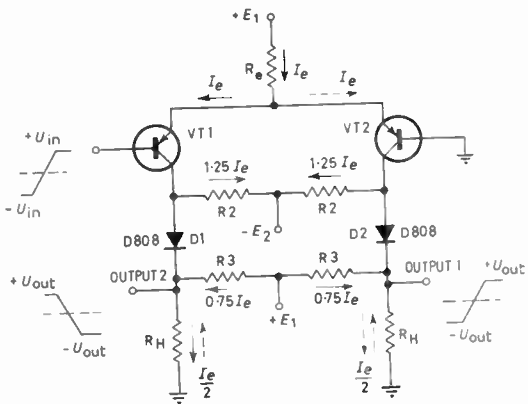


Fig. 2. Single-transistor-type current switch.

output signal identical to the input signal is shown in Fig. 2. The circuit action is as follows:

When the input signal is positive, VT1 is 'off' and VT2 conducts. Consequently, the current generator R_2-E_2 requiring current of value $1{\cdot}25\ I_e$ takes the full emitter current I_e and an additional $0{\cdot}25\ I_e$ from the current source E_1-R_3. The remaining current of value $0{\cdot}5\ I_e$ from the E_1-R_3 source flows through R_H swinging output 1 positive, and also in the collector circuit of VT1 a current $0{\cdot}5\ I_e$ flows through R_H from earth and together with $0{\cdot}75\ I_e$ current from E_1-R_3 source makes the total current $1{\cdot}25\ I_e$ to flow through D1 towards $-E_2$. Hence output 2 swings negative. When the input signal swings negative, relative to earth potential, the triode VT1 is 'on' and VT2 'off'. Output 2 swings positive and output 1 negative relative to earth potential. It may be noted that the signals are identical and the circuit can drive another p-n-p switch.

The component design equations are as follows:

$$R_e = \frac{(E_1-U_{eb})}{I_e},$$

$$R_2 = \frac{(E_2-E_K)+U_{out}}{\alpha . I_e + I_{D(min)}},$$

$$R_3 = \frac{E_1+U_{out}}{\alpha . I_e - I_{D(min)}} \qquad ......(13)$$

together with the condition for good stability of the circuit, namely

$$E_2 \gg E_K \quad \text{and} \quad E_1 \gg +U_{out} \qquad ......(14)$$

$I_{D(min)}$ is the minimum reverse current required to maintain the Zener diode in the breakdown region.

These equations give a typical set of circuit component values as follows: $E_K = 9$ V, $I_e = 6$ mA, taking $I_{D(min)} = 0{\cdot}25 I_e = 1{\cdot}5$ mA, $E_1 = 6$ V, $E_2 = 30$ V, $R_e = 950$ ohms, $R_2 = 2{\cdot}3$ kilohms and $R_3 = 1{\cdot}4$ kilohms.

An exact circuit analysis gives the following results:

$$U^+_{out} = R_H \frac{E_1 R_2 + \alpha_1 I_e R_3 R_2 - (E_2-E_K)R_3}{R_3 R_2 + R_3 R_H + R_2 R_H} \qquad ......(15)$$

$$U^-_{out} = R_H \frac{(E_2-E_K)R_3 - E_1 R_2}{R_3 R_2 + R_3 R_H + R_2 R_H} \qquad ......(16)$$

The 'd.c. worst case design' consideration gives the following conditions to be satisfied for reliable operation of the circuit:

$$U^+_{out}(1 \pm \delta U^+_{out}) \geq U_{in(min)} \qquad ......(17)$$

and

$$U^-_{out}(1 \pm \delta U^-_{out}) \geq U_{in(min)} \qquad ......(18)$$

where

$$\delta U^+_{out} = \frac{\left[2\delta R\left(\frac{E_1}{R_3}+\alpha I_e\right)R_2 R_H R_3 + \delta E_1\left(\frac{E_1}{R_3}+\alpha I_e\right) + (\delta E_2 . E_2 + \delta E_k E_k)R_3 R_H\right]}{E_1 R_2 R_H + \alpha I_e R_3 R_2 R_H - (E_2-E_k)R_3 R_H} \qquad ......(19)$$

and

$$\delta U^-_{out} = \frac{[2\delta R(E_2-E_k)R_H R_3 + R_3 R_H(E_2\delta E_2 - E_k \delta E_k) + E_1 R_2 R_H \delta E_1]}{(E_2-E_k)R_H R_3 - E_1 R_2 R_H} \qquad ......(20)$$

These equations again indicate very low tolerance limits allowed for the circuit resistances and the voltage sources. The equation also shows that for minimum effect of the power-supply fluctuation on the working of the circuit the condition in eqn. (14) must be satisfied.

The stability of this circuit is also poor. Consider that E_2 decreases in magnitude. Then, U^-_{out} will decrease and this will affect the pyramiding factor of the circuit and also the switching speed of the loading

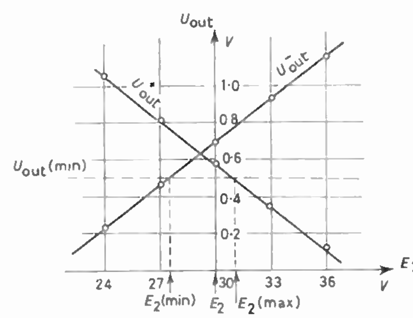


Fig. 3. Variation of output voltage with change in E_2.

circuits. Again, a decrease in E_2 increases U^+_{out}. This again spoils the symmetry of the output signal about the reference potential. The variation of U^+_{out} and U^-_{out} with change in E_2 is shown in Fig. 3. The large variations obtained in U_{out} with voltage fluctuations in E_2 can be seen from these curves.

3. Modified Current Switching Circuits and their D.C. Circuit Analysis

The tolerance specifications of the components and power supplies forming the current generators in the circuits described above are very rigid and also the number of components used in these circuits are large. The major reason for the critical behaviour of these circuits is the accurate current distribution arrangements needed in the collector circuits for achieving a two-level, symmetrical output signal which is a basic requirement for the c.m.l. circuit operation. Certain modifications of these circuits are described below and these modifications very much ease the tolerance requirement and also minimize the complexity of the c.m.l. circuits. The basic technique in these modified circuits is to derive the base reference potential from the collector and emitter voltage sources so that any change in output signal levels due to the changes in these voltage sources is also reflected in the base reference potentials in such a way as to compensate the changes in the output signal levels.

3.1. *Modified Complementary Transistor Circuits*

Such p-n-p and n-p-n switching circuits are shown in Figs. 4(*a*) and 4(*b*). For the p-n-p circuit we have

$$I_e = \frac{E_1 - E_B - U_{eb}}{R_e} \qquad \ldots\ldots(21)$$

$$E_B = E'_K - \frac{E'_K + E'_1}{R_{D2} + R_{D1}} \cdot R_{D2} \qquad \ldots\ldots(22)$$

$$U^+_{out} = \alpha . I_e . R_H \qquad \ldots\ldots(23)$$

and

$$U^-_{out} = 0. \qquad \ldots\ldots(24)$$

For a n-p-n circuit

$$I_e = \frac{E'_B + E'_1 - U_{eb}}{R_e} \qquad \ldots\ldots(25)$$

$$E'_B = \frac{E_1 R_{D2}}{R_{D1} + R_{D2}} \qquad \ldots\ldots(26)$$

$$U^+_{out} = E'_K \qquad \ldots\ldots(27)$$

$$U^-_{out} = E'_K - \tfrac{1}{2}\alpha . I_e . R_H \qquad \ldots\ldots(28)$$

For a given value of I_e and U_K the design of the circuit is carried as follows:

$$E_B = E_K \qquad E'_B = 0{\cdot}5 I_e \alpha . R_H \qquad \ldots\ldots(29)$$

then

$$E'_K = E_K + 0{\cdot}5 I_e \alpha . R_H \qquad \ldots\ldots(30)$$

and for keeping the same value of R_e in both the circuits we must have

$$|E_1| \sim |E'_1| = E_K \qquad \ldots\ldots(31)$$

The resistances R_{D1} and R_{D2} are chosen such that

$$\frac{E_1}{R_{D1} + R_{D2}} \gg \Sigma I_b \qquad \ldots\ldots(32)$$

and

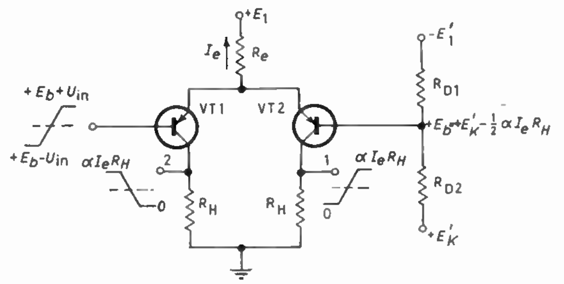
$$\frac{E'_1 + E'_K}{R_{D1} + R_{D2}} \gg \Sigma I_b \qquad \ldots\ldots(33)$$

where ΣI_b represents the sum of the base currents of all the switches which use the common potential-divider network for reference base voltages.

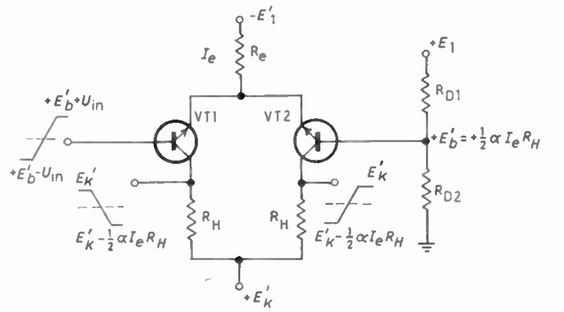
Typical component and voltage source values of such a circuit obtained from the above given equation are as follows:

$E_K = 9$ V, $I_e = 6$ mA, $U_{out} = \pm 0{\cdot}6$ V
$E_1 = 20$ V, $E'_K = 9$ V, $E'_1 = 11$ V,
$E_B = 8{\cdot}4$ V, $E'_B = +0{\cdot}6$ V
$R_e = 2$ kilohms, $R_H = 240$ ohms,
$R_{D1} = 3{\cdot}8$ kilohms, $R_{D2} = 200$ ohms

It is interesting to note that the circuit does not place any restriction on the fluctuations of the power supplies. For example, let E_1 increase, with it increases I_e and also the output voltage U^+_{out}. But as the base reference voltage E'_B of the n-p-n switch also increases with E, variation in E_1 has little effect on the circuit stability and the switching speed. Variations of U^+_{out} with change in E_1 is shown in Fig. 5. The wide margin of allowed variation in E_1 is clearly



(*a*) Modified p-n-p current switch.



(*b*) Modified n-p-n current switch.

Fig. 4.

seen in the figure. It can be shown that variations in other voltage-source values have little effect on the circuit operation.

As noted above a common potential-divider arrangement gives the reference base potential E'_B. Further, this potential-divider arrangement can be common for a large number of identical switches. Thus compared with the circuit in Fig. 1, the circuit is considerably simplified, it is more economical and it allows a larger variation (more than 10%) in its circuit components and voltage source values.

3.2. *Modified Single Transistor Type Circuits*

Such a circuit for a p-n-p type transistor is shown in Fig. 6(*a*). In this circuit the base reference voltage E_{ref} is obtained from a potential divider consisting of a Zener diode and resistors R3 and R4 connected across the voltage source E_2. In this circuit when VT1 conducts a current $I_{L(on)}$ is drawn through the load resistance R_H of VT1, while a relatively larger current $I_{L(off)}$ is drawn through the load resistance of R_H of VT2. The circuit analysis gives the following relations:

$$I_e = \frac{E_1 + E_{ref} - U_{eb}}{R_e} \qquad \ldots\ldots(34)$$

$$I_{L(off)} = \frac{E_2 - E_D}{R_H + R_2} \qquad \ldots\ldots(35)$$

$$U^-_{out} = I_{L(off)} . R_H \qquad \ldots\ldots(36)$$

$$I_{L(on)} = I_{L(off)} - \alpha . I_e \qquad \ldots\ldots(37)$$

$$U^+_{out} = I_{L(on)} . R_H \qquad \ldots\ldots(38)$$

$$E_{ref} = -\frac{E_2 - E_D}{R_3 + R_4} . R_4 = \tfrac{1}{2}(U^+_{out} + U^-_{out}) \ldots\ldots(39)$$

'D.c. worst case' analysis gives the following relations:

$$U^-_{out} = U^-_{out}(1 \pm \delta U^-_{out}) \qquad \ldots\ldots(40)$$

where
$$U^-_{out} = \frac{(E_2 - E_D)(R_H R_3 - R_2 R_4)}{R_H R_3 + R_H R_4 + R_2 R_3 + R_2 R_4} \qquad \ldots\ldots(41)$$

and
$$\delta U^-_{out} = \frac{R_H R_3\{(E_2 - E_D)2\delta R + E_2\delta E_2 + E_D\delta E_D\} + R_2 R_4\{(E_2 - E_D)2\delta R - E_2\delta E_2 - E_D\delta E_D\}}{R_H . R_3 + R_H . R_4 + R_2 . R_3 + R_2 . R_4} \qquad \ldots\ldots(42)$$

Hence to minimize the value of δU^-_{out} we must have

$$E_2 \gg E_D \qquad \ldots\ldots(43)$$

Some conclusion is reached from the 'd.c. worst case' analysis for U^+_{out}. The condition to be satisfied by the circuit for d.c. stable and fast transient operation are

$$U^{+,-}_{out}(1 - \delta U^{+,-}_{out}) - U_{ref} \,|\, \geq |U_{in(min)}| \ldots\ldots(44)$$

where $U_{in(min)}$ is the value of minimum allowable input voltage swing for the proper operation of the switch. The other condition for the proper operation of the breakdown diode is as follows:

$$I_{L(on)} = I_{L(off)} - \alpha I_e \geq I_{D(min)} \qquad \ldots\ldots(45)$$

Equation (45) gives the value of R_2.

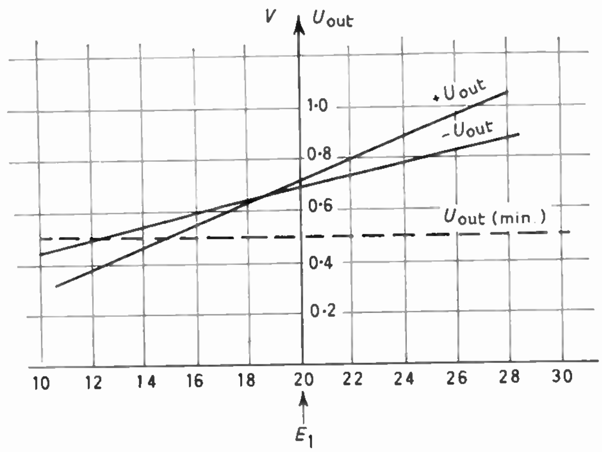


Fig. 5. Variation of output voltage with change in E_1.

The reference voltage can be calculated from eqn. (39). It must be noted that the circuit R3, D3, R4 is common for a number of similar switches and hence the conditions to be satisfied for proper operations are

$$\frac{E_2}{R_3 + R_4} \gg \Sigma I_b \qquad \ldots\ldots(46)$$

and
$$R_3 \gg R_4 \qquad \ldots\ldots(47)$$

From (46) and (47) the value of R_4 and R_3 can be calculated. A typical value of R_4 lies between 50 to 75 ohms.

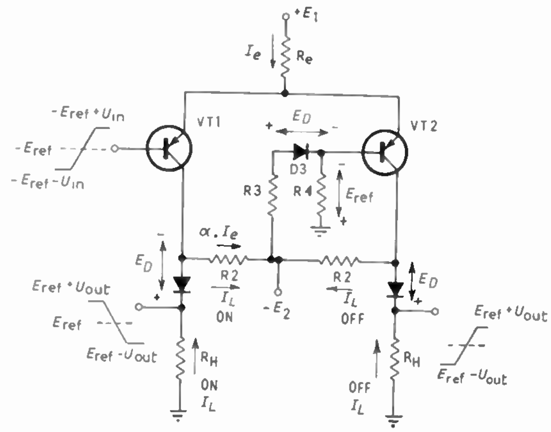
The circuit component and voltage source tolerances are very large (more than 10%) as calculated from eqn. (42). Compared to the circuit in Fig. 2 this circuit is much more simple, stable and economical.

For a given value of I_e = 6 mA and U_K = 9 V, we get the typical values of the components (for the given values of E_D = 9 V and $I_{D(min)}$ = 2 mA for Zener diodes D808) as follows: E_1 = 6 V, E_2 = −30 V, E_e = 1·1 kilohms, R_2 = 2·2 kilohms, R_H = 200 ohms, R_3 = 1·1 kilohms, R_4 = 60 ohms.

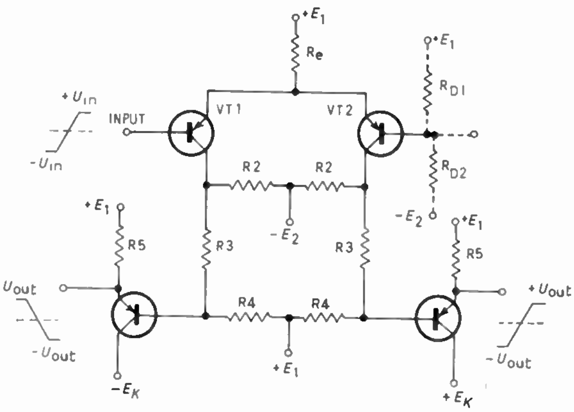
In all the circuits described above the collector currents through the two transistors VT1 and VT2 are not equal. We have

$$I_{K1} - I_{K2} = \alpha_e . I_H . \frac{R_H}{R_e}$$

and also we get the ratio of output voltage to the input



(*a*) Modified single-transistor-type current switch.



(*b*) Single-transistor-type current switch with break-down diodes eliminated.

Fig. 6.

voltage for a c.m.l. circuit as

$$\frac{U_{out}}{U_{in}} = \frac{R_H}{R_e} \qquad \ldots\ldots(48)$$

Both the eqns. (41) and (42) indicate that for proper operation of the switch we must have

$$R_e \gg R_H \qquad \ldots\ldots(49)$$

and this is the relation used for the selection of the resistance R_e. R_e cannot be very high because this would call for a higher value of E_1 to get the same emitter current. At the same time a large value of E_1 avoids the circuit running into saturation region for a given signal input,[8] and hence a compromise value of R_e is selected.

In Fig. 6(*b*) a circuit (suggested by Ryabtsev and Lubovitch of A.H.C.C.P.) is given and this circuit eliminates the use of Zener diodes. This type of circuit may be very useful and economical for the diode current-switching circuits[7] as the emitter-follower output circuits can drive a large number of diode level-setting chains. The base-reference potential may be obtained from a potential divider $R_{d1} - R_{d2}$ common to a large number of identical switches. Typical values of such a circuit are:

$E_1 = 6\,\text{V}$, $E_2 = -10\,\text{V}$,

$E_K = -4\,\text{V}$, $\pm U_{out} = 0{\cdot}8\,\text{V}$,

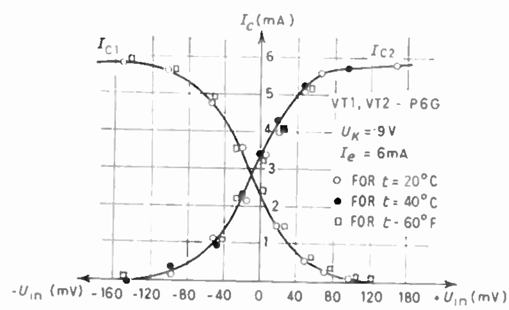
$R_e = 680$ ohms, $R_2 = 430$ ohms,

$R_3 = 570$ ohms, $R_4 = 750$ ohms,

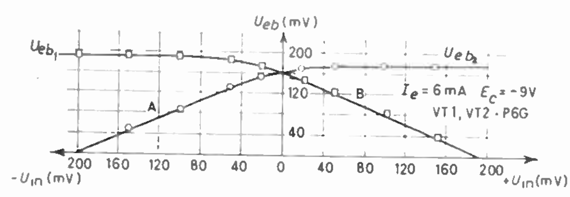
$R_5 = 1$ kilohms.

4. Static Characteristics of a C.M.L. Switch and Choice of Signal Voltage Levels

The static and dynamic behaviour of the c.m.l. switch will now be described and this discussion is common for all the circuits described above. Typical static characteristics for a c.m.l. switch are shown in Fig. 7(*a*) and Fig. 7(*b*). We see that the operation of the circuit can be divided into three states: (1) VT1 'on' and VT2 'off', (2) VT1 'off' and VT2 'on', and (3) transient state in which both VT1 and VT2 are partially conducting. In the first two states the input impedance of the switch is extremely high, but in the transient state the input generator is heavily loaded as will be seen below. It may be also noted from the static characteristics that a value of input voltage $\pm 0{\cdot}1$ V is sufficient to switch the circuit from one state to another. However, at higher temperatures it was shown[1] that for reliable switching of current from one transistor to another a minimum input voltage $U_{in(min)} = \pm 0{\cdot}4$ V is necessary. It will be shown below that nearly the same input voltage is sufficient to have the highest speed of switch-over from one state to the other. A maximum allowed value of the output voltage is obtained as follows: consider a



(*a*) Static characteristics of a current switch.



(*b*) Variation of emitter-base voltage with input voltage.

Fig. 7.

transistor VT1′ connected in parallel with VT1 in a c.m.l. switch. Let VT1 be 'off' and VT1′ conducting. Then the base-emitter voltage of VT1 will be $2\,U_{in} - U_{eb(on)}$ and if the maximum reverse base-emitter voltage for the triode be $U_{eb(max)}$ then we must have

$$2U_{in} \leqslant U_{eb(max)} + U_{eb(on)} \qquad \ldots\ldots(50)$$

For P402, $U_{eb(max)}$ is 1·3 V and U_{eb} is 0·3 V and hence the maximum value of input voltage $\pm U_{in(max)}$ will be $\pm 0{\cdot}8$ V. Therefore, the normal output signal for c.m.l. switches is chosen to be $\pm 0{\cdot}6$ V.

5. Pyramiding Factor of the C.M.L. Switches

Consider that there are $(M-1)$ transistors connected in parallel with the transistor, VT1 and n transistor bases connected as loads at the output terminal of the transistor VT1. Then the pyramiding factor of the two types of circuits can be calculated, the basic requirement being that the output voltage should in no case be less than its minimum value $U_{out(min)}$.

5.1. Single Transistor Type Circuit

Consider that the M input transistors are 'off' and hence the loading n transistors will be 'on'. The condition to be satisfied at the output terminal is as follows:

$$\frac{U_{out} - U_{out(min)}}{R_H} \geq M\,.\,I_{co(max)} + n\frac{I_e}{\alpha_{e(min)}} - nI_{co(max)} \qquad \ldots\ldots(51)$$

Equation (51) gives for $I_e = 6$ mA, $U_{in(max)} = 0{\cdot}4$ V, $R_H = 240$ ohms, $\alpha_{e(min)} = 20$ and $I_{co(max)} = 100 \times 10^{-6}$ A, at a maximum temperature of 60° C, the permitted combinations of n and M being as in Table 1. Combination $M = 4$, $n = 3$ may be chosen as this value of n cannot be exceeded as will be seen from transient behaviour consideration of this circuit.

Table 1

n	1	2	3	3
M	10	7	4	1

5.2. Complementary-transistors Circuit

In this case the value of $I_{co(max)}$ of the n-p-n transistors at the highest temperature allowed for p-n-p transistors is negligibly small. Hence there is no upper limit of M for the n-p-n transistors. The upper limit of M for a p-n-p circuit driving a n-p-n circuit is given by

$$\frac{U_{out} - U_{out(min)}}{R_H} \geq M\,.\,I_{co(max)} \qquad \ldots\ldots(52)$$

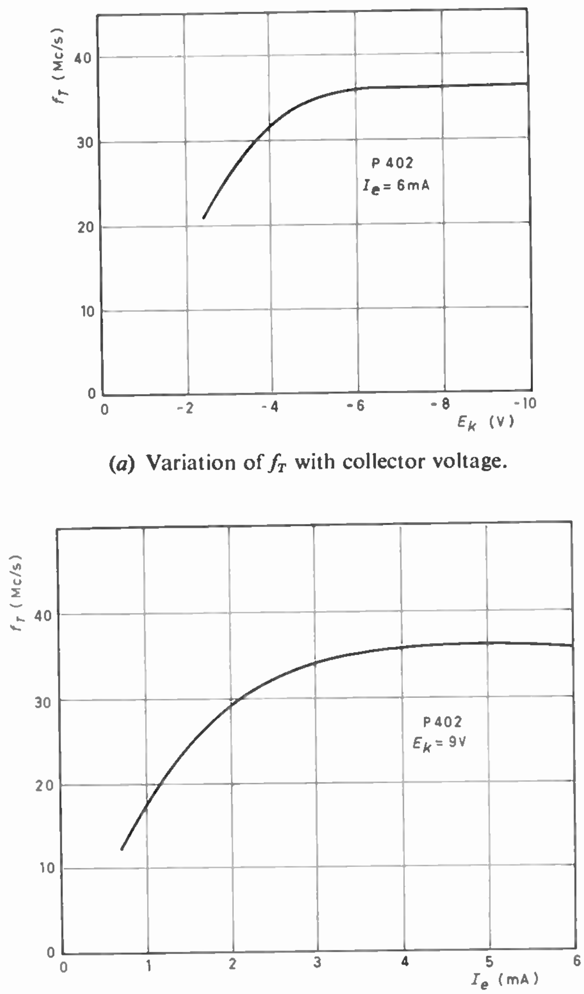
and M turns out to be equal to 8. The maximum limit of n is determined at the room temperature itself by the condition

$$\frac{U_{out} - U_{out(min)}}{R_H} \geq \frac{I_e\,.\,n}{\alpha_{e(min)}} \qquad \ldots\ldots(53)$$

and n comes out to be equal to 3 as for the single-transistor type circuit.

6. Choice of U_k, I_e and R_H

Choice of the collector voltage U_k, emitter current I_e and load resistance R_H is governed by the conflicting requirements of the frequency-transient response considerations and the transistor power-dissipation considerations. To keep the latter at a minimum, I_e and U_k should be kept minimum, but the values of I_e and U_k should be above certain minimum values and these can be found from the variation of f_T with I_e and U_k as shown in Fig. 8(*a*) and 8(*b*). Another advantage of using a high value of U_k is the decreased collector capacitance associated with it. The value of R_H should be large enough to give the required output



(*a*) Variation of f_T with collector voltage.

(*b*) Variation of f_T with emitter current.

Fig. 8.

voltage swing $\pm U_{out}$ and should be small enough from the transient considerations given below. The values chosen may be: $U_K = 6$ to 9 V, $I_e = 4$ to 9 mA and $R_H = 200$ to 300 ohms.

7. Transient Processes in Current-switching Circuits

The transient processes in the c.m.l. circuit as seen on the oscilloscope screen are shown in Fig. 9. In this figure the variation with time of the following quantities is shown:

U_{in} = input voltage at the base terminal of transistor VT1.

i_{b2} = base current of transistor VT2.

U_{eb2} = emitter base voltage of VT2.

$I_{c1,2}$ = collector currents of the transistors VT1 and VT2.

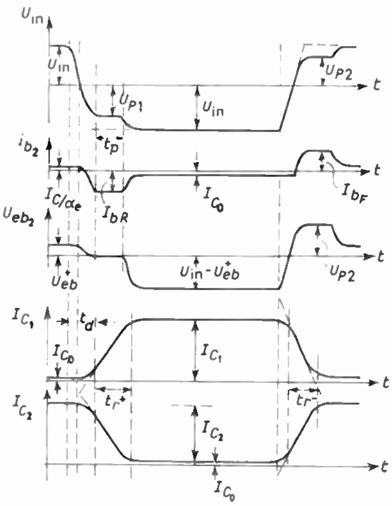


Fig. 9. Transient processes in a current switch.

In the analysis it has been assumed that the rise-time of the input voltage is much smaller than the rise-time of the output voltage. Transient processes may be most conveniently described with the aid of a transistor equivalent circuit shown in Fig. 10.

When the input voltage varies from $+U_{in}$ towards zero volts, VT1 remains 'off' and the depletion-layer capacitance C_{eT1} is charged through the generator resistance R_H and the base resistance $r_{bb'}$. As the input voltage swings towards its negative value $-U_{in}$, VT1

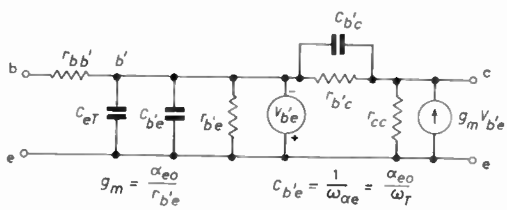


Fig. 10. H.f. transistor equivalent circuit.

conducts at the threshold input voltage $-U_{p1}$, i.e. the capacitance $C_{b'e}$ is charged through the resistance $(R_H + r_{bb'})$. During this process the input circuit is heavily loaded due to the large base current flowing for a time determined by the inherent base-charging speed (i.e. inherent frequency response) of the transistor. This heavy loading is seen as a distinct plateau in U_{in} and U_{eb} waveforms. As soon as the base-charging process in VT1 is over the capacitance C_{eT2} is charged to the final emitter node voltage $U_{in} - U_{eb1}$.

When the input voltage swings from its negative value $-U_{in}$ towards zero, no change occurs in the circuit as the input source sees a high input impedance of an emitter follower; hence the input voltage rises much faster towards the positive threshold value $+U_{p2}$ and consequently the delay time associated with the fall of I_{c1} or the rise of I_{c2} is smaller than the delay associated with rise-time of I_{c1}. As the input swings towards its positive value U_{in}, VT2 conducts and starts the base-charge removal process in VT1. Again during this period the input circuit is heavily loaded. As soon as the base of VT1 is completely free of charge, the input voltage rises towards its final value $+U_{in}$ with the time-constant $C_{eT}(r_{bb'}, +R_H)$.

7.1. *Charge Transfer in the Base Regions*

When the input voltage swings towards its negative value $-U_{in}$, a forward base current flows whose value is determined by the external resistances in the circuit. Now at the emitter mode we have $I_e = I_{e1} + I_{e2}$ at all times. Hence a diversion of forward base current $I_{b(F)}$ in VT1 means that an equal amount of current is reduced in VT2 and the injected hole level now corresponds to $I_e - I_{b(F)}$ at this moment. To keep the neutrality of the base region a number of electrons corresponding to the forward current $I_{b(F)}$ have to rush out of the base lead at this instant and this constitutes a reverse base current $I_{b(R)}$ equal in magnitude to the forward base current in VT1. Thus

$$I_{b(F)} = I_{b(R)} = \frac{U_{in}}{R_H + r_{bb'}} \qquad \text{......(54)}$$

where R_H is the input load resistance and $r_{bb'}$, is the ohmic base resistance of the transistor. Thus the emitter base region of one transistor acts as a low impedance path for the forward and reverse current drives of the other transistor irrespective of the fact that VT1 is star-connected, i.e. has resistances connected in all three of its leads. The forward and reverse base currents reduce the rise and fall times of the collector current in their respective circuits.

7.2. *Charge Transition Time Interval t_p*

The charge variation in the base region can be written as:

$$\frac{dQ}{dt} = I_e - I_k - \frac{Q}{\tau_e}$$

$$= I_b - \frac{Q}{\tau_e} \qquad \ldots\ldots(55)$$

where τ_e is the effective life-time of the minority carriers in the base règion and is equal to $1/\omega\alpha_e$. For steady state condition $dQ/dt = 0$ and charge in the base region is given by

$$Q = I_b . \tau_e = I_e . \tau_T \qquad \ldots\ldots(56)$$

where $\tau_T = 1/\omega_T$. τ_T is also the time of transit required for the emitter pulse to reach the collector.

Now the base charge in VT2 is being removed at the rate

$$Q_r = \int_0^t I_{b(R)}\, dt \qquad \ldots\ldots(57)$$

and the base charge is removed at this rate until the current level falls from I_e to $I_{b(R)}$, i.e. the amount of charge removed will be $(I_e - I_{b(R)})$. The time required to achieve this is

$$t_p = \frac{1}{I_{b(R)}} . \tau_T(I_e - I_{b(R)})$$

$$= \tau_T\left(\frac{I_e}{I_{b(R)}} - 1\right) \qquad \ldots\ldots(58)$$

With the transistors P6G ($\tau_T = 0{\cdot}1 \times 10^{-6}$ seconds) and P402 ($\tau_T = 3 \times 10^{-9}$ seconds) we get, for values of $I_e = 6$ mA and $I_{b(R)} = 2$ mA, $t_p = 0{\cdot}2 \times 10^{-6}$ seconds and 6×10^{-9} seconds respectively, these figures agreeing fairly well with the experimental observations. It may be noted that the time interval t_p is independent of collector circuit parameters. The variations of t_p and hence $I_{b(R)}$ as a function of input voltage U_{in} and emitter current I_e have been verified experimentally.

7.3. *Plateau Voltage U_p*

In the emitter-base circuits of the two transistors the following relation holds good in steady state conditions

$$U_{in} = U_{eb1} + U_{eb2} \qquad \ldots\ldots(59)$$

But in the transient region when the input voltage reaches the value $U_{in} = -2U_{eb}$, the transistor VT1 has across its emitter-base diode junction voltage $+U_{eb}$ and as the current is decreasing in VT2 its emitter-to-base voltage instantly falls to zero but cannot become negative till its base region is completely free of charge, i.e. for the duration t_p. Consequently, the input voltage is locked at the value

$$U_{in} = U_p = -2U_{eb} \qquad \ldots\ldots(60)$$

for the time interval t_p, after which the emitter node can take its final negative value equal to $(-U_{in} + U_{eb1})$ as shown in the figure.

Transition from state II to state I is identical and all the above given reasoning is equally valid.

7.4. *Rise-time and Fall-time of the Collector Currents*

The build-up of the collector current consists of two parts: (1) Initial delay t_d for the collector current to rise up to 10% of its final value and (2) the rise-time t_r for the collector current to rise up to 90% of its final value.

7.4.1. Initial delay t_d

The initial delay time consists of two parts: (i) propagation delay time in the base equal to[9]

$$t_{d1} = 0{\cdot}04/f_T \qquad \ldots\ldots(61)$$

and (ii) the time required to charge the base-emitter barrier capacitance C_{eT} and the associate stray capacitances C_N at the emitter mode junction up to the threshold voltage U_p. This time was calculated graphically[10] and was found to be:

$$t_{d2} = 0{\cdot}9(R_{II} + r_{bb'})(C_{eT} + C_N) \qquad \ldots\ldots(62)$$

Total input capacitance of VT1 was measured and found to be equal to 50 pF for the P402 switch and 200 pF for the P6G switch; hence with $R_{II} = 240$ ohms the delay t_{d2} will be $11{\cdot}25 \times 10^{-9}$ seconds for the P402 switch and about $0{\cdot}045 \times 10^{-6}$ seconds for the P6G switch. These values are confirmed by the experimental observations. It may be noted that the values of t_{d1} for the P402 switch are negligibly small, but assume relatively large values for the P6G switch. The total delay time for the P6G is about $0{\cdot}07 \times 10^{-6}$ seconds and is the sum of the delay times t_{d1} and t_{d2}.

7.4.2. Rise-time of the collector current

Let us consider the transient processes occurring in the transistor VT1. It has resistances in all the three leads and is driven by a generator providing a voltage E_b at the base terminal and having an internal impedance R_{II}. The base current of the triode will be given by

$$I_b = I_b(0) - \beta I_c \qquad \ldots\ldots(63)$$

where $$I_b(0) = \frac{E_b}{[R_b + R_e(1+\alpha_e)]} \qquad \ldots\ldots(64)$$

β is the feedback factor equal to $\dfrac{R_e}{R_e + R_b}$

R_e = total series resistance in the emitter lead

R_b = total series resistance in the base lead and is equal to $R_{II} + r_{bb}'$.

Substituting eqn. (63) in eqn. (55) we get

$$\frac{dQ}{dt} + \left(\frac{1+\alpha_e\beta}{\tau_e}\right)Q - I_b(0) = 0 \qquad \ldots\ldots(65)$$

Solving this equation with the initial condition $Q(0) = 0$ we get:

$$q(t) = I_b(0)\frac{\tau_e}{1+\alpha_e\beta}\left\{1 - \exp\left[\frac{-t(1+\alpha_e\beta)}{\tau_e}\right]\right) \qquad \ldots\ldots(66)$$

The collector current has a finite value αI_e limited by the emitter current source alone and we get the rise-time from (66) as:

$$t_r = \frac{I_e}{1+\alpha_e\beta}\ln\frac{\alpha_e I_b(0)}{\alpha_e I_b(0)-\alpha I_e(1+\beta\alpha_e)} \quad \ldots\ldots(67)$$

and taking into consideration the effect[11] of collector capacitance C_c which is assumed to have a constant-value as the collector voltage swing is small, we get

$$t_r = \left\{\frac{\tau_e}{1+\alpha_e\beta}+(1+\alpha_e)R_{II}C_c\right\}\times$$

$$\times\ln\frac{\alpha_e I_b(0)}{\alpha_e I_b(0)-\alpha I_e(1+\beta\alpha_e)} \quad \ldots\ldots(68)$$

Now for sufficiently large input voltage drive we will have

$$\alpha_e I_b(0) \gg \alpha I_e(1+\beta\alpha_e)$$

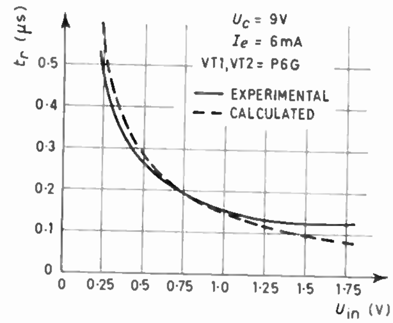
and hence

$$t_r \simeq \frac{\tau_e}{1+\alpha_e\beta}\,\frac{\alpha . I_e(1+\beta\alpha_e)}{\alpha_e . I_b(0)}+\frac{R_{II}C_c I_e(1+\beta\alpha_e)}{\alpha . I_b(0)} \quad \ldots\ldots(69)$$

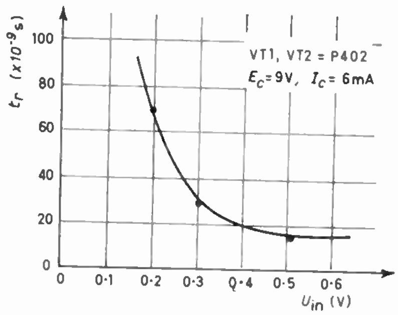
We have seen in the previous section that in c.m.l.-circuits during the transient process there is a low-impedance path available at the emitter node point and hence for this circuit we get $R_e = 0$, i.e. $\beta = 0$, and $I_b(0)$ is given by eqn. (54). Hence eqn. (69) will reduce to

$$t_r = \frac{\alpha I_e}{\omega_T}\cdot\frac{R_{II}+r_{bb'}}{U_{in}}+\frac{I_e R_{II} C_c(R_{II}+r_{bb'})}{\alpha . U_{in}} \quad \ldots\ldots(70)$$

where ω_T is the characteristic frequency of the transistor and is equal to $\alpha_e\omega_{\alpha e}$. Thus we get an expression for rise-time directly in terms of the generator voltage and impedance and small-signal parameters. The results obtained experimentally are shown in Fig. 11 and agree fairly well with the calculated values. The collector capacitance C_c may be considered as equal to $(C_c + C_s)$ when C_s is the stray capacitance, as the resistance in the base circuit is quite small as compared with the emitter circuit resistance. The effect of the load resistance R_{II} and the total collector capacitance C_c is prominent in the case of the P402 switch, but is of little importance for the P6G because of its low ω_T as compared with that of v.h.f. P402 device. The experimental results are shown in Fig. 12. It can be seen that the maximum value of R_{II} allowed is 300 ohms; above this value the rise-time worsens.



(*a*) Variation of t_r with input voltage for l.f. transistors.



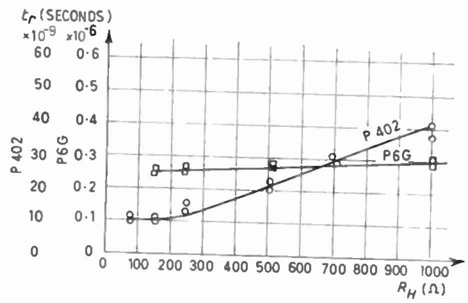
(*b*) Variation of t_r with input voltage for v.h.f. transistors.

Fig. 11.

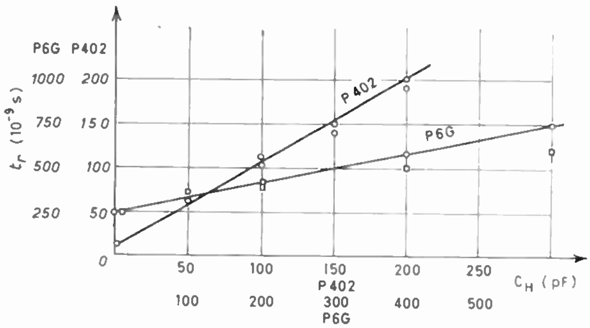
As seen in the previous section the reverse base current $I_{b(R)}$ has the same magnitude as the forward base current $I_{b(F)}$ and consequently the fall-time will be represented by the same expression as the rise-time. This has also been verified experimentally.

8. Pyramiding Factor and Delay per Stage of the C.M.L. Switch

The delay per stage with a different pyramiding factor n for the transistors P6G, P402 and P403 is shown in the Fig. 13. The delay was averaged for three identically-loaded c.m.l. switches connected in cascade. It can be seen that their delay increases with increased pyramiding factor as the output current available is shared by the bases of the loading switches.



(*a*) Variation of t_r with load resistance.

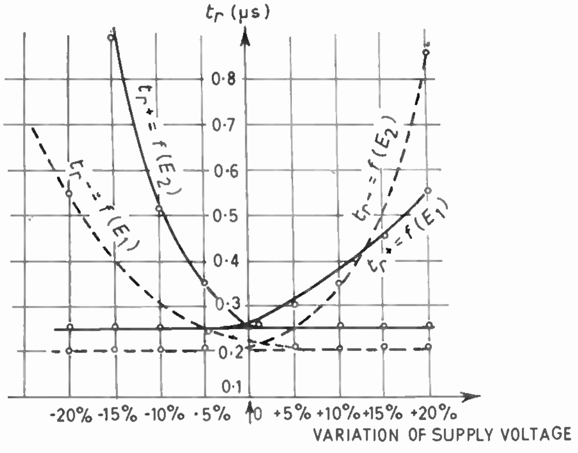


(*b*) Variation of t_r with load capacitance.

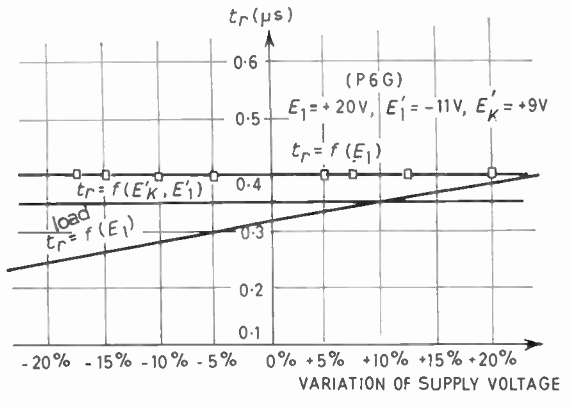
Fig. 12.

But the most interesting fact is that the delay was nearly the same for P402 and P403 switches though they have different characteristics f_T. The only conclusion which can be drawn is that the delay was dominated by the base-emitter capacitances which is nearly the same for these transistors. Hence it is necessary to use devices of lower C_{eT} if a smaller delay per stage is to be achieved. Variation of rise-time with power supply fluctuation is shown in Fig. 14(*a*) for the circuit given in Fig. 2 and in Fig. 14(*b*) for the circuit of Fig. 4. Again the figures show the superiority of the modified circuits and also the necessity of keeping signal symmetry about the base reference potential.

Effect of finite rise-time of the input waveform on the output rise-time of the switch is shown in Fig. 15. For very small input rise-times, the output rise-time of the switch is completely determined by the frequency and output time-constant $R_{II}C_c$ as given by eqn. (70). But the output rise-time increases linearly with the input signal slope value. This also can be explained from eqn. (70) as with decrease in the value of U_{in}, the rise-time will increase correspondingly.



(*a*) Variation of t_r with supply voltage for current switch.



(*b*) Variation of t_r with supply voltage for modified current switch.

Fig. 14.

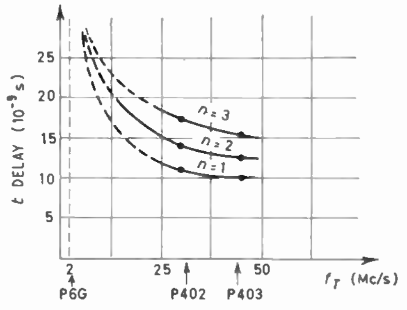


Fig. 13. Variation of delay time with f_T.

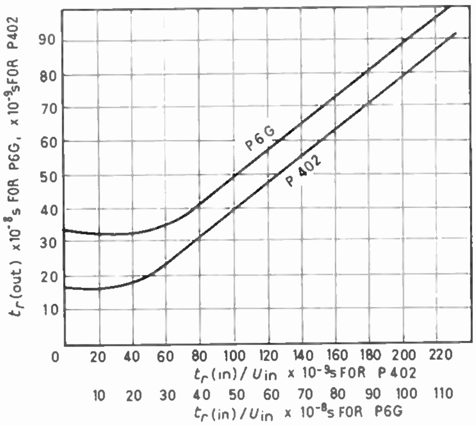


Fig. 15. Variation of delay time with rise time of input signal.

9. Applications

9.1. *C.M.L. Circuit as a Pulse-generating Circuit*

When a sinusoidal voltage is fed into a typical c.m.l. flip-flop circuit, we get a square wave symmetrical about the reference potential at the output of the circuit. The pulse repetition frequency of the output pulse is the same as the input frequency and the rise-time of the pulse depends upon the input frequency of the sinusoidal waveform. This relationship and the circuit arrangement is shown in Figs. 16 and 17. Good square wave output pulses are obtained from 100 c/s up to a few megacycles. The curve (*b*) in Fig. 17 corresponds to the rise-time of the square wave obtained from a c.m.l. switch driven by a sinusoidal input source.

9.2. *C.M.L. Logic Circuit*

In concluding we will consider typical half-adders based on the basic circuits described above. These are shown in Fig. 18(*a*) and (*b*). We see that each half-adder contains 5 transistors and this is not a very large number when compared with half-adders based on other types of logic circuits, e.g. d.c.t.l. logic. At the same time the current-switching circuits give tremendously higher speeds and very good reliability compared

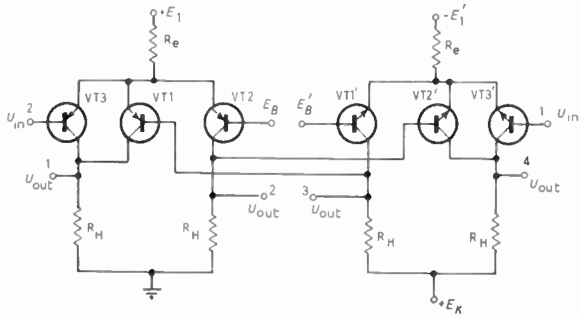


Fig. 16. Flip-flop circuit.

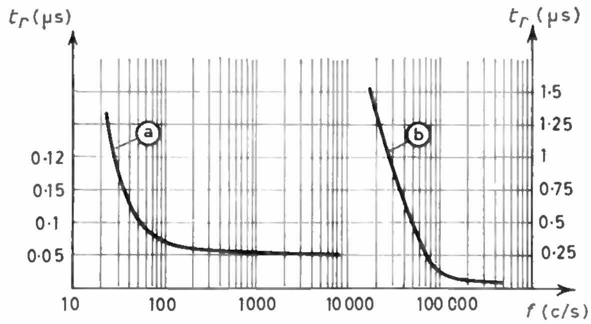
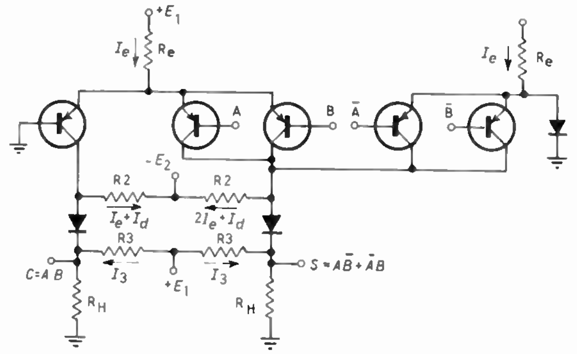
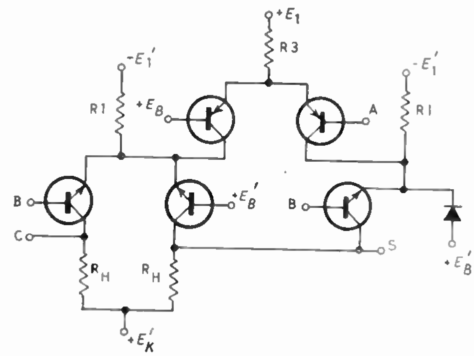


Fig. 17. Rise time of output square wave as a function of input sine wave frequency: (*a*) for flip-flop. (*b*) for current switch.



(*a*) Single-transistor-type half-adder.



(*b*) Complementary-transistor-type half-adder.

Fig. 18.

with other logic circuits. In particular its immunity from the manufacturer's spread of transistor parameters and also its high stability with temperature might lead to its application in other fields than computers.

10. References

1. Y. N. Bapat, "High speed computer switching circuits", *J. Inst. Telecomm. Engrs*, **8,** No. 1, pp. 55–60, January 1962.
2. H. S. Yourke, "Millimicrosecond transistor current switching circuits", *Trans. Inst. Radio Engrs (Circuit Theory)*, **CT-4,** No. 3, pp. 229–40, September 1957.
3. C. M. Campbell, "New configurations in nonsaturating complementary current switching circuits", *Semiconductor Products*, **2,** No. 7, July 1959.
4. Y. N. Bapat and I. L. Kaganov, "A high speed semiconductor switching and trigger circuit", *Electrochestvo (USSR)*, **81,** No. 6, pp. 76–82, June 1960.
5. Y. N. Bapat and I. L. Kaganov, "High speed p-n-p to n-p-n logical circuits for computers", Radio-Electronic Engineering Research Reports, M.E.I. (USSR), Vol. 34, pp. 5–15, 1961.
6. D. B. Jarvis and others, "Transistor current switching and routing techniques", *Trans. Inst. Radio Engrs (Electronic Computers)*, **EC-9,** No. 3, pp. 302–8, September 1960.
7. F. K. Buelow, "Improvements to current switching", *Trans. Inst. Radio Engrs (Electronic Computers)*, **EC-9,** No. 4, pp. 415–8, December 1960.
8. J. L. Walsch, Section 15, "Handbook of Semiconductor Electronics ", Ed. L. P. Hunter, 2nd Ed. (McGraw-Hill, New York, 1962.)
9. R. F. Shea, and others, "Transistor Circuit Engineering", 1st Ed., p. 301, (John Wiley, New York, 1957).
10. R. P. Nanavati and R. A. Johnson, "Turn-on delay time and its prediction", *Proc. Natl. Electronics Conf.*, **14,** 1958.
11. T. W. Easley, "The effect of collector capacity on the transient response of junction transistors", *Trans. Inst. Radio Engrs (Electron Devices)*, **ED-4,** No. 1, pp. 6–14, January 1957.

11. Appendix 1: H.F. Parameters of the Transistors used in the C.M.L. Circuits

Transistor type	P6G	P402	P403
Measured value of			
f_α	2×10^6 c/s	100×10^6 c/s	500×10^6 c/s
f_{osc1}	$4{\cdot}5 \times 10^6$ c/s	60×10^6 c/s	100×10^6 c/s
f_T	2×10^6 c/s	35×10^6 c/s	45×10^6 c/s
$r_{bb'}$	80 ohms	75 ohms	50 ohms
C_c	50 pF	10 pF	5 pF
$t_{r(\min)}$	150×10^{-9}s	10×10^{-9}s	$<5 \times 10^{-9}$s
Calculated			
$t_{r(\min)} = 0{\cdot}325/f_T$			
	162×10^{-9}s	$9{\cdot}3 \times 10^{-9}$s	$7{\cdot}2 \times 10^{-9}$s
$t_{r(\min)} = 0{\cdot}325/f_\alpha$			
	162×10^{-9}s	$3{\cdot}25 \times 10^{-9}$s	$0{\cdot}65 \times 10^{-9}$s

Manuscript first received by the Institution on 16th April 1963 and in final form on 2nd December 1963. (Paper No. 904.)

U.D.C. 621.37/8: 681.828

Electronic Melodic Instruments

By

K. A. MACFADYEN, M.Sc.†

Presented at a meeting of the Electro-Acoustics Group in London on 1st May 1963.

Summary: An historical survey of electronic melodic instruments requiring a performer is given, together with an appraisal of their relationship to the art of music. It is argued that the most promising role for this class of instrument is in the spheres of amateur music and the teaching of music rather than as a successor to traditional instruments. The principles of design of a new instrument for the former class of service are discussed.

1. Introduction

The subject of electronic musical instruments is now so large that in a single paper one could cover it only in a very superficial way. Accordingly the present discussion will be confined to the class of instrument requiring a performer and producing only a single melodic line. It will be necessary to refer only briefly to the other main classes of electronic instruments, firstly polyphonic or chordal instruments, of which the electronic organ is the most familiar, and secondly, the instruments from which the constituents of taped electronic music (of the type of *musique concrète*) are obtained. In this music composer and performer are one, and no interpreter is needed, so it lies beyond our present scope.

2. Historical Survey

The idea of the electronic musical instrument was the result of the great advances made during the last century in our knowledge in two fields: acoustics and electrodynamics. The most remarkable of the early instruments, Cahill's 'Telharmonium',[1] (described in 1906), although never completed, is of interest in that it embodied the principle of *harmonic synthesis*. As Mersenne[2] pointed out over three centuries ago (and Helmholtz[3] amply confirmed) a note of rich tone quality can be regarded as a mixture of fundamental and harmonics. Cahill reversed this argument and synthesized rich tones from their component harmonics generated electrically by a multitude of rotary electric generators. Difficulties in making the resulting music audible delayed progress until the invention of the loudspeaker just after the First World War.

The invention of radio telephony had by that time given us oscillators of easily-controlled pitch and some of the earliest attempts to exploit the musical possibilities of the situation were made by a German, J. Mager. Better known, however, is the work of a Russian of French descent, Léon Thérémin, whose 'Etherophone' was demonstrated in the principal capitals of the world during the 'twenties. The instrument generated a heterodyne note the frequency and intensity of which were both varied by hand-capacitance. Rudimentary tone control was also provided. Great interest was aroused at the time but it was generally felt that the playing technique made accurate intonation and clear attack very difficult if not impossible. Improvements were made and the instrument was at one time marketed by the Radio Corporation of America.

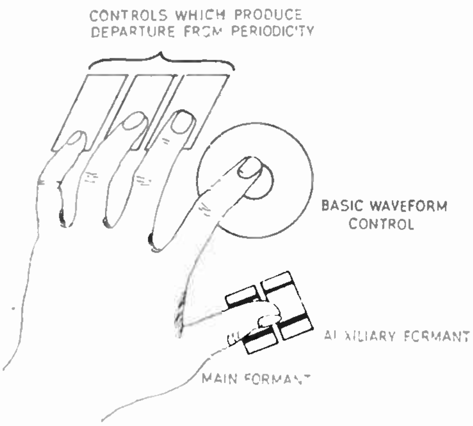
Trautwein in Germany and Martenot in France eased matters for the performer by providing some form of finger-board or keyboard. Their instruments have undergone changes over the years, but the original Trautonium used a gas-tube relaxation oscillator, with filter circuits to modify the tone colour. Playing was done by fingering a horizontally-stretched conductor so that detached notes, *legato* passages and *staccato* playing were equally possible. Developments of the instrument are in use in the studios where taped electronic composition is carried out.

Martenot's instrument (Fig. 1), sometimes called the *Ondes Musicales*, has undergone considerable development since its introduction in 1928. The model currently available is controlled either by a 6-octave keyboard or by a finger-ring moving over a dummy keyboard and operating a continuous frequency-control device by an endless cord; the latter control is used for *glissando* or *portamento* playing. No sound is heard, however, unless a sprung 'attack key', operated by the player's left hand, is pressed. The attack of every note can thus be controlled to any degree from percussive to gradual, and the loudness is similarly governed. Tone control circuits and several loudspeakers are provided, some of the latter adding coloration and reverberation, by simple mechanical means, to the tone. The instrument in expert hands can give good imitations of many traditional solo instruments (strings, woodwind, brass, xylophone, etc.) as well as having an enormous repertoire of unheard-of sounds ranging from the simple to the extremely *macabre*.

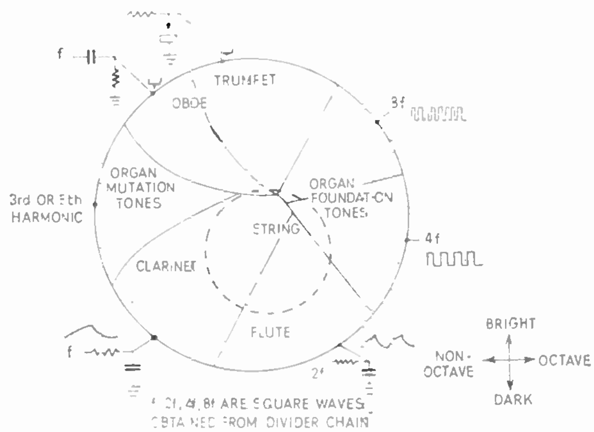
† Department of Physics, University of Birmingham.

Fig. 1. Martenot's *Ondes Musicales*.

Elaboration of the control means was carried even further by LeCaine who now directs the electronic music laboratory of the National Research Council of Canada. About 1947 he evolved what he now satirically calls the 'electronic sackbut'[4] to indicate that, like its namesake, it is now defunct. Working basically on the radio-frequency heterodyne principle, it utilized a keyboard in which downward pressure controlled loudness and sideways pressure pitch, so that a vibrato or a glissando could be produced. The player's left hand was employed in continuously controlling the tone quality by a remarkable set of controls illustrated in Fig. 2. The circular segmented plate, shown enlarged in Fig. 2(*b*), enabled the player instantaneously to vary the basic harmonic structure of the note by moving, with one finger, a circular capacitively coupled electrode (dotted line) over the plate. These signals were fed to two *formant circuits*,



(*a*) Controls.



(*b*) Waveform control.

Fig. 2. LeCaine's 'electronic sackbut'[4].

i.e. resonant circuits of which the resonant frequencies and the damping could be varied by movements of the left thumb. More will be said about formants later. Also controlled by the left hand was a device for relieving the mathematically accurate repetition of the beat-frequency oscillator by feeding a coloured noise signal to an appropriate point in the circuit. When heavily applied this device caused the note to degenerate into the harsh rasping sound produced (when required) by jazz trumpeters. The instrument could be used for simulating traditional instruments or for producing entirely novel tones. In both of these spheres its performance was very striking.

These three typical instruments, the Trautonium, the Martenot and the 'sackbut', clearly require in the performer a considerable knowledge and experience as well as a very specialized playing technique. During and since the Second World War several melodic instruments requiring only elementary playing technique have been marketed. These instruments exploit the keyboard technique which is possessed, in at least a rudimentary form, by a large section of the population. The 'Solovox',[5] of American origin, is a three-octave portable keyboard instrument with a large number of pre-set controls of tone quality, rate of attack, pitch-range, vibrato, etc., but only one continuously controlled feature, namely the loudness. This limitation is imposed with the object of enabling the performer to accompany himself on the pianoforte with his left hand whilst playing the Solovox with his right. The volume is controlled by a knee-lever. The circuit of the instrument is directed towards producing novel tones rather than traditional ones.

A simpler instrument, the 'Clavioline',[5] of French origin, is more imitative and can give quite pleasing reproductions of several traditional instruments. The tone generator is a special form of multivibrator controlled in frequency by a single grid-leak consisting of a chain of high-stability resistors. Key-contacts select the operative length of this chain and so decide the note sounded. The generator is followed by an ingenious if somewhat enigmatic tone-forming network and an attack-control valve also worked by key-contacts. A vibrato is obtainable by modulating the multivibrator anode-supply with a resistance-capacitance oscillator of suitably low frequency (about 6 c/s). A conventional power amplifier and loudspeaker complete the instrument.

Although the Clavioline can simulate a number of woodwind instruments fairly well its performance in brass and string is considerably less satisfactory. Even the woodwind tone is soon found to lack the variety of its traditional counterpart and, as we shall see, this is attributable to the reduction of continuously adjustable controls, in the interests of self-accompaniment, to a single crescendo lever.

3. The Impact of Electronics on Music

Even as long ago as 1929 electronic musical instruments were sufficiently developed for a famous conductor to hail them as the liberators of music from the limitations of traditional musical instruments and to predict that in a few years traditional instruments would disappear. It may seem strange that over 30 years later practically all Western music still relies on traditional instruments; but the reason, though a subtle one, is sufficient to guarantee the continuance of this state of affairs for many years to come. It is that our enjoyment of music, i.e. our reception of its full 'message' (one which cannot be put into words), is dependent, at least in part, on the associations of the classes of instruments we hear. Thus, the sound of the trumpet is so strongly associated, in the experiences of us all, with the idea of 'stiffening the sinews and summoning up the blood' that a part of the 'message' of the music is received almost as soon as the instrument is heard. Again the subtle tone of the violin suggests tension, not only because its lightest inflexions mirror the nervous movements of the player, but because we know the instrument itself as compounded of tensions, emitting slight musical notes even when merely handled. So we could go on, to recognize in the horn the suggestions of hunting, and in the flute the insubstantiality of human breath. I do not say that these associations are fully conscious, but, as we know, unconscious associations profoundly affect our reception or rejection of ideas and impressions.

How is all this related to the impact of electronic musical instruments or Western music? Simply in this way: that if a composer wishes to make use of the 'vocabulary' of associations possessed by his future audiences he writes for traditional instruments. If, on the other hand, like the devotees of the serial system of composition, he is prepared to go to great lengths to avoid associations, he will employ electronic tone-production as a means of eschewing familiar sounds; *but he will not then be satisfied with electronic instruments requiring performers.* With the facilities of tape recording available he will eliminate the last human element and compose in electronic sounds recorded directly on tape. Composers such as Eimert and Stockhausen[6] have made their attitude on this issue abundantly clear.

It seems, then, that the 'performer' type of electronic instrument falls between two stools in the world of professional, public music. But there is another world: that of the amateur musician. The greatest danger in music today is that these two worlds may fall apart. The split has already begun between the *avant-garde* composers, each writing for an exclusive coterie on the one hand, and the perplexed musical public on the other. A vigorous and active *performing* public is the best safeguard against the decay of music by fragmentation into coteries. The performer comes into sufficiently close contact with a composition to size it up with a shrewdness denied to the mere listener.

4. Electronics and the Amateur Musician

During many years of active amateur music-making of various kinds, I have often wondered what electronics can do to encourage active participation in good music by 'serious' amateurs, by which term I do not mean excessively solemn amateurs, but amateurs who wish to take part in music for its own sake and not for gain or fame. What then, is the background of the average amateur? The ability to read music and play on an instrument is acquired only by training from an early age—about seven years—and the majority who attempt such a course are directed to the pianoforte. In one way this is a good thing because the keyboard, from its very nature, can implant the basic concepts of scale, key-relationship, harmony and counterpoint far more effectively than any non-keyboard instrument can. On the other hand, the popular image of a pianist is one who can hold audiences spellbound by his unaided efforts, and most embryo pianists, finding after several years' work that they do not even approximate to this image, give up in despair. The more discerning, perceiving that team-work can often achieve what individual efforts cannot, turn to accompanying or to chamber music, only to find that the technical demands are at least as great as those of works for pianoforte solo. Many a disappointed pianist, contrasting the simplicity of the single line of notes of the violinist's part with his own score cluttered with chords, cross-rhythms and counterpoint, has abandoned the pianoforte in favour of an orchestral instrument, only to find that it is as hard to produce a single, really pleasing note on it as to play a whole passage on his forsaken keyboard. But encouraged by the prospect of participation in great music he battles on, wrestling with the vagaries of instruments and the difficulties of a musical notation bristling with the absurdities that are the litter of four centuries of haphazard growth. He discovers, for instance, that if he aspires to be a 'cellist, he must be at home in any one of three different clefs, or that if he takes up the horn he must be able to transpose at sight. To the aspiring professional these and other difficulties of traditional instruments form a challenge, and the overcoming of them a source of satisfaction; and since the years of practice needed to overcome them run concurrently with a growing knowledge of music and a maturing musical outlook one cannot say that the time is wasted. But with the amateur the situation is very different. Limitation of time and energy is likely to tell in the end and, disappointed, he gives up. But his latent keyboard training is still with him.

It seems that here we have the perfect field for the cultivation of the electronic musical instrument. In an imitative form capable of deputizing for a dozen types of traditional instrument it would be a passport to participation in an enormous repertoire of music, from the suave string fantasies of Orlando Gibbons to the witty woodwind suites of Milhaud and Ibert. Why, then, do amateurs not flock to purchase keyboard instruments, such as Martenots or Claviolines, so that they can take their places in string quartets, wind trios or school orchestras?

Firstly, there is the high cost of these instruments. Furthermore, the Martenot, as we have seen, requires a specialized playing technique which has to be acquired before any progress can be made, and the six-octave keyboard makes the instrument difficult to transport. The Clavioline, on the other hand, is readily transportable, but, designed as it is with a view to self-accompaniment at the piano, provides the amateur with meagre occupation for his left hand and limits him to a range of loudness with none of that variation in tone quality which, as we shall see, is an essential feature in the tones of traditional instruments.

It appears, then, that the problem of designing an electronic instrument for the amateur with some keyboard training is to strike a balance between simplicity of control, with its attendant monotony, and great elaboration with its heavy demands on manipulative skill.

5. Principles of Design

The design of a suitable electronic instrument takes us back to the study of instrumental tone. Whilst there is no doubt that a *single complex tone* can be synthesized from the correct proportions of fundamental and harmonics, the generalization so commonly made, namely, that the tone of an *instrument* is characterized by such a 'harmonic recipe', is totally false. This has been repeatedly proved by the harmonic analysis of instrumental tones. Not only are neighbouring notes on the same violin astonishingly different in harmonic recipe but the same note undergoes radical changes of harmonic structure when transferred from one violin to another.[7] The same is true of other instruments. Moreover, the changes in harmonic content from *pianissimo* to *fortissimo* are often enormous. These results are not only given by harmonic analysis but can be verified by anyone with a sensitive ear. Having disposed of the idea of the 'fixed recipe', one is forced to ask: By what acoustic feature of its tone do we recognize an instrument? Instruments can seldom be identified by sounding a single note; it is always necessary to play a musical phrase. This fact suggests that we look for features common to a series of notes, and it also draws attention to the transients which occur at the attack and the release of a note and between successive notes. It can be demonstrated that, deprived of their transients, single notes of suitable pitch and strength played on a flute and a trumpet, although sounding vaguely different, have none of the characteristic quality of the instruments and cannot be identified with certainty.† As soon as the transients are heard the identities of the instruments are revealed beyond doubt.

Studies of the mechanisms of tone production in traditional instruments provide further clues to the characteristic features of instrumental tone. Acoustically most instruments conform to the following general pattern:

(i) Production of a periodic disturbance having a particular characteristic of amplitude *versus* harmonic number for a given intensity of playing. There may also be resonances of fixed frequency, at which any harmonic is intensified.

(ii) Filtering of the resultant signal by a filter retuned for each note.

(iii) Further filtering by mechanical or acoustic systems of fixed resonances and definite attenuation characteristics.

For example, in woodwind instruments (i) is represented by the periodic puffs of air released by the reed and 'moulded' by its free vibrations, (ii) by the transmission properties of the sounding part of the tube, and (iii) by the remainder of the tube with its associated mismatch of acoustical impedances to the atmosphere. In stringed instruments (iii) includes the cavity and body resonances. In the clarinet, which has a quarter-wave resonator, the even harmonics are largely suppressed at stage (ii). Stages (i) and (ii) are almost invariably coupled in frequency.

This brief survey suggests that in electronic simulation of instruments the essential elements are:

(*a*) A generator of periodic pulses.

(*b*) A means of instantaneously controlling the progressive attenuation of harmonics with rising frequency.

(*c*) A device for eliminating or weakening the even harmonics.

(*d*) The provision of resonances of appropriate frequencies and damping factors (formants).

(*e*) The control of transients.

The properties of the human ear are such that (*d*) is of surprisingly high importance; surprising, that is, until one remembers that the basis of speech is the recognition of formants both in vowels, in which the resonant cavities are those of the mouth, and in consonants, in which cavities formed by the lips and teeth are responsible for differences, e.g. between sibilants and other fricatives. Similarly, it is the formants in viola tone which enable us to differentiate it from a violin, even when these two instruments play identical musical passages.

† This was shown at the meeting with recorded sounds. The experiment has often been done with the actual instruments and players concealed behind a curtain. A loud noise is used to mask the beginnings and ends of the notes.

6. Demonstration Instrument

An instrument based on these principles has been constructed (Fig. 3) with a view to the kind of amateur use already described. The basic note-generator is a

Fig. 3. Electronic melodic instrument demonstrated at the lecture. The three controls for the left hand are for loudness, depth of vibrato and 'muting', respectively. The panel controls affect formants, harmonic series (natural or even numbers), attack, vibrato speed, pitch range, etc.

form of multivibrator. This type of oscillator is chosen partly because its output (periodic pulses) is rich in harmonics and partly because the frequency f is given by an expression of the form

$$f = \frac{1}{CRF(V)}$$

where C represents a capacitance, R a resistance and $F(V)$ a function of various potentials in the circuit. One of these factors may be controlled by the keyboard, another by an octave-shift key and a third by tuning or transposing devices. The octave-shift key increases the range from three to five octaves, giving a range from 'cello C (65·25 c/s) to the top C of the descant recorder (2088 c/s). The facility for retuning the entire keyboard over a wide range is an essential one because many traditional instruments are transposing instruments, i.e. their natural scale is not C major. The score is therefore printed in such a key as to correct for this peculiarity of the instrument.

Assuming that the player of a keyboard instrument is unable to transpose at sight, a feat beyond most amateurs, he cannot use, for example, a clarinet score unless the keyboard instrument is first re-tuned to the clarinet scale (usually B flat, a whole tone below the pianoforte). With a multivibrator circuit this can readily be done.

Figure 4 shows the principle of the instrument. B represents the multivibrator with its keyboard B_1 and

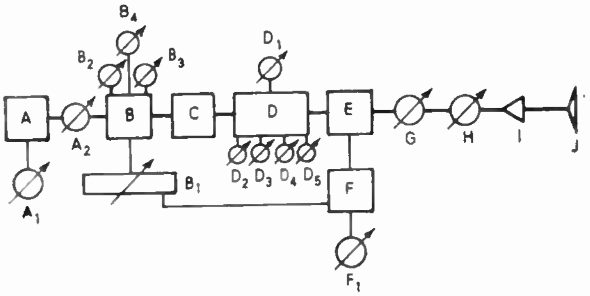


Fig. 4. Block diagram of electronic melodic instrument.

other controls B_2-B_4. A is the vibrato generator with frequency-control A_1 and amplitude-control A_2, which is one of the three controls continuously operated by the player's left hand (Fig. 4). The signal passes to a bistable multivibrator C which changes it to a series to pulses of alternately opposite sign; only odd harmonics are now present. However, the next stage, a tuned amplifier D, has a bias control D_1, which permits the negative pulses to be suppressed, restoring the even harmonics when required. The characteristic 'hollow' and 'full' tones of the clarinet and oboe respectively can thus be provided. The amplifier D is a pentode with a resistance and two anti-resonant LC circuits (the formant circuits) associated with it. Resonant frequencies and damping are controlled by D_2-D_5. The signal is next para-phased and applied to a carefully balanced push-pull stage of triode amplification with a specially designed transformer connecting the anodes. A control signal derived from key-contacts through the medium of an attack-valve E and a timing circuit F controlled by F_1 is applied to the common grid-biasing point of the push-pull pair so that the onset of each note can be made sharp or gentle as required. The object of the push-pull feature is to balance out the control signal and so reduce as far as possible the 'keying thump' likely to appear with control circuits of this kind. G is an ordinary volume control in continuous operation by the player and H a three-stage RC high-frequency attenuation circuit also continuously controlled by the player. As this control affects the tone in much the same way as the application of a mute to a violin I provisionally refer to it as a 'muting control', feeling that it is important to choose a name suggestive to the musician even at the risk of a collision with the nomenclature of radio engineers. It affects the rate of attenuation with rising frequency to any degree from zero to 18 dB per octave, and is used not only to produce attenuation characteristic of a particular instrument but also to give the variations in tone of any one instrument. I and J are a power amplifier and loudspeaker respectively.

7. Manner of Performance

Since the keyboard is standard and can be tuned to the natural scale of the instrument to be simulated, the reading of most instrumental 'parts' presents no problem to a player with rudimentary training. It will, however, be found necessary to cultivate a special touch, the aim being to leave a very short time between the release of a key and the depression of the following one to allow the 'attack' circuit to work. 'Overlap' leads to a hard attack. The settings of vibrato speed, formant frequencies and intensities, attack speed, etc., can be made once and for all for each instrument to be simulated. For example, oboe tone needs one strong resonance at 1800 c/s, sharp attack, a relatively slow vibrato (if used) and about 6 dB per octave of 'muting' (see Section 6). Violin tone calls for two weak resonances at about 4500 c/s and 300 c/s respectively, a slower attack and faster vibrato than the oboe, continuous operation of the volume control to simulate bowing and variation of muting to suit the pitch of the note. The need for the latter springs from the fact that the four strings of a violin have progressively changing acoustic matching to the atmosphere.

Continuous use of the muting control is also necessary in simulating certain woodwind and brass instruments, in which the upper harmonics become markedly more intense with increasing loudness. This is, in fact, the secret of the 'splash' in trumpet tone and the alternation of 'edginess' and sweetness in the clarinet.

Enough has been said to show that, while the beginner can produce a tolerably good effect, the instrument provides scope for very much more lively

I	II
Oboe†	Trumpet†
Cor anglais	Trombone
Oboe d'amore	Horn
Bassoon†	Viola da gamba†
Clarinet†	
Bass clarinet	III
Recorders (descant, treble†, tenor, bass)	Violin
	Viola
Flute	'Cello†

† These were demonstrated at the meeting.

and life-like performances in the hands of a skilled player who can operate the left-hand controls with the necessary judgment.

In the form demonstrated, the instrument can give good simulations of the first group of instruments listed opposite but is progressively less satisfactory in Groups II and III.

8. Further Progress

The experience gained in playing the instrument in various chamber and small orchestral groups has led to an improved principle which, however, has not yet been incorporated in an instrument suitable for demonstration. The chief defects of the existing instrument are the impossibility of *portamento* (the discreet 'slide' between notes which is a distinctive feature of string playing) and the presence of the transient (resembling the sound 'plop') at the end of each note. No matter how perfect the functioning of the balanced attack circuit, the fact that the note has to be silenced before the oscillator stops or changes frequency sets an upper limit to the time that can be allowed for the functioning of the attack circuit.

In most traditional instruments the note either dies away gradually or changes more or less continuously to the new note. This behaviour provided the suggestion for the new instrument[8] which is now under development. The keyboard influences the frequency of a continuously-running multivibrator by voltage control instead of resistance control; a suitable resistance-capacitance time-constant circuit enables *portamenti* to be played by a simple movement of the left hand.

Other improvements lie in the direction of combining the four basic controls: volume, muting, vibrato and portamento, into fewer units and redistributing them between the two hands.

9. Other Spheres of Usefulness

In addition to use by the serious amateur as a deputy instrument in chamber music and other combinations, this instrument could give valuable service as an aid to practice. School orchestras, for example, cannot usually obtain the service of a cor anglais or bass clarinet player until the 'dress rehearsal'. The electronic deputy would clearly be a great help at earlier rehearsals. Again, in schools of music it would enable singers to become accustomed to the sound of a clarinet or trumpet *obbligato* when the actual instruments are unavailable.

Electronic instruments are more often attractive to musicians in a subsidiary role than when they dominate the scene. An example is the success of electronic pedal units for pipe organs in circles where a full electronic organ would not be acceptable. The present development may be regarded as fulfilling a subsidiary role acceptably rather than a dominant one in a less satisfactory way.

10. References

1. T. Cahill, "Electrical World" (New York, 1906).
2. M. Mersenne, "Harmonie Universelle" (Paris, 1636).
3. H. L. F. Helmholtz, "On the Sensations of Tone" (Translated by Ellis) (Longmans, London, 1930 and Dover Publications Inc., New York, 1954).
4. H. LeCaine, "Electronic music", *Proc. Inst. Radio Engrs*, **44**, No. 4, p. 457, April, 1956.
5. A. Douglas, "The Electronic Musical Instrument Manual" (Pitman, London, 1957).
6. K. Stockhausen, "Die Reihe, No. 1—Electronic Music" (Universal Editions Ltd., 1958).
7. A. Wood, "Acoustics", Chapter 8 (Blackie, London, 1960).
8. British Patent No. 932,303.

Manuscript first received by the Institution on 9th September 1963 and in final form on 11th March 1964. (Paper No. 905/EA14.)

DISCUSSION

Under the chairmanship of Mr. E. D. Parchment

Mr. R. L. West: I was most impressed with the sounds produced by Mr. Macfadyen's instrument. I thought the oboe reproduction well-nigh perfect and the others generally better than all previous imitations I have come across.

I did miss the Doppler effect due to the performer's usual body movement and wonder whether a suitable control would be possible or desirable to simulate this. The overall effect may well be complicated and be made up of a combination of directional pattern, Doppler and interference effects due to local and not-so-local reflections.

Quite a number of instruments are characterized by non-harmonic overtones, and I wonder also whether any thought has been given to this problem.

Whilst the actual quality of string tone was good, I did feel it sounded rather mechanical due to the frequency being constant, or if with vibrato, too regular. I consider frequency regularity has been one of the outstanding objectionable features of all electronic instruments to date. As this instrument is so good in all other respects, would it not be worthwhile trying to overcome this feature? It might be possible to inject low-frequency noise into a suitable point in the circuit so as to produce a slight random frequency wobble as in natural instruments.

Professor Thurston Dart: Mr. Macfadyen has devised the most effective electronic melody instrument I have yet heard. I would like to know whether the instrument is designed to play in equal temperament (which means that it will always be what a professional musician would call 'out of tune'), and whether its rate of vibrato can be altered at the player's will.

I suggest that Mr. Macfadyen is mistaken in supposing that professional musicians like their instruments to be difficult to master. In my view, the reason why 'traditional' instruments are preferred is that they alone enable the musician to make full use of the delicate and complex feed-back mechanism which links what he hears with how he plays. This mechanism, hard to analyse and harder still to investigate, is at the root of a performer's skill; a violinist with what is known as 'good tone' can produce this good tone even on an inferior instrument, though less easily than on a fine one.

Professor Alfred Nieman: The quality of the sound of Mr. Macfadyen's instrument represents quite an achievement. There are two points I would like to raise. Firstly, there is the need for notes to merge into each other as a singer would create a 'line' or a crotch, and not to sound like 'blobs', typical of percussion or undifferentiated instruments, for example a piano and an organ respectively. I think it is important and Mr. Macfadyen recognized that too. I think his instrument might achieve this.

The other point I would make is the complexity of finding the correct places with the left hand. Could not this be made much simpler? If Mr. Macfadyen had a system of marking the best positions for certain instrumental sounds, even to putting the names down, this would then be very much easier.

I am not prepared to go into mathematical or aesthetic arguments. Personally, I am all for new ideas, and this is surely only just a beginning. What I think Mr. Macfadyen should *also* concentrate on is genuine electronic music, that is generators for making new music, not only playing the old generators with new sounds.

Captain J. D. M. Robinson: I have a point to make which refers essentially to the amateur field of application. One of the major social problems of the day is the increasing reliance for entertainment to be 'laid on' in various synthetic forms requiring little, if any, personal mental or physical effort, and the consequential loss of ability to make one's own entertainment. If, as appears to be the case, you have designed a musical instrument which brings the art of playing within reach of young people who could not otherwise attempt it, you will have struck a major blow for the better character training and development of young people, particularly if the instrument could be produced at a reasonable price.

Mr. G. R. Pontzen: I congratulate Mr. Macfadyen on his achievement of producing an instrument of such high quality. Considering the problems involved in the emulation of the sounds of conventional musical instruments, I was particularly impressed with the imitations of the oboe and the cello, demonstrated on the instrument's internal loudspeaker. In my opinion, the tonal quality of the recordings which have been played do not by any means convey the versatile capabilities of the instrument. Considerable advantage might be gained in most cases if recordings were made by feeding the output of the instrument directly into the tape recorder (especially if suitable artificial reverberation could be added), thus avoiding the influence of the particular characteristics of the loudspeaker, microphone and room acoustics involved in normal tape recording techniques.

AUTHOR'S REPLY

I agree with Mr. West that regularity of tone is a defect in this type of instrument. I personally do not think, however, that the Doppler effect plays a significant part in modifying the aural effect of traditional instruments although I agree that directional properties and changes in the standing-wave pattern in the room are probably important. I have experimented with arrangements to disturb the perfect regularity of my tone-generator. In one case I injected filtered white noise at a suitable point in the circuit, but the effect was merely to reproduce a disagreeable roughening of the tone. It is likely that I used a band of too high a frequency. The band of frequencies needed to produce the effect that Mr. West desires probably lies between 0·1 and 2 cycles per second, and there are consequently difficulties in deriving this from white noise.

In reply to Professor Thurston Dart I must admit that the instrument is tuned in equal temperament and is thus 'out of tune' as he suggests, though no more so than a piano or an organ. At the cost of some extra difficulty in performance it would be technically possible to incorporate a pitch adjustment to enable sufficiently sensitive players to modify their intervals whilst playing, in order to accord with other instrumentalists of equal skill. The vibrato speed can be varied, though in the interests of simplicity this control has been sited on the panel as a pre-set adjustment and is not very easy to operate while playing.

The production of a melodic line rather than a series of notes, as described by Professor Nieman, is a feature in which the new version of the instrument, described at the end of the paper, should be greatly superior to the one demonstrated. The freedom of control of tone, loudness and depth of vibrato by the left hand is one of the most important features of the instrument, and this would be lost if these controls were pre-set as Prof. Nieman suggests. The panel controls could, without much loss of flexibility, be replaced by an automatic setting device such as a punched card appropriate to each instrument to be simulated.

I heartily agree with Capt. Robinson's views and I am doing all I can to have the instrument produced as cheaply as is consistent with the aims I have set out in the paper.

It is true, as Mr. Pontzen suggests, that direct electrical recording would avoid certain difficulties, such as the problem of securing perfect balance with other instruments, a problem which is largely due to the directional properties of the loudspeaker. It would, however, be a rather artificial procedure to record directly since the essential aim of the instrument is to enable the player to take part in concerted playing. The recordings played during the lecture were made merely because it was inconvenient to bring other instrumentalists to the lecture room to demonstrate the blending qualities of the electronic instrument.

U.D.C. 621.372.4: 621.375

Simulation by a Single Operational Amplifier of Third-order Transfer Functions having a Pole at the Origin

By

L. K. WADHWA,

M.Sc., M.S. (Elect. Eng.)†

Summary: The paper discusses a method for the simulation of third-order linear transfer functions, having a simple pole at the origin, with the aid of only one operational amplifier and a two-terminal network consisting entirely of resistors and capacitors. The design equations and conditions of physical realizability are given. Some of the salient advantages, disadvantages and limitations of this technique of simulation are briefly mentioned.

1. Introduction

Second-order transfer function realization with the aid of one operational amplifier and RC passive networks and employing single-path feedback has been widely reported in published literature.[1] Bridgeman and Brennen[2] and Paul[3] have discussed the simulation of initially-relaxed second-order transfer functions with positive real constant and adjustable coefficients with one operational amplifier and multi-path feedback RC networks. The author has discussed [4-9] the simulation of initially-relaxed third-order linear systems of the type

$$F(s) = -\frac{b_2 s^2 + b_1 s + b_0}{a_3 s^3 + a_2 s^2 + a_1 s + 1}$$

that is, systems having three non-zero poles lying entirely on the left-hand side of the s-plane. However, in a number of practical problems, such as the design of control systems and filters, or in the study and solution of certain transfer functions on an electronic differential analyser it becomes necessary to simulate third-order linear systems of the type

$$F(s) = -\frac{b_0(b_2 s^2 + b_1 s + 1)}{s(a_2 s^2 + a_1 s + 1)} \qquad \ldots\ldots(1)$$

that is, systems having a simple pole at the origin and two non-zero real or complex poles on the left-hand side of the s-plane.

The purpose of this paper is to discuss the simulation of the general third-order linear system characterized by eqn. (1) and its two other particular cases,

$$b_1 = b_2 = 0, \quad \text{and} \quad b_0 > 0 \qquad \ldots\ldots(1a)$$

$$b_2 = 0, \quad \text{and} \quad b_1 > 0, \quad b_0 > 0 \qquad \ldots\ldots(1b)$$

with the aid of only one operational amplifier and a two-terminal multi-path feedback network consisting entirely of resistors and capacitors.

† Defence Research & Development Organization, New Delhi, India.

2. Third-order System Realization

A basic network for the simulation of third-order linear systems is shown in Fig. 1 and its transfer function has been shown to be[4]

$$\frac{E_0}{E_1} = -\frac{Y_1 Y_3 Y_5}{\begin{aligned} &Y_6(Y_1+Y_2+Y_8)(Y_3+Y_4+Y_5+Y_7)+ \\ &+Y_3Y_6(Y_4+Y_5+Y_7)+ \\ &+Y_5Y_7(Y_1+Y_2+Y_3+Y_8)+Y_3Y_5Y_8 \end{aligned}} \qquad \ldots\ldots(2)$$

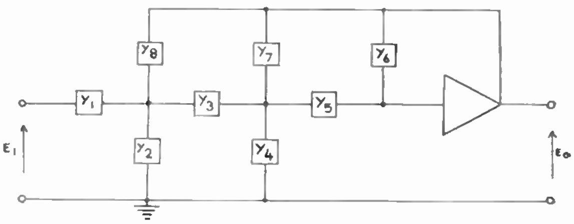


Fig. 1. Network for the simulation of third-order systems.

It is possible to simulate the system of eqn. (1) with the network of Fig. 1 provided the admittances (Y) are suitably chosen and a and b are positive real constants. Further, on examination of eqn. (2), it should also be obvious that the network of Fig. 1 with

$$\left.\begin{aligned} Y_4 &= Y_8 = 0 \\ Y_2 &= Y_4 = 0 \\ Y_2 &= Y_7 = 0 \\ Y_7 &= Y_8 = 0 \end{aligned}\right\} \qquad \ldots\ldots(3)$$

can also simulate the system of eqn. (1), as the three nodes in the network are still preserved.

It will be quite impracticable to discuss, in a single paper, all the circuits that can result from the consideration of all the four alternatives at eqn. (3). It is intended in this paper to discuss only the circuits resulting from the alternative

$$Y_4 = Y_8 = 0 \qquad \ldots\ldots(4)$$

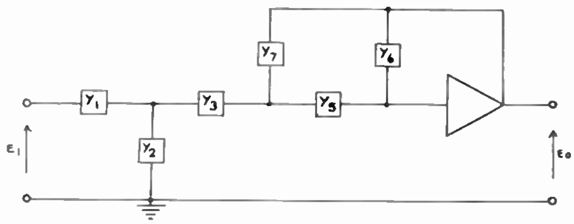


Fig. 2. Simplified circuit of Fig. 1.

A block diagram of the reduced network is shown in Fig. 2 and its transfer function, in view of eqns. (2) and (4), will be

$$\frac{E_0}{E_1} = -\frac{Y_1 Y_3 Y_5}{Y_6(Y_1+Y_2)(Y_3+Y_5+Y_7) + Y_3 Y_6(Y_5+Y_7) + Y_5 Y_7(Y_1+Y_2+Y_3)} \qquad \ldots\ldots(5)$$

2.1. $b_1 = b_2 = 0$; *and* $b_0 > 0$

The transfer function for this particular case of the general third-order system of eqn. (1) reduces to

$$F_1(s) = -\frac{b_0}{s(a_2 s^2 + a_1 s + 1)} \qquad \ldots\ldots(6)$$

The system of eqn. (6) can be simulated with the network of Fig. 2 provided

$$\left.\begin{array}{l} Y_2 = sc_2, \quad Y_6 = sc_6, \quad Y_7 = sc_7 \\ Y_1 = Y_3 = Y_5 = \dfrac{1}{R} \end{array}\right\} \qquad \ldots\ldots(7)$$

Substitution of eqn. (7) into eqn. (5) on simplification, and comparison with eqn. (6) gives

$$b_0 = \frac{1}{3T_6 + 2T_7} \qquad \ldots\ldots(8)$$

$$a_1 = \frac{2T_2 T_6 + T_2 T_7 + 2T_6 T_7}{3T_6 + 2T_7} \qquad \ldots\ldots(9)$$

$$a_2 = \frac{T_2 T_6 T_7}{3T_6 + 2T_7} \qquad \ldots\ldots(10)$$

where

$$T_n = Rc_n \qquad \ldots\ldots(11)$$

Now the network of Fig. 2 is physically realizable only if the values of T_2, T_6, T_7 obtained as the solution of eqns. (8)–(10) are positive real. Since the methods of obtaining these physical realizability conditions have been already discussed elsewhere[4–9] no attempt will be made to work these out here; these conditions will simply be mentioned for every circuit.

The circuit of Fig. 2 can simulate the system of eqn. (6) with a set of Y's as given in eqn. (7), provided that

$$I^3 - 27J^2 < 0 \qquad \ldots\ldots(12)$$

and either $\Delta > 0$

$$\frac{2}{9b_0} > \sqrt[3]{A} + \sqrt[3]{B} \qquad \ldots\ldots(13a)$$

or $\Delta < 0$

$$\frac{1}{9b_0} > \left\{ \begin{array}{l} \sqrt{\dfrac{-p}{3}}\cos\dfrac{\phi}{3} \\ -\sqrt{\dfrac{-p}{3}}\cos\left(\dfrac{\pi}{3} - \dfrac{\phi}{3}\right) \\ -\sqrt{\dfrac{-p}{3}}\cos\left(\dfrac{\pi}{3} + \dfrac{\phi}{3}\right) \end{array} \right. \qquad \ldots\ldots(13b)$$

where

$$I = 9a_2 b_0^2 - \tfrac{3}{2}b_0(a_1 + 7a_2 b_0) + \tfrac{1}{12}(3a_1 b_0 + 1)^2$$

$$J = \begin{vmatrix} 9b_0^2, & -\tfrac{3}{2}b_0, & \tfrac{1}{6}(3a_1 b_0+1) \\ -\tfrac{3}{2}b_0, & \tfrac{1}{6}(3a_1 b_0+1), & -\tfrac{1}{4}(a_1+7a_2 b_0) \\ \tfrac{1}{6}(3a_1 b_0+1), & -\tfrac{1}{4}(a_1+7a_2 b_0), & a_2 \end{vmatrix}$$

$$\Delta = 4p^3 + 27q^2$$

$$p = \frac{a_1}{3b_0} - \frac{1}{3}\left(\frac{1}{3b_0}\right)^2$$

$$q = -\frac{a_2}{3b_0} + \frac{a_1}{27b_0^2} - \frac{2}{27}\left(\frac{1}{3b_0}\right)^3$$

$$A = -\frac{q}{2} + \sqrt{\frac{\Delta}{108}}$$

$$B = -\frac{q}{2} - \sqrt{\frac{\Delta}{108}}$$

$$\phi = \tan^{-1}\left[\frac{-\sqrt{-\Delta}}{q\sqrt{27}}\right]$$

The circuit component values may be determined with the aid of

$$9b_0^2 T_6^4 - 6b_0 T_6^3 + (3a_1 b_0 + 1)T_6^2 - (a_1 + 7a_2 b_0)T_6 + a_2 = 0 \qquad \ldots\ldots(14)$$

$$T_7 = \frac{1}{2}\left[\frac{1}{b_0} - 3T_6\right] \qquad \ldots\ldots(15)$$

and eqn. (10). Having thus determined T_2, T_6, T_7 and choosing arbitrarily a convenient value for any one of the capacitors the remaining component values may be determined with the aid of eqn. (11).

The design procedure would be first to check if inequality (12) and either inequalities (13a) or (13b) are satisfied, the fulfilment of which means that the circuit is physically realizable. The next step then

Table 1

1	2	3	4	5	6	7	8	9	10	11	12	13
Y Parameters / Circuit Case	Y_1	Y_2	Y_3	Y_5	Y_6	Y_7	b_0	b_1	b_2	a_1/b_0	a_2/b_0	Design equations and conditions of physical realizability are given under circuit
I	$\frac{1}{R}$	sc_2	$\frac{1}{R}$	$\frac{1}{R}$	sc_6	sc_7	$\frac{1}{3T_6 + 2T_7}$	0	0	$2T_2T_6 + T_2T_7 + 2T_6T_7$	$T_2T_6T_7$	I
II(a)	$sc_1 + \frac{1}{R}$	$\frac{1}{\alpha R}$	$\frac{1}{R}$	$\frac{1}{R}$	sc_6	sc_7	$\frac{\alpha}{(3\alpha + 2)T_6 + (2\alpha + 1)T_7}$	T_1	0	$2T_1T_6 + T_1T_7 + \left(2 + \frac{1}{\alpha}\right) T_6T_7$	$T_1T_6T_7$	II(a)
II(b)	$\frac{1}{R}$	$\frac{1}{\alpha R}$	$sc_3 + \frac{1}{R}$	$\frac{1}{R}$	sc_6	sc_7	$\frac{\alpha}{(3\alpha + 2)T_6 + (2\alpha + 1)T_7}$	T_3	0	$\left(2 + \frac{1}{\alpha}\right) T_3T_6 + T_3T_7 + \left(2 + \frac{1}{\alpha}\right) T_6T_7$	$T_3T_6T_7$	II(b)
III(a)	$sc_1 + \frac{1}{R}$	$\frac{1}{\alpha R}$	$sc_3 + \frac{1}{R}$	$\frac{1}{R}$	sc_6	sc_7	$\frac{\alpha}{(3\alpha + 2)T_6 + (2\alpha + 1)T_7}$	$T_1 + T_3$	T_1T_3	$2b_1T_6 + b_1T_7 + \frac{1}{\alpha} T_3T_6 + \left(2 + \frac{1}{\alpha}\right) T_6T_7$	$b_2T_6 + b_1T_6T_7$	III(a)
III(b)	$sc_1 + \frac{1}{R}$	$\frac{1}{\alpha R}$	$\frac{1}{R}$	$sc_5 + \frac{1}{R}$	sc_6	sc_7	$\frac{\alpha}{(3\alpha + 2)T_6 + (2\alpha + 1)T_7}$	$T_1 + T_5$	T_1T_5	$2b_1T_6 + T_1T_7 + \frac{1}{\alpha} T_5T_6 + \left(2 + \frac{1}{\alpha}\right) T_5T_7 + \left(2 + \frac{1}{\alpha}\right) T_6T_7$	$b_2(T_6 + T_7) + T_1T_6T_7$	III(b)
III(c)	$\frac{1}{R}$	$\frac{1}{\alpha R}$	$sc_3 + \frac{1}{R}$	$sc_5 + \frac{1}{R}$	sc_6	sc_7	$\frac{\alpha}{(3\alpha + 2)T_6 + (2\alpha + 1)T_7}$	$T_3 + T_5$	T_3T_5	$b_1 \left(2 + \frac{1}{\alpha}\right) T_6 + \left(2 + \frac{1}{\alpha}\right) T_6T_7 + T_3T_7 + \left(2 + \frac{1}{\alpha}\right) T_5T_7$	$b_2(T_6 + T_7) + T_3T_6T_7$	III(c)

would be to solve eqn. (14) and obtain the component values with the aid of eqns. (15), (10) and (11).

Since the design procedure is similar for all the circuits, only the design equations and the physical realizability conditions will be given for each circuit.

For convenience, the network parameters and equations for each circuit are summarized in Table 1; the design equations and the conditions of physical realizability for the corresponding circuits are given in the text.

2.2. $b_2 = 0;\ b_1 > 0,\ b_0 > 0$

The transfer function of eqn. (1) reduces to

$$F_2(s) = -\frac{b_0(b_1 s+1)}{s(a_2 s^2 + a_1 s+1)} \qquad \ldots\ldots(16)$$

It will be evident, from inspection of eqn. 5, that within the restrictions imposed by the choice of the network (shown in Fig. 2) two basic circuits with certain RC combinations are possible and these are listed in case II(a) and (b) of Table I. It may, however, be mentioned that the choice of $Y_2 = 1/\alpha R$ and $Y_3 = Y_5 = 1/R$ has been arbitrarily made. One may choose with equal justification either Y_3 or Y_5 or Y_3 and Y_5 as $1/\alpha R$. In other words, since the number of constants (or constraints) in eqn. (16) is four, whereas the number of parameters in the circuit of Fig. 2 is seven, the choice of resistor values is required to be made so as to reduce the number of circuit parameters to match the number of constants in the given equation. This requirement offers considerable latitude in the choice of intended design values of the resistors as a result of which a number of circuits are possible which are essentially a variation of the two basic circuits listed in case II (a) and (b). Each circuit for its physical realizability, however, will require satisfaction of a set of conditions which will be different for every circuit.

2.2.1. Case II(a)

For this particular choice of RC combination the design equations are

$$4b_0^2 b_1^2 T_6^4 - b_0(2a_1 b_1 - a_2)T_6^3 + a_2(4b_0 b_1+1)T_6^2 - - a_1 a_2 T_6 + a_2^2 = 0 \quad \ldots\ldots(17)$$

$$\alpha = \frac{a_2 T_6}{-2b_0 b_1^2 T_6^2 + (a_1 b_1 - 2a_2)T_6 - a_2 b_1} \quad \ldots(18)$$

together with those listed in columns 8–12. The conditions of physical realizability are

$$\min\left[a_1 b_1,\ \frac{b_1}{12b_0},\ \frac{1}{b_0 b_1}\left(\frac{a_1}{2} - \frac{a_2}{b_1}\right)^2\right] > 2a_2$$

and either $\lambda_2 > \mu_2 > \lambda_1 > \mu_1$

or $\mu_2 > \lambda_2 > \mu_1 > \lambda_1$

or $I^3 - 27J^2 > 0$, $\mu_2 > \lambda_2 > \lambda_1 > \mu_1$

or $I^3 - 27J^2 < 0$, $\lambda_2 > \mu_2 > \mu_1 > \lambda_1$, $\mu_2 > T_6' > \mu_1$

where

$$(\lambda_1, \lambda_2) = \frac{b_1 \mp \sqrt{b_1^2 - 24a_2 b_0 b_1}}{6b_0 b_1}$$

$$(\mu_1, \mu_2) = \frac{(a_1 b_1 - 2a_2) \mp \sqrt{(a_1 b_1 - 2a_2)^2 - 8a_2 b_0 b_1^3}}{4b_0 b_1^2}$$

$$I = 4a_2^2 b_0^2 b_1^2 - \tfrac{1}{4}a_1 a_2 b_0(2a_1 b_1 - a_2) + + \tfrac{1}{12}a_2^2(4b_0 b_1+1)^2$$

$$J = \begin{vmatrix} 4b_0^2 b_1^2, & -\frac{b_0}{4}(2a_1 b_1 - a_2), & \frac{a_2}{6}(4b_0 b_1+1) \\ -\frac{b_0}{4}(2a_1 b_1 - a_2), & \frac{a_2}{6}(4b_0 b_1+1), & -\tfrac{1}{4}a_1 a_2 \\ \frac{a_2}{6}(4b_0 b_1+1), & -\tfrac{1}{4}a_1 a_2, & a_2^2 \end{vmatrix}$$

and, T_6' is the positive real root of eqn. (17).

2.2.2. Case II(b)

The design equations are

$$b_0^2 b_1^2 T_6^4 - b_0(2a_1 b_1 - a_2 - b_1^2)T_6^3 + a_2(2b_0 b_1+1)T_6^2 - - a_1 a_2 T_6 + a_2^2 = 0 \quad \ldots(19)$$

$$\alpha = \frac{2b_0 b_1 T_6^2 + a_2}{-3b_0 b_1 T_6^2 + b_1 T_6 - 2a_2} \qquad \ldots\ldots(20)$$

The conditions of physical realizability are

$$\min\left[a_1 b_1,\ \frac{b_1}{12b_0},\ \frac{1}{b_0 b_1}\left(\frac{a_1}{2} - \frac{a_2}{b_1}\right)^2\right] > 2a_2$$

and either $\lambda_2 > \mu_2 > \lambda_1 > \mu_1$

or $\mu_2 > \lambda_2 > \mu_1 > \lambda_1$

or $(2a_1 b_1 - a_2 - b_1^2) > 0$, $I^3 - 27J^2 > 0$, $\mu_2 > \lambda_2 > \lambda_1 > \mu_1$

or $(2a_1 b_1 - a_2 - b_1^2) > 0$, $I^3 - 27J^2 < 0$, $\lambda_2 > \mu_2 > \mu_1 > \lambda_1$, $\mu_2 > T_6' > \mu_1$

where

$$(\lambda_1, \lambda_2) = \frac{b_1 \mp \sqrt{b_1^2 - 24a_2 b_0 b_1}}{6b_0 b_1}$$

$$(\mu_1, \mu_2) = \frac{(a_1 b_1 - 2a_2) \mp \sqrt{(a_1 b_1 - 2a_2)^2 - 8a_2 b_0 b_1^3}}{4b_0 b_1^2}$$

$$I = a_2^2 b_0^2 b_1^2 - \tfrac{1}{4} a_1 a_2 b_0 (2a_1 b_1 - a_2 - b_1^2) + \tfrac{1}{12} a_2^2 (2b_0 b_1 + 1)$$

$$J = \begin{vmatrix} b_0^2 b_1^2 & -\tfrac{1}{4} b_0 (2a_1 b_1 - a_2 - b_1^2), & \tfrac{1}{6} a_2 (2b_0 b_1 + 1) \\ -\tfrac{1}{4} b_0 (2a_1 b_1 - a_2 - b_1^2), & \tfrac{1}{6} a_2 (2b_0 b_1 + 1), & -\tfrac{1}{4} a_1 a_2 \\ \tfrac{1}{6} a_2 (2b_0 b_1 + 1), & -\tfrac{1}{4} a_1 a_2, & a_2^2 \end{vmatrix}$$

and, T_6' is the positive real root of eqn. (19).

2.3. $b_0 > 0,\ b_1 > 0,\ b_2 > 0$

It can be seen by inspection of eqn. (5) that the network of Fig. 2 will require a minimum of four capacitors for it to be able to simulate the system of eqn. (1) and that three such circuits are possible.

2.3.1. Case III(a)

The design equations are

$$T_1 = \tfrac{1}{2}[b_1 \pm \sqrt{b_1^2 - 4b_2}] \qquad \text{......(21)}$$

$$T_3 = \tfrac{1}{2}[b_1 \mp \sqrt{b_1^2 - 4b_2}] \qquad \text{......(22)}$$

$$P_4 T_6^4 + P_3 T_6^3 + P_2 T_6^2 + P_1 T_6 - a_2^2 b_1 = 0 \qquad \text{...(23)}$$

$$\alpha = \frac{(b_0 T_3^2 T_6 + a_2) T_6}{-A_2 T_6^2 + A_1 T_6 - a_2 b_1} \qquad \text{......(24)}$$

The conditions of physical realizability are

$$b_1^2 \geqslant 4b_2$$

$$\min\left[b_1(a_1 + b_0 b_2), \frac{A_1^2}{2b_1 A_2}\right] > 2a_2$$

$$\min\left[\frac{a_2}{b_0 b_2}, \mu_2\right] > T_6' > \mu_1$$

and,

either

$$\Delta_1 > 0$$
$$\Delta_2 < 0$$
$$\lambda_2 > T_6' > \lambda_1$$

or

$$\Delta_1 > 0$$
$$\Delta_2 > 0$$

and, either $\lambda_2 > T_6' > \lambda_1 > p_2 > p_1$

or $\lambda_2 > \lambda_1 > p_2 > T_6' > p_1$

or $p_2 > T_6' > \lambda_2 > \lambda_1 > p_1$

or $p_2 > \lambda_2 > \lambda_1 > T_6' > p_1$

or $p_2 > T_6' > p_1 > \lambda_2 > \lambda_1$

or $p_2 > p_1 > \lambda_2 > T_6' > \lambda_1$

or $\lambda_2 > T_6' > p_2 > p_1 > \lambda_1$

or $\lambda_2 > p_2 > p_1 > T_6' > \lambda_1$

or $p_2 > T_6' > \lambda_2 > p_1 > \lambda_1$

or $p_2 > \lambda_2 > p_1 > T_6' > \lambda_1$

or $\lambda_2 > T_6' > p_2 > \lambda_1 > p_1$

or $\lambda_2 > p_2 > \lambda_1 > T_6' > p_1$

where

$$A_1 = b_1(a_1 + b_0 b_2) - 2a_2$$

$$A_2 = 2b_0(b_1^2 - b_2)$$

$$P_1 = a_2(A_1 + 2a_2 + b_0 b_1 b_2)$$

$$P_2 = 2a_2 b_0 T_3^2 - A_2 a_2 - A_1 b_0 b_2 - 2a_2 b_0 b_1^2 - a_2(b_1 + 2b_0 b_2)$$

$$P_3 = b_0[2A_1 b_1 + A_2 b_2 + 3a_2 b_1 - T_3^2(b_1 + 2b_0 b_2)]$$

$$P_4 = b_0 b_1 (3T_3^2 b_0 - 2A_2)$$

$$(\lambda_1, \lambda_2) = \frac{(b_1 + 2b_0 b_1) \mp \sqrt{\Delta_1}}{6b_0 b_1}$$

$$(\mu_1, \mu_2) = \frac{A_1 \mp \sqrt{A_1^2 - 4A_2 a_2 b_1}}{2A_2}$$

$$(p_1, p_2) = \frac{b_0 b_2 \mp \sqrt{\Delta_2}}{4b_0 b_1}$$

$$\Delta_1 = (b_1 + 2b_0 b_2)^2 - 24a_2 b_0 b_1$$

$$\Delta_2 = b_0^2 b_2^2 - 8a_2 b_0 b_1$$

and, T_6' is the positive real root of eqn. (23).

2.3.2. Case III(b)

The design equations are

$$T_1 = \tfrac{1}{2}[b_1 \pm \sqrt{b_1^2 - 4b_2}]$$

$$T_5 = \tfrac{1}{2}[b_1 \mp \sqrt{b_1^2 - 4b_2}]$$

$$-b_0^2 T_1^3 T_7^4 + P_3 T_7^3 + P_2 T_7^2 + P_1 T_7 + a_2(A_0 T_5 - 2B_0) = 0 \qquad \text{...(25)}$$

$$\alpha = \frac{a_2(T_7 + T_5)}{-b_0 T_1^2 T_7^2 + B_1 T_7 - B_0} \qquad \text{......(26)}$$

The conditions of physical realizability are

$$b_1^2 \geqslant 4b_2$$

$$\frac{a_2}{b_0 b_2} > T_7'$$

together with

either

$$A_0 > 0$$
$$\Delta_1 > 0$$
$$\Delta_2 > 0$$

and, either $\lambda_2 > T_7' > \lambda_1 > p_2 > p_1$
or $\lambda_2 > \lambda_1 > p_2 > T_7' > p_1$
or $p_2 > T_7' > \lambda_2 > \lambda_1 > p_1$
or $p_2 > \lambda_2 > \lambda_1 > T_7' > p_1$
or $p_2 > T_7' > p_1 > \lambda_2 > \lambda_1$
or $p_2 > p_1 > \lambda_2 > T_7' > \lambda_1$
or $\lambda_2 > T_7' > p_2 > p_1 > \lambda_1$
or $\lambda_2 > p_2 > p_1 > T_7' > \lambda_1$
or $p_2 > T_7' > \lambda_2 > p_1 > \lambda_1$
or $p_2 > \lambda_2 > p_1 > T_7' > \lambda_1$
or $\lambda_2 > T_7' > p_2 > \lambda_1 > p_1$
or $\lambda_2 > p_2 > \lambda_1 > T_7' > p_1$

or
$$\left.\begin{aligned} A_0 &> 0\\ \Delta_1 &> 0\\ \Delta_2 &< 0\\ \lambda_2 > T_7' &> \lambda_1 \end{aligned}\right|$$

or
$$\left.\begin{aligned} A_0 &> 0\\ \Delta_1 &< 0\\ \Delta_2 &> 0\\ p_2 > T_7' &> p_1 \end{aligned}\right|$$

or
$$\left.\begin{aligned} A_0 &< 0\\ \Delta_2 &> 0 \end{aligned}\right|$$

and, either $\lambda_2 > T_7' > p_2 > p_1$
or $\lambda_2 > p_2 > p_1 > T_7'$
or $p_2 > T_7' > \lambda_2 > p_1$
or $p_2 > \lambda_2 > p_1 > T_7'$
or $p_2 > T_7' > p_1 > \lambda_2$
or $p_2 > p_1 > \lambda_2 > T_7'$

or
$$\left.\begin{aligned} A_0 &< 0\\ \Delta_2 &< 0\\ \lambda_2 &> T_7' \end{aligned}\right|$$

and, also

either
$$\left.\begin{aligned} B_0 &> 0\\ B_1 &> 0\\ \Delta_3 &> 0\\ \mu_2 > T_7' &> \mu_1 \end{aligned}\right|$$

or
$$\left.\begin{aligned} B_0 &< 0\\ \mu_2 &> T_7' \end{aligned}\right|$$

where

$$\begin{aligned}
A_0 &= 3a_2 - b_2\\
B_0 &= 2a_2 b_1 - a_1 b_2\\
B_1 &= T_1(a_1 + b_0 b_2) - 2a_2\\
P_1 &= a_2(A_0 + 2B_1) + B_0 b_0 b_2 - a_2 T_5(T_1 + b_0 b_2)\\
P_2 &= 2a_2 b_0 b_2 - 2a_2 b_0 T_1^2 - B_0 T_1 b_0 -\\
&\qquad - a_2(T_1 + b_0 b_2) - b_0 b_2 B_1\\
P_3 &= b_0 T_1^2(a_1 + 2b_0 b_2)
\end{aligned}$$

$$(\lambda_1, \lambda_2) = \frac{(T_1 + b_0 b_2) \mp \sqrt{\Delta_1}}{4b_0 T_1}$$

$$(\mu_1, \mu_2) = \frac{B_1 \mp \sqrt{\Delta_3}}{2b_0 T_1^2}$$

$$(p_1, p_2) = \frac{b_0 b_2 \mp \sqrt{\Delta_2}}{2b_0 T_1}$$

$$\begin{aligned}
\Delta_1 &= (T_1 + b_0 b_2)^2 - 8A_0 T_1 b_0\\
\Delta_2 &= b_0^2 b_2^2 - 8a_2 b_0 T_1\\
\Delta_3 &= B_1^2 - 4B_0 T_1^2 b_0
\end{aligned}$$

and, T_7' is the positive real root of eqn. (25).

2.3.3. Case III(c)

The design equations are

$$T_3 = \tfrac{1}{2}[b_1 \pm \sqrt{b_1^2 - 4b_2}]$$

$$T_5 = \tfrac{1}{2}[b_1 \mp \sqrt{b_1^2 - 4b_2}]$$

$$\begin{aligned}
-b_0^2 T_3^3 T_7^4 + a_1 b_0 T_3^2 T_7^3 + P_2 T_7^2 + P_1 T_7 +\\
+ a_2(A_0 b_1 + 2B_0) = 0 \quad \ldots(27)
\end{aligned}$$

$$\alpha = \frac{B_2 T_7 + a_2 b_1}{-b_0 T_3^2 T_7^2 + B_1 T_7 + B_0} \qquad \ldots\ldots(28)$$

The conditions of physical realizability are

$$\left.\begin{aligned} b_1^2 &\geqslant 4b_2\\ \frac{a_2}{b_0 b_2} &> T_7' \end{aligned}\right|$$

together with

either
$$\left.\begin{aligned} A_0 &> 0\\ \Delta_1 &> 0\\ \Delta_2 &> 0 \end{aligned}\right|$$

and, either $\lambda_2 > T_7' > \lambda_1 > p_2 > p_1$
or $\lambda_2 > \lambda_1 > p_2 > T_7' > p_1$
or $p_2 > T_7' > \lambda_2 > \lambda_1 > p_1$
or $p_2 > \lambda_2 > \lambda_1 > T_7' > p_1$
or $p_2 > T_7' > p_1 > \lambda_2 > \lambda_1$
or $p_2 > p_1 > \lambda_2 > T_7' > \lambda_1$

or $\lambda_2 > T_7' > p_2 > p_1 > \lambda_1$

or $\lambda_2 > p_2 > p_1 > T_7' > \lambda_1$

or $p_2 > T_7' > \lambda_2 > p_1 > \lambda_1$

or $p_2 > \lambda_2 > p_1 > T_7' > \lambda_1$

or $\lambda_2 > T_7' > p_2 > \lambda_1 > p_1$

or $\lambda_2 > p_2 > \lambda_1 > T_7' > p_1$

or $A_0 > 0$
$\Delta_1 > 0$
$\Delta_2 < 0$
$\lambda_2 > T_7' > \lambda_1$

or $A_0 > 0$
$\Delta_1 < 0$
$\Delta_2 > 0$
$p_2 > T_7' > p_1$

or $A_0 < 0$
$\Delta_2 > 0$

and, either $\lambda_2 > T_7' > p_2 > p_1$

or $\lambda_2 > p_2 > p_1 > T_7'$

or $p_2 > T_7' > p_1 > \lambda_2$

or $p_2 > p_1 > \lambda_2 > T_7'$

or $p_2 > T_7' > \lambda_2 > p_1$

or $p_2 > \lambda_2 > p_1 > T_7'$

or $A_0 < 0$
$\Delta_2 < 0$
$\lambda_2 > T_7'$

and, also either $B_0 < 0$
$B_1 > 0$
$B_2 > 0$
$\Delta_3 > 0$
$\mu_2 > T_7' > \mu_1$

or $B_0 > 0$
$B_2 > 0$
$\mu_2 > T_7'$

or $B_0 < 0$
$B_1 > 0$
$B_2 < 0$
$\Delta_3 > 0$

and, either $T_7' > \mu_2 > \mu_1 > q$

or $\mu_2 > \mu_1 > T_7 > q$

or $T_7' > q > \mu_2 > \mu_1$

or $q > \mu_2 > T_7' > \mu_1$

or $T_7' > \mu_2 > q > \mu_1$

or $\mu_2 > q > T_7' > \mu_1$

or $B_0 > 0$
$B_2 < 0$

and, either $T_7' > q > \mu_2$

or $q > \mu_2 > T_7'$

or $T_7' > \mu_2 > q$

or $\mu_2 > q > T_7'$

or $B_0 < 0$
$B_2 < 0$
$T_7' > q$

and, either $B_1 < 0$

or $B_1 > 0$
$\Delta_3 < 0$

where

$$A_0 = 3a_2 - b_2$$
$$B_0 = a_1 b_2 - 2a_2 b_1$$
$$B_1 = T_3(a_1 + b_0 b_2) - 2a_2$$
$$B_2 = a_2 - b_0 b_2 T_3$$
$$P_1 = 2a_2 B_1 + A_0 B_2 - b_0 b_2 B_0 - a_2 b_1(b_0 b_2 + T_3)$$
$$P_2 = b_0 T_3(2a_2 b_1 + B_0) - 2a_2 b_0 T_3^2 - b_0 b_2 B_1 - B_2(b_0 b_2 + T_3)$$

$$(\lambda_1, \lambda_2) = \frac{(b_0 b_2 + T_3) \mp \sqrt{\Delta_1}}{4b_0 T_3}$$

$$(\mu_1, \mu_2) = \frac{B_1 \mp \sqrt{\Delta_3}}{2b_0 T_3^2}$$

$$(p_1, p_2) = \frac{b_0 b_2 \mp \sqrt{\Delta_2}}{2b_0 T_3}$$

$$q = -\frac{a_2 b_1}{B_2}$$

$$\Delta_1 = (b_0 b_2 + T_3)^2 - 8A_0 b_0 T_3$$
$$\Delta_2 = b_0^2 b_2^2 - 8a_2 b_0 T_3$$
$$\Delta_3 = B_1^2 + 4b_0 T_3^2 B_0$$

and, T_7' is the positive real root of eqn. (27).

3. Conclusions

Since the basic network has been assumed *a priori*, the simulation of any particular type of the third-order linear system is possible subject to the fulfilment of a set of conditions which are listed for every one of the circuits presented in the text.

This method of simulation, however, has certain disadvantages, the most important ones being that the coefficient of the differential equation cannot be as easily varied, the initial conditions cannot be conveniently introduced and the method requires a little more time for calculating the circuit parameter values. But because it effects savings in active and passive circuit elements and increases the reliability and accuracy of the simulator and reduces its expense and size the method is considered to possess some merit.

4. Acknowledgment

The author wishes to thank the Director of Electronics, Defence Research and Development Organization, for his permission to publish this paper.

5. References

1. H. D. Huskey and G. A. Korn, "Computer Handbook" (McGraw-Hill, New York, 1962).
2. A. Bridgeman and R. Brennen, "Simulation of transfer functions using only one amplifier", *I.R.E. Wescon Conv. Rec.*, **5**, Part IV, pp. 273–8, 1957.
3. R. J. A. Paul, "Simulation of rational transfer functions with adjustable coefficients", *Proc. Instn Elect. Engrs*, **110**, No. 4, pp. 671–9, April 1963.
4. L. K. Wadhwa, "Simulation of third-order systems with one operational amplifier", *Proc. Natl Inst. Sci. (India)*, **29A**, No. 2, pp. 213–25, March 1963.
5. L. K. Wadhwa, "Simulation of third-order systems with a simple lead using one operational amplifier", *Proc. Inst. Radio Engrs*, **50**, No. 4, p. 465, April 1962 (Letter).
6. L. K. Wadhwa, "Simulation of third-order systems with a double lead using one operational amplifier", *Proc. Natl Inst. Sci. (India)*, **29A**, No. 1, pp. 46–64, January 1963.
7. L. K. Wadhwa, "The simulation of third-order systems with a leading time constant by a single operational amplifier", Sixth Technical Convention of Institution of Telecommunications Engineers, New Delhi, 8th December 1962.
8. L. K. Wadhwa and H. S. Rao, "One operational amplifier simulates third-order systems with a lead and a time constant", *J. Instn Telecomm. Engrs*, **8**, No. 6, pp. 281–7, November 1962.
9. L. K. Wadhwa, "Simulation of general third-order systems by a single operational amplifier", Indian Science Congress, New Delhi, October 1963, Session 50. (To be published in *Proc. Natl Inst. Sci. (India)*.)

Manuscript first received by the Institution on 2nd January 1964 and in final form on 3rd March 1964. (Paper No. 906/C71.)

U.D.C. 621.316.86:621.391.822:661.665

Noise in Silicon Carbide Non-linear Resistors

By

A. G. J. HOLT, Ph.D., Dip.El. (*Associate Member*),†

P. L. BAINBRIDGE, B.Sc.†

AND

F. W. STEPHENSON, B.Sc.†

Summary: This paper deals with the noise arising in excess of thermal noise when a steady current is passed through a silicon-carbide non-linear resistor. Curves are given showing the variation of excess noise with current for four different resistors. Pulses of an amplitude considerably greater than the r.m.s. value of the excess noise are found to occur. These are photographed and their amplitude is measured.

1. Introduction

Much work has already been done on the study of noise in linear resistors.[1-6] It is the purpose of this paper to give the results of some measurements of excess current noise and noise bursts which occur when current is passed through silicon-carbide non-linear resistors.

2. General Properties of Silicon-carbide Resistors

The electrical properties of silicon carbide and the voltage/current characteristics of silicon-carbide resistors have been described by Ashworth, Needham and Sillars.[7]

The relation between current and voltage for these resistors can be expressed by the equation:

$$I = HV^{\alpha}$$

Here I = current through the resistor,

V = voltage drop across the resistor,

H and α are constants for a given resistor. (One manufacturer quotes typical values of α as being in the range 2 to 6.)

This relation is empirical and approximate; nevertheless it is sufficiently accurate for many purposes over a range of nearly 1000 : 1 in current.

3. Noise in Silicon-carbide Resistors

When these resistors are used in small-signal circuits the question of the electrical fluctuations generated when current passes through the resistors can become important.

Measurements of the excess noise produced were carried out for a range of steady currents and the results were compared with the noise produced by the carbon composition resistors having the same ohmic value as the nonlinear resistor at each current value. Figure 1 shows both results plotted on the same graph.

These measurements were carried out with an apparatus having a bandwidth of 3 kc/s centred at 85 kc/s.

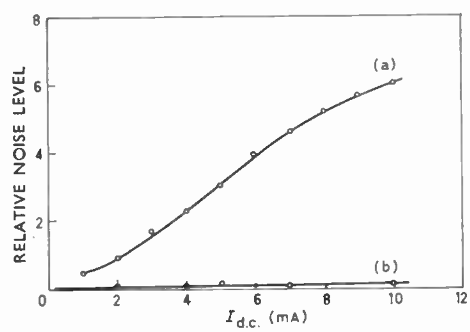


Fig. 1. Graph of excess noise vs steady current.

It is evident that the excess noise generated by the silicon carbide resistor increases rapidly when the current through it is increased. Over most of the range of measurement the mean squared value of noise is greater than that produced by the carbon composition resistors by a factor of at least 20. When it is remembered that carbon composition resistors generate more excess noise than any of the other types of resistor in common use the importance to circuit designers of noting the effects of noise arising in the silicon-carbide resistors is underlined.

In view of the very noisy character of these non-linear resistors it was decided to measure the noise generated when steady currents of the order of 1 mA or less were passing through them.

Figure 2 shows the voltage/current relationship obtained for four resistors. The specimens are numbered, 1–4, the latter being a solid disk, whilst the others are annular disks. The dimensions are given in Table 1.

† Electrical Engineering Department, University of Newcastle upon Tyne.

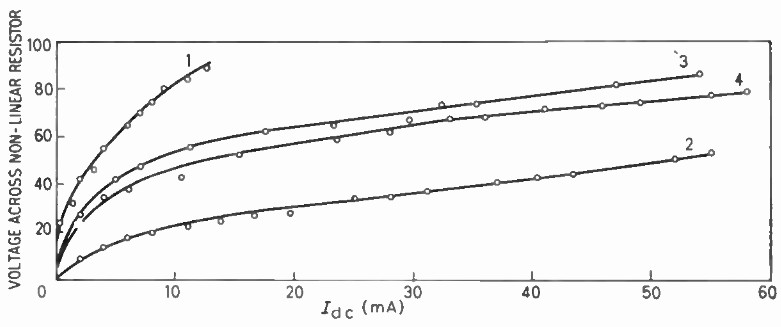


Fig. 2. Voltage/current relationship of four silicon-carbide non-linear resistors.

Table 1

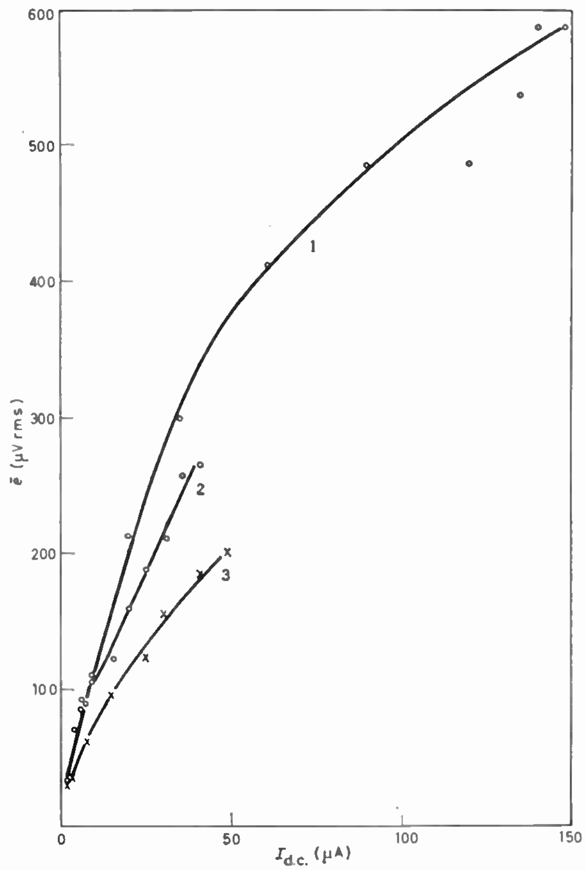
Resistor No.	Outer Diam.	Inner Diam.	Thickness
1, 2 and 3	1·0 in	0·25 in	0·031 in
4	1·0 in		0·031 in

Noise measurements were carried out at frequencies in the audio range; the amplifier in the measuring apparatus had a maximum voltage gain of 60 dB, the gain being 3 dB below the maximum value at 2 and at 20 kc/s. The noise of the amplifier was negligible when compared with that arising in the non-linear resistors.

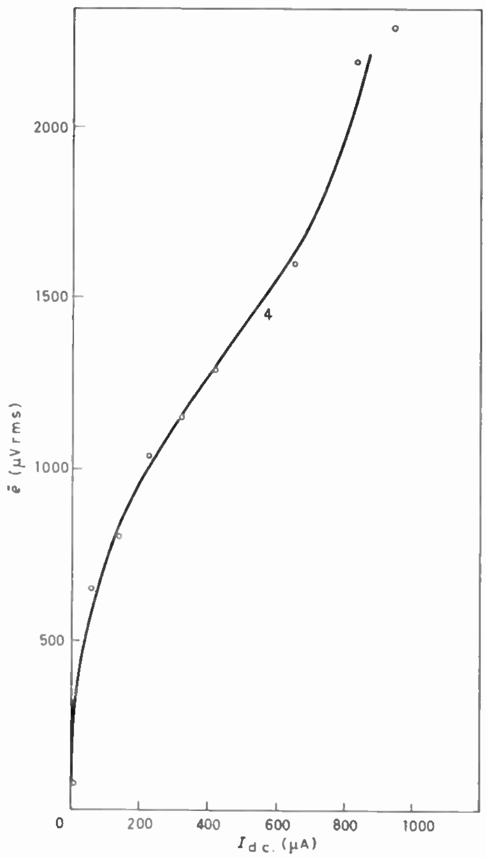
4. Results

4.1. *R.M.S. Noise*

Curves showing the r.m.s. noise voltage generated by the non-linear resistors nos. 1, 2 and 3 when passing steady currents in the range zero to 150 μA appear in Fig. 3(*a*).

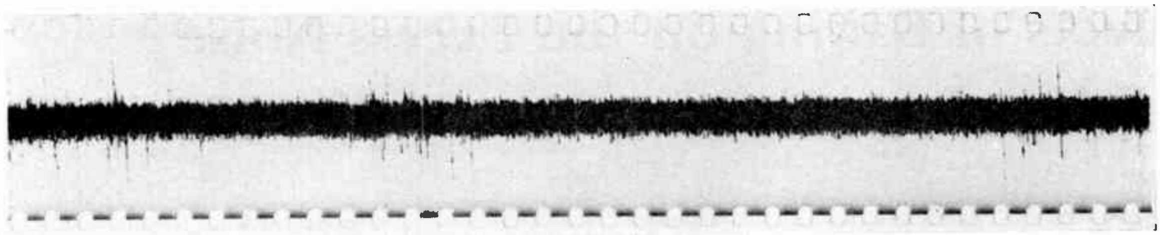


(*a*) Resistors 1, 2 and 3, steady currents up to 150 μA.

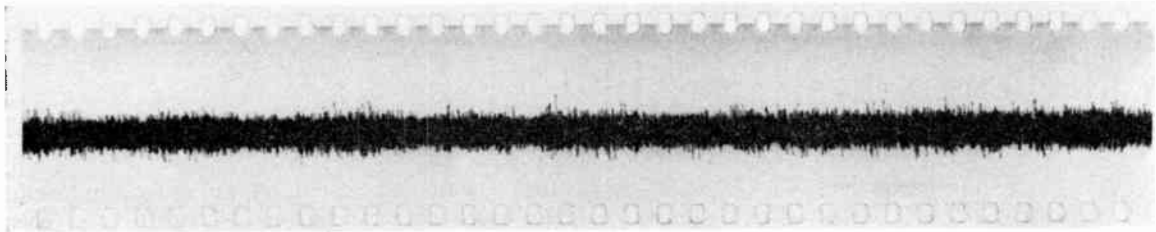


(*b*) Resistor 4, steady current up to 1 mA.

Fig. 3. Graph of r.m.s. noise voltage generated by non-linear resistors.



(*a*) Bursts of 350 → 500 μV peak-to-peak for an r.m.s. voltage of 104 μV and current 9·4 μA.



(*b*) Bursts of 160 → 210 μV peak-to-peak for an r.m.s, voltage of 94 μV and current 15·2 μA.

Fig. 4. Oscilloscope traces showing 'bursts'.

These curves are of the same general form as the steady voltage/current curves for these three resistors. The increasing noise for increasing current, even of the very small magnitude shown here, makes it clear that excess noise arising in silicon-carbide resistors is of such a magnitude that it may make an important contribution to the total noise output from an amplifier if a resistor of this type is used in the early stages of the amplifier.

A curve for the r.m.s. noise produced by the resistor no. 4 is shown in Fig. 3(*b*). Here the noise tends to increase more rapidly for higher current values than was the case for resistors 1, 2 and 3. For steady currents of less than 600 μA however, the curve shape is the same as that obtained for resistors nos. 1, 2 and 3.

4.2. *Bursts*

These are irregularly-spaced pulses of an amplitude much greater than the r.m.s. value of the excess current noise and which are superimposed upon it. This phenomenon is known to occur in carbon resistors and semiconductor diodes and has been studied by Pay.[8]

It was found that two of the four silicon-carbide resistors tested produced bursts of considerably greater amplitude than the r.m.s. excess noise. Figures 4(*a*) and (*b*) show the large bursts produced by resistor no. 1 and, in contrast, the comparatively burst-free noise output from resistor no. 3.

5. Acknowledgments

The authors express their thanks to Professor R. L. Russell, Professor of Electrical Engineering in the University of Newcastle upon Tyne, for his encouragement and for the use of the facilities of the Department.

6. References

1. J. B. Johnson, "Thermal agitation of electricity in conductors", *Phys. Rev.*, **32**, p. 97, 1928.
2. H. Nyquist, "Thermal agitation of electric charge in conductors", *Phys. Rev.*, **32**, p. 110, 1928.
3. J. Bernamont, "Fluctuations de potential aux bornes d'un conducteur metallique de faible volume parcouru par un courant", *Ann. de Phys.*, 7, p. 71, 1937.
4. D. A. Bell and K. Chong, "Current noise in composition resistors", *Wireless Engineer*, **31**, p. 142, 1954.
5. P. L. Kirby and G. P. Sibilia, "Investigations of current noise in fixed resistors", *Proceedings of Symposium on Noise in Fixed Resistors*, Malvern, March 1959, p. 1.
6. D. A. Bell, "Electrical Noise", (Van Nostrand, London, 1960).
7. F. Ashworth, W. Needham and R. W. Sillars, "Silicon carbide non-ohmic resistors", *J. Instn Elect. Engrs*, **93**, Part 1, p. 385, 1946.
8. R. G. Pay, "Excess Noise in Semiconductors", M.Sc. Thesis, University of Birmingham, 1956.

Manuscript first received by the Institution on 11th January 1963 and in final form on 30th May 1963. (Contribution No. 76.)

U.D.C. 621.316.86: 621.391.822

Effects of Baking on the Excess Noise Produced by Carbon Composition Resistors

By

A. G. J. HOLT, Ph.D., Dip.El.
(*Associate Member*)†,

P. L. BAINBRIDGE, B.Sc.†

AND

F. W. STEPHENSON, B.Sc.†

Summary: The results of experiments carried out to show the effects of baking process on the noise in excess of thermal noise produced in carbon composition resistors are given. Resistors of ½ watt and 1 watt rating, having various nominal resistance values, were tested and curves are given showing the variation in noise with steady current.

1. Introduction

It has been found‡ that the effect of baking certain carbon resistors rated at ½ watt is to reduce the noise arising in the resistors when they carry a steady current.

This paper reports the effect of baking on the noise in excess of thermal noise produced by carbon resistors when a controlled current is passed through them. The baking process used for this work consists in placing the resistors in an oven at a temperature of the order of 110° C for a period of 120 hours. After baking, the resistors were removed from the oven and coated with a quick-drying varnish in order to protect them from the effects of moisture. When they had cooled to room temperature a current was passed through them and the excess noise generated in the resistors was measured.

One group of resistors was tested with an apparatus having a frequency range centred at 85 kc/s and a bandwidth of 3 kc/s between the frequencies at which the response was 3 dB below the maximum. The other group was tested on an apparatus having a bandwidth extending from 1·7 to 16 kc/s (between points 3 dB below maximum).

Details of the physical size and wattage ratings of the resistors tested are given in Table 1.

In all the resistors tested the change in resistance produced by the baking process was small.

Table 1

	Rating (watts)	Length (inches)	Diameter (inches)
Group A (insulated)	½	0·45	0·22
	1	0·71	0·25
Group B (insulated)	½	0·4	0·15
	1	0·56	0·21

2. Results

The low-frequency measuring apparatus was used to obtain the curves shown in Figs. 1, 2 and 3 for the variation of excess noise with steady current through the resistors in Group A. These curves are typical

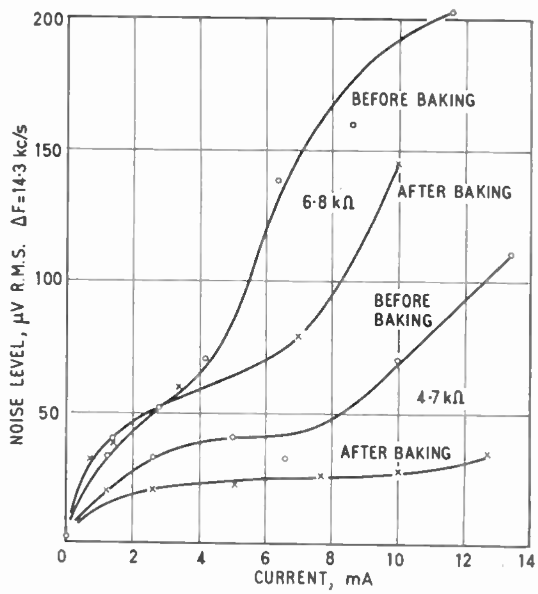


Fig. 1. Variation of noise level with current. Group A. ½ watt, 4·7 kΩ and 6·8 kΩ. Measuring amplifier bandwidth 14·3 kc/s.

† Department of Electrical Engineering, University of Newcastle upon Tyne.

‡ W. A. Ostaff, "An accelerated aging and coating procedure for lowering current noise in carbon composition resistors", (Letter), *Proc. Inst. Radio Engrs*, **45**, pp. 691–2, May 1957.

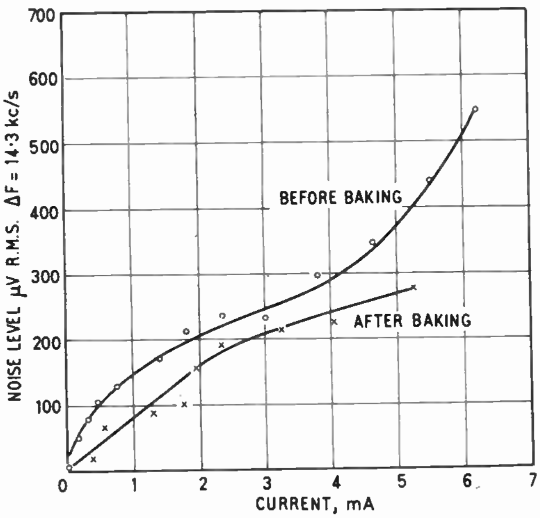


Fig. 2. Variation of noise level with current. As Fig. 1 but for 22 kΩ resistors.

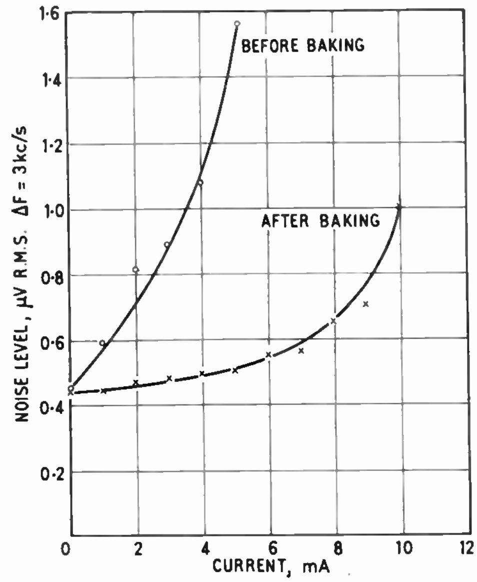


Fig. 4. Variation of noise level with current. Group B. Measuring amplifier bandwidth 3 kc/s. ½ watt, 4·7 kΩ.

of the results obtained. Figures 1 and 2 show that for the ½-watt resistors the excess noise measured after the baking process is less than that measured before baking. It was not possible to predict from an examination of the curve obtained before baking what the degree of reduction in the excess noise brought about by the baking process would be. However, some reduction was usually obtained after the baking had been carried out.

When 1-watt resistors were tested it was found that, although the excess noise characteristics were changed by baking, a reduction was much less likely to occur with these resistors than with the ½-watt specimens. Figure 3 shows a typical curve for which the excess noise after baking is slightly greater than that produced before the baking process for small values of steady current but is less than the pre-baking noise when the steady current is increased.

Some results of measurements made using the high-frequency apparatus are given in Figs. 4 and 5. These curves show the excess noise produced by ½-watt resistors from Group B before and after baking. It is evident that an appreciable noise reduction has been brought about by baking. Measure-

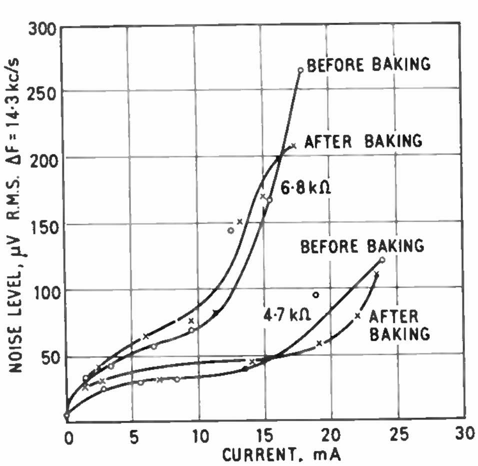


Fig. 3. Variation of noise level with current. As Fig. 1 but for 1 watt resistors.

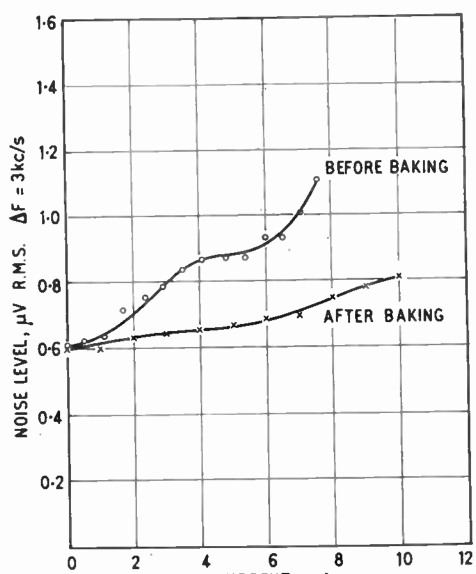


Fig. 5. Variation of noise level with current. As Fig. 4 but for 10 kΩ resistors.

ments were made on 1-watt resistors in Group B with results similar to those obtained for 1-watt resistors in Group A. The excess noise characteristics were changed by baking but a reduction in the noise generated for a particular value of steady current was not always obtained.

It was found that bursts of noise pulses occurred when current passed through the resistors, the pulse amplitude becoming greater when the steady current was increased. When the excess noise arising in resistors which had been baked was observed on an oscillograph it was found that noise pulses were still present. However it was frequently observed that they appeared when larger steady currents were passed through the resistor than were required to give bursts before the baking process had been carried out.

3. Conclusions

It is evident from the results obtained that the excess current noise generated in carbon composition resistors of $\frac{1}{2}$-watt rating can be considerably reduced by baking. This effect can be explained as follows.

The resistors are made from carbon, together with fillers and bonding resins. During the manufacturing process resistors are 'cured' by heating them for a short time. The baking was carried out at a lower temperature but for a period considerably longer than the curing during manufacture. Hence it is to be expected that further curing of the bonding resins and fillers will occur during baking with the result that contacts between the carbon particles in the resistor will be improved. This will reduce the excess noise produced by the resistors.

Moisture present in the resistors will be driven out by the baking and kept out by the layer of insulating varnish. This must result in some improvement in the contacts between adjacent carbon particles in the bulk of the resistor and also in the contacts with the connecting wires.

The small improvement obtained with the 1-watt resistors can be explained because the greater volume of these resistors would be expected to require baking for a longer period or at a higher temperature than the $\frac{1}{2}$-watt specimens to produce a reliable reduction in noise. It is hoped to carry out tests of a longer duration on 1-watt resistors at a later date.

4. Acknowledgments

The authors have pleasure in thanking Professor R. L. Russell, Professor of Electrical Engineering in the University of Newcastle upon Tyne, for his encouragement and for the use of the facilities of the Department.

It is also a pleasure to acknowledge stimulating discussions with Mr. C. A. Wolsey and Mr. K. Stark of the Morganite Resistors Limited of Jarrow, and Dr. P. L. Kirby and Mr. S. Osborne of Welwyn Electric Limited of Bedlington.

Manuscript first received by the Institution on 3rd September 1963 and in revised form on 4th January 1964. (Contribution No. 77.)

U.D.C. 534.844/5: 621.395.625.3

Correlation Techniques in Studio Testing

By

A. N. BURD, B.Sc.†

Presented at a meeting of the Electro-Acoustics Group in London on 12th February 1964.

Summary: Correlation techniques allow the separation of signals which have arrived by paths of different transit times or originated from different sources. This ability is of assistance in certain types of acoustic testing in studios or other enclosures.

The construction of an analogue correlator to derive the correlation function between two signals is described. Its application both in laboratory and in field measurements has confirmed and supplemented results obtained by other techniques. Sound reduction measurements and the identification of flanking paths have proved to be the area in which the technique has the greatest value. Details of laboratory and field measurements are included in the paper.

List of Symbols

c	Velocity of sound in air
$f_1(t), f_2(t)$	Time varying functions
M	Area density of a panel
P_i	Incident sound pressure
P_r	Reflected sound pressure
P_t	Transmitted sound pressure
r	Amplitude reflection coefficient of a surface
RC	Integrating time of low-pass filter
T	Integrating time in derivation of correlation function
$\phi_{21}(\tau)$	Cross correlation function
ρ_0	Density of air
τ	Delay time
ω	Angular frequency
$\Delta\omega$	Bandwith common to input signals to correlator

1. Introduction

In many respects broadcasting studios represent the most stringent requirements in acoustical design. It is necessary to provide a suitable acoustical environment to stimulate the people involved and it is further necessary that the programme shall be clearly intelligible to the listeners. Until the fuller introduction of stereophonic broadcasting a single chain exists between the programme source and the listener, who is thus prevented from using his binaural facility to differentiate between wanted and unwanted sounds; for example no separation can be obtained between a speaker's voice and extraneous noises or excessive reverberation.

† British Broadcasting Corporation, Research Department, Kingswood Warren, Surrey.

This is not the place to discuss in detail the points which are implicit in these requirements but they necessitate a close control over the reverberation times of enclosures at all frequencies, the provision of acoustic isolation between neighbouring areas, the reduction of noise arising within the studios and similar measures.

Many techniques have been developed to provide such information and a notable addition to the range of equipment was made in 1955 when K. W. Goff[1, 2] described the construction and application to acoustics of an analogue electronic correlator. A correlator of basically the same type has been constructed in the B.B.C. Research Department and has been operated intermittently in the intervening years. Some of the results obtained will be used to illustrate the application and limitations of the equipment.

The theory of correlation is a familiar tool to the statistician who uses it to express the degree of dependence of one set of variables upon a second set. In time series the correlation which chiefly comes into play is that between a certain sequence of values and the same or another sequence of values after a shift in time. Where we are dealing with continuous data we consider the time function f(t) and we define the cross-correlation function between two such time series with a time delay applied to one as

$$\phi_{21}(\tau) = \lim_{T\to\infty} \frac{1}{T}\int_0^T f_1(t)\, f_2(t-\tau)\, dt \quad \ldots.(1)$$

If the correlation function has a maximum at some delay τ_1 then it will be found to have a finite value for a range of delay times around τ_1. The range of delay times is inversely proportional to the bandwidth common to both input signals. Thus, provided the integrating time of the equipment is large compared with this time interval, the error introduced by a

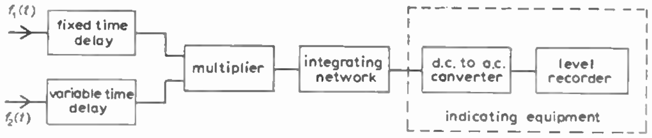


Fig. 1. Block schematic of correlator.

finite time integration will be small. The correlation function is taken as

$$\phi_{21}(\tau) = \frac{1}{T}\int_0^T f_1(t)\, f_2(t-\tau)\, dt \quad \text{for } T \gg 1/\Delta\omega \qquad \ldots.(2)$$

where $\Delta\omega$ is the bandwidth common to both $f_1(t)$ and $f_2(t)$.

The correlator shown in schematic form in Fig. 1 performs the operations defined in (2) directly on the input signals.

A simple low-pass filter of suitable time-constant carries out the integration. The method chosen for the multiplication of the functions is that known as a quarter-squaring multiplier which derives the product from the relationship

$$f_1(t).f_2(t-\tau) = \tfrac{1}{4}\{[f_1(t)+f_2(t-\tau)]^2-[f_1(t)-f_2(t-\tau)]^2\}$$

which can be easily verified by binomial expansion of the squares.

The construction of the correlator is described in greater detail in the Appendix, particular attention being paid to the points at which it differs from that described by Goff.

2. The Operation of the Correlator

2.1. *The Effects of Filtering on the Correlation Function*

The autocorrelation function of a pure tone has the same periodicity as the tone and therefore, since there will be a maximum each time the path difference between the signals is a multiple of a wavelength, no information on the time delay corresponding to different path lengths can be obtained. The autocorrelation function of a random signal of infinite bandwidth has a finite value only at zero relative delay. When the bandwidth of one of the signals applied to the correlator input is limited, the correlation function takes the form of a damped oscillation in which the period of the oscillation is that of the centre frequency of the pass-band while the overall width of the pattern is a function of the bandwidth of the signal.

Information can now be derived on the behaviour of this band of frequencies but the ability to discriminate between delay times is restricted by the necessity to resolve adjacent groups of peaks. Even when it is possible to see that two groups of peaks exist, the peak heights may be modified by the skirts of the adjacent group. Thus by adjusting the bandwidth a compromise between frequency discrimination and path length information can be achieved.

2.2. *Limitations in the Application of the Correlator*

There remain two major limitations to the application of the correlator in studio testing. The first is the physical size of the equipment which could not readily be transported to studio centres around the country. This limitation was overcome in the first instance by the use of the lines which interconnect most B.B.C. studios. Measurements were successfully carried out on one occasion lasting several days in which this method was employed. However, the second limitation is the time involved in these measurements, which remains excessive since it can take at least 30 minutes to correlate over a time delay corresponding to a path length of 40 ft at a rate suitable for a high signal/noise ratio. Such periods of time as would be required for measurements in several frequency ranges are not normally available in studios and we have therefore had to rely on recordings of test signals which can be brought back to the laboratory for analysis. A twin-track tape recorder enables the phase relationship between the two channels of information to be maintained to an adequate degree. It is not necessary to have a continuous recording of the noise, a loop of duration greater than the integrating time of the correlator being sufficient; this is replayed continuously to the analysis equipment.

Since most high quality loudspeakers contain cross-over networks which will provide unknown phase shifts at mid-frequencies, it has become general practice to use such a speaker for the lower frequency measurements only and a horn-loaded pressure unit for the high frequencies. Recordings are therefore made with each loudspeaker and the necessary microphone positions.

3. Application of the Correlator to Sound Reduction Measurements

The sound transmission coefficient of a partition is defined as the ratio of the energy transmitted through the partition to the energy incident on it, and the sound reduction factor is

$$10 \log_{10} (1/\text{sound transmission coefficient})$$

For the simple case of a free travelling sound wave the sound reduction factor is $20 \log_{10} (P_i/P_t)$ where P_i is the incident sound pressure and P_t the transmitted sound pressure.

In the case of a single mass-controlled limp partition, the sound reduction factor can be approximated at normal incidence by the expression $20 \log_{10} (\omega M/2\rho_0 c)$ where ω is 2π times the frequency,

M is the area density of the partition, $\rho_0 c$ the characteristic impedance of air. For a given material this expression will be seen to increase by 6 dB for each doubling of the frequency or for a given frequency to increase by 6 dB for each doubling of the mass.

The use of a correlator enables a determination of the incident and transmitted sound pressures to be made under conditions where conventional measurements are impossible or at least would give misleading answers. The ability to differentiate between sound which has arrived by paths of different lengths enables us to carry out measurements on samples of limited size since the energy which has been transmitted through the sample can be separated from that which has come round the edges.

An alternative method has been proposed by Professor Raes[3] in which a short pulse is radiated and picked up by microphones on each side of the partition. The signals, suitably amplified, are exhibited on an oscilloscope and the amplitude of the pulses arriving with different transit times measured from photographs of the traces. This method is equivalent to correlation and for a given bandwidth of signal has the same ability to separate paths of different lengths in the absence of extraneous noise. The signal/noise ratio obtainable in correlation is, however, considerably better than with the short pulse method.

A loudspeaker driven from a random noise generator is used as the correlation source, and Fig. 2 shows two possible ways of operating the equipment. The first uses two microphones, the electrical outputs of which, after suitable amplification, are applied to the correlator. As the time delay is varied a maximum of the correlation function will be obtained for each delay corresponding to a path traversed by sound between the two microphones. An improved method in which the output of a single microphone is correlated against the electrical input to the loudspeaker has considerable advantages in protecting against spurious reflections within the source room and, of course, simplifies the equipment necessary.

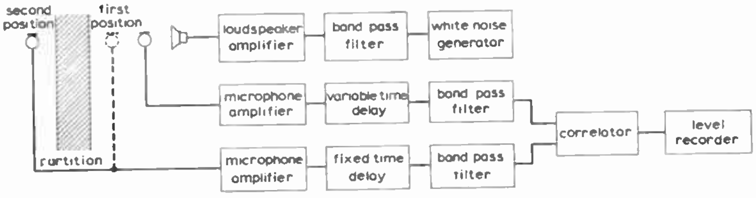
3.1. *Measurements on a Small Panel*

The sound reduction factor of a limp panel of lead-loaded p.v.c. was determined both by the correlator and by conventional measurements. The sample was of limited size (6 ft 6 in × 2 ft 9 in) and correlator results could be obtained only for normal incidence where the line joining loudspeaker and microphone is normal to the panel.

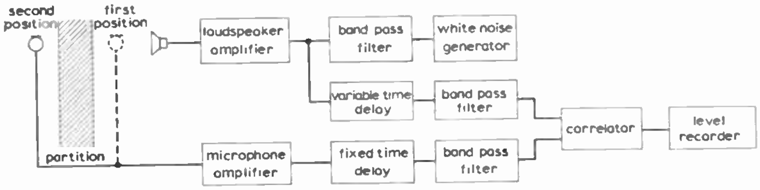
A microphone was placed approximately 30 in from a horn-loaded pressure unit and a series of correlator traces made in different octave bands. The test panel was then placed between the loudspeaker and the microphone and the procedure repeated. After allowing for changes of microphone amplifier gain, the change in height of the correlation peak which arrived after the same time delay as found previously represented the change of sound pressure level at that point.

In fact the material proved to be of too high a sound reduction factor for straightforward measurements, the reverberant energy making recognition of the direct peaks impossible. The panel was fitted loosely into a doorway to reduce the reverberant energy and the results obtained are shown in Fig. 3 (lower curve). For comparison purposes the theoretical mass law values are shown on the same graph.

In Fig. 4, measurements made by the conventional method are compared with the same theoretical value. In this case the door-shaped panel had to be securely



(*a*) Correlation of two microphone outputs.



(*b*) Correlation of loudspeaker input and microphone output.

Fig. 2. Block schematic of correlator measurements.

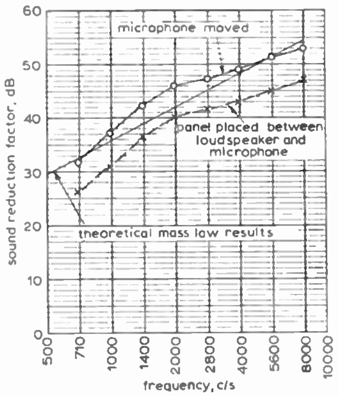


Fig. 3. Sound reduction factor of lead-loaded p.v.c. membrane as determined by correlator measurement.

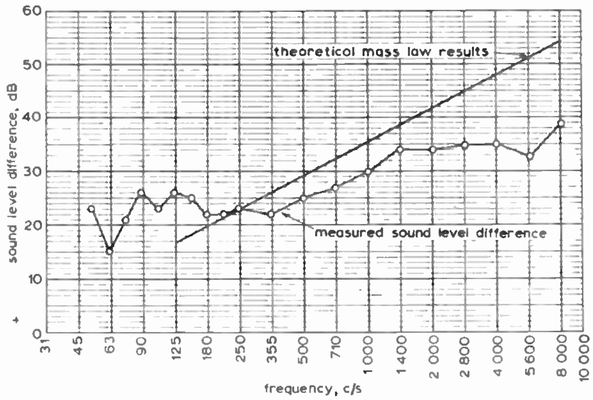


Fig. 4. Sound level difference given by lead-loaded p.v.c. membrane.

fastened in a door opening. A diffuse sound field existed on the source side of the panel and theory would indicate that for random incidence on a mass-controlled panel the results would lie 6–10 dB below the normal incidence case. In the region from 350 c/s to 2 kc/s this result holds fairly well; above 2 kc/s the reduction of the measured values below those expected could still be due to the difficulties of providing adequate seals around the edges of the sample.

If the partition which is to be measured is not portable, then a determination of the incident energy has to be made with the microphone placed first on the source side. At low frequencies difficulties will arise in the separation of the direct and reflected sound and it will become necessary to increase the spacing of the microphone from the wall.

In addition, a factor due to the divergence of the energy from the loudspeaker enters. In typical cases a difference of 6–8 dB in sound pressure might be found between the two microphone positions. The sound pressure incident on the panel will be between the two values found. The measurements on the lead-loaded p.v.c. panel were repeated in this way and the results were found to be greater by the expected amount than those measured previously (see upper curve in Fig. 3).

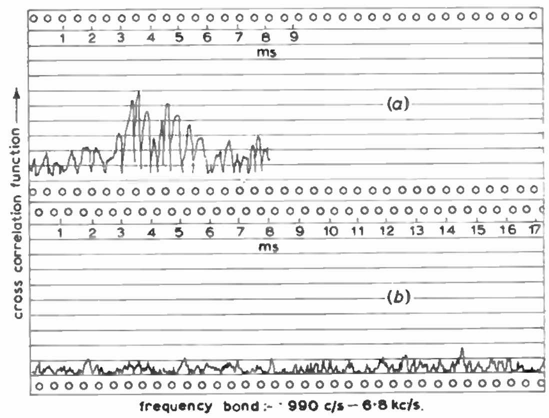
For octave bandwidths of sound, results can be obtained at and above frequencies for which the wavelength of sound in air is half the minimum dimension of the panel being measured; above this frequency it is possible to separate sound transmitted through the panel from that which has passed round the edges of the sample.

3.2. *Measurements on an Experimental Partition*

A further experimental application of the correlator was in the determination of the maximum possible sound reduction indices which could be obtained from a complex partition.

A double skin construction of expanded metal lath and plaster suspended by springs from a central framework was erected in an existing enclosure to determine the sound reduction indices of which it was capable. On completion, the conventional measurements gave poor results and it was subjectively apparent that many flanking paths existed. The correlator was used with broad bands of noise radiated from a loudspeaker and picked up by a microphone which was placed first directly in front of the loudspeaker and subsequently on the distant side of the wall.

High values of the sound reduction index were obtained in these measurements, no directly transmitted energy being apparent on the traces shown in Fig. 5. Minimum values were deduced for the sound reduction index and are shown in Fig. 6. They indicate that the experimental wall was behaving much as could be expected from a partition of this type.



(*a*) Trace with microphone in front of wall (microphone amplifier gain 80–44 dB).

(*b*) Trace with microphone behind wall (microphone amplifier gain 80 dB).

Fig. 5. Examples of correlator traces.

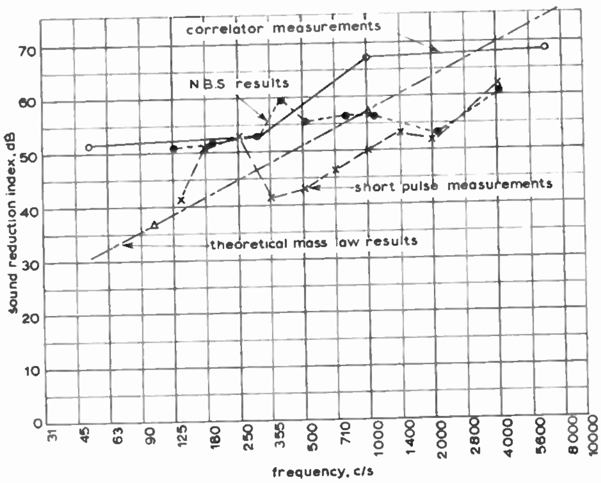


Fig. 6. Sound reduction indices for partition wall.

The results obtained by correlation were confirmed in this case by short pulse measurements (Fig. 6).

It is obvious that one shortcoming of both these measurements is that they provide results at normal incidence only. No indication would therefore be given of lower values of the sound reduction index that might exist at other angles of incidence and would tend to reduce the sound reduction index as the term is normally defined and used. Such an effect was postulated by Cremer[4] and is known as the coincidence effect. At a frequency for which the speed of flexural waves in the surface is equal to the speed of progression of the incident wave front along the surface, resonance occurs and a low sound reduction index is found. The lowest frequency at which this occurs, the critical frequency, corresponds to sound at grazing incidence exciting the flexural waves; for all higher frequencies there will exist some angle of incidence of the acoustic waves which will satisfy the above condition.

The coincidence effect was demonstrated in the laboratory for a $\frac{3}{16}$-in sheet of aluminium which was sealed into an opening. A sound field which was approximately diffuse was set up on the source side and the sound pressure level on the other side measured at a reasonable distance to define the angle from which the radiation was received. Figure 7 shows the results obtained with warble tone of centre frequency 4 kc/s where a lobe B is found at an angle of 57 deg. There is also a lobe A in the forward direction but this was principally due to a higher incident energy in this direction. The measurements were repeated with two other frequencies and from the measured coincidence angles the critical frequency for this thickness of aluminium sheet was calculated. The results are shown in Table 1 and may be seen to be reasonably self-consistent and to agree with a theoretical value derived from the constants of the material.

Table 1

Critical frequencies calculated from measured coincidence angles for $\frac{3}{16}$-in aluminium sheet (compare with 2·4 kc/s critical frequency calculated from constants of material).

Frequency	Measured Coincidence Angle	Critical Frequency
4·0 kc/s	57°	2·83 kc/s
4·8 kc/s	52°	2·88 kc/s
5·6 kc/s	44°	2·75 kc/s

Measurements were made with the correlator at the coincidence angle for an octave band of white noise centred on 5·6 kc/s and showed a sound reduction factor which was 5 dB below that for normal incidence measured close to the panel and 10 dB lower at the distance used in the previous measurements.

With this confirmation that a correlator could indicate a coincidence frequency, measurements were repeated on the experimental wall using five angles of incidence and four frequency bands. No evidence was obtained for the existence of a coincidence effect.

If a true measure of the sound reduction index of a partition were required, it would be necessary to measure at several angles of incidence and to integrate the energy transmission expected at each angle.

As previously, results for octave bands can be obtained at or above frequencies for which the wavelength of sound is half the minimum dimension of the panel being measured. Additional limitations affect low frequency measurements due to an inability to separate direct and reflected sound.

3.3. *Measurements in a Studio*

A major factor in the design of a broadcasting studio is the provision of adequate acoustic isolation between the studio, its associated control cubicle and surrounding areas. Failure to achieve sufficient protection can give rise to 'howl round' in the case of a

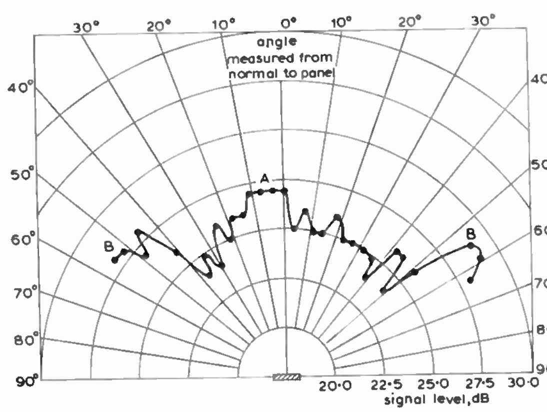


Fig. 7. Polar diagram of radiation from a $\frac{3}{16}$-in aluminium sheet.

studio and the cubicle where the programme is being monitored on a loudspeaker, or interfering noise from surrounding areas.

Conventional measurements involve the determination of the sound pressure level on either side of the common wall which frequently contains doors, windows, ventilation trunking and other possible weak links. When values of sound level reduction lower than those expected from the type of construction are found, some indication may be given by this form of measurement as to where the leakage is occurring, but generally the reverberant energy is sufficient to make this impossible. Correlation will enable the energy associated with the sound arriving by different routes with different delays to be determined and it may prove possible to associate particular time delays with certain of the physical features of the construction. If measurements are carried out with at least two sets of microphone positions, the additional information often permits the recognition of a flanking path with a fair degree of confidence.

Measurements of this kind were carried out in an enclosure converted to a film dubbing suite by mounting a partition containing a double door and double-glazed window across the centre. When the suite was tested after construction a low frequency oscillation or 'howl round' was found to build up between the microphone in the studio on one side of the partition and the loudspeaker at the other side which was used as a control cubicle. Conventional sound reduction measurements showed low values at low frequencies but of course, gave little idea of the position of the leakage. Recordings for correlation analysis were made at both sides of the room and with high and low microphone positions on the transmitted side. The results which are shown in Fig. 8 compare conventional measurements with that set of correlator results showing the earliest arrival of the largest amount of energy. The agreement at lower frequencies has proved better than would be expected but the principal achievement of the technique was shown in the difference between sets of traces. Those obtained when the loudspeaker and microphone were at the right-hand side of the cubicle showed a flanking path long compared with the direct transmission. When the loudspeaker and microphone were transferred to the left-hand side, the main path became the shortest.

Physical examination of the suspect corner subsequently showed that the leaves of the partition had not been carried solidly on to the structural walls of the building. There existed a path of low insulation passing through the acoustic treatment and around the edge of the partition.

In the search for flanking paths, it is often found that the most direct distances involved do not correspond to the measured time delays. This may be due to the

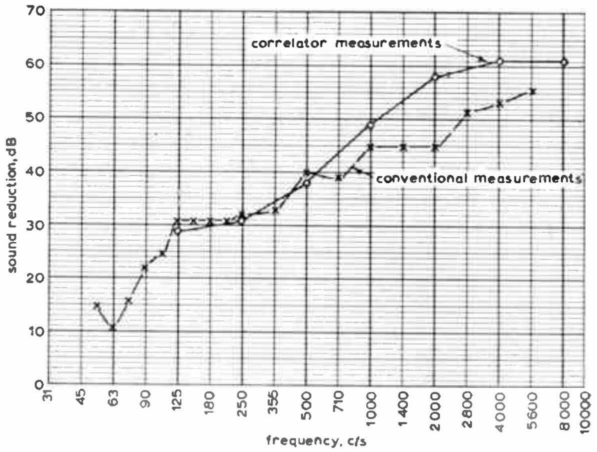


Fig. 8. Level difference produced by a partition.

directionality of the loudspeaker, where the energy emitted at right angles to the axis is normally considerably less than that along the axis. Thus a wave which has been reflected across the room and through an opening may still contain more energy than one travelling directly to the opening.

4. Application of the Correlator to Absorption Measurements

In broadcasting studios it is considered essential that the reverberation time of the enclosure should be adjusted to a value which is determined by its use and volume and that the reverberation time-frequency characteristic shall be flat at this value. To meet these requirements, suitable absorptive treatment is mounted on the surface of the room. It frequently occurs that when acceptance tests are made the reverberation time does not achieve its designed value over part or the whole of the frequency range. It would be of great assistance in the correction of such faults if they could be attributed definitely to particular absorbers. Attempts have therefore been made to determine the reflection coefficient of a surface *in situ* using a correlator.

The accuracy obtainable in these measurements is limited, but provided care is taken to control such factors as can be controlled, an indication of the reflection coefficient and hence the absorption of the surface can be obtained.

Level stability of the white-noise generator was found difficult to ensure due, apparently, to minor variations in the emission from the cathode of the thyratron used in the equipment for generating the white noise. A magnet attached to one side of the valve to reduce fluctuations in the electron beam eliminated most of the major changes but there remained short-term fluctuations of the order of ½ dB. Since this variation occurs in both signals, a change of 1 dB in the correlated output is possible. This corresponds to a change of 10% in the measured reflection coefficient and hence to an uncertainty of 20% in the corresponding absorption coefficient.

In all practical cases it would be necessary to determine for a particular loudspeaker the fall of pressure with distance and it seemed simplest to reduce the experiment to a comparison of the reflection coefficient of the absorbing surface with that of a reflecting surface. The loudspeaker and microphone were set on the normal to the surface and a sufficient distance from it to enable the direct and reflected peaks of the correlation function to be separated. A sheet of plate glass held against a wall was taken to be a perfect reflector at the frequencies at which these measurements were possible.

An increase in the ratio of direct to reflected sound when the absorber is substituted for the plate glass is due to absorption at the surface. The reflection coefficient is therefore calculable and the absorption coefficient, being $(1-r^2)$ where r is the pressure reflection coefficient, can be calculated.

An octave band centred on 1 kc/s was found to contain the lowest frequencies at which measurements could be made. The microphone-wall spacing was increased to 2 ft and in order to maintain a reasonably low direct to reflected signal ratio the loudspeaker to microphone spacing had to be increased to 3 ft. Spurious reflections from ceiling or walls prevented a lower frequency from being employed.

4.1. *Measurements on a Proprietary Absorber*

In a recent case a studio showed a rise of reverberation time at high frequencies, the subjective effect of which was accentuated by a fall at 250–350 c/s. Doubt was cast on the behaviour of the ceiling tiles which, together with a carpet, comprised practically all the high frequency absorption. Reflection coefficient measurements were made at normal incidence using the correlator, a sample of the tiles being mounted against the wall of a large room and a loudspeaker and microphone placed in line in front of it.

The results are shown in Fig. 9 together with the quoted absorption coefficients obtained by the National Physical Laboratory. The N.P.L. results are random incidence coefficients determined under standard conditions in a reverberation room and so are not strictly comparable. The other results shown for comparison were obtained by mounting the tile against the end of an impedance tube and determining the pressure distribution in a standing wave set up between a loudspeaker and the tile. These results are strictly comparable with those obtained by the correlator, being obtained at normal incidence. The greater frequency range in which the measurements could be made enabled a verification of the fact that

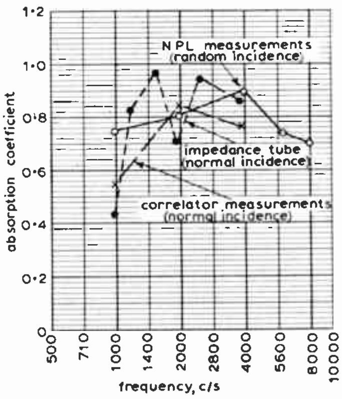


Fig. 9. Absorption coefficients of an acoustic tile.

the unexpectedly high absorption in the 250–350 c/s region was produced by the tiles in combination with the deep air space behind a false ceiling construction.

The correlator gave reasonable results which could be obtained without demounting the absorber. However, the restricted frequency range must limit very severely the applications of the technique.

5. Application to Measurements in a Ventilating System

On occasions the need has arisen to determine the noise developed by a ventilating fan or to determine the reduction of this noise produced by silencing elements within the ventilation system. This implies measurement made in the moving air stream of the fan noise in one case and possibly of an external noise source in the other. In view of the high air-speeds employed in high-velocity ventilation systems (say 3000 ft/min), a streamlined probe microphone was constructed to carry out measurements in such a system but failed because of the noise generated by the air stream.

The correlator provides a possible means of measuring such noise if two microphones are placed side by side in the air stream. The noise generated by the wind at each microphone should be incoherent and the correlated output should only be obtained for noise generated up or downstream of the measuring point.

6. Conclusions

Correlation techniques provide a means of analysing the total resultant sound at a point into components arriving by different paths. This ability enables laboratory measurements of the sound transmission characteristics to be made for limited size panels or those which are not efficiently sealed into openings. The results show reasonable agreement with values predicted by theory or those measured by other techniques. Within the laboratory and in studio measurements, it proves possible to separate and in some cases identify weak points in structures which provide low insulation flanking paths for acoustic signals.

It is possible to measure the absorption coefficient of materials *in situ* but only in a restricted frequency range.

7. Acknowledgments

The author wishes to acknowledge his gratitude to his colleagues for assistance in this work and to the Director of Engineering of the British Broadcasting Corporation for permission to publish the results.

8. References

1. K. W. Goff, "Analog electronic correlator for acoustic measurements", *J. Acoust. Soc. Amer.*, 27, pp. 223–36, March 1955.
2. K. W. Goff, "Application of correlation techniques to some acoustic measurements", *J. Acoust. Soc. Amer.*, 27, pp. 236–46, March 1955.
3. A. C. Raes, "Tentative method for the measurement of sound transmission losses in unfinished buildings", *J. Acoust. Soc. Amer.*, 27, pp. 98–102, January 1955.
4. L. Cremer, "Die wissenschaftlichen Grundlagen der Raumakustik", "Wellentheoretische Raumakustik", Vol. III, p. 173 (S. Hirzel Verlag, Leipzig 1950).

9. Appendix

The Component Parts of the Correlator

9.1. *Time Delays*

The time delays required correspond to the time for the propagation of an acoustic signal over the distances normally found in architectural acoustics, and delays up to 100 ms have proved adequate for most measurements. It is not possible to reduce the delay below the minimum value which is obtained when recording and replay heads are touching; this amounts to 60–70 ms at the tape speeds used in the present equipment. Since very short or zero time delays are frequently required, it is necessary to provide two complete recording channels with a variable delay in one channel and a fixed delay in the other. The relative delay between two channels may then be varied through zero delay up to the maximum value required.

The correlator output is very sensitive to phase variations between the input signals and requires extreme stability in the generation of the time delays. A maximum excursion of the time delay from its mean value of not more than 1/10th of a period of the highest frequency component common to both inputs of the correlator is proposed by Goff. For a time delay of 100 ms and a 10 kc/s upper frequency limit, the fluctuation in time delay would have to be less than 0·01%. In order to achieve these values, Goff utilized a rotating drum coated with magnetic material as his storage medium. The large mass involved assisted in the elimination of wow and flutter components and the use of a drum ensured that all points on the surface moved at the same speed.

However, the mechanical requirements in the construction of such a system were very stringent. Recording, replay and erase heads were maintained out of contact with the surface of the drum to eliminate wear; this necessitated a clearance between the surface of the drum and the heads which varied by less than $\pm 0{\cdot}0001$ in to maintain the high frequency response.

In view of the difficulties involved in the construction of such a delay unit, it was decided in the B.B.C. Research Department to modify an experimental tape recording system. The early version recorded the signals on a loop of tape but it proved impossible to prepare a loop in which the joint did not cause a variation of speed every time it passed a head, guide or similar obstacle. It was therefore decided to use reels of tape and to simplify the layout to reduce the numbers of guides. Figure 10 shows the drive system

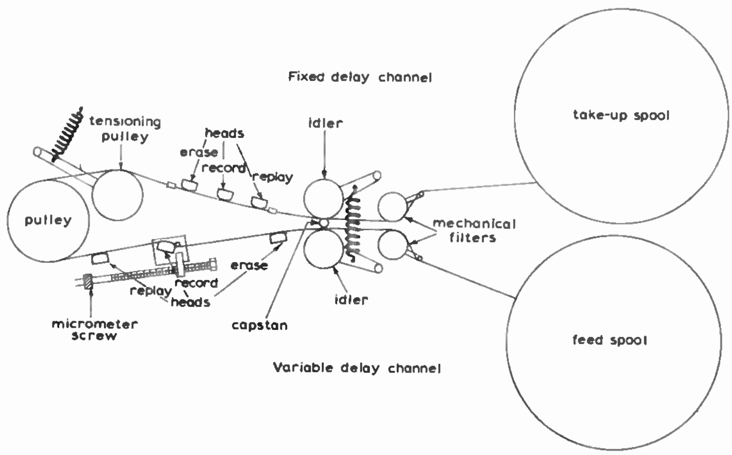


Fig. 10. Magnetic recording tape delay unit.

which is employed. It permits the isolation of the working length of tape from variations of torque arising both in the take-up and in the feed spools. The tape is driven by a capstan at two points and passes over a pulley between these two points. The two sets of recording heads are mounted on the two sides of the loop so formed.

In order to eliminate variations of tape speed due to inaccuracy in the construction or mounting of the idlers, the system would normally be arranged so that the tape contacts the capstan rather than the idler before the driving point is reached. The standards of accuracy required in the construction of the pulleys involved in the loop should be comparable with those in the capstan.

The low-frequency components of the speed fluctuation are principally due to the motor-capstan assembly and have a fundamental frequency corresponding to the rotational speed of the motor. This could be reduced below the present value of 0·05% r.m.s. with improved types of motor which are capable of 0·01%.

The high-frequency components of speed fluctuation are principally developed by friction between the recording tape and stationary points on the system. In order to reduce this flutter to the minimum possible, the heads and guides are cleaned regularly. There exists only one fixed guide around which the tape is pulled, the remaining guides merely setting the height of the tape. If there are significant variations in the width of the recording tape these guides, which must be manufactured to a close tolerance, can still give rise to variable friction.

Additional mechanical filters are incorporated outside the loop to reduce further the variations of torque arising in the motor driving the take-up spool and the brakes limiting the speed of the feed spool. They are of the conventional type having a spring-loaded arm and a flywheel with the tape wrapped round it sufficiently to avoid slip. The resonance frequency of the filter is of the order of 4–5 c/s and is below most disturbing frequencies. Sufficient damping is provided to prevent oscillation if the system is excited by a transient.

It is not possible to specify a figure of wow and flutter which would be necessary on a loop configuration such as that described here to reduce sufficiently phase variations at the input to the correlator. The measured values are of the order of 0·06% r.m.s. but vary according to which channel is measured, the time for which the system has run, the length of the variable delay, and the relative amount of tape on feed and take-up spool. A more meaningful measurement is a direct comparison of the phase at the two replay amplifiers, and this is found to be approximately one-tenth of a cycle at 4 kc/s. These factors limit the equipment described here to use at frequencies below 4 kc/s although qualitative results may be obtained at higher frequencies.

9.2. *The Multiplier*

In the equipment at present in use the multiplier is of the same form as that described by Goff and known as a quarter-squaring multiplier. A type of valve was chosen which could be arranged to give an output proportional to the square of the input. Great care must be taken in the selection of the valves, but even so this form of multiplier is sensitive to changes of mains voltage and to temperature variations.

An attempt has been made by the author's colleagues to utilize a multiplier based on the Hall effect by which a voltage is developed across the breadth of the conductor when a current flows along its length and a magnetic field exists at right angles to both these directions. Indium arsenide, a semi-conductor, has proved suitable for this application having a small temperature coefficient for the Hall effect and low resistance plates giving a greater output for a given dissipation. However, Hall effect multipliers themselves tend to drift due to heating when the maximum drive current is passed into the plate. No experience has yet been gained in the use of the new multiplier.

9.3. *Integrator and Indicating Device*

Since the definition of the correlation function requires the limiting value of the integral as the integrating time tends to infinity, an approximation only can be obtained in a finite time. A low-pass RC filter performs the integration and this requires a time of 3 *RC* seconds to come within 5% ($\frac{1}{2}$ dB) of an abrupt change in the correlation function. Four integrating times are available and may be selected to provide the optimum response under given working conditions.

The d.c. output of the integrator modulates linearly a 1-kc/s square wave, the carrier being suppressed. This signal is amplified and drives a high-speed level recorder. In order to maintain a simple relationship between the results plotted by the level recorder and the delay time (or path length with which it is directly related) the level recorder is arranged to drive the movable head of the variable delay channel at a suitable speed through its gear box; movement of this head is controlled by a micrometer screw thread.

Manuscript received by the Institution on 5th December 1963. (Paper No. 907/EA15.)

Radio Engineering Overseas . . .

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

IONOSPHERIC CHANGES DURING I.G.Y.

From the ionospheric observations by the Max Planck Institute, which were carried out at a chain of stations along the meridian 15° E during the International Geophysical Year, the mean behaviour of the critical frequency of the F2-layer, of the transmission factor, and of the maximum usable frequency (m.u.f.) has been derived as functions of the geomagnetical inclination ('latitude profiles') for days with corpuscular disturbances. For this purpose the mean deviations from the monthly median value on days with disturbances are established for the individual stations ($C_p \geqslant 1 \cdot 5$). A systematic geographical distribution of zones with positive and negative deviations, which depends on the season. As a rule positive deviations will prevail near the equator, negative deviations at higher latitudes. The results can be brought to qualitative agreement with the F2-layer theory devised by Martyn. For a quantitative check, electron profiles would be required which are not yet available.

"Ionization variations in the F2-layer of the ionosphere with latitude caused by corpuscular light disturbances from the sun", H. D. Volkmann. *Archiv der Elektrischen Übertragung*, **18**, No. 1, pp. 14–24, January 1964.

SPEECH RECOGNITION

A spoken digit recognizer has been designed by examining sound spectrograms of typical utterances in the Japanese language. Recognition is made according to the statistical parametric decision procedure based on Bayes decision rule applied to a set of eight parameters concerning the formant structures. The extraction of the parameters and the decision-making are carried out by simple transistor circuits which give 99·7% recognition for 1000 utterances of one male speaker and 97·9% for 1000 utterances of 20 male speakers.

"Spoken digit recognizer for Japanese language", K. Nagata, Y. Kato and S. Chiba. *N.E.C. Research and Development (Tokyo)*, No. 6, pp. 76–80, December 1963. (In English.)

LOW-FREQUENCY OSCILLATOR

For a particular series of measurements on magnetic amplifiers an 80-c/s oscillator was needed capable of delivering 2 W at a voltage of 50 V, a special requirement being that under severe non-linear loading the voltage should show no more than 0·01% distortion and fluctuate in amplitude by no more than 0·1%. This requirement implied that the internal resistance should be very low (0·6 ohm).

In the solution adopted by a Dutch company the instrument is divided into an oscillating section and an output stage incorporating a control system. The oscillator delivers an alternating voltage of 10 V at 80 c/s (reference voltage) which is constant to within 0·02% and shows no more than 0·003% distortion. The control system compares part of the output voltage with the reference voltage and drives the output stage (two EL 86 pentodes in a single-ended push-pull arrangement). The high loop gain enables the internal resistance of the output stage to be reduced to the required low value. The distortion in the amplifying stages of the control system is kept very low.

"A low-frequency oscillator with very low distortion under non-linear loading", G. Klein and J. J. Zaalberg van Zelst. *Philips Technical Review*, **25**, No. 1, pp. 22–30, 1963/4. (In English.)

GROUP SYNCHRONIZATION FOR P.C.M.

Group synchronization systems with digital feedback which are particularly applicable to pulse-code modulation systems have been studied by Japanese engineers. Group synchronization is used to determine the time origin of a sequence of binary systems by finding a specified deterministic pattern from a randomly modulated binary sequence. A statistical analysis has been made on the recovery process of synchronization which is a discrete-time random process with finite transitions. A new system with faster recovery characteristics has been developed. The optimal framing pattern for the resetting sequence system will almost always recover synchronism in one frame if the number of the framing digits is moderately large. The stability of the synchronizer against mis-hunting is improved by using a circuit to lock the hunting by one frame.

"Group synchronization for digital transmission systems", T. Sekimoto and H. Kaneko. *N.E.C. Research and Development (Tokyo)*, No. 6, pp. 32–43, December 1963. (In English.)

CRYSTALS FOR BANDPASS FILTERS

The properties of quartz crystals for electrical filter networks in the range 60 kc/s to 100 Mc/s have been studied by a German engineer. The knowledge of the data of the equivalent electric circuit, and freedom from unwanted modes in a specified frequency range, are the essential conditions for filter crystals. At low frequencies, the unwanted modes are produced by coupling between various modes, while for high frequency thickness modes the unwanted modes are due to overtones. For thickness-shear modes of plates with a diameter-thickness ratio greater than 40, the unwanted modes are a function of the electrode radius which also determines the data of the parameters of the equivalent electric circuit. Curves for thickness modes operating in the range 10 to 100 Mc/s show the electrode-diameter necessary for the suppression of unwanted modes to 40 dB and the resultant motional capacitance and resistance.

"Crystal modes for bandpass filters", R. Bechmann. *Archiv der Elektrischen Übertragung*, **18**, No. 2, pp. 129–36, February 1964.